RECRUITMENT IN TRANS OF GENE EDITING SYSTEM COMPONENTS

SEQUENCE LISTING

The instant application contains a Sequence Listing which has been submitted electronically in XML format and is hereby incorporated by reference in its entirety. Said XML copy, created on September 2, 2022, is named V2065-7030WO_SL.xml and is 15,727,041 bytes in size.

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No. 63/242,003, filed September 8, 2021. The contents of the aforementioned applications are hereby incorporated by reference in their entirety.

BACKGROUND

Integration of a nucleic acid of interest into a genome occurs at low frequency and with little site specificity, in the absence of a specialized protein to promote the insertion event. Some existing approaches, like CRISPR/Cas9, are more suited for small edits that rely on host repair pathways, and are less effective at integrating longer sequences. Other existing approaches, like Cre/loxP, require a first step of inserting a loxP site into the genome and then a second step of inserting a sequence of interest into the loxP site. There is a need in the art for improved compositions (e.g., proteins and nucleic acids) and methods for inserting, altering, or deleting sequences of interest in a genome.

SUMMARY OF THE INVENTION

This disclosure relates to novel compositions, systems and methods for altering a genome at one or more locations in a host cell, tissue or subject, *in vivo* or *in vitro*. In particular, the invention features compositions, systems and methods for inserting, altering, or deleting sequences of interest in a host genome.

As demonstrated in this disclosure, Applicants have discovered compositions and mechanisms for enabling editing sequences of interest in a host genome by delivering gene modifying polypeptide, or a polynucleotide encoding such polypeptide, in conjunction with separate RNA template elements, including a trans template RNA element. The present disclosure relates, in part, to association of a trans template RNA to a gene modifying polypeptide:sgRNA:target genomic DNA complex by two or more interactions. Without wishing to be bound by theory, it is has been found that such association by way of

two or more interactions or points of anchoring can achieve high rewriting activity, e.g., for achieving single or several nucleotide long edits. As described herein, examples of two of more interactions include, for example, 1) an RRS:RBP interaction, typically between the gene modifying polypeptide and the 3' end of the trans template, and 2) a 5' end block Cas9 scaffold and spacer to target DNA interaction (mediated via an additional gene modifying polypeptide). This configuration exemplifies exemplary interactions that together anchor a trans template RNA to a gene modifying polypeptide:sgRNA:target genomic DNA complex to enable rewriting. It is contemplated that the RRS:RBP interaction is critical in the absence of the 5' end block spacer. It is further contemplated that the presence of both an RRS" RBD interaction and a 5' end block spacer can provide high rewriting activity and the presence of the 5' end block spacer rescues rewriting activity observed with a trans template having a weaker RRS:RBP interaction.

Features of the compositions or methods can include one or more of the following enumerated embodiments.

1. A template RNA comprising:

a) a heterologous object sequence comprising a mutation region to introduce a mutation into a target nucleic acid sequence (wherein optionally the heterologous object sequence comprises, from 5' to 3', a post-edit homology region, the mutation region, and a pre-edit homology region), and

b) a primer binding site sequence (PBS sequence) that binds a first portion of the target nucleic acid sequence, wherein first portion is in the first strand of the target nucleic acid sequence, and wherein the PBS sequence is 3' of the heterologous object sequence, and

c) an RBD recruitment site (RRS), wherein the RRS is 3' of the PBS sequence or 5' of the heterologous object sequence.

2. A template RNA comprising:

a) a heterologous object sequence comprising a mutation region to introduce a mutation into a target nucleic acid sequence (wherein optionally the heterologous object sequence comprises, from 5' to 3', a post-edit homology region, the mutation region, and a pre-edit homology region), and

b) a primer binding site sequence (PBS sequence) that binds a first portion of the target nucleic acid sequence, wherein first portion is in the first strand of the target nucleic acid sequence, and wherein the PBS sequence is 3' of the heterologous object sequence, and

c) an RBD recruitment site (RRS), wherein optionally the RRS is situated between the PBS sequence and the heterologous object sequence, or within the heterologous object sequence (e.g., between the pre-edit homology region and the mutation region).

3. The template RNA of embodiment 1 or 2, which further comprises an end block sequence, e.g., an end block sequence of Table **41** or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto.

4. The template RNA of any of the preceding embodiments, which comprises an end block 5' of the heterologous object sequence.

5. The template RNA of any of the preceding embodiments, which comprises an end block 3' of the PBS sequence, and optionally wherein the RRS is situated between the end block and the PBS sequence.

6. The template RNA of any of the preceding embodiments, which comprises a first end block sequence 3' of the PBS sequence and a second end block sequence 5' of the heterologous object sequence.

7. The template RNA of any of embodiments 3-6, wherein the end block sequence is 5' of the heterologous object sequence and the RRS is 3' of the PBS sequence.

8. The template RNA of any of embodiments 3-6, wherein the end block sequence is 3' of the PBS sequence and the RRS is 5' of the heterologous object sequence.

9. The template RNA of any of the preceding embodiments, wherein the RRS has a sequence according to Table 40 or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto, or the reverse complement thereof.

10. The template RNA of any of the preceding embodiments, which comprises a plurality of RRSs, e.g., a tandem array of 2, 3, 4, 5, or 10 RRSs.

11. The template RNA of any if the preceding embodiments, wherein the PBS sequence is 5 – 1000 nt in length.

12. The template RNA of any if the preceding embodiments, wherein the PBS sequence comprises 8-17 nucleotides, e.g., 8-17 nucleotides of 100% identity to the target nucleic acid sequence.

13. The template RNA of any of the preceding embodiments wherein the pre-edit homology region comprises up to 30 nucleotides, e.g., up to 20 nucleotides, e.g., up to 20 nucleotides of 100% identity to the target nucleic acid sequence.

14. The template RNA of any of embodiments 1-12, which does not comprise a post-edit homology region.

15. The template RNA of any of the preceding embodiments wherein the post-edit homology region comprises 5-1000, 5-500 nucleotides, e.g., 5-500 nucleotides of 100% identity to the target nucleic acid sequence.

16. The template RNA of any embodiments 114, which does not comprise a post-edit homology region.

17. The template RNA of any of the preceding embodiments, wherein the mutation region is configured to produce an insertion, a deletion, or a substitution in the target nucleic acid.

18. The template RNA of any of the preceding embodiments, which further comprises:

a gRNA spacer that is complementary to a different portion (e.g., a third portion) of the target nucleic acid sequence, e.g., wherein the different portion (e.g., third portion) is on the first strand of the target nucleic acid sequence; and

a gRNA scaffold.

19. The template RNA of embodiment 18, wherein the gRNA spacer is 5' of the heterologous object sequence.

20. The template RNA of embodiment 18 or 19, wherein the gRNA scaffold is situated between the gRNA spacer and the heterologous object sequence.

21. The template RNA of any of embodiments 18-20 wherein the gRNA spacer and the PBS sequence bind the same strand of the target nucleic acid sequence.

22. The template RNA of any of embodiments 18-21 wherein the gRNA spacer, the heterologous object sequence, and the PBS sequence bind the same strand of the target nucleic acid sequence.

23. The template RNA of any of embodiments 1-8, which does not comprise a gRNA spacer or a gRNA scaffold.

24. The template RNA of any of the preceding embodiments, which comprises a linker of up to 20 nucleotides between the RRS and the PBS sequence.

25. A gene modifying polypeptide comprising:

a reverse transcriptase (RT) domain; and

a DNA binding domain (DBD) that binds to a target nucleic acid sequence and is heterologous to the RT domain (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain); and

a RNA-binding domain (RBD) that is heterologous to the DBD and the RT domain.

26. A gene modifying polypeptide comprising:

a reverse transcriptase (RT) domain; and

a DNA binding domain (DBD) that binds to a target nucleic acid sequence and is heterologous to the RT domain (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain); and

a RNA-binding domain (RBD) that is heterologous to the DBD and the RT domain, wherein the domains are arranged, in an N-terminal to C-terminal direction:

a) DBD, RT domain, RBD;
b) RT domain, DBD, RBD;
c) RBD, DBD, RT domain;
d) RBD, RT domain, DBD;
e) DBD, RBD, RT domain; or
f) RT domain, RBD, DBD.

27. A gene modifying polypeptide comprising:

a reverse transcriptase (RT) domain; and

a DNA binding domain (DBD) that binds to a target nucleic acid sequence and is heterologous to the RT domain (e.g., a Cas nickase domain, e.g., a Cas9 nickase domain); and

a plurality (e.g., 2, 3, 4, or 5) RNA-binding domains (RBD) that are heterologous to the DBD and the RT domain.

28. The gene modifying polypeptide of embodiment 27, wherein the RBD has an amino acid sequence according to Table 31, or at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

29. The gene modifying polypeptide of any of of the preceding embodiments wherein the plurality of RBDs have the same amino acid sequence as each other.

30. The gene modifying polypeptide of any of the preceding embodiments, wherein the plurality of RBDs have different amino acid sequences from each other.

31. The gene modifying polypeptide of any of the preceding embodiments, wherein the DBD has an amino acid sequence according to Table 7 or 8, or at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

32. The gene modifying polypeptide of any of the preceding embodiments, wherein the RT domain is from a retrovirus, or a polypeptide domain having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acids sequence identity thereto.

33. The gene modifying polypeptide of any of the preceding embodiments, wherein the RT domain has an amino acid sequence according to Table 6, or at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

34. The gene modifying polypeptide of any of the preceding embodiments, wherein the gene modifying polypeptide comprises a linker.

35. The gene modifying polypeptide of any of the preceding embodiments, wherein the linker comprises a sequence according to Table 10, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

36. The gene modifying polypeptide of embodiment 34 or 35, wherein the linker is disposed between the DBD and the RT domain, the RT domain and the RBD, or between the RBD and the DBD.

37. The gene modifying polypeptide of any of the preceding embodiments, wherein the gene modifying polypeptide comprises, in an N-terminal to C-terminal direction:

a) the DBD, a first linker, the RT domain, a second linker, the RBD;
b) the RT domain, a first linker, the DBD, a second linker, the RBD;
c) the RBD, a first linker, the DBD, a second linker, the RT domain;
d) RBD, a first linker, RT domain, a second linker, DBD;
e) the DBD, a first linker, the RBD, a second linker, the RT domain; or
f) the RT domain, a first linker, the RBD, a second linker, the DBD.

38. The gene modifying polypeptide of any of the preceding embodiments, which was produced by intein-mediated fusion of an N-terminal portion comprising an intein-N domain and a C-terminal portion comprising an intein-C domain.

39. A polypeptide system (e.g., a polypeptide complex) comprising:

a) a reverse transcriptase (RT) domain; and

b) a DNA binding domain (DBD) that binds to a target nucleic acid sequence and is heterologous to the RT domain (e.g., a Cas domain, e.g., a Cas9 domain, e.g., a Cas9 nickase domain); and

c) a RNA-binding domain (RBD) that is heterologous to the DBD and the RT domain,

wherein at least 2 of (e.g., all of) (a), (b), and (c) are in separate polypeptides, e.g., separate polypeptides that noncovalently form a complex.

40. The polypeptide system of embodiment 39, wherein complex formation is mediated by a first dimerization domain that binds a second, compatible dimerization domain.

41. The polypeptide system of embodiment 40, wherein complex formation is mediated by a third dimerization domain that binds a fourth, compatible dimerization domain.

42. The polypeptide system of any of embodiments 39-41, wherein:

the RBD is operably linked (e.g., via a linker) to a first dimerization domain;

the DBD is operably linked (e.g., via a linker) to a second dimerization domain that binds the first dimerization domain;

the DBD is operably linked (e.g., via a linker) to a third dimerization domain; and

the RT domain is operably linked (e.g., via a linker) to a fourth dimerization domain that binds the third dimerization domain.

43. The polypeptide system of any of embodiments 39-42 wherein the first and second dimerization domains are: chemical- induced dimerization domains, light-induced dimerization domains, antibody-peptide dimerization domains, or coiled coil dimerization domains.

44. The polypeptide system of any of embodiments 39-43, wherein the third and fourth dimerization domains are: chemical- induced dimerization domains, light-induced dimerization domains, antibody-peptide dimerization domains, or coiled coil dimerization domains.

45. The polypeptide system of any of embodiments 39-44wherein the first dimerization domain and the second dimerization domain are each present in a plurality of copies, e.g., 2, 3, 4, 5, 10, 15, 20, or 30 copies.

46. The polypeptide system of any of embodiments 39-45, wherein the third dimerization domain and the fourth dimerization domain are each present in a plurality of copies, e.g., 2, 3, 4, 5, 10, 15, 20, or 30 copies.

47. The polypeptide system of any of embodiments 39-46, wherein the first dimerization domain and the second dimerization domain have the same sequence (e.g., wherein the first dimerization domain and the second dimerization domain form a homodimer).

48. The polypeptide system of any of embodiments 39-47 wherein the third dimerization domain and the fourth dimerization domain have the same sequence (e.g., wherein the third dimerization domain and the fourth dimerization domain form a homodimer).

49. The polypeptide system of any of embodiments 39-48wherein the first dimerization domain and the second dimerization domain have different sequences (e.g., wherein the first dimerization domain and the second dimerization domain form a heterodimer).

50. The polypeptide system of any of embodiments 39-49 wherein the third dimerization domain and the fourth dimerization domain have different sequences (e.g., wherein the third dimerization domain and the fourth dimerization domain form a hetero dimer).

51. The polypeptide system of any of embodiments 39-50 wherein the DBD is operably linked to one or more additional DBDs, wherein optionally the additional DBDs have the same sequence as the DBD.

52. The polypeptide system of any of embodiments 39-51 wherein the RBD has an amino acid sequence according to Table 31, or at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

53. The polypeptide system of any of embodiments 39-52, wherein the plurality of RBDs have the same amino acid sequence as each other.

54. The polypeptide system of any of embodiments 39-52 wherein the plurality of RBDs have different amino acid sequences from each other.

55. The polypeptide system of any of embodiments 39-54 wherein the DBD has an amino acid sequence according to Table 7 or 8, or at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

56. The polypeptide system of any of embodiments 39-55, wherein the RT domain is from a retrovirus, or a polypeptide domain having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acids sequence identity thereto.

57. The polypeptide system of any of embodiments 39-56 wherein the RT domain has an amino acid sequence according to Table 6, or at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

58. The polypeptide system of any of embodiments 39-57 wherein each linker independently comprises a sequence according to Table 10, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

59. A nucleic acid or a plurality of nucleic acids encoding the polypeptides of any of the systems of embodiment 39-57.

60. A system comprising:

a template RNA of any of embodiments 1-24;

a gene modifying polypeptide of any of embodiments 25-38 or the polypeptide system of any of embodiments 39-58; and

a first gRNA comprising:

a gRNA spacer that binds a second portion of the target nucleic acid sequence, wherein the second portion is one the second strand of the target nucleic acid sequence; and

a gRNA scaffold that binds the DBD of the gene modifying polypeptide or the polypeptide system.

61. The system of embodiment 60, wherein the template RNA does not comprise a gRNA spacer or a gRNA scaffold.

62. The system of embodiment 60 or 61, wherein the gRNA spacer binds to a region of the target nucleic acid sequence that is within about 5, 10, 15, 20, 25, 30, or 40 nucleotides of the region of the target nucleic acid sequence bound by the PBS sequence.

63. The system of any of embodiments 60-62, which further comprises:

a second Cas protein (e.g., a dead Cas protein) and

a second gRNA comprising:

a gRNA spacer that binds the first strand of the target nucleic acid at a location 3' of the location bound by the PBS sequence, and

a gRNA scaffold that binds the second Cas protein.

64. The system of embodiment 63, wherein the second Cas protein is a dead Cas protein (e.g., a dead Cas9 protein) or a Cas nickase protein (e.g., a Cas9 nickase protein)

65. The system of embodiment 63, wherein the gRNA spacer of the second gRNA has a length of at least 18 nucleotides (e.g., 18-28 nucleotides, e.g., 18-21 nucleotides) and the second Cas protein is a dead Cas protein.

66. The system of embodiment 63, wherein the gRNA spacer of the second gRNA has a length of 17 nucleotides or less (e.g., 14-17 nucleotides), wherein optionally the second Cas protein is a Cas nickase protein.

67. The system of embodiment 60, wherein the template RNA further comprises:

a gRNA spacer that is complementary to a third portion of the target nucleic acid sequence wherein the third portion is on the first strand of the target nucleic acid sequence; and

a gRNA scaffold.

68. The system of embodiment 67, wherein the gRNA scaffold binds the DBD of the gene modifying polypeptide or the polypeptide system.

69. The system of embodiment 67 or 68, wherein the gRNA spacer has a length of 17 nucleotides or less.

70. The system of any of embodiments 60-69, wherein the gRNA spacer of the template RNA induces nicking of the template nucleic acid, e.g., at the second strand of the target nucleic acid sequence.

71. The system of any of embodiments 60-69, wherein the gRNA spacer of the template RNA does not induce nicking of the template nucleic acid.

72. A system comprising:

i) a template RNA of any of embodiments 1-24 (e.g., a template RNA of embodiment 23);

ii) a first polypeptide comprising:

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain); and

a RNA-binding domain (RBD) that is heterologous to the DBD, wherein the RBD binds the RRS of the template RNA;

iii) a first gRNA comprising:

a gRNA spacer that directs the DBD of the first polypeptide to a second portion of the target nucleic acid sequence, wherein the second portion of the target nucleic acid sequence is on the second strand of the nucleic acid sequence; and

a gRNA scaffold that binds the DBD of the first polypeptide;

iv) a second polypeptide comprising:

an RT domain, and

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain), that is heterologous to the RT domain, and wherein the DBD of the second polypeptide has a different sequence from the DBD of the first polypeptide; and

v) a second gRNA comprising:

a gRNA spacer that directs the DBD of the second polypeptide to a third portion of the target nucleic acid sequence, wherein the third portion is on the first strand of the target nucleic acid, and

a gRNA scaffold that binds the DBD of the second polypeptide.

73. The system of embodiment 72, wherein the DBD of the second polypeptide comprises a Cas nickase domain or a dead Cas domain.

74. The system of embodiment 72, wherein the gRNA spacer of the second RNA induces nicking of the template nucleic acid, e.g., at the second strand of the target nucleic acid sequence.

75. The system of embodiment 72, wherein the gRNA spacer of the second RNA does not induce nicking of the template nucleic acid.

76. The system of embodiment 72, wherein the first gRNA does not detectably bind to the DBD of the second polypeptide.

77. The system of embodiment 72, wherein the second gRNA does not detectably bind to the DBD of the first polypeptide.

78. A system comprising:

i) a template RNA of any of embodiments 1-24 wherein the template RNA comprises:

a gRNA spacer that is complementary to a third portion of the target nucleic acid sequence wherein the third portion is on the first strand of the target nucleic acid sequence; and

a gRNA scaffold;

ii) a first polypeptide comprising:

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain); and

a RNA-binding domain (RBD) that is heterologous to the DBD, wherein the RBD binds the RRS of the template RNA;

iii) a first gRNA comprising:

a gRNA spacer that directs the DBD of the first polypeptide to a second portion of the target nucleic acid sequence, wherein the second portion of the target nucleic acid sequence is on the second strand of the nucleic acid sequence; and

a gRNA scaffold that binds the DBD of the first polypeptide; and

iv) a second polypeptide comprising:

an RT domain, and

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain), that is heterologous to the RT domain, and wherein the DBD of the second polypeptide has a different sequence from the DBD of the first polypeptide, and wherein the gRNA scaffold of the template RNA binds the DBD of the second polypeptide.

79. The system of embodiment 78, wherein the DBD of the second polypeptide comprises a Cas nickase domain or a dead Cas domain.

80. The system of embodiment 78, wherein the gRNA spacer of the template RNA induces nicking of the template nucleic acid, e.g., at the second strand of the target nucleic acid sequence.

81. The system of embodiment 78, wherein the gRNA spacer of the template RNA does not induce nicking of the template nucleic acid.

82. The system of any of embodiments 78-, wherein the first gRNA does not detectably bind to the DBD of the second polypeptide.

83. The system of any of embodiments 78-82, wherein the gRNA of the template RNA does not detectably bind to the DBD of the first polypeptide.

84. A polypeptide system comprising:

a first polypeptide comprising:

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain);

a RNA-binding domain (RBD) that is heterologous to the DBD; and

optionally, a linker disposed between the DBD and the RBD; and

a second polypeptide comprising:

an RT domain, and

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain), that is heterologous to the RT domain; and

optionally, a linker disposed between the RT domain and the DBD.

85. The template RNA or system of any of embodiments 1-24 or 60-83, wherein the target nucleic acid sequence is a target gene, enhancer, or promoter.

86. The template RNA of system of embodiment 85wherein the target nucleic acid sequence is a human target gene, human enhancer, or human promoter.

87. The system or polypeptide system of any of the preceding embodiments, wherein the RBD has a sequence of Table 31, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto.

88. A method for modifying a target nucleic acid in a cell (e.g., a human cell), the method comprising contacting the cell with the system of any one of embodiments60-83, or nucleic acid encoding the same, thereby modifying the target nucleic acid.

89. The method of embodiment 88, wherein presence of the second polypeptide, compared to an otherwise similar system lacking the second polypeptide, results in one or more of:

increased unwinding of the target nucleic acid;

increased number of target nucleic acids that are modified;

increased length of insertion into the target nucleic acid; or

reduced MMR activity at the target nucleic acid.

90. The method of embodiment 88 or 89, wherein the cell is in vivo or ex vivo.

91. A template RNA comprising:

a) a heterologous object sequence comprising a mutation region to introduce a mutation into a target nucleic acid sequence (wherein optionally the heterologous object sequence comprises, from 5' to 3', a post-edit homology region, the mutation region, and a pre-edit homology region), and

b) a primer binding site sequence (PBS sequence) that binds a first portion of the target nucleic acid sequence, wherein first portion is in the first strand of the target nucleic acid sequence, and wherein the PBS sequence is 3' of the heterologous object sequence, and

c) an RBD recruitment site (RRS), wherein the RRS is 3' of the PBS sequence or 5' of the heterologous object sequence.

92. The template RNA of embodiment 91, wherein the RRS comprises the RRS of a template sequence as listed in Table S4, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto.

93. The template RNA of embodiment 91 or 92, which further comprises an end block sequence, e.g., an end block sequence of Table 41, or comprising a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto.

94. The template RNA of embodiment 93, wherein the end block sequence is 5' of the heterologous object sequence (e.g., located at the 5' end of the template RNA), optionally wherein the RRS is 3' of the PBS sequence.

95. The template RNA of embodiment 94, wherein the end block sequence comprises a gRNA scaffold.

96. The template RNA of embodiment 95, wherein the gRNA scaffold is chosen from Table 41.

97. The template RNA of embodiment 95, wherein the gRNA scaffold is a Cas9 scaffold.

98. The template RNA of any of embodiments 93-97, wherein the end block sequence comprises a gRNA spacer, e.g., positioned at the 5' end of the end block (e.g., 5' of the gRNA scaffold and/or positioned at the 5' end of the template RNA).

99. The template RNA of any of embodiments 94-98, wherein the gRNA spacer is a pro-spacer (e.g., as described herein).

100. The template RNA of embodiment 98, wherein the end block binds to a DNA binding domain, e.g., of a gene modifying polypeptide (e.g., as described herein).

101. The template RNA of embodiment 100, wherein the gene modifying polypeptide bound to the end block does not create a nick in the second strand of the target nucleic acid sequence.

102. The template RNA of any of embodiments 98-101, wherein the gRNA spacer binds to a second portion of the first strand of the target nucleic acid sequence located 3' relative to the first portion of the target nucleic acid sequence.

103. The template RNA of embodiment 102, wherein the 5' end of the portion of the first strand bound by the gRNA spacer is between 10-20, 20-30, 30-40, 40-50, 50-100, 100-150, or 150-200 nucleotides from the 3' end of the first portion.

104. The template RNA of any of embodiments 98-103, wherein:

(i) the gRNA spacer has a length of less than or equal to 17 nucleotides, e.g., about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or 17 nucleotides;

(ii) the gRNA spacer has 100% complementarity to the second portion on the first strand of the target nucleic acid sequence; and/or

(iii) the gRNA spacer directs nicking activity by a Cas domain..

105. The template RNA of embodiment 104, wherein:

(i) the gRNA spacer has a length of less than or equal to 17 nucleotides, e.g., about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or 17 nucleotides; and

(ii) the gRNA spacer has 100% complementarity to the second portion on the first strand of the target nucleic acid sequence.

106. The template RNA of embodiment 104, wherein:

(ii) the gRNA spacer has 100% complementarity to the second portion on the first strand of the target nucleic acid sequence; and

(iii) the gRNA spacer directs nicking activity by a Cas domain.

107. The template RNA of any of embodiments 93-106, wherein the end block sequence is 3' of the PBS sequence and/or the RRS (e.g., located at the 3' end of the template RNA), optionally wherein the RRS is 5' of the heterologous object sequence.

108. The template RNA of embodiment 107, wherein the end block sequence comprises GGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC (SEQ ID NO: 18,101), an end block sequence of Table 41, or comprising a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity to any thereof.

109. The template RNA of any of embodiments 93-108, wherein the end block sequence comprises an aptamer.

110. The template RNA of any of embodiments 93-109, wherein the end block sequence is capable of binding to an RNA aptamer-binding protein (e.g., an RNA aptamer-binding protein attached to a gene modifying polypeptide, e.g., at the DBD).

111. The template RNA of any of embodiments 93-110, wherein the end block comprises one or more hairpins (e.g., 1, 2, 3, 4, or 5 hairpins).

112. The template RNA of any of embodiments 93-111, wherein the end block comprises an ePEG end block.

113. The template RNA of any of embodiments 91-92, further comprising:

a 5' end block sequence, e.g., an end block sequence of Table 41, or comprising a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto, wherein the 5' end block sequence is 5' of the heterologous object sequence (e.g., located at the 5' end of the template RNA), optionally wherein the RRS is 3' of the PBS sequence; and

a 3' end block sequence, e.g., an end block sequence of Table 41 or the sequence GGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC (SEQ ID NO: 18,101), or comprising a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity to any thereof, wherein the 3' end block sequence is 3' of the PBS sequence and/or the RRS (e.g., located at the 3' end of the template RNA), optionally wherein the RRS is 5' of the heterologous object sequence.

114. The template RNA of any of the preceding embodiments, wherein the RRS comprises an MS2 sequence.

115. The template RNA of any of the preceding embodiments, wherein the RRS binds to an MCP polypeptide.

116. The template RNA of any of the preceding embodiments, wherein the RRS comprises a PP7 sequence.

117. The template RNA of any of the preceding embodiments, wherein the RRS and the PBS are separated by a region having of length of about 5-10, 10-15, or 15-20 nucleotides (e.g., about 8 nucleotides or about 16 nucleotides).

118. The template RNA of any of the preceding embodiments, wherein the RRS has a sequence according to Table 40 or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto.

119. The template RNA of any of the preceding embodiments, which comprises a plurality of RRSes (e.g., identical or different RRSes), e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 RRSes, e.g., a tandem array of 2, 3, 4, 5, or 10 RRSs.

120. The template RNA of embodiment 119, wherein the plurality of RRSes each comprises an MS2 sequence.

121. The template RNA of embodiment 119 or 120, wherein the plurality of RRSes comprises 4 repeats of the MS2 sequence.

122. The template RNA of any of the preceding embodiments, wherein the PBS sequence comprises 8-17 nucleotides, e.g., 8-17 nucleotides of 100% identity to the target nucleic acid sequence.

123. The template RNA of embodiment 122, wherein the PBS sequence has a length of about 8, 13, or 17 nucleotides.

124. The template RNA of embodiment 122, wherein the PBS sequence has a length of about 13 nucleotides.

125. The template RNA of any of the preceding embodiments, wherein the pre-edit homology region comprises up to 20 nucleotides, e.g., up to 20 nucleotides of 100% identity to the target nucleic acid sequence.

126. The template RNA of any of the preceding embodiments, wherein the post-edit homology region comprises 5-500 nucleotides, e.g., 5-500 nucleotides of 100% identity to the target nucleic acid sequence.

127. The template RNA of any of the preceding embodiments, wherein the post-edit homology region comprises 10-20, 20-30, 30-40, 40-50, 50-60, or 60-70 nucleotides, e.g., about 12 nucleotides or about 63 nucleotides.

128. The template RNA of embodiment 127, wherein the post-edit homology region comprises one or more (e.g., 1, 2, 3, 4, or 5) single nucleotide substitutions, e.g., at approximately regular intervals (e.g., spaced about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides apart).

129. The template RNA of any of the preceding embodiments, wherein the mutation region is configured to produce an insertion, a deletion, or a substitution in the target nucleic acid.

130. The template RNA of any of the preceding embodiments, wherein the gRNA spacer is complementary to a different portion (e.g., a third portion) of the target nucleic acid sequence, e.g., wherein the different portion (e.g., third portion) is on the first strand of the target nucleic acid sequence.

131. The template RNA of embodiment 130, wherein the gRNA spacer is 5' of the heterologous object sequence.

132. The template RNA of embodiment 130 or 131, wherein the gRNA scaffold is situated between the gRNA spacer and the heterologous object sequence.

133. The template RNA of any of embodiments 130-132 wherein the gRNA spacer and the PBS sequence bind the same strand of the target nucleic acid sequence.

134. The template RNA of any of embodiments 130-133 wherein the gRNA spacer, the heterologous object sequence, and the PBS sequence bind the same strand of the target nucleic acid sequence.

135. The template RNA of any of embodiments 91-129, which does not comprise a gRNA spacer or a gRNA scaffold.

136. The template RNA of any of the preceding embodiments, which comprises a linker of up to 20 nucleotides between the RRS and the PBS sequence.

137. The template RNA of any of the preceding embodiments, wherein the template RNA is linear.

138. The template RNA of any of the preceding embodiments, wherein the template RNA is circular.

139. A gene modifying polypeptide comprising:

a reverse transcriptase (RT) domain; and

a DNA binding domain (DBD) that binds to a target nucleic acid sequence and is heterologous to the RT domain (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain); and

a RNA-binding domain (RBD) that is heterologous to the DBD and the RT domain,

wherein the domains are arranged, in an N-terminal to C-terminal direction:

g) DBD, RT domain, RBD;

h) RT domain, DBD, RBD;

i) RBD, DBD, RT domain;

j) RBD, RT domain, DBD;

k) DBD, RBD, RT domain; or

l) RT domain, RBD, DBD.

140. The gene modifying polypeptide of embodiment 139, further comprising one or more (e.g., 1, 2, 3, or 4) additional RBDs (e.g., one or more additional copies of the RBD, e.g., adjacent to the RBD).

141. A gene modifying polypeptide comprising:

a reverse transcriptase (RT) domain; and

a DNA binding domain (DBD) that binds to a target nucleic acid sequence and is heterologous to the RT domain (e.g., a Cas nickase domain, e.g., a Cas9 nickase domain); and

a plurality (e.g., 2, 3, 4, or 5) RNA-binding domains (RBD) that are heterologous to the DBD and the RT domain.

142. The gene modifying polypeptide of any of the preceding embodiments, wherein the RBD comprises an amino acid sequence according to Table 31 or the amino acid sequence of the RBD of a gene modifying polypeptide as listed in any of Tables S1-S3, or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

143. The gene modifying polypeptide of any of the preceding embodiments, wherein the plurality of RBDs have the same amino acid sequence as each other.

144. The gene modifying polypeptide of any of the preceding embodiments, wherein the plurality of RBDs have different amino acid sequences from each other.

145. The gene modifying polypeptide of any of the preceding embodiments, wherein the DBD comprises an amino acid sequence according to Table 7 or 8 or the amino acid sequence of the DBD of a gene modifying polypeptide as listed in any of Tables S1-S3, or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

146. The gene modifying polypeptide of any of the preceding embodiments, wherein the RT domain is from a retrovirus, or a polypeptide domain having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acids sequence identity thereto.

147. The gene modifying polypeptide of any of the preceding embodiments, wherein the RT domain comprises an amino acid sequence according to Table 6 or the amino acid sequence of the RT domain of a gene modifying polypeptide as listed in any of Tables S1-S3, or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

148. The gene modifying polypeptide of any of the preceding embodiments, wherein:

(a) the RBD comprises an amino acid sequence of the RBD of a gene modifying polypeptide as listed in any of Tables S1-S3, or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto;

(b) the DBD comprises an amino acid sequence of the DBD of said gene modifying polypeptide listed in any of Tables S1-S3, or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and

(c) the RT domain comprises an amino acid sequence of the RT domain of said gene modifying polypeptide listed in any of Tables S1-S3, or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

149. The gene modifying polypeptide of any of the preceding embodiments, wherein the gene modifying polypeptide comprises a linker.

150. The gene modifying polypeptide of embodiment 149, wherein the linker is 2-5 amino acids in length (e.g., 4 amino acids in length).

151. The gene modifying polypeptide of embodiment 149, wherein the linker is 5-10 amino acids in length (e.g., 8 amino acids in length).

152. The gene modifying polypeptide of embodiment 149, wherein the linker is 10-20 amino acids in length (e.g., 16 amino acids in length).

153. The gene modifying polypeptide of any of embodiments 149-152, wherein the linker comprises a sequence according to Table 10, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

154. The gene modifying polypeptide of any of embodiments 149-153, wherein the linker is disposed between the DBD and the RT domain, the RT domain and the RBD, or between the RBD and the DBD.

155. The gene modifying polypeptide of any of embodiments 149-154, which comprises a first linker and a second linker, wherein:

(i) the first linker is disposed between the DBD and the RT domain and the second linker is disposed between the RT domain and the RBD;

(ii) the first linker is disposed between the DBD and the RBD and the second linker is disposed between the RBD and RT domain; or

(iii) the first linker is disposed between the RT domain and the DBD and the second linker is disposed between the DBD and RBD.

156. The gene modifying polypeptide of any of the preceding embodiments, wherein the gene modifying polypeptide comprises, in an N-terminal to C-terminal direction:

g) the DBD, a first linker, the RT domain, a second linker, the RBD;
h) the RT domain, a first linker, the DBD, a second linker, the RBD;
i) the RBD, a first linker, the DBD, a second linker, the RT domain;
j) RBD, a first linker, RT domain, a second linker, DBD;
k) the DBD, a first linker, the RBD, a second linker, the RT domain; or
l) the RT domain, a first linker, the RBD, a second linker, the DBD.

157. The gene modifying polypeptide of any of the preceding embodiments, which was produced by intein-mediated fusion of an N-terminal portion comprising an intein-N domain and a C-terminal portion comprising an intein-C domain.

158. The gene modifying polypeptide of any of the preceding embodiments, wherein the DBD comprises a Cas domain, e.g., a Cas9 domain, e.g., a Cas9 nickase domain (e.g., as described herein).

159. The gene modifying polypeptide embodiment 158, wherein the Cas domain is a dCas9 domain.

160. The gene modifying polypeptide embodiment 158, wherein the Cas domain is an nCas9 domain.

161. The gene modifying polypeptide of any of the preceding embodiments, wherein the RT domain comprises an AVIRE domain (e.g., as described herein, e.g., an AVIRE RT domain as listed in Table 6), or an amino acid sequence have at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

162. The gene modifying polypeptide of embodiment 161, wherein the PBS sequence has a length of greater than 8 nucleotides, e.g., about 9, 10, 11, 12, 13, 14, 15, 16, or 17 nucleotides.

163. The gene modifying polypeptide of any of the preceding embodiments, wherein the RT domain comprises an MLVMS domain (e.g., as described herein, e.g., an MLVMS RT domain as listed in Table 6), or an amino acid sequence have at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity thereto.

164. The gene modifying polypeptide of any of the preceding embodiments, wherein the RT domain comprises a retrotransposon RT domain.

165. The gene modifying polypeptide of any of the preceding embodiments, wherein the domains are arranged, in an N-terminal to C-terminal direction:

a) DBD, RT domain, RBD;
b) RT domain, DBD, RBD;
c) RBD, DBD, RT domain;
d) RBD, RT domain, DBD;
e) DBD, RBD, RT domain; or
f) RT domain, RBD, DBD.

166. The gene modifying polypeptide of embodiment 165, further comprising one or more (e.g., 1, 2, 3, or 4) additional RBDs (e.g., one or more additional copies of the RBD, e.g., adjacent to the RBD).

167. The gene modifying polypeptide of embodiment 165 or 166, further comprising one or more additional RT domains (e.g., one or more additional copies of the RT domain, e.g., adjacent to the RT domain).

168. The gene modifying polypeptide of embodiment 167, wherein one or more of the additional RT domains comprises an AVIRE domain (e.g., as described herein).

169. The gene modifying polypeptide of embodiment 167 or 168, wherein one or more of the additional RT domains comprises an MLVMS domain (e.g., as described herein).

170. The gene modifying polypeptide of any of the preceding embodiments, further comprising an RNA aptamer-binding domain.

171. The gene modifying polypeptide of embodiment 170, wherein the DBD is attached to the RNA aptamer-binding domain, e.g., via a linker.

172. A polypeptide system (e.g., a polypeptide complex) comprising:

a) a reverse transcriptase (RT) domain; and

b) a DNA binding domain (DBD) that binds to a target nucleic acid sequence and is heterologous to the RT domain (e.g., a Cas domain, e.g., a Cas9 domain, e.g., a Cas9 nickase domain); and

c) a RNA-binding domain (RBD) that is heterologous to the DBD and the RT domain,

wherein at least 2 of (e.g., all of) (a), (b), and (c) are in separate polypeptides, e.g., separate polypeptides that noncovalently form a complex.

173. The polypeptide system of embodiment 172, wherein the RT domain and the DBD are in separate polypeptides.

174. The polypeptide system of embodiment 172, wherein the RT domain and the RBD are in separate polypeptides.

175. The polypeptide system of embodiment 172, wherein complex formation is mediated by a first dimerization domain that binds a second, compatible dimerization domain.

176. The polypeptide system of embodiment 172, wherein complex formation is mediated by a third dimerization domain that binds a fourth, compatible dimerization domain.

177. The polypeptide system of any of embodiments 172-176, wherein:

the RBD is operably linked (e.g., via a linker) to a first dimerization domain;

the DBD is operably linked (e.g., via a linker) to a second dimerization domain that binds the first dimerization domain;

the DBD is operably linked (e.g., via a linker) to a third dimerization domain; and

the RT domain is operably linked (e.g., via a linker) to a fourth dimerization domain that binds the third dimerization domain.

178. The polypeptide system of any of embodiments 172-177, wherein the first and second dimerization domains are: chemical- induced dimerization domains, light-induced dimerization domains, antibody-peptide dimerization domains, or coiled coil dimerization domains.

179. The polypeptide system of any of embodiments 172-178, wherein the third and fourth dimerization domains are: chemical- induced dimerization domains, light-induced dimerization domains, antibody-peptide dimerization domains, or coiled coil dimerization domains.

180. The polypeptide system of any of embodiments 172-179, wherein the first dimerization domain and the second dimerization domain are each present in a plurality of copies, e.g., 2, 3, 4, 5, 10, 15, 20, or 30 copies.

181. The polypeptide system of any of embodiments 172-180, wherein the third dimerization domain and the fourth dimerization domain are each present in a plurality of copies, e.g., 2, 3, 4, 5, 10, 15, 20, or 30 copies.

182. The polypeptide system of any of embodiments 172-181, wherein the first dimerization domain and the second dimerization domain have the same sequence (e.g., wherein the first dimerization domain and the second dimerization domain form a homodimer).

183. The polypeptide system of any of embodiments 172-182, wherein the third dimerization domain and the fourth dimerization domain have the same sequence (e.g., wherein the third dimerization domain and the fourth dimerization domain form a homodimer).

184. The polypeptide system of any of embodiments 172-181, wherein the first dimerization domain and the second dimerization domain have different sequences (e.g., wherein the first dimerization domain and the second dimerization domain form a heterodimer).

185. The polypeptide system of any of embodiments 172-184, wherein the third dimerization domain and the fourth dimerization domain have different sequences (e.g., wherein the third dimerization domain and the fourth dimerization domain form a hetero dimer).

186. The polypeptide system of any of embodiments 172-185, wherein the DBD is operably linked to one or more additional DBDs, wherein optionally the additional DBDs have the same sequence as the DBD.

187. The polypeptide system of any of embodiments 172-186, wherein the RBD has an amino acid sequence according to Table 31, or at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

188. The polypeptide system of any of embodiments 172-187, wherein the plurality of RBDs have the same amino acid sequence as each other.

189. The polypeptide system of any of embodiments 172-188, wherein the plurality of RBDs have different amino acid sequences from each other.

190. The polypeptide system of any of embodiments 172-189, wherein the DBD has an amino acid sequence according to Table 31, or at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

191. The polypeptide system of any of embodiments 172-190, wherein the RT domain is from a retrovirus, or a polypeptide domain having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acids sequence identity thereto.

192. The polypeptide system of any of embodiments 172-191, wherein the RT domain has an amino acid sequence according to Table 6, or at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

193. The polypeptide system of any of embodiments 172-192, wherein each linker independently comprises a sequence according to Table 10, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

194. A nucleic acid or a plurality of nucleic acids encoding the polypeptides of any of the systems of embodiment 172-193.

195. A system comprising:

a template RNA of any of embodiments 91-138;

a gene modifying polypeptide, e.g., a gene modifying polypeptide of any of embodiments 139-171, or a polypeptide system, e.g., a polypeptide system of any of embodiments 172-193; and

a first gRNA comprising:

a gRNA spacer that binds a second portion of the target nucleic acid sequence, wherein the second portion is one the second strand of the target nucleic acid sequence; and

a gRNA scaffold that binds the DBD of the gene modifying polypeptide or the polypeptide system.

196. The system of embodiment 195, wherein the gRNA scaffold of the first gRNA has the same protein binding specificity as the gRNA sequence of the template RNA.

197. The system of embodiment 196, wherein the gRNA sequence of the template RNA binds to a first copy of a gene modifying polypeptide (e.g., at the DBD of the gene modifying polypeptide), and the gRNA scaffold of the first gRNA binds to a second copy of the gene modifying polypeptide (e.g., at the DBD of the gene modifying polypeptide).

198. The system of embodiment 195, wherein the template RNA does not comprise a gRNA spacer or a gRNA scaffold.

199. The system of embodiment 195 or 198, wherein the gRNA spacer binds to a region of the target nucleic acid sequence that is within about 5, 10, 15, 20, 25, 30, or 40 nucleotides of the region of the target nucleic acid sequence bound by the PBS sequence.

200. The system of any of embodiments 195-199, which further comprises:
a second Cas protein (e.g., a dead Cas protein) and
a second gRNA comprising:
a gRNA spacer that binds the first strand of the target nucleic acid at a location 3' of the location bound by the PBS sequence, and
a gRNA scaffold that binds the second Cas protein.

201. The system of embodiment 200, wherein the second Cas protein is a dead Cas protein (e.g., a dead Cas9 protein) or a Cas nickase protein (e.g., a Cas9 nickase protein)

202. The system of embodiment 200, wherein the gRNA spacer of the second gRNA has a length of at least 18 nucleotides (e.g., 18-28 nucleotides, e.g., 18-21 nucleotides) and the second Cas protein is a dead Cas protein.

203. The system of embodiment 200, wherein the gRNA spacer of the second gRNA has a length of 17 nucleotides or less (e.g., 14-17 nucleotides), wherein optionally the second Cas protein is a Cas nickase protein.

204. The system of embodiment 195, wherein the template RNA further comprises:
a gRNA spacer that is complementary to a third portion of the target nucleic acid sequence wherein the third portion is on the first strand of the target nucleic acid sequence; and
a gRNA scaffold.

205. The system of embodiment 204, wherein the gRNA scaffold binds the DBD of the gene modifying polypeptide or the polypeptide system.

206. The system of embodiment 204 or 205, wherein the gRNA spacer has a length of 17 nucleotides or less.

207. The system of any of embodiments 195-206, wherein the gRNA spacer of the template RNA induces nicking of the template nucleic acid, e.g., at the second strand of the target nucleic acid sequence.

208. The system of any of embodiments 195-206, wherein the gRNA spacer of the template RNA does not induce nicking of the template nucleic acid.

209. A system comprising:

i) a template RNA of any of embodiments 91-138 (e.g., a template RNA of embodiment 16);

ii) a first polypeptide comprising:

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain); and

a RNA-binding domain (RBD) that is heterologous to the DBD, wherein the RBD binds the RRS of the template RNA;

iii) a first gRNA comprising:

a gRNA spacer that directs the DBD of the first polypeptide to a second portion of the target nucleic acid sequence, wherein the second portion of the target nucleic acid sequence is on the second strand of the nucleic acid sequence; and

a gRNA scaffold that binds the DBD of the first polypeptide;

iv) a second polypeptide comprising:

an RT domain, and

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain), that is heterologous to the RT domain, and wherein the DBD of the second polypeptide has a different sequence from the DBD of the first polypeptide; and

v) a second gRNA comprising:

a gRNA spacer that directs the DBD of the second polypeptide to a third portion of the target nucleic acid sequence, wherein the third portion is on the first strand of the target nucleic acid, and

a gRNA scaffold that binds the DBD of the second polypeptide.

210. The system of embodiment 209, wherein the DBD of the second polypeptide comprises a Cas nickase domain or a dead Cas domain.

211. The system of embodiment 209, wherein the gRNA spacer of the second RNA induces nicking of the template nucleic acid, e.g., at the second strand of the target nucleic acid sequence.

212. The system of embodiment 209, wherein the gRNA spacer of the second RNA does not induce nicking of the template nucleic acid.

213. The system of embodiment 209, wherein the first gRNA does not detectably bind to the DBD of the second polypeptide.

214. The system of embodiment 209, wherein the second gRNA does not detectably bind to the DBD of the first polypeptide.

215. A system comprising:

i) a template RNA of any of the preceding embodiments, wherein the template RNA comprises:

a gRNA spacer that is complementary to a third portion of the target nucleic acid sequence wherein the third portion is on the first strand of the target nucleic acid sequence; and

a gRNA scaffold;

ii) a first polypeptide comprising:

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain); and

a RNA-binding domain (RBD) that is heterologous to the DBD, wherein the RBD binds the RRS of the template RNA;

iii) a first gRNA comprising:

a gRNA spacer that directs the DBD of the first polypeptide to a second portion of the target nucleic acid sequence, wherein the second portion of the target nucleic acid sequence is on the second strand of the nucleic acid sequence; and

a gRNA scaffold that binds the DBD of the first polypeptide; and

iv) a second polypeptide comprising:

an RT domain, and

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain), that is heterologous to the RT domain, and wherein the DBD of the second polypeptide has a different sequence from the DBD of the first polypeptide, and wherein the gRNA scaffold of the template RNA binds the DBD of the second polypeptide.

216. The system of embodiment 215, wherein the DBD of the second polypeptide comprises a Cas nickase domain or a dead Cas domain.

217. The system of embodiment 215, wherein the gRNA spacer of the template RNA induces nicking of the template nucleic acid, e.g., at the second strand of the target nucleic acid sequence.

218. The system of embodiment 215, wherein the gRNA spacer of the template RNA does not induce nicking of the template nucleic acid.

219. The system of any of embodiments 215-218, wherein the first gRNA does not detectably bind to the DBD of the second polypeptide.

220. The system of any of embodiments 215-219, wherein the gRNA of the template RNA does not detectably bind to the DBD of the first polypeptide.

221. A polypeptide system comprising:

a first polypeptide comprising:

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain);

a RNA-binding domain (RBD) that is heterologous to the DBD; and

optionally, a linker disposed between the DBD and the RBD; and

a second polypeptide comprising:

an RT domain, and

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain), that is heterologous to the RT domain; and

optionally, a linker disposed between the RT domain and the DBD.

222. The template RNA or system of any of the preceding embodiments, wherein the target nucleic acid sequence is a target gene, enhancer, or promoter.

223. The template RNA of system any of the preceding embodiments, wherein the target nucleic acid sequence is a human target gene, human enhancer, or human promoter.

224. The system or polypeptide system of any of the preceding embodiments, wherein the RBD has a sequence of Table 31, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto.

225. A method for modifying a target nucleic acid in a cell (e.g., a human cell), the method comprising contacting the cell with the system of any one of the preceding embodiments, or nucleic acid encoding the same, thereby modifying the target nucleic acid.

226. The method of embodiment 225, wherein presence of the second polypeptide, compared to an otherwise similar system lacking the second polypeptide, results in one or more of:

increased unwinding of the target nucleic acid;

increased number of target nucleic acids that are modified;

increased length of insertion into the target nucleic acid; or

reduced MMR activity at the target nucleic acid.

227. The method of any of embodiments 225 and 226, wherein the cell is in vivo or ex vivo.

In one aspect, the disclosure relates to a system for modifying DNA, comprising (a) a nucleic acid encoding a gene modifying polypeptide capable of target primed reverse transcription, the polypeptide comprising (i) a reverse transcriptase domain and (ii) a Cas9 nickase that binds DNA and has endonuclease activity, and (b) a template RNA comprising (i) a gRNA spacer that is complementary to a first portion of a human gene, (ii) a gRNA scaffold that binds the polypeptide, (iii) a heterologous object sequence comprising a mutation region, and (iv) a primer binding site (PBS) sequence comprising at least 3, 4, 5, 6, 7, or 8 bases of 100% homology to a target DNA strand at the 3′ end of the template RNA.

The gRNA spacer may comprise at least 15 bases of 100% homology to the target DNA at the 5′ end of the template RNA. The template RNA may further comprise a PBS sequence comprising at least 5 bases of at least 80% homology to the target DNA strand. The template RNA may comprise one or more chemical modifications.

The domains of the gene modifying polypeptide may be joined by a peptide linker. The polypeptide may comprise one or more peptide linkers. The gene modifying polypeptide may further comprise a nuclear localization signal. The polypeptide may comprise more than one nuclear localization signal, e.g., multiple adjacent nuclear localization signals or one or more nuclear localization signals in different regions of the polypeptide, e.g., one or more nuclear localization signals in the N-terminus of the

polypeptide and one or more nuclear localization signals in the C-terminus of the polypeptide. The nucleic acid encoding the gene modifying polypeptide may encode one or more intein domains.

Introduction of the system into a target cell may result in insertion of at least 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 500, or 1000 base pairs of exogenous DNA. Introduction of the system into a target cell may result in deletion, wherein the deletion is less than 2, 3, 4, 5, 10, 50, or 100 base pairs of genomic DNA upstream or downstream of the insertion. Introduction of the system into a target cell may result in substitution, e.g., substitution of 1, 2, or 3 nucleotides, e.g., consecutive nucleotides.

The heterologous object sequence may be at least 5, 10, 25, 50, 100, 150, 200, 250, 300, 400, 500, 600, or 700 base pairs.

In one aspect, the disclosure relates to a pharmaceutical composition comprising the system described above and a pharmaceutically acceptable excipient or carrier, wherein the pharmaceutically acceptable excipient or carrier is selected from the group consisting of a plasmid vector, a viral vector, a vesicle, and a lipid nanoparticle. In one aspect, the disclosure relates to a pharmaceutical composition comprising the system described above and multiple pharmaceutically acceptable excipients or carriers, wherein the pharmaceutically acceptable excipients or carriers are selected from the group consisting of a plasmid vector, a viral vector, a vesicle, and a lipid nanoparticle, e.g., where the system described above is delivered by two distinct excipients or carriers, e.g., two lipid nanoparticles, two viral vectors, or one lipid nanoparticle and one viral vector. The viral vector may be an adeno-associated virus (AAV).

In one aspect, the disclosure relates to a host cell (e.g., a mammalian cell, e.g., a human cell) comprising the system described above.

The system may be introduced *in vivo*, *in vitro*, *ex vivo*, or *in situ*. The nucleic acid of (a) may be integrated into the genome of the host cell. In some embodiments, the nucleic acid of (a) is not integrated into the genome of the host cell. In some embodiments, the heterologous object sequence is inserted at only one target site in the host cell genome. The heterologous object sequence may be inserted at two or more target sites in the host cell genome, e.g., at the same corresponding site in two homologous chromosomes or at two different sites on the same or different chromosomes. The heterologous object sequence may encode a mammalian polypeptide, or a fragment or a variant thereof. The components of the system may be delivered on 1, 2, 3, 4, or more distinct nucleic acid molecules. The system may be introduced into a host cell by electroporation or by using at least one vehicle selected from a plasmid vector, a viral vector, a vesicle, and a lipid nanoparticle.

BRIEF DESCRIPTION OF THE DRAWINGS

The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

FIG. 1 is a series of diagrams showing components of an exemplary trans gene modifying system. The exemplary system comprises three components: (1) a gene modifying polypeptide, (2) a template RNA, and (3) a gRNA. The gene modifying polypeptide includes a nickase Cas9 (nCas9), an RNA binding domain (RBD), and a polymerase (in this example a retroviral reverse transcriptase (RT)). The template contains an RBD recruitment site (RRS), a primer binding site sequence (PBS sequence) (Priming) and a heterologous object sequence (template region), as well as an end protection/ end block sequence that (a) protects the structure from exonucleases, and/or (b) terminates the RT due to the secondary structure. The third component is a gRNA. In a fully assembled trans gene modifying reaction, the gRNA associates with the nCas9 of the gene modifying polypeptide, and directs the polypeptide to the DNA. The nCas9 then introduces a nick into the DNA. The RBD of the polypeptide recruits the template to the site of the nick through its interaction with the RRS on the template RNA. The Cas9 induced nick results in a 3' flap, that can anneal to the PBS sequence of the template RNA. The RT can then reverse transcribe the template until it hits the end protection structure. The highly structured end protection will terminate the reverse transcription. Cellular repair processes will incorporate the edited strand into the genome.

FIGS. 2A-2B are a series of diagrams showing exemplary polypeptides that can be used in a trans gene modifying system as described herein. There are several ways by which a polypeptide containing an nCas9-RT-RBD can be assembled: (A) by direction fusion, (B) by using either intein or dimerization (homo or hetero) domains that covalently or non-covalently assemble the full polypeptide, respectively. (A) In a direct fusion approach, a linker connects the nCas9 with the RPD, which in turn is connected through a linker with the RT (e.g., as shown). Exemplary possible configurations are listed in the panel below Fig. 2A, and RBDs /linkers are listed in a separate table. The RBP can be present once or multiple (e.g., n=1-5) times. (B) The polypeptide can also be assembled using various intein or dimerization domains. In some instances, the nCas9 is linked to a dimerization domain (FD#1), and the RPD is linked to its partner dimerization domain. The nCas9 is linked to a second dimerization domain (FD2), while the RT is linked to its partner. The dimerization domain can either result in covalent linkage (e.g., when using inteins), or in non-covalent assembly of the polypeptide (e.g., using chemical or light induced dimerization). Two dimerization reactions are utilized, upon which a polypeptide complex is assembled.

Exemplary possible variations are described herein (e.g., intein dimerization domains, chemically-induced dimerization domains, light-induced dimerization domains, antibody-peptide dimerization domains, coiled-coil dimerization domains). The dimerization domains can be present once or multiple (n=1-30) times, e.g., as tandem repeats.

FIGS 3A-3C are a series of diagrams showing an exemplary template RNA and subregions thereof. (A) Schematic of an exemplary template RNA. This template includes (3' to 5') of one or several (n=1-10) RRS at the 3' end, a linker, followed by a PBS sequence (priming) (8-17 nts), followed by a heterologous object sequence (template). The template region contains, in some embodiments, a pre-edit homology region (0-20 nts), the mutation region having a desired modification to the genome (e.g., an insertion, deletion, or point mutation(s)), and a post-edit homology region (e.g., n=5-500 nts). Lastly, an end protection/ end block sequence is present at the 5' end of the template RNA. Exemplary possible configurations are listed in the panel below Fig. 3A. (B) Exemplary variations for the various template RNA components are listed. Exemplary sequences for such components are described herein. (C) Schematic of an exemplary template RNA wherein the RRS is situated between the pre-edit homology region and the mutation region.

FIGS. 4A-4B are a series of diagrams showing, among other things, increased unwinding of a target nucleic acid, as well as engagement and modulation of a second strand of the target nucleic acid, e.g., to increase gene modifying efficiency and/or to permit long insertions. There are several ways in which the second strand can be engaged in the context of trans gene modification. (A) In one exemplary configuration, a second Cas9-gRNA complex can be introduced in trans. This second Cas9 complex can be, for example, a nickase Cas9 (nCas9) to direct a nick on the second strand . This nick could be used to initiate second strand synthesis after the RT reaction, and/or to signal to the cell endogenous Mismatch repair system that the first (edited) strand should be maintained and copied. Alternatively, the Cas9 can be, for example, a catalytically inactive (dead) Cas9 (dCas9). Without wishing to be bound by theory, in some embodiments this would unwind the DNA and could facilitate the repair of especially longer insertions. The Cas9 in this scenario can be of the same or orthogonal species as the Cas9 present in the trans rewriting polypeptide. (B) In an alternate configuration, the second strand modulation is recruited by the template RNA, by using a gRNA (full or partial) as an end structure. This gRNA can either be a full gRNA with a scaffold and a 20nt spacer, or a partial gRNA with a scaffold and a spacer of 17 or fewer nucleotides. A full gRNA will engage the polypeptide complex and can position the nick from the nCas9 in the polypeptide complex to the second strand. Placement of this nick could be used to initiate second strand synthesis after the RT reaction, and/or to signal to the cell endogenous mismatch repair system that the first (edited) strand should be maintained and copied. A spacer region (e.g., having a length of less

than or equal to 17 nucleotides, e.g., about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or 17 nucleotides) can lead to binding of the polypeptide complex, but will not result in a nick. This would unwind the DNA and may facilitate the repair of insertions (e.g., longer insertions).

FIGS. 5A-5B are a series of diagrams showing further exemplary configurations for engagement and modulation of a second strand of the target nucleic acid, e.g., to increase gene modifying efficiency and/or to permit long insertions. In these alternative configurations, the nCas9 is fused to only the RBD. The gRNA associated with the nCas9-RBD polypeptide recruits it to the DNA, and the nCas9 introduces a nick. The RBD recruits the template RNA. The configurations further comprise a second polypeptide complex consisting of a Cas9 (e.g., nickase or dead Cas9) fused to the RT domain. This second complex can associate with the DNA in the following ways: (A) by using a second gRNA, or (B) by using a gRNA present in the 5' end of the template RNA. In both scenarios, the gRNA can include a full 20 nts spacer to direct cleavage, or a spacer having a length of less than or equal to 17 nucleotides (e.g., about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or 17 nucleotides) to unwind the DNA without introducing a nick.

FIG. 6A is a diagram showing exemplary driver configurations.

FIG. 6B is a diagram showing exemplary template nucleic acid configurations.

FIG. 7A is a diagram showing an exemplary assay for analyzing rewriter activity in cells.

FIG. 7B is a graph showing rewriting activity for exemplary gene modifying polypeptides comprising a first exemplary RT domain or a second RT domain, as indicated.

FIG. 8 is a diagram showing rewriting activity of exemplary gene modifying systems.

FIG. 9 is a diagram showing rewriting activity of exemplary gene modifying systems.

FIG. 10 is a series of graphs showing rewriting activity for exemplary gene modifying systems.

FIGS. 11A-11B are a series of graphs showing rewriting activity for exemplary gene modifying systems.

DETAILED DESCRIPTION

Definitions

The term "expression cassette," as used herein, refers to a nucleic acid construct comprising nucleic acid elements sufficient for the expression of the nucleic acid molecule of the instant invention.

A "gRNA spacer", as used herein, refers to a portion of a nucleic acid that has complementarity to a target nucleic acid and can, together with a gRNA scaffold, target a Cas protein to the target nucleic acid.

A "gRNA scaffold", as used herein, refers to a portion of a nucleic acid that can bind a Cas protein and can, together with a gRNA spacer, target the Cas protein to the target nucleic acid. In some embodiments, the gRNA scaffold comprises a crRNA sequence, tetraloop, and tracrRNA sequence.

A "gene modifying polypeptide", as used herein, refers to a polypeptide comprising a retroviral reverse transcriptase, or a polypeptide comprising an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity to a retroviral reverse transcriptase, which is capable of integrating a nucleic acid sequence (e.g., a sequence provided on a template nucleic acid) into a target DNA molecule (e.g., in a mammalian host cell, such as a genomic DNA molecule in the host cell). In some embodiments, the gene modifying polypeptide is capable of integrating the sequence substantially without relying on host machinery. In some embodiments, the gene modifying polypeptide integrates a sequence into a random position in a genome, and in some embodiments, the gene modifying polypeptide integrates a sequence into a specific target site. In some embodiments, a gene modifying polypeptide includes one or more domains that, collectively, facilitate 1) binding the template nucleic acid, 2) binding the target DNA molecule, and 3) facilitate integration of the at least a portion of the template nucleic acid into the target DNA. Gene modifying polypeptides include both naturally occurring polypeptides as well as engineered variants of the foregoing, e.g., having one or more amino acid substitutions to the naturally occurring sequence. Gene modifying polypeptides also include heterologous constructs, e.g., where one or more of the domains recited above are heterologous to each other, whether through a heterologous fusion (or other conjugate) of otherwise wild-type domains, as well as fusions of modified domains, e.g., by way of replacement or fusion of a heterologous sub-domain or other substituted domain. Exemplary gene modifying polypeptides, and systems comprising them and methods of using them, that can be used in the methods provided herein are described, e.g., in PCT/US2021/020948, which is incorporated herein by reference with respect to gene modifying polypeptides that comprise a retroviral reverse transcriptase domain. In some embodiments, a gene modifying polypeptide integrates a sequence into a gene. In some embodiments, a gene modifying polypeptide integrates a sequence into a sequence outside of a gene. A "gene modifying system," as used herein, refers to a system comprising a gene modifying polypeptide and a template nucleic acid.

The term "domain" as used herein refers to a structure of a biomolecule that contributes to a specified function of the biomolecule. A domain may comprise a contiguous region (e.g., a contiguous sequence) or distinct, non-contiguous regions (e.g., non-contiguous sequences) of a biomolecule. Examples of protein domains include, but are not limited to, an endonuclease domain, a DNA binding

domain, a reverse transcription domain; an example of a domain of a nucleic acid is a regulatory domain, such as a transcription factor binding domain. In some embodiments, a domain (e.g., a Cas domain) can comprise two or more smaller domains (e.g., a DNA binding domain and an endonuclease domain).

The term "end block sequence," as used herein, refers to an RNA sequence having a secondary structure that impairs reverse transcription and/or impairs exonuclease activity. In some instances, an end block sequence comprises a stem-loop sequence.

As used herein, the term "exogenous", when used with reference to a biomolecule (such as a nucleic acid sequence or polypeptide) means that the biomolecule was introduced into a host genome, cell or organism by the hand of man. For example, a nucleic acid that is as added into an existing genome, cell, tissue or subject using recombinant DNA techniques or other methods is exogenous to the existing nucleic acid sequence, cell, tissue or subject.

As used herein, "first strand" and "second strand", as used to describe the individual DNA strands of target DNA, distinguish the two DNA strands based upon which strand the reverse transcriptase domain initiates polymerization, e.g., based upon where target primed synthesis initiates. The first strand refers to the strand of the target DNA upon which the reverse transcriptase domain initiates polymerization, e.g., where target primed synthesis initiates. The second strand refers to the other strand of the target DNA. First and second strand designations do not describe the target site DNA strands in other respects; for example, in some embodiments the first and second strands are nicked by a polypeptide described herein, but the designations 'first' and 'second' strand have no bearing on the order in which such nicks occur.

A "genomic safe harbor site" (GSH site) is a site in a host genome that is able to accommodate the integration of new genetic material, e.g., such that the inserted genetic element does not cause significant alterations of the host genome posing a risk to the host cell or organism. A GSH site generally meets 1, 2, 3, 4, 5, 6, 7, 8 or 9 of the following criteria: (i) is located >300kb from a cancer-related gene; (ii) is >300kb from a miRNA/other functional small RNA; (iii) is >50kb from a 5′ gene end; (iv) is >50kb from a replication origin; (v) is >50kb away from any ultraconservered element; (vi) has low transcriptional activity (i.e. no mRNA +/- 25 kb); (vii) is not in a copy number variable region; (viii) is in open chromatin; and/or (ix) is unique, with 1 copy in the human genome. Examples of GSH sites in the human genome that meet some or all of these criteria include (i) the adeno-associated virus site 1 (AAVS1), a naturally occurring site of integration of AAV virus on chromosome 19; (ii) the chemokine (C-C motif) receptor 5 (CCR5) gene, a chemokine receptor gene known as an HIV-1 coreceptor; (iii) the human ortholog of the mouse Rosa26 locus; (iv) the ribosomal DNA ("rDNA") locus. Additional GSH sites are known and described, e.g., in Pellenz et al. epub August 20, 2018 (https://doi.org/10.1101/396390).

The term "heterologous," as used herein to describe a first element in reference to a second element means that the first element and second element do not exist in nature disposed as described. For example, a heterologous polypeptide, nucleic acid molecule, construct or sequence refers to (a) a polypeptide, nucleic acid molecule or portion of a polypeptide or nucleic acid molecule sequence that is not native to a cell in which it is expressed, (b) a polypeptide or nucleic acid molecule or portion of a polypeptide or nucleic acid molecule that has been altered or mutated relative to its native state, or (c) a polypeptide or nucleic acid molecule with an altered expression as compared to the native expression levels under similar conditions. For example, a heterologous regulatory sequence (e.g., promoter, enhancer) may be used to regulate expression of a gene or a nucleic acid molecule in a way that is different than the gene or a nucleic acid molecule is normally expressed in nature. In another example, a heterologous domain of a polypeptide or nucleic acid sequence (e.g., a DNA binding domain of a polypeptide or nucleic acid encoding a DNA binding domain of a polypeptide) may be disposed relative to other domains or may be a different sequence or from a different source, relative to other domains or portions of a polypeptide or its encoding nucleic acid. In certain embodiments, a heterologous nucleic acid molecule may exist in a native host cell genome, but may have an altered expression level or have a different sequence or both. In other embodiments, heterologous nucleic acid molecules may not be endogenous to a host cell or host genome but instead may have been introduced into a host cell by transformation (e.g., transfection, electroporation), wherein the added molecule may integrate into the host genome or can exist as extra-chromosomal genetic material either transiently (e.g., mRNA) or semi-stably for more than one generation (e.g., episomal viral vector, plasmid or other self-replicating vector).

As used herein, "insertion" of a sequence into a target site refers to the net addition of DNA sequence at the target site, e.g., where there are new nucleotides in the heterologous object sequence with no cognate positions in the unedited target site. In some embodiments, a nucleotide alignment of the PBS sequence and heterologous object sequence to the target nucleic acid sequence would result in an alignment gap in the target nucleic acid sequence.

As used herein, a "deletion" generated by a heterologous object sequence in a target site refers to the net deletion of DNA sequence at the target site, e.g., where there are nucleotides in the unedited target site with no cognate positions in the heterologous object sequence. In some embodiments, a nucleotide alignment of the PBS sequence and heterologous object sequence to the target nucleic acid sequence would result in an alignment gap in the molecule comprising the PBS sequence and heterologous object sequence.

The term "inverted terminal repeats" or "ITRs" as used herein refers to AAV viral cis-elements named so because of their symmetry. These elements promote efficient multiplication of an AAV genome. It is hypothesized that the minimal elements for ITR function are a Rep-binding site (RBS; 5′-

GCGCGCTCGCTCGCTC-3′ for AAV2) and a terminal resolution site (TRS; 5′-AGTTGG-3′ for AAV2) plus a variable palindromic sequence allowing for hairpin formation. According to the present invention, an ITR comprises at least these three elements (RBS, TRS, and sequences allowing the formation of an hairpin). In addition, in the present invention, the term "ITR" refers to ITRs of known natural AAV serotypes (e.g. ITR of a serotype 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or 11 AAV), to chimeric ITRs formed by the fusion of ITR elements derived from different serotypes, and to functional variants thereof. "Functional variant" refers to a sequence presenting a sequence identity of at least 80%, 85%, 90%, preferably of at least 95% with a known ITR and allowing multiplication of the sequence that includes said ITR in the presence of Rep proteins.

The term "mutation region," as used herein, refers to a region in a template RNA having one or more sequence difference relative to the corresponding sequence in a target nucleic acid. The sequence difference may comprise, for example, a substitution, insertion, frameshift, or deletion.

The term "mutated" when applied to nucleic acid sequences means that nucleotides in a nucleic acid sequence are inserted, deleted, or changed compared to a reference (e.g., native) nucleic acid sequence. A single alteration may be made at a locus (a point mutation), or multiple nucleotides may be inserted, deleted, or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. A nucleic acid sequence may be mutated by any method known in the art.

"Nucleic acid molecule" refers to both RNA and DNA molecules including, without limitation, complementary DNA ("cDNA"), genomic DNA ("gDNA"), and messenger RNA ("mRNA"), and also includes synthetic nucleic acid molecules, such as those that are chemically synthesized or recombinantly produced, such as RNA templates, as described herein. The nucleic acid molecule can be double-stranded or single-stranded, circular, or linear. If single-stranded, the nucleic acid molecule can be the sense strand or the antisense strand. Unless otherwise indicated, and as an example for all sequences described herein under the general format "SEQ ID NO:," "nucleic acid comprising SEQ ID NO:1" refers to a nucleic acid, at least a portion which has either (i) the sequence of SEQ ID NO:1, or (ii) a sequence complimentary to SEQ ID NO:1. The choice between the two is dictated by the context in which SEQ ID NO:1 is used. For instance, if the nucleic acid is used as a probe, the choice between the two is dictated by the requirement that the probe be complementary to the desired target. Nucleic acid sequences of the present disclosure may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more naturally occurring nucleotides with an analog, inter-nucleotide modifications such as uncharged linkages (for example, methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (for example, phosphorothioates,

phosphorodithioates, etc.), pendant moieties, (for example, polypeptides), intercalators (for example, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (for example, alpha anomeric nucleic acids, etc.). Also included are chemically modified bases (see, for example, Table 13), backbones (see, for example, Table 14), and modified caps (see, for example, Table 15). Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of a molecule, e.g., peptide nucleic acids (PNAs). Other modifications can include, for example, analogs in which the ribose ring contains a bridging moiety or other structure such as modifications found in "locked" nucleic acids (LNAs). In various embodiments, the nucleic acids are in operative association with additional genetic elements, such as tissue-specific expression-control sequence(s) (e.g., tissue-specific promoters and tissue-specific microRNA recognition sequences), as well as additional elements, such as inverted repeats (e.g., inverted terminal repeats, such as elements from or derived from viruses, e.g., AAV ITRs) and tandem repeats, inverted repeats/direct repeats, homology regions (segments with various degrees of homology to a target DNA), untranslated regions (UTRs) (5′, 3′, or both 5′ and 3′ UTRs), and various combinations of the foregoing. The nucleic acid elements of the systems provided by the invention can be provided in a variety of topologies, including single-stranded, double-stranded, circular, linear, linear with open ends, linear with closed ends, and particular versions of these, such as doggybone DNA (dbDNA), closed-ended DNA (ceDNA).

As used herein, a "gene expression unit" is a nucleic acid sequence comprising at least one regulatory nucleic acid sequence operably linked to at least one effector sequence. A first nucleic acid sequence is operably linked with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter or enhancer is operably linked to a coding sequence if the promoter or enhancer affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be contiguous or non-contiguous. Where necessary to join two protein-coding regions, operably linked sequences may be in the same reading frame.

The terms "host genome" or "host cell", as used herein, refer to a cell and/or its genome into which protein and/or genetic material has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell and/or genome, but to the progeny of such a cell and/or the genome of the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein. A host genome or host cell may be an isolated cell or cell line grown in culture, or genomic material isolated

from such a cell or cell line, or may be a host cell or host genome which composing living tissue or an organism. In some instances, a host cell may be an animal cell or a plant cell, e.g., as described herein. In certain instances, a host cell may be a mammalian cell, a human cell, avian cell, reptilian cell, bovine cell, horse cell, pig cell, goat cell, sheep cell, chicken cell, or turkey cell. In certain instances, a host cell may be a corn cell, soy cell, wheat cell, or rice cell.

As used herein, "operative association" describes a functional relationship between two nucleic acid sequences, such as a 1) promoter and 2) a heterologous object sequence, and means, in such example, the promoter and heterologous object sequence (e.g., a gene of interest) are oriented such that, under suitable conditions, the promoter drives expression of the heterologous object sequence. For instance, a template nucleic acid carrying a promoter and a heterologous object sequence may be single-stranded, e.g., either the (+) or (-) orientation. An "operative association" between the promoter and the heterologous object sequence in this template means that, regardless of whether the template nucleic acid will be transcribed in a particular state, when it is in the suitable state (e.g., is in the (+) orientation, in the presence of required catalytic factors, and NTPs, etc.), it is accurately transcribed. Operative association applies analogously to other pairs of nucleic acids, including other tissue-specific expression control sequences (such as enhancers, repressors and microRNA recognition sequences), IR/DR, ITRs, UTRs, or homology regions and heterologous object sequences or sequences encoding a retroviral RT domain.

The term "primer binding site sequence" or "PBS sequence," as used herein, refers to a portion of a template RNA capable of binding to a region comprised in a target nucleic acid sequence. In some instances, a PBS sequence is a nucleic acid sequence comprising at least 3, 4, 5, 6, 7, or 8 bases with 100% identity to the region comprised in the target nucleic acid sequence. In some embodiments the primer region comprises at least 5, 6, 7, 8 bases with 100% identity to the region comprised in the target nucleic acid sequence. Without wishing to be bound by theory, in some embodiments when a template RNA comprises a PBS sequence and a heterologous object sequence, the PBS sequence binds to a region comprised in a target nucleic acid sequence, allowing a reverse transcriptase domain to use that region as a primer for reverse transcription, and to use the heterologous object sequence as a template for reverse transcription.

As used herein, a "stem-loop sequence" refers to a nucleic acid sequence (e.g., RNA sequence) with sufficient self-complementarity to form a stem-loop, e.g., having a stem comprising at least two (e.g., 3, 4, 5, 6, 7, 8, 9, or 10) base pairs, and a loop with at least three (e.g., four) base pairs. The stem may comprise mismatches or bulges.

As used herein, a "tissue-specific expression-control sequence" means nucleic acid elements that increase or decrease the level of a transcript comprising the heterologous object sequence in a target tissue in a tissue-specific manner, e.g., preferentially in on-target tissue(s), relative to off-target tissue(s).

In some embodiments, a tissue-specific expression-control sequence preferentially drives or represses transcription, activity, or the half-life of a transcript comprising the heterologous object sequence in the target tissue in a tissue-specific manner, e.g., preferentially in an on-target tissue(s), relative to an off-target tissue(s). Exemplary tissue-specific expression-control sequences include tissue-specific promoters, repressors, enhancers, or combinations thereof, as well as tissue-specific microRNA recognition sequences. Tissue specificity refers to on-target (tissue(s) where expression or activity of the template nucleic acid is desired or tolerable) and off-target (tissue(s) where expression or activity of the template nucleic acid is not desired or is not tolerable). For example, a tissue-specific promoter drives expression preferentially in on-target tissues, relative to off-target tissues. In contrast, a microRNA that binds the tissue-specific microRNA recognition sequences is preferentially expressed in off-target tissues, relative to on-target tissues, thereby reducing expression of a template nucleic acid in off-target tissues. Accordingly, a promoter and a microRNA recognition sequence that are specific for the same tissue, such as the target tissue, have contrasting functions (promote and repress, respectively, with concordant expression levels, i.e., high levels of the microRNA in off-target tissues and low levels in on-target tissues, while promoters drive high expression in on-target tissues and low expression in off-target tissues) with regard to the transcription, activity, or half-life of an associated sequence in that tissue.

Table of Contents

1) Introduction
2) Gene modifying systems
    a) Polypeptide components of gene modifying systems
        i) Writing domain
        ii) Endonuclease domains and DNA binding domains
            (1) Gene modifying polypeptides comprising Cas domains
            (2) TAL Effectors and Zinc Finger Nucleases
        iii) Linkers
        iv) Localization sequences for gene modifying systems
        v) Evolved Variants of Gene Modifying Polypeptides and Systems
        vi) Inteins
        vii) Additional domains
    b) Template nucleic acids
        i) gRNA spacer and gRNA scaffold
        ii) Heterologous object sequence
        iii) PBS sequence

iv) Exemplary Template Sequences
c) gRNAs with inducible activity
d) Circular RNAs and Ribozymes in Gene Modifying Systems
e) Target Nucleic Acid Site
f) Second strand nicking
3) Production of Compositions and Systems
4) Therapeutic Applications
5) Administration and Delivery
a) Tissue Specific Activity/Administration
i) Promoters
ii) microRNAs
b) Viral vectors and components thereof
c) AAV Administration
d) Lipid Nanoparticles
6) Kits, Articles of Manufacture, and Pharmaceutical Compositions
7) Chemistry, Manufacturing, and Controls (CMC)

Introduction

This disclosure relates to methods compositions for targeting, editing, modifying or manipulating a DNA sequence (e.g., inserting a heterologous object sequence into a target site of a mammalian genome) at one or more locations in a DNA sequence in a cell, tissue or subject, e.g., *in vivo* or *in vitro*. The heterologous object DNA sequence may include, e.g., a substitution, a deletion, an insertion, e.g., a coding sequence, a regulatory sequence, or a gene expression unit.

This disclosure relates, in part, to anchoring of a trans template RNA to a gene modifying polypeptide:sgRNA:target genomic DNA complex by two or more interactions. Without wishing to be bound by theory, it is contemplated that such anchoring can achieve high rewriting activity, e.g., for achieving single or several nucleotide long edits. For example, 1) an RRS:RBP interaction and 2) a 5' end block Cas9 scaffold and spacer to target DNA interaction (mediated via an additional gene modifying polypeptide) represent exemplary interactions that together anchor a trans template RNA to a gene modifying polypeptide:sgRNA:target genomic DNA complex to enable rewriting. It is contemplated that the RRS:RBP interaction is critical in the absence of the 5' end block spacer. It is further contemplated that the presence of both can provide high rewriting activity and the presence of the 5' end block spacer in combination with a weaker RRS:RBP interaction rescues rewriting activity.

The disclosure also provides methods for treating disease using reverse transcriptase-based systems for altering a genomic DNA sequence of interest, e.g., by inserting, deleting, or substituting one or more nucleotides into/from the sequence of interest.

The disclosure provides, in part, methods for treating disease using a gene modifying system comprising a gene modifying polypeptide component and a template nucleic acid (e.g., template RNA) component. In some embodiments, a gene modifying system can be used to introduce an alteration into a target site in a genome. In some embodiments, the gene modifying polypeptide component comprises a writing domain (e.g., a reverse transcriptase domain), a DNA-binding domain, and an endonuclease domain (e.g., nickase domain). In some embodiments, the template nucleic acid (e.g., template RNA) comprises a sequence (e.g., a gRNA spacer) that binds a target site in the genome (e.g., that binds to a second strand of the target site), a sequence (e.g., a gRNA scaffold) that binds the gene modifying polypeptide component, a heterologous object sequence, and a PBS sequence. Without wishing to be bound by theory, it is thought that the template nucleic acid (e.g., template RNA) binds to the second strand of a target site in the genome, and binds to the gene modifying polypeptide component (e.g., localizing the polypeptide component to the target site in the genome). It is thought that the endonuclease (e.g., nickase) of the gene modifying polypeptide component cuts the target site (e.g., the first strand of the target site), e.g., allowing the PBS sequence to bind to a sequence adjacent to the site to be altered on the first strand of the target site. It is thought that the writing domain (e.g., reverse transcriptase domain) of the polypeptide component uses the first strand of the target site that is bound to the complementary sequence comprising the PBS sequence of the template nucleic acid as a primer and the heterologous object sequence of the template nucleic acid as a template to, e.g., polymerize a sequence complementary to the heterologous object sequence. Without wishing to be bound by theory, it is thought that selection of an appropriate heterologous object sequence can result in substitution, deletion, and/or insertion of one or more nucleotides at the target site.

Gene modifying systems

In some embodiments, a gene modifying system described herein comprises: (A) a gene modifying polypeptide or a nucleic acid encoding the gene modifying polypeptide, wherein the gene modifying polypeptide comprises (i) a reverse transcriptase domain, and either (x) an endonuclease domain that contains DNA binding functionality or (y) an endonuclease domain and separate DNA binding domain; and (B) a template RNA. A gene modifying polypeptide, in some embodiments, acts as a substantially autonomous protein machine capable of integrating a template nucleic acid sequence into a target DNA molecule (e.g., in a mammalian host cell, such as a genomic DNA molecule in the host cell), substantially without relying on host machinery. For example, the gene modifying protein may comprise

a DNA-binding domain, a reverse transcriptase domain, and an endonuclease domain. In some embodiments, the DNA-binding function may involve an RNA component that directs the protein to a DNA sequence, e.g., a gRNA spacer. In other embodiments, the gene modifying polypeptide may comprise a reverse transcriptase domain and an endonuclease domain. The RNA template element of a gene modifying system is typically heterologous to the gene modifying polypeptide element and provides an object sequence to be inserted (reverse transcribed) into the host genome. In some embodiments, the gene modifying polypeptide is capable of target primed reverse transcription. In some embodiments, the gene modifying polypeptide is capable of second-strand synthesis.

In some embodiments, a gene modifying system described herein comprises a gene modifying polypeptide comprising the amino acid sequence, or a functional portion thereof, of an exemplary gene modifying polypeptide as listed in any of Tables S1-S3, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a nucleic acid molecule encoding the gene modifying polypeptide. In some embodiments, a gene modifying system described herein comprises a gene modifying polypeptide comprising the amino acid sequence of an RT domain of an exemplary gene modifying polypeptide as listed in any of Tables S1-S3, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a nucleic acid molecule encoding the gene modifying polypeptide. In some embodiments, a gene modifying system described herein comprises a gene modifying polypeptide comprising the amino acid sequence of a DBD of an exemplary gene modifying polypeptide as listed in any of Tables S1-S3, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a nucleic acid molecule encoding the gene modifying polypeptide. In some embodiments, a gene modifying system described herein comprises a gene modifying polypeptide comprising the amino acid sequence of an RBD of an exemplary gene modifying polypeptide as listed in any of Tables S1-S3, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a nucleic acid molecule encoding the gene modifying polypeptide. In some embodiments, a gene modifying system described herein comprises a gene modifying polypeptide comprising the amino acid sequence of the RT domain, DBD, and RBD of an exemplary gene modifying polypeptide as listed in any of Tables S1-S3, or amino acid sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a nucleic acid molecule encoding the gene modifying polypeptide.

In some embodiments, a gene modifying system described herein comprises a gene modifying polypeptide comprising the amino acid sequence, or a functional portion thereof, of an exemplary gene modifying polypeptide as listed in Table S1, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a nucleic acid molecule encoding the gene

modifying polypeptide. In some embodiments, a gene modifying system described herein comprises a gene modifying polypeptide comprising the amino acid sequence of an RT domain of an exemplary gene modifying polypeptide as listed in Table S1, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a nucleic acid molecule encoding the gene modifying polypeptide. In some embodiments, a gene modifying system described herein comprises a gene modifying polypeptide comprising the amino acid sequence of a DBD of an exemplary gene modifying polypeptide as listed in Table S1, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a nucleic acid molecule encoding the gene modifying polypeptide. In some embodiments, a gene modifying system described herein comprises a gene modifying polypeptide comprising the amino acid sequence of an RBD of an exemplary gene modifying polypeptide as listed in Table S1, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a nucleic acid molecule encoding the gene modifying polypeptide. In some embodiments, a gene modifying system described herein comprises a gene modifying polypeptide comprising the amino acid sequence of the RT domain, DBD, and RBD of an exemplary gene modifying polypeptide as listed in Table S1, or amino acid sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a nucleic acid molecule encoding the gene modifying polypeptide.

In some embodiments, a gene modifying system described herein comprises a gene modifying polypeptide comprising the amino acid sequence, or a functional portion thereof, of an exemplary gene modifying polypeptide as listed in Table S2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a nucleic acid molecule encoding the gene modifying polypeptide. In some embodiments, a gene modifying system described herein comprises a gene modifying polypeptide comprising the amino acid sequence of an RT domain of an exemplary gene modifying polypeptide as listed in Table S2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a nucleic acid molecule encoding the gene modifying polypeptide. In some embodiments, a gene modifying system described herein comprises a gene modifying polypeptide comprising the amino acid sequence of a DBD of an exemplary gene modifying polypeptide as listed in Table S2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a nucleic acid molecule encoding the gene modifying polypeptide. In some embodiments, a gene modifying system described herein comprises a gene modifying polypeptide comprising the amino acid sequence of an RBD of an exemplary gene modifying polypeptide as listed in Table S2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a nucleic acid molecule encoding the gene modifying polypeptide. In some embodiments, a gene modifying system described herein comprises a

gene modifying polypeptide comprising the amino acid sequence of the RT domain, DBD, and RBD of an exemplary gene modifying polypeptide as listed in Table S2, or amino acid sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a nucleic acid molecule encoding the gene modifying polypeptide.

In some embodiments, a gene modifying system described herein comprises a gene modifying polypeptide comprising the amino acid sequence, or a functional portion thereof, of an exemplary gene modifying polypeptide as listed in Table S3, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a nucleic acid molecule encoding the gene modifying polypeptide. In some embodiments, a gene modifying system described herein comprises a gene modifying polypeptide comprising the amino acid sequence of an RT domain of an exemplary gene modifying polypeptide as listed in Table S3, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a nucleic acid molecule encoding the gene modifying polypeptide. In some embodiments, a gene modifying system described herein comprises a gene modifying polypeptide comprising the amino acid sequence of a DBD of an exemplary gene modifying polypeptide as listed in Table S3, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a nucleic acid molecule encoding the gene modifying polypeptide. In some embodiments, a gene modifying system described herein comprises a gene modifying polypeptide comprising the amino acid sequence of an RBD of an exemplary gene modifying polypeptide as listed in Table S3, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a nucleic acid molecule encoding the gene modifying polypeptide. In some embodiments, a gene modifying system described herein comprises a gene modifying polypeptide comprising the amino acid sequence of the RT domain, DBD, and RBD of an exemplary gene modifying polypeptide as listed in Table S3, or amino acid sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or a nucleic acid molecule encoding the gene modifying polypeptide.

In some embodiments, a gene modifying system described herein comprises a template RNA comprising a nucleic acid sequence as listed in Table S4, or a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying system described herein comprises a template RNA comprising a 5' end block sequence of a template sequence as listed in Table S4, or a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying system described herein comprises a template RNA comprising a PBS sequence of a template sequence as listed in Table S4, or a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%,

98%, or 99% identity thereto. In some embodiments, a gene modifying system described herein comprises a template RNA comprising a linker sequence of a template sequence as listed in Table S4, or a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying system described herein comprises a template RNA comprising one or more (e.g., 1, 2, 3, or 4) RRS sequences of a template sequence as listed in Table S4, or nucleic acid sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying system described herein comprises a template RNA comprising a 3' end block sequence of a template sequence as listed in Table S4, or a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying system described herein comprises a template RNA comprising one or more (e.g., 1, 2, 3, or 4) of (e.g., in 5' to 3' order) a 5' end block sequence, optionally a PBS sequence, one or more (e.g., 1, 2, 3, or 4) RRS sequences, and a 3' end block sequence of a template sequence as listed in Table S4, or nucleic acid sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

In some embodiments the gene modifying system is combined with a second polypeptide. In some embodiments, the second polypeptide may comprise an endonuclease domain. In some embodiments, the second polypeptide may comprise a polymerase domain, e.g., a reverse transcriptase domain. In some embodiments, the second polypeptide may comprise a DNA-dependent DNA polymerase domain. In some embodiments, the second polypeptide aids in completion of the genome edit, e.g., by contributing to second-strand synthesis or DNA repair resolution.

A functional gene modifying polypeptide can be made up of unrelated DNA binding, reverse transcription, and endonuclease domains. This modular structure allows combining of functional domains, e.g., dCas9 (DNA binding), MMLV reverse transcriptase (reverse transcription), FokI (endonuclease). In some embodiments, multiple functional domains may arise from a single protein, e.g., Cas9 or Cas9 nickase (DNA binding, endonuclease).

In some embodiments, a gene modifying polypeptide includes one or more domains that, collectively, facilitate 1) binding the template nucleic acid, 2) binding the target DNA molecule, and 3) facilitate integration of the at least a portion of the template nucleic acid into the target DNA. In some embodiments, the gene modifying polypeptide is an engineered polypeptide that comprises one or more amino acid substitutions to a corresponding naturally occurring sequence. In some embodiments, the gene modifying polypeptide comprises two or more domains that are heterologous relative to each other, e.g., through a heterologous fusion (or other conjugate) of otherwise wild-type domains, or well as fusions of modified domains, e.g., by way of replacement or fusion of a heterologous sub-domain or other substituted domain. For instance, in some embodiments, one or more of: the RT domain is heterologous

to the DBD; the DBD is heterologous to the endonuclease domain; or the RT domain is heterologous to the endonuclease domain.

In some embodiments, a template RNA molecule for use in the system comprises, from 5′ to 3′ (1) a gRNA spacer; (2) a gRNA scaffold; (3) heterologous object sequence (4) a primer binding site (PBS) sequence. In some embodiments:

(1) Is a gRNA spacer of ~18-22 nt, e.g., is 20 nt

(2) Is a gRNA scaffold comprising one or more hairpin loops, e.g., 1, 2, of 3 loops for associating the template with a Cas domain, e.g., a nickase Cas9 domain. In some embodiments, the gRNA scaffold comprises the sequence, from 5′ to 3′, GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA AGTGGGACCGAGTCGGTCC (SEQ ID NO: 8).

(3) In some embodiments, the heterologous object sequence is, e.g., 7-74, e.g., 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, or 70-80 nt or, 80-90 nt in length. In some embodiments, the first (most 5′) base of the sequence is not C.

(4) In some embodiments, the PBS sequence that binds the target priming sequence after nicking occurs is e.g., 3-20 nt, e.g., 7-15 nt, e.g., 12-14 nt. In some embodiments, the PBS sequence has 40-60% GC content.

In some embodiments, a second gRNA associated with the system may help drive complete integration. In some embodiments, the second gRNA may target a location that is 0-200 nt away from the first-strand nick, e.g., 0-50, 50-100, 100-200 nt away from the first-strand nick. In some embodiments, the second gRNA can only bind its target sequence after the edit is made, e.g., the gRNA binds a sequence present in the heterologous object sequence, but not in the initial target sequence.

In some embodiments, a gene modifying system described herein is used to make an edit in HEK293, K562, U2OS, or HeLa cells. In some embodiment, a gene modifying system is used to make an edit in primary cells, e.g., primary cortical neurons from E18.5 mice.

In some embodiments, a gene modifying polypeptide as described herein comprises a reverse transcriptase or RT domain (e.g., as described herein) that comprises a MoMLV RT sequence or variant thereof. In embodiments, the MoMLV RT sequence comprises one or more mutations selected from D200N, L603W, T330P, T306K, W313F, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, L435G, N454K, H594Q, D653N, R110S, and K103L. In embodiments, the MoMLV RT sequence comprises a combination of mutations, such as D200N, L603W, and T330P, optionally further including T306K and/or W313F.

In some embodiments, an endonuclease domain (e.g., as described herein) comprises nCAS9, e.g., comprising the H840A mutation.

In some embodiments, the heterologous object sequence (e.g., of a system as described herein) is about 1-50, 50-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 900-1000, or more, nucleotides in length.

In some embodiments, the RT and endonuclease domains are joined by a flexible linker, e.g., comprising the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGSS (SEQ ID NO: 6).

In some embodiments, the endonuclease domain is N-terminal relative to the RT domain. In some embodiments, the endonuclease domain is C-terminal relative to the RT domain.

In some embodiments, the system incorporates a heterologous object sequence into a target site by TPRT, e.g., as described herein.

In some embodiments, a gene modifying polypeptide comprises a DNA binding domain. In some embodiments, a gene modifying polypeptide comprises an RNA binding domain. In some embodiments, the RNA binding domain comprises an RNA binding domain of B-box protein, MS2 coat protein, dCas, or an element of a sequence of a table herein. In some embodiments, the RNA binding domain is capable of binding to a template RNA with greater affinity than a reference RNA binding domain.

In some embodiments, a gene modifying system is capable of producing an insertion into the target site of at least 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 nucleotides (and optionally no more than 500, 400, 300, 200, or 100 nucleotides). In some embodiments, a gene modifying system is capable of producing an insertion into the target site of at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 nucleotides (and optionally no more than 500, 400, 300, 200, or 100 nucleotides). In some embodiments, a gene modifying system is capable of producing an insertion into the target site of at least 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5 or 10 kilobases (and optionally no more than 1, 5, 10, or 20 kilobases). In some embodiments, a gene modifying system is capable of producing a deletion of at least 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides (and optionally no more than 500, 400, 300, or 200 nucleotides). In some embodiments, a gene modifying system is capable of producing a deletion of at least 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides (and optionally no more than 500, 400, 300, or 200 nucleotides). In some embodiments, a gene modifying system is capable of producing a deletion of at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides (and optionally no more than 500, 400, 300, or 200 nucleotides). In some embodiments, a gene modifying system is capable of producing a deletion of at least 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1,

1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5 or 10 kilobases (and optionally no more than 1, 5, 10, or 20 kilobases). In some embodiments, a gene modifying system is capable of producing a substitution into the target site of at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, or 100 or more nucleotides. In some embodiments, a gene modifying system is capable of producing a substitution in the target site of 1-2, 2-3, 3-4, 4-5, 5-10, 10-15, 15-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, or 90-100 nucleotides.

In some embodiments, the substitution is a transition mutation. In some embodiments, the substitution is a transversion mutation. In some embodiments, the substitution converts an adenine to a thymine, an adenine to a guanine, an adenine to a cytosine, a guanine to a thymine, a guanine to a cytosine, a guanine to an adenine, a thymine to a cytosine, a thymine to an adenine, a thymine to a guanine, a cytosine to an adenine, a cytosine to a guanine, or a cytosine to a thymine.

In some embodiments, an insertion, deletion, substitution, or combination thereof, increases or decreases expression (e.g. transcription or translation) of a gene. In some embodiments, an insertion, deletion, substitution, or combination thereof, increases or decreases expression (e.g. transcription or translation) of a gene by altering, adding, or deleting sequences in a promoter or enhancer, e.g. sequences that bind transcription factors. In some embodiments, an insertion, deletion, substitution, or combination thereof alters translation of a gene (e.g. alters an amino acid sequence), inserts or deletes a start or stop codon, alters or fixes the translation frame of a gene. In some embodiments, an insertion, deletion, substitution, or combination thereof alters splicing of a gene, e.g. by inserting, deleting, or altering a splice acceptor or donor site. In some embodiments, an insertion, deletion, substitution, or combination thereof alters transcript or protein half-life. In some embodiments, an insertion, deletion, substitution, or combination thereof alters protein localization in the cell (e.g. from the cytoplasm to a mitochondria, from the cytoplasm into the extracellular space (e.g. adds a secretion tag)). In some embodiments, an insertion, deletion, substitution, or combination thereof alters (e.g. improves) protein folding (e.g. to prevent accumulation of misfolded proteins). In some embodiments, an insertion, deletion, substitution, or combination thereof, alters, increases, decreases the activity of a gene, e.g. a protein encoded by the gene.

Exemplary gene modifying polypeptides, and systems comprising them and methods of using them are described, e.g., in PCT/US2021/020948, which is incorporated herein by reference with respect to retroviral RT domains, including the amino acid and nucleic acid sequences therein.

Exemplary gene modifying polypeptides and retroviral RT domain sequences are also described, e.g., in International Application No. PCT/US21/20948 filed March 4, 2021, e.g., at Table 30, Table 31, and Table 44 therein; the entire application is incorporated by reference herein with respect to retroviral

RTs, e.g., in said sequences and tables. Accordingly, a gene modifying polypeptide described herein may comprise an amino acid sequence according to any of the Tables mentioned in this paragraph, or a domain thereof (e.g., a retroviral RT domain), or a functional fragment or variant of any of the foregoing, or an amino acid sequence having at least 70%, 80%, 85%, 90%, 95%, or 99% identity thereto.

In some embodiments, a polypeptide for use in any of the systems described herein can be a molecular reconstruction or ancestral reconstruction based upon the aligned polypeptide sequence of multiple homologous proteins. In some embodiments, a reverse transcriptase domain for use in any of the systems described herein can be a molecular reconstruction or an ancestral reconstruction, or can be modified at particular residues, based upon alignments of reverse transcriptase domains from the same or different sources. A skilled artisan can, based on the Accession numbers provided herein, align polypeptides or nucleic acid sequences, e.g., by using routine sequence analysis tools as Basic Local Alignment Search Tool (BLAST) or CD-Search for conserved domain analysis. Molecular reconstructions can be created based upon sequence consensus, e.g. using approaches described in Ivics et al., Cell 1997, 501 – 510 ; Wagstaff et al., Molecular Biology and Evolution 2013, 88-99.

Polypeptide components of gene modifying systems

In some embodiments, the gene modifying polypeptide possesses the functions of DNA target site binding, template nucleic acid (e.g., RNA) binding, DNA target site cleavage, and template nucleic acid (e.g., RNA) writing, e.g., reverse transcription. In some embodiments, each functions is contained within a distinct domain. In some embodiments, a function may be attributed to two or more domains (e.g., two or more domains, together, exhibit the functionality). In some embodiments, two or more domains may have the same or similar function (e.g., two or more domains each independently have DNA-binding functionality, e.g., for two different DNA sequences). In other embodiments, one or more domains may be capable of enabling one or more functions, e.g., a Cas9 domain enabling both DNA binding and target site cleavage. In some embodiments, the domains are all located within a single polypeptide. In some embodiments, a first domain is in one polypeptide and a second domain is in a second polypeptide. For example, in some embodiments, the sequences may be split between a first polypeptide and a second polypeptide, e.g., wherein the first polypeptide comprises a reverse transcriptase (RT) domain and wherein the second polypeptide comprises a DNA-binding domain and an endonuclease domain, e.g., a nickase domain. As a further example, in some embodiments, the first polypeptide and the second polypeptide each comprise a DNA binding domain (e.g., a first DNA binding domain and a second DNA binding domain). In some embodiments, the first and second polypeptide may be brought together post-translationally via a split-intein to form a single gene modifying polypeptide.

In some aspects, a gene modifying polypeptide described herein comprises (e.g., a system described herein comprises a gene modifying polypeptide that comprises): 1) a Cas domain (e.g., a Cas nickase domain, e.g., a Cas9 nickase domain); 2) a reverse transcriptase (RT) domain of Table 1, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, or 99% identity thereto, wherein the RT domain is C-terminal of the Cas domain; and a linker disposed between the RT domain and the Cas domain, wherein the linker has a sequence from the same row of Table 1 as the RT domain, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, or 99% identity thereto.

In some embodiments, the RT domain has a sequence with 100% identity to the RT domain of Table 1 and the linker has a sequence with 100% identity to the linker sequence from the same row of Table 1 as the RT domain. In some embodiments, the Cas domain comprises a sequence of Table 8, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto. In some embodiments, the gene modifying polypeptide comprises an amino acid sequence according to any of SEQ ID Nos: 1-3332 in the sequence listing, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, or 99% identity thereto.

In some embodiments, a gene modifying polypeptide described herein comprises the amino acid sequence, or a functional portion thereof, of an exemplary gene modifying polypeptide as listed in any of Tables S1-S3, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide described herein comprises the amino acid sequence of an RT domain of an exemplary gene modifying polypeptide as listed in any of Tables S1-S3, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide described herein comprises the amino acid sequence of a DBD of an exemplary gene modifying polypeptide as listed in any of Tables S1-S3, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide described herein comprises the amino acid sequence of an RBD of an exemplary gene modifying polypeptide as listed in any of Tables S1-S3, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide described herein comprises the amino acid sequence of the RT domain, DBD, and RBD of an exemplary gene modifying polypeptide as listed in any of Tables S1-S3, or amino acid sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

In some embodiments, a gene modifying polypeptide described herein comprises the amino acid sequence, or a functional portion thereof, of an exemplary gene modifying polypeptide as listed in Table S1, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide described herein comprises the

amino acid sequence of an RT domain of an exemplary gene modifying polypeptide as listed in Table S1, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide described herein comprises the amino acid sequence of a DBD of an exemplary gene modifying polypeptide as listed in Table S1, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide described herein comprises the amino acid sequence of an RBD of an exemplary gene modifying polypeptide as listed in Table S1, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide described herein comprises the amino acid sequence of the RT domain, DBD, and RBD of an exemplary gene modifying polypeptide as listed in Table S1, or amino acid sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

In some embodiments, a gene modifying polypeptide described herein comprises the amino acid sequence, or a functional portion thereof, of an exemplary gene modifying polypeptide as listed in Table S2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide described herein comprises the amino acid sequence of an RT domain of an exemplary gene modifying polypeptide as listed in Table S2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide described herein comprises the amino acid sequence of a DBD of an exemplary gene modifying polypeptide as listed in Table S2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide described herein comprises the amino acid sequence of an RBD of an exemplary gene modifying polypeptide as listed in Table S2, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide described herein comprises the amino acid sequence of the RT domain, DBD, and RBD of an exemplary gene modifying polypeptide as listed in Table S2, or amino acid sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

In some embodiments, a gene modifying polypeptide described herein comprises the amino acid sequence, or a functional portion thereof, of an exemplary gene modifying polypeptide as listed in Table S3, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide described herein comprises the amino acid sequence of an RT domain of an exemplary gene modifying polypeptide as listed in Table S3,

or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide described herein comprises the amino acid sequence of a DBD of an exemplary gene modifying polypeptide as listed in Table S3, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide described herein comprises the amino acid sequence of an RBD of an exemplary gene modifying polypeptide as listed in Table S3, or an amino acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide described herein comprises the amino acid sequence of the RT domain, DBD, and RBD of an exemplary gene modifying polypeptide as listed in Table S3, or amino acid sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

In some embodiments, a gene modifying polypeptide described herein comprises a DBD, RT domain, and one or more RBDs (e.g., as described herein).

In certain embodiments, the gene modifying polypeptide comprises, in N-terminal to C-terminal order, a DBD (e.g., a Cas domain, e.g., a Cas9 domain, e.g., as described herein), one or more (e.g., 1, 2, 3, or 4) RBDs, and an RT domain. In embodiments, the DBD and the N-terminal RBD are connected by a linker (e.g., as described herein). In embodiments, the C-terminal RBD and the RT domain are connected by a linker (e.g., as described herein).

In certain embodiments, the gene modifying polypeptide comprises, in N-terminal to C-terminal order, an RT domain, one or more (e.g., 1, 2, 3, or 4) RBDs, and a DBD (e.g., a Cas domain, e.g., a Cas9 domain, e.g., as described herein). In embodiments, the RT domain and the N-terminal RBD are connected by a linker (e.g., as described herein). In embodiments, the C-terminal RBD and the DBD are connected by a linker (e.g., as described herein).

In certain embodiments, the gene modifying polypeptide comprises, in N-terminal to C-terminal order, a DBD (e.g., a Cas domain, e.g., a Cas9 domain, e.g., as described herein), an RT domain, and one or more (e.g., 1, 2, 3, or 4) RBDs. In embodiments, the DBD and RT domain are connected by a linker (e.g., as described herein). In embodiments, the RT domain and the the N-terminal RBD are connected by a linker (e.g., as described herein).

In some embodiments, the gene modifying polypeptide comprises an N-terminal methionine residue.

In some embodiments, the gene modifying polypeptide comprises one or more nuclear localization sequences (NLSes), e.g., as described herein.

In some embodiments, the gene modifying polypeptide comprises a GG amino acid sequence between the Cas domain and the linker, an AG amino acid sequence between the RT domain and the second NLS, and/or a GG amino acid sequence between the linker and the RT domain. In some embodiments, the gene modifying polypeptide comprises a sequence of SEQ ID NO: 4000 which comprises the first NLS and the Cas domain, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto. In some embodiments, the gene modifying polypeptide comprises a sequence of SEQ ID NO: 4001 which comprises the second NLS, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto.

Exemplary N-terminal NLS-Cas9 domain

MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHP
IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVN
TEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNR
EKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDR
EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPE
NIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD
MYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY
KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY
SVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR
KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGG (SEQ ID NO: 4000)

Exemplary C-terminal sequence comprising an NLS

AGKRTADGSEFEKRTADGSEFESPKKKAKVE (SEQ ID NO: 4001)

Gene modifying domain (RT Domain)

In certain aspects of the present invention, the gene modifying domain of the gene modifying system possesses reverse transcriptase activity and is also referred to as a reverse transcriptase domain (a RT domain). In some embodiments, the RT domain comprises an RT catalytic portion and RNA-binding region (e.g., a region that binds the template RNA).

In some embodiments, a nucleic acid encoding the reverse transcriptase is altered from its natural sequence to have altered codon usage, e.g. improved for human cells. In some embodiments the reverse transcriptase domain is a heterologous reverse transcriptase from a retrovirus. In some embodiments, the RT domain comprising a gene modifying polypeptide has been mutated from its original amino acid sequence, e.g., has at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 substitutions. In some embodiments, the RT domain is derived from the RT of a retrovirus, e.g., HIV-1 RT, Moloney Murine Leukemia Virus (MMLV) RT, avian myeloblastosis virus (AMV) RT, or Rous Sarcoma Virus (RSV) RT.

In some embodiments, the retroviral reverse transcriptase (RT) domain exhibits enhanced stringency of target-primed reverse transcription (TPRT) initiation, e.g., relative to an endogenous RT domain. In some embodiments, the RT domain initiates TPRT when the 3 nt in the target site immediately upstream of the first strand nick, e.g., the genomic DNA priming the RNA template, have at least 66% or 100% complementarity to the 3 nt of homology in the RNA template. In some embodiments, the RT domain initiates TPRT when there are less than 5 nt mismatched (e.g., less than 1, 2, 3, 4, or 5 nt mismatched) between the template RNA homology and the target DNA priming reverse transcription. In some embodiments, the RT domain is modified such that the stringency for mismatches in priming the TPRT reaction is increased, e.g., wherein the RT domain does not tolerate any mismatches or tolerates fewer mismatches in the priming region relative to a wild-type (e.g., unmodified) RT domain. In some embodiments, the RT domain comprises a HIV-1 RT domain. In embodiments, the HIV-1 RT domain initiates lower levels of synthesis even with three nucleotide mismatches relative to an alternative RT domain (e.g., as described by Jamburuthugoda and Eickbush J Mol Biol 407(5):661-672 (2011); incorporated herein by reference in its entirety). In some embodiments, the RT domain forms a dimer (e.g., a heterodimer or homodimer). In some embodiments, the RT domain is monomeric. In some embodiments, an RT domain, naturally functions as a monomer or as a dimer (e.g., heterodimer or homodimer). In some embodiments, an RT domain naturally functions as a monomer, e.g., is derived from a virus wherein it functions as a monomer. In embodiments, the RT domain is selected from an RT domain from murine leukemia virus (MLV; sometimes referred to as MoMLV) (e.g., P03355), porcine endogenous retrovirus (PERV) (e.g., UniProt Q4VFZ2), mouse mammary tumor virus (MMTV) (e.g.,

UniProt P03365), Mason-Pfizer monkey virus (MPMV) (e.g., UniProt P07572), bovine leukemia virus (BLV) (e.g., UniProt P03361), human T-cell leukemia virus-1 (HTLV-1) (e.g., UniProt P03362), human foamy virus (HFV) (e.g., UniProt P14350), simian foamy virus (SFV) (e.g., UniProt P23074), or bovine foamy/syncytial virus (BFV/BSV) (e.g., UniProt O41894), or a functional fragment or variant thereof (e.g., an amino acid sequence having at least 70%, 80%, 90%, 95%, or 99% identity thereto). In some embodiments, an RT domain is dimeric in its natural functioning. In some embodiments, the RT domain is derived from a virus wherein it functions as a dimer. In embodiments, the RT domain is selected from an RT domain from avian sarcoma/leukemia virus (ASLV) (e.g., UniProt A0A142BKH1), Rous sarcoma virus (RSV) (e.g., UniProt P03354), avian myeloblastosis virus (AMV) (e.g., UniProt Q83133), human immunodeficiency virus type I (HIV-1) (e.g., UniProt P03369), human immunodeficiency virus type II (HIV-2) (e.g., UniProt P15833), simian immunodeficiency virus (SIV) (e.g., UniProt P05896), bovine immunodeficiency virus (BIV) (e.g., UniProt P19560), equine infectious anemia virus (EIAV) (e.g., UniProt P03371), or feline immunodeficiency virus (FIV) (e.g., UniProt P16088) (Herschhorn and Hizi *Cell Mol Life Sci* 67(16):2717-2747 (2010)), or a functional fragment or variant thereof (e.g., an amino acid sequence having at least 70%, 80%, 90%, 95%, or 99% identity thereto). Naturally heterodimeric RT domains may, in some embodiments, also be functional as homodimers. In some embodiments, dimeric RT domains are expressed as fusion proteins, e.g., as homodimeric fusion proteins or heterodimeric fusion proteins. In some embodiments, the RT function of the system is fulfilled by multiple RT domains (e.g., as described herein). In further embodiments, the multiple RT domains are fused or separate, e.g., may be on the same polypeptide or on different polypeptides.

In some embodiments, a gene modifying system described herein comprises an integrase domain, e.g., wherein the integrase domain may be part of the RT domain. In some embodiments, an RT domain (e.g., as described herein) comprises an integrase domain. In some embodiments, an RT domain (e.g., as described herein) lacks an integrase domain, or comprises an integrase domain that has been inactivated by mutation or deleted. In some embodiment, a gene modifying system described herein comprises an RNase H domain, e.g., wherein the RNase H domain may be part of the RT domain. In some embodiments, the RNase H domain is not part of the RT domain and is covalently linked via a flexible linker. In some embodiments, an RT domain (e.g., as described herein) comprises an RNase H domain, e.g., an endogenous RNAse H domain or a heterologous RNase H domain. In some embodiments, an RT domain (e.g., as described herein) lacks an RNase H domain. In some embodiments, an RT domain (e.g., as described herein) comprises an RNase H domain that has been added, deleted, mutated, or swapped for a heterologous RNase H domain. In some embodiments, the polypeptide comprises an inactivated endogenous RNase H domain. In some embodiments, an endogenous RNase H domain from one of the other domains of the polypeptide is genetically removed such that it is not included in the polypeptide,

e.g., the endogenous RNase H domain is partially or completely truncated from the comprising domain. In some embodiments, mutation of an RNase H domain yields a polypeptide exhibiting lower RNase activity, e.g., as determined by the methods described in Kotewicz et al. *Nucleic Acids Res* 16(1):265-277 (1988) (incorporated herein by reference in its entirety), e.g., lower by at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% compared to an otherwise similar domain without the mutation. In some embodiments, RNase H activity is abolished.

In some embodiments, an RT domain is mutated to increase fidelity compared to an otherwise similar domain without the mutation. For instance, in some embodiments, a YADD or YMDD motif in an RT domain (e.g., in a reverse transcriptase) is replaced with YVDD. In embodiments, replacement of the YADD or YMDD or YVDD results in higher fidelity in retroviral reverse transcriptase activity (e.g., as described in Jamburuthugoda and Eickbush J Mol Biol 2011; incorporated herein by reference in its entirety).

In some embodiments, a gene modifying polypeptide described herein comprises an RT domain having an amino acid sequence according to Table 6, or a sequence having at least 70%, 80%, 85%, 90%, 95%, 97%, 98%, or 99% identity thereto. In some embodiments, a nucleic acid described herein encodes an RT domain having an amino acid sequence according to Table 6, or a sequence having at least 70%, 80%, 85%, 90%, 95%, 97%, 98%, or 99% identity thereto.

Table 6: Exemplary reverse transcriptase domains from retroviruses

RT Name	RT amino acid sequence
AVIRE_P03360	TAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQY PITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPT VPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGF KNSPTLFDEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRV SGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGTIGYCRLWIPGFA ELAQPLYAATRGGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKG VLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLES LLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTST RPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTK ALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGLLTAGGKAIKNAPEILALLTAVWLPKRV AVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATIS
AVIRE_P03360_3mut	TAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQY PITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPT VPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGF KNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRV SGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGTIGYCRLWIPGFA ELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKG VLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLES LLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTST RPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTK ALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKR VAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATIS

AVIRE_P03360_3mutA	TAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQY PITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPT VPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGF KNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRV SGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFA ELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKG VLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLES LLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTST RPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTK ALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKR VAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATIS
BAEVM_P10272	TVSLQDEHRLFDIPVTTSLPDVWLQDFPQAWAETGGLGRAKCQAPIIIDLKPTAVPVSIKQYP MSLEAHMGIRQHIIKFLELGVLRPCRSPWNTPLLPVKKPGTQDYRPVQDLREINKRTVDIHPT VPNPYNLLSTLKPDYSWYTVLDLKDAFFCLPLAPQSQELFAFEWKDPERGISGQLTWTRLPQ GFKNSPTLFDEALHRDLTDFRTQHPEVTLLQYVDDLLLAAPTKKACTQGTRHLLQELGEKGY RASAKKAQICQTKVTYLGYILSEGKRWLTPGRIETVARIPPPRNPREVREFLGTAGFCRLWIPG FAELAAPLYALTKESTPFTWQTEHQLAFEALKKALLSAPALGLPDTSKPFTLFLDERQGIAKG VLTQKLGPWKRPVAYLSKKLDPVAAGWPPCLRIMAATAMLVKDSAKLTLGQPLTVITPHTL EAIVRQPPDRWITNARLTHYQALLLDTDRVQFGPPVTLNPATLLPVPENQPSPHDCRQVLAET HGTREDLKDQELPDADHTWYTDGSSYLDSGTRRAGAAVVDGHNTIWAQSLPPGTSAQKAEL IALTKALELSKGKKANIYTDSRYAFATAHTHGSIYERRGLLTSEGKEIKNKAEIIALLKALFLP QEVAIIHCPGHQKGQDPVAVGNRQADRVARQAAMAEVLTLATEPDNTSHIT
BAEVM_P10272_3mut	TVSLQDEHRLFDIPVTTSLPDVWLQDFPQAWAETGGLGRAKCQAPIIIDLKPTAVPVSIKQYP MSLEAHMGIRQHIIKFLELGVLRPCRSPWNTPLLPVKKPGTQDYRPVQDLREINKRTVDIHPT VPNPYNLLSTLKPDYSWYTVLDLKDAFFCLPLAPQSQELFAFEWKDPERGISGQLTWTRLPQ GFKNSPTLFNEALHRDLTDFRTQHPEVTLLQYVDDLLLAAPTKKACTQGTRHLLQELGEKGY RASAKKAQICQTKVTYLGYILSEGKRWLTPGRIETVARIPPPRNPREVREFLGTAGFCRLWIPG FAELAAPLYALTKPSTPFTWQTEHQLAFEALKKALLSAPALGLPDTSKPFTLFLDERQGIAKG VLTQKLGPWKRPVAYLSKKLDPVAAGWPPCLRIMAATAMLVKDSAKLTLGQPLTVITPHTL EAIVRQPPDRWITNARLTHYQALLLDTDRVQFGPPVTLNPATLLPVPENQPSPHDCRQVLAET HGTREDLKDQELPDADHTWYTDGSSYLDSGTRRAGAAVVDGHNTIWAQSLPPGTSAQKAEL IALTKALELSKGKKANIYTDSRYAFATAHTHGSIYERRGWLTSEGKEIKNKAEIIALLKALFLP QEVAIIHCPGHQKGQDPVAVGNRQADRVARQAAMAEVLTLATEPDNTSHIT
BAEVM_P10272_3mutA	TVSLQDEHRLFDIPVTTSLPDVWLQDFPQAWAETGGLGRAKCQAPIIIDLKPTAVPVSIKQYP MSLEAHMGIRQHIIKFLELGVLRPCRSPWNTPLLPVKKPGTQDYRPVQDLREINKRTVDIHPT VPNPYNLLSTLKPDYSWYTVLDLKDAFFCLPLAPQSQELFAFEWKDPERGISGQLTWTRLPQ GFKNSPTLFNEALHRDLTDFRTQHPEVTLLQYVDDLLLAAPTKKACTQGTRHLLQELGEKGY RASAKKAQICQTKVTYLGYILSEGKRWLTPGRIETVARIPPPRNPREVREFLGKAGFCRLFIPG FAELAAPLYALTKPSTPFTWQTEHQLAFEALKKALLSAPALGLPDTSKPFTLFLDERQGIAKG VLTQKLGPWKRPVAYLSKKLDPVAAGWPPCLRIMAATAMLVKDSAKLTLGQPLTVITPHTL EAIVRQPPDRWITNARLTHYQALLLDTDRVQFGPPVTLNPATLLPVPENQPSPHDCRQVLAET HGTREDLKDQELPDADHTWYTDGSSYLDSGTRRAGAAVVDGHNTIWAQSLPPGTSAQKAEL IALTKALELSKGKKANIYTDSRYAFATAHTHGSIYERRGWLTSEGKEIKNKAEIIALLKALFLP QEVAIIHCPGHQKGQDPVAVGNRQADRVARQAAMAEVLTLATEPDNTSHIT
BLVAU_P25059	GVLDAPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKP NGAWRFVHDLRVTNALTKPIPALSPGPPDLTAIPTHLPHIICLDLKDAFFQIPVEDRFRSYFAFT LPTPGGLQPHRRFAWRVLPQGFINSPALFERALQEPLRQVSAAFSQSLLVSYMDDILYVSPTEE QRLQCYQTMAAHLRDLGFQVASEKTRQTPSPVPFLGQMVHERMVTYQSLPTLQISSPISLHQL QTVLGDLQWVSRGTPTTRRPLQLLYSSLKGIDDPRAIIHLSPEQQQGIAELRQALSHNARSRY NEQEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQAQALS SYAKTILKYYHNLPKTSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLVTRAEVFLTP QFSPEPIPAALCLFSDGAARRGAYCLWKDHLLDFQAVPAPESAQKGELAGLLAGLAAAPPEP LNIWVDSKYLYSLLRTLVLGAWLQPDPVPSYALLYKSLLRHPAIFVGHVRSHSSASHPIASLN NYVDQL

BLVAU_P25059_2mut	GVLDAPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKP NGAWRFVHDLRVTNALTKPIPALSPGPPDLTAIPTHLPHIICLDLKDAFFQIPVEDRFRSYFAFT LPTPGGLQPHRRFAWRVLPQGFINSPALFQRALQEPLRQVSAAFSQSLLVSYMDDILYVSPTE EQRLQCYQTMAAHLRDLGFQVASEKTRQTPSPVPFLGQMVHERMVTYQSLPTLQISSPISLHQ LQTVLGDLQWVSRGTPTTRRPLQLLYSSLKPIDDPRAIIHLSPEQQQGIAELRQALSHNARSRY NEQEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQAQALS SYAKTILKYYHNLPKTSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLVTRAEVFLTP QFSPEPIPAALCLFSDGAARRGAYCLWKDHLLDFQAVPAPESAQKGELAGLLAGLAAAPPEP LNIWVDSKYLYSLLRTLVLGAWLQPDPVPSYALLYKSLLRHPAIFVGHVRSHSSASHPIASLN NYVDQL
BLVJ_P03361	GVLDTPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKP NGAWRFVHDLRATNALTKPIPALSPGPPDLTAIPTHPPHIICLDLKDAFFQIPVEDRFRFYLSFT LPSPGGLQPHRRFAWRVLPQGFINSPALFERALQEPLRQVSAAFSQSLLVSYMDDILYASPTEE QRSQCYQALAARLRDLGFQVASEKTSQTPSPVPFLGQMVHEQIVTYQSLPTLQISSPISLHQLQ AVLGDLQWVSRGTPTTRRPLQLLYSSLKRHHDPRAIIQLSPEQLQGIAELRQALSHNARSRYN EQEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQTQALSS YAKPILKYYHNLPKTSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLITRAEVFLTPQ FSPDPIPAALCLFSDGATGRGAYCLWKDHLLDFQAVPAPESAQKGELAGLLAGLAAAPPEPV NIWVDSKYLYSLLRTLVLGAWLQPDPVPSYALLYKSLLRHPAIVVGHVRSHSSASHPIASLNN YVDQL
BLVJ_P03361_2mut	GVLDTPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKP NGAWRFVHDLRATNALTKPIPALSPGPPDLTAIPTHPPHIICLDLKDAFFQIPVEDRFRFYLSFT LPSPGGLQPHRRFAWRVLPQGFINSPALFNRALQEPLRQVSAAFSQSLLVSYMDDILYASPTEE QRSQCYQALAARLRDLGFQVASEKTSQTPSPVPFLGQMVHEQIVTYQSLPTLQISSPISLHQLQ AVLGDLQWVSRGTPTTRRPLQLLYSSLKRHHDPRAIIQLSPEQLQGIAELRQALSHNARSRYN EQEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQTQALSS YAKPILKYYHNLPKTSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLITRAEVFLTPQ FSPDPIPAALCLFSDGATGRGAYCLWKDHLLDFQAVPAPESAQKGELAGLLAGLAAAPPEPV NIWVDSKYLYSLLRTWVLGAWLQPDPVPSYALLYKSLLRHPAIVVGHVRSHSSASHPIASLN NYVDQL
BLVJ_P03361_2mutB	GVLDTPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKP NGAWRFVHDLRATNALTKPIPALSPGPPDLTAPPTHPPHIICLDLKDAFFQIPVEDRFRFYLSFT LPSPGGLQPHRRFAWRVLPQGFINSPALFQRALQEPLRQVSAAFSQSLLVSYMDDILYASPTEE QRSQCYQALAARLRDLGFQVASEKTSQTPSPVPFLGQMVHEQIVTYQSLPTLQISSPISLHQLQ AVLGDLQWVSRGTPTTRRPLQLLYSSLKRHHDPRAIIQLSPEQLQGIAELRQALSHNARSRYN EQEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQTQALSS YAKPILKYYHNLPKTSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLITRAEVFLTPQ FSPDPIPAALCLFSDGATGRGAYCLWKDHLLDFQAVPAPESAQKGELAGLLAGLAAAPPEPV NIWVDSKYLYSLLRTWVLGAWLQPDPVPSYALLYKSLLRHPAIVVGHVRSHSSASHPIASLN NYVDQL
FFV_O93209	MDLLKPLTVERKGVKIKGYWNSQADITCVPKDLLQGEEPVRQQNVTTIHGTQEGDVYYVNL KIDGRRINTEVIGTTLDYAIITPGDVPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKEELKQ LFEKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVLI QKESTMNTPVYPVPKPNGRWRMVLDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLS NGFWAHPIVPEDYWITAFTWQGKQYCWTVLPQGFLNSPGLFTGDVVDLLQGIPNVEVYVDD VYISHDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLENI TAPTTLKQLQSILGLLNFARNFIPDFTELIAPLYALIPKSTKNYVPWQIEHSTTLETLITKLNGAE YLQGRKGDKTLIMKVNASYTTGYIRYYNEGEKKPISYVSIVFSKTELKFTELEKLLTTVHKGL LKALDLSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDPQMPAL KDLPAVDTGKDNKKHPSNFQHIFYTDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWS ISLGNHTAQFAEIAAFEFALKKCLPLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNRKKP LKHISKWKSVADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQASFKVH
FFV_O93209_2mut	MDLLKPLTVERKGVKIKGYWNSQADITCVPKDLLQGEEPVRQQNVTTIHGTQEGDVYYVNL KIDGRRINTEVIGTTLDYAIITPGDVPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKEELKQ LFEKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVLI QKESTMNTPVYPVPKPNGRWRMVLDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLS

	NGFWAHPIVPEDYWITAFTWQGKQYCWTVLPQGFLNSPGLFNGDVVDLLQGIPNVEVYVDD VYISHDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLENI TAPTTLKQLQSILGLLNFARNFIPDFTELIAPLYALIPKSPKNYVPWQIEHSTTLETLITKLNGAE YLQGRKGDKTLIMKVNASYTTGYIRYYNEGEKKPISYVSIVFSKTELKFTELEKLLTTVHKGL LKALDLSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDPQMPAL KDLPAVDTGKDNKKHPSNFQHIFYTDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWS ISLGNHTAQFAEIAAFEFALKKCLPLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNRKKP LKHISKWKSVADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQASFKVH
FFV_O93209_2mutA	MDLLKPLTVERKGVKIKGYWNSQADITCVPKDLLQGEEPVRQQNVTTIHGTQEGDVYYVNL KIDGRRINTEVIGTTLDYAIITPGDVPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKEELKQ LFEKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVLI QKESTMNTPVYPVPKPNGRWRMVLDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLS NGFWAHPIVPEDYWITAFTWQGKQYCWTVLPQGFLNSPGLFNGDVVDLLQGIPNVEVYVDD VYISHDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLENI TAPTTLKQLQSILGKLNFARNFIPDFTELIAPLYALIPKSPKNYVPWQIEHSTTLETLITKLNGAE YLQGRKGDKTLIMKVNASYTTGYIRYYNEGEKKPISYVSIVFSKTELKFTELEKLLTTVHKGL LKALDLSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDPQMPAL KDLPAVDTGKDNKKHPSNFQHIFYTDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWS ISLGNHTAQFAEIAAFEFALKKCLPLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNRKKP LKHISKWKSVADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQASFKVH
FFV_O93209-Pro	VPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRP HKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVLIQKESTMNTPVYPVPKPNGRWRMV LDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGK QYCWTVLPQGFLNSPGLFTGDVVDLLQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKE AGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLENITAPTTLKQLQSILGLLNFARNFIP DFTELIAPLYALIPKSTKNYVPWQIEHSTTLETLITKLNGAEYLQGRKGDKTLIMKVNASYTTG YIRYYNEGEKKPISYVSIVFSKTELKFTELEKLLTTVHKGLLKALDLSMGQNIHVYSPIVSMQN IQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDPQMPALKDLPAVDTGKDNKKHPSNFQHIF YTDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFALKKCL PLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNRKKPLKHISKWKSVADLKRLRPDVVVT HEPGHQKLDSSPHAYGNNLADQLATQASFKVH
FFV_O93209-Pro_2mut	VPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRP HKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVLIQKESTMNTPVYPVPKPNGRWRMV LDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGK QYCWTVLPQGFLNSPGLFNGDVVDLLQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKE AGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLENITAPTTLKQLQSILGLLNFARNFIP DFTELIAPLYALIPKSPKNYVPWQIEHSTTLETLITKLNGAEYLQGRKGDKTLIMKVNASYTTG YIRYYNEGEKKPISYVSIVFSKTELKFTELEKLLTTVHKGLLKALDLSMGQNIHVYSPIVSMQN IQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDPQMPALKDLPAVDTGKDNKKHPSNFQHIF YTDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFALKKCL PLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNRKKPLKHISKWKSVADLKRLRPDVVVT HEPGHQKLDSSPHAYGNNLADQLATQASFKVH
FFV_O93209-Pro_2mutA	VPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRP HKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVLIQKESTMNTPVYPVPKPNGRWRMV LDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGK QYCWTVLPQGFLNSPGLFNGDVVDLLQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKE AGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLENITAPTTLKQLQSILGKLNFARNFIP DFTELIAPLYALIPKSPKNYVPWQIEHSTTLETLITKLNGAEYLQGRKGDKTLIMKVNASYTTG YIRYYNEGEKKPISYVSIVFSKTELKFTELEKLLTTVHKGLLKALDLSMGQNIHVYSPIVSMQN IQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDPQMPALKDLPAVDTGKDNKKHPSNFQHIF YTDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFALKKCL PLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNRKKPLKHISKWKSVADLKRLRPDVVVT HEPGHQKLDSSPHAYGNNLADQLATQASFKVH
FLV_P10273	TLQLEEEYRLFEPESTQKQEMDIWLKNFPQAWAETGGMGTAHCQAPVLIQLKATATPISIRQY PMPHEAYQGIKPHIRRMLDQGILKPCQSPWNTPLLPVKKPGTEDYRPVQDLREVNKRVEDIHP TVPNPYNLLSTLPPSHPWYTVLDLKDAFFCLRLHSESQLLFAFEWRDPEIGLSGQLTWTRLPQ

	GFKNSPTLFDEALHSDLADFRVRYPALVLLQYVDDLLLAAATRTECLEGTKALLETLGNKGY RASAKKAQICLQEVTYLGYSLKDGQRWLTKARKEAILSIPVPKNSRQVREFLGTAGYCRLWI PGFAELAAPLYPLTRPGTLFQWGTEQQLAFEDIKKALLSSPALGLPDITKPFELFIDENSGFAK GVLVQKLGPWKRPVAYLSKKLDTVASGWPPCLRMVAAIAILVKDAGKLTLGQPLTILTSHPV EALVRQPPNKWLSNARMTHYQAMLLDAERVHFGPTVSLNPATLLPLPSGGNHHDCLQILAE THGTRPDLTDQPLPDADLTWYTDGSSFIRNGEREAGAAVTTESEVIWAAPLPPGTSAQRAELI ALTQALKMAEGKKLTVYTDSRYAFATTHVHGEIYRRRGLLTSEGKEIKNKNEILALLEALFLP KRLSIIHCPGHQKGDSPQAKGNRLADDTAKKAATETHSSLTVLP
FLV_P10273_3mut	TLQLEEEYRLFEPESTQKQEMDIWLKNFPQAWAETGGMGTAHCQAPVLIQLKATATPISIRQY PMPHEAYQGIKPHIRRMLDQGILKPCQSPWNTPLLPVKKPGTEDYRPVQDLREVNKRVEDIHP TVPNPYNLLSTLPPSHPWYTVLDLKDAFFCLRLHSESQLLFAFEWRDPEIGLSGQLTWTRLPQ GFKNSPTLFNEALHSDLADFRVRYPALVLLQYVDDLLLAAATRTECLEGTKALLETLGNKGY RASAKKAQICLQEVTYLGYSLKDGQRWLTKARKEAILSIPVPKNSRQVREFLGTAGYCRLWI PGFAELAAPLYPLTRPGTLFQWGTEQQLAFEDIKKALLSSPALGLPDITKPFELFIDENSGFAK GVLVQKLGPWKRPVAYLSKKLDTVASGWPPCLRMVAAIAILVKDAGKLTLGQPLTILTSHPV EALVRQPPNKWLSNARMTHYQAMLLDAERVHFGPTVSLNPATLLPLPSGGNHHDCLQILAE THGTRPDLTDQPLPDADLTWYTDGSSFIRNGEREAGAAVTTESEVIWAAPLPPGTSAQRAELI ALTQALKMAEGKKLTVYTDSRYAFATTHVHGEIYRRRGWLTSEGKEIKNKNEILALLEALFL PKRLSIIHCPGHQKGDSPQAKGNRLADDTAKKAATETHSSLTVLP
FLV_P10273_3mutA	TLQLEEEYRLFEPESTQKQEMDIWLKNFPQAWAETGGMGTAHCQAPVLIQLKATATPISIRQY PMPHEAYQGIKPHIRRMLDQGILKPCQSPWNTPLLPVKKPGTEDYRPVQDLREVNKRVEDIHP TVPNPYNLLSTLPPSHPWYTVLDLKDAFFCLRLHSESQLLFAFEWRDPEIGLSGQLTWTRLPQ GFKNSPTLFNEALHSDLADFRVRYPALVLLQYVDDLLLAAATRTECLEGTKALLETLGNKGY RASAKKAQICLQEVTYLGYSLKDGQRWLTKARKEAILSIPVPKNSRQVREFLGKAGYCRLFIP GFAELAAPLYPLTRPGTLFQWGTEQQLAFEDIKKALLSSPALGLPDITKPFELFIDENSGFAKG VLVQKLGPWKRPVAYLSKKLDTVASGWPPCLRMVAAIAILVKDAGKLTLGQPLTILTSHPVE ALVRQPPNKWLSNARMTHYQAMLLDAERVHFGPTVSLNPATLLPLPSGGNHHDCLQILAET HGTRPDLTDQPLPDADLTWYTDGSSFIRNGEREAGAAVTTESEVIWAAPLPPGTSAQRAELIA LTQALKMAEGKKLTVYTDSRYAFATTHVHGEIYRRRGWLTSEGKEIKNKNEILALLEALFLP KRLSIIHCPGHQKGDSPQAKGNRLADDTAKKAATETHSSLTVLP
FOAMV_P14350	MNPLQLLQPLPAEIKGTKLLAHWNSGATITCIPESFLEDEQPIKKTLIKTIHGEKQQNVYYVTF KVKGRKVEAEVIASPYEYILLSPTDVPWLTQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLK TLFVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVL TPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDL ANGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPALFTADVVDLLKEIPNVQVYVD DIYLSHDDPKEHVQQLEKVFQILLQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKTK LLNITPPKDLKQLQSILGLLNFARNFIPNFAELVQPLYNLIASAKGKYIEWSEENTKQLNMVIE ALNTASNLEERLPEQRLVIKVNTSPSAGYVRYYNETGKKPIMYLNYVFSKAELKFSMLEKLLT TMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDK TLPELKHIPDVYTSSQSPVKHPSQYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVL NQWSIPLGNHTAQMAEIAAVEFACKKALKIPGPVLVITDSFYVAESANKELPYWKSNGFVNN KKKPLKHISKWKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN
FOAMV_P14350_2mut	MNPLQLLQPLPAEIKGTKLLAHWNSGATITCIPESFLEDEQPIKKTLIKTIHGEKQQNVYYVTF KVKGRKVEAEVIASPYEYILLSPTDVPWLTQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLK TLFVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVL TPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDL ANGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPALFNADVVDLLKEIPNVQVYVD DIYLSHDDPKEHVQQLEKVFQILLQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKTK LLNITPPKDLKQLQSILGLLNFARNFIPNFAELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIE ALNTASNLEERLPEQRLVIKVNTSPSAGYVRYYNETGKKPIMYLNYVFSKAELKFSMLEKLLT TMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDK TLPELKHIPDVYTSSQSPVKHPSQYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVL NQWSIPLGNHTAQMAEIAAVEFACKKALKIPGPVLVITDSFYVAESANKELPYWKSNGFVNN KKKPLKHISKWKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN
FOAMV_P143	MNPLQLLQPLPAEIKGTKLLAHWNSGATITCIPESFLEDEQPIKKTLIKTIHGEKQQNVYYVTF KVKGRKVEAEVIASPYEYILLSPTDVPWLTQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLK

50_2mutA	TLFVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVL TPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDL ANGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPALFNADVVDLLKEIPNVQVYVD DIYLSHDDPKEHVQQLEKVFQILLQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKTK LLNITPPKDLKQLQSILGKLNFARNFIPNFAELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIE ALNTASNLEERLPEQRLVIKVNTSPSAGYVRYYNETGKKPIMYLNYVFSKAELKFSMLEKLLT TMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDK TLPELKHIPDVYTSSQSPVKHPSQYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVL NQWSIPLGNHTAQMAEIAAVEFACKKALKIPGPVLVITDSFYVAESANKELPYWKSNGFVNN KKKPLKHISKWKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN
FOAM_P14350-Pro	VPWLTQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKI RPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLTPQNSTMNTPVYPVPKPDGRWR MVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLTAFTWQ GKQYCWTRLPQGFLNSPALFTADVVDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQIL LQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKTKLLNITPPKDLKQLQSILGLLNFAR NFIPNFAELVQPLYNLIASAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTS PSAGYVRYYNETGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYS PIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSQSPVKHPS QYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEF ACKKALKIPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLKHISKWKSIAECLSMKP DITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN
FOAM_P14350-Pro_2mut	VPWLTQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKI RPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLTPQNSTMNTPVYPVPKPDGRWR MVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLTAFTWQ GKQYCWTRLPQGFLNSPALFNADVVDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQIL LQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKTKLLNITPPKDLKQLQSILGLLNFAR NFIPNFAELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTS PSAGYVRYYNETGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYS PIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSQSPVKHPS QYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEF ACKKALKIPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLKHISKWKSIAECLSMKP DITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN
FOAM_P14350-Pro_2mutA	VPWLTQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKI RPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLTPQNSTMNTPVYPVPKPDGRWR MVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLTAFTWQ GKQYCWTRLPQGFLNSPALFNADVVDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQIL LQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKTKLLNITPPKDLKQLQSILGKLNFAR NFIPNFAELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTS PSAGYVRYYNETGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYS PIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSQSPVKHPS QYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEF ACKKALKIPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLKHISKWKSIAECLSMKP DITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN
GALV_P21414	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQY PMSKEAREGIRPHIQKFLDLGVLVPCRSPWNTPLLPVKKPGTNDYRPVQDLREINKRVQDIHP TVPNPYNLLSSLPPSYTWYSVLDLKDAFFCLRLHPNSQPLFAFEWKDPEKGNTGQLTWTRLP QGFKNSPTLFDEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYEDCKKGTQKLLQELSKLG YRVSAKKAQLCQREVTYLGYLLKEGKRWLTPARKATVMKIPVPTTPRQVREFLGTAGFCRL WIPGFASLAAPLYPLTKESIPFIWTEEHQQAFDHIKKALLSAPALALPDLTKPFTLYIDERAGV ARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIAS HSLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILA EETGTRRDLEDQPLPGVPTWYTDGSSFITEGKRRAGAPIVDGKRTVWASSLPEGTSAQKAEL VALTQALRLAEGKNINIYTDSRYAFATAHIHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLP RRVAIIHCPGHQRGSNPVATGNRRADEAAKQAALSTRVLAGTTKP

GALV_P21414_3mut	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQY PMSKEAREGIRPHIQKFLDLGVLVPCRSPWNTPLLPVKKPGTNDYRPVQDLREINKRVQDIHP TVPNPYNLLSSLPPSYTWYSVLDLKDAFFCLRLHPNSQPLFAFEWKDPEKGNTGQLTWTRLP QGFKNSPTLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYEDCKKGTQKLLQELSKLG YRVSAKKAQLCQREVTYLGYLLKEGKRWLTPARKATVMKIPVPTTPRQVREFLGTAGFCRL WIPGFASLAAPLYPLTKPSIPFIWTEEHQQAFDHIKKALLSAPALALPDLTKPFTLYIDERAGVA RGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASH SLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAE ETGTRRDLEDQPLPGVPTWYTDGSSFITEGKRRAGAPIVDGKRTVWASSLPEGTSAQKAELV ALTQALRLAEGKNINIYTDSRYAFATAHIHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPR RVAIIHCPGHQRGSNPVATGNRRADEAAKQAALSTRVLAGTTKP
GALV_P21414_3mutA	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQY PMSKEAREGIRPHIQKFLDLGVLVPCRSPWNTPLLPVKKPGTNDYRPVQDLREINKRVQDIHP TVPNPYNLLSSLPPSYTWYSVLDLKDAFFCLRLHPNSQPLFAFEWKDPEKGNTGQLTWTRLP QGFKNSPTLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYEDCKKGTQKLLQELSKLG YRVSAKKAQLCQREVTYLGYLLKEGKRWLTPARKATVMKIPVPTTPRQVREFLGKAGFCRL FIPGFASLAAPLYPLTKPSIPFIWTEEHQQAFDHIKKALLSAPALALPDLTKPFTLYIDERAGVA RGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASH SLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAE ETGTRRDLEDQPLPGVPTWYTDGSSFITEGKRRAGAPIVDGKRTVWASSLPEGTSAQKAELV ALTQALRLAEGKNINIYTDSRYAFATAHIHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPR RVAIIHCPGHQRGSNPVATGNRRADEAAKQAALSTRVLAGTTKP
HTL1A_P03362	AVLGLEHLPRPPQISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRF IHDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQTIDLRDAFFQIPLPKQFQPYFAFTVPQQCNY GPGTRYAWKVLPQGFKNSPTLFEMQLAHILQPIRQAFPQCTILQYMDDILLASPSHEDLLLLSE ATMASLISHGLPVSENKTQQTPGTIKFLGQIISPNHLTYDAVPTVPIRSRWALPELQALLGEIQ WVSKGTPTLRQPLHSLYCALQRHTDPRDQIYLNPSQVQSLVQLRQALSQNCRSRLVQTLPLL GAIMLTLTGTTTVVFQSKEQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLC QTIHHNISTQTFNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALMP VFTLSPVIINTAPCLFSDGSTSRAAYILWDKQILSQRSFPLPPPHKSAQRAELLGLLHGLSSARS WRCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHTNLPDPIS RLNALTDALLITPVLQL
HTL1A_P03362_2mut	AVLGLEHLPRPPQISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRF IHDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQTIDLRDAFFQIPLPKQFQPYFAFTVPQQCNY GPGTRYAWKVLPQGFKNSPTLFQMQLAHILQPIRQAFPQCTILQYMDDILLASPSHEDLLLLSE ATMASLISHGLPVSENKTQQTPGTIKFLGQIISPNHLTYDAVPTVPIRSRWALPELQALLGEIQ WVSKGTPTLRQPLHSLYCALQPHTDPRDQIYLNPSQVQSLVQLRQALSQNCRSRLVQTLPLL GAIMLTLTGTTTVVFQSKEQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLC QTIHHNISTQTFNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALMP VFTLSPVIINTAPCLFSDGSTSRAAYILWDKQILSQRSFPLPPPHKSAQRAELLGLLHGLSSARS WRCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHTNLPDPIS RLNALTDALLITPVLQL
HTL1A_P03362_2mutB	AVLGLEHLPRPPQISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRF IHDLRATNSLTIDLSSSSPGPPDLSSPPTTLAHLQTIDLRDAFFQIPLPKQFQPYFAFTVPQQCNY GPGTRYAWKVLPQGFKNSPTLFQMQLAHILQPIRQAFPQCTILQYMDDILLASPSHEDLLLLSE ATMASLISHGLPVSENKTQQTPGTIKFLGQIISPNHLTYDAVPTVPIRSRWALPELQALLGEIQ WVSKGTPTLRQPLHSLYCALQPHTDPRDQIYLNPSQVQSLVQLRQALSQNCRSRLVQTLPLL GAIMLTLTGTTTVVFQSKEQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLC QTIHHNISTQTFNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALMP VFTLSPVIINTAPCLFSDGSTSRAAYILWDKQILSQRSFPLPPPHKSAQRAELLGLLHGLSSARS WRCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHTNLPDPIS RLNALTDALLITPVLQL
HTL1C_P14078	AVLGLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRF IHDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQTIDLKDAFFQIPLPKQFQPYFAFTVPQQCNY GPGTRYAWRVLPQGFKNSPTLFEMQLAHILQPIRQAFPQCTILQYMDDILLASPSHADLQLLS EATMASLISHGLPVSENKTQQTPGTIKFLGQIISPNHLTYDAVPKVPIRSRWALPELQALLGEIQ

	WVSKGTPTLRQPLHSLYCALQRHTDPRDQIYLNPSQVQSLVQLRQALSQNCRSRLVQTLPLL GAIMLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLC QTIHHNISTQTFNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTTAPLAPVKALMPV FTLSPVIINTAPCLFSDGSTSQAAYILWDKHILSQRSFPLPPPHKSAQRAELLGLLHGLSSARSW RCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHTNLPDPISRL NALTDALLITPVLQL
HTL1C_P14078_2mut	AVLGLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRF IHDLRATNSLTIDLSSSSPGPPDLSSLPTTLAHLQTIDLKDAFFQIPLPKQFQPYFAFTVPQQCNY GPGTRYAWRVLPQGFKNSPTLFQMQLAHILQPIRQAFPQCTILQYMDDILLASPSHADLQLLS EATMASLISHGLPVSENKTQQTPGTIKFLGQIISPNHLTYDAVPKVPIRSRWALPELQALLGEIQ WVSKGTPTLRQPLHSLYCALQPHTDPRDQIYLNPSQVQSLVQLRQALSQNCRSRLVQTLPLL GAIMLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLC QTIHHNISTQTFNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTTAPLAPVKALMPV FTLSPVIINTAPCLFSDGSTSQAAYILWDKHILSQRSFPLPPPHKSAQRAELLGLLHGLSSARSW RCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHTNLPDPISRL NALTDALLITPVLQL
HTL1L_P0C211	GLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHD LRATNSLTVDLSSSSPGPPDLSSLPTTLAHLQTIDLKDAFFQIPLPKQFQPYFAFTVPQQCNYGP GTRYAWKVLPQGFKNSPTLFEMQLASILQPIRQAFPQCVILQYMDDILLASPSPEDLQQLSEAT MASLISHGLPVSQDKTQQTPGTIKFLGQIISPNHITYDAVPTVPIRSRWALPELQALLGEIQWVS KGTPTLRQPLHSLYCALQGHTDPRDQIYLNPSQVQSLMQLQQALSQNCRSRLAQTLPLLGAI MLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTI HHNISIQTFNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALTPVFTL SPIIINTAPCLFSDGSTSQAAYILWDKHILSQRSFPLPPPHKSAQQAELLGLLHGLSSARSWHCL NIFLDSKYLYHYLRTLALGTFQGKSSQAPFQALLPRLLAHKVIYLHHVRSHTNLPDPISKLNAL TDALLITPIL
HTL1L_P0C211_2mut	GLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHD LRATNSLTVDLSSSSPGPPDLSSLPTTLAHLQTIDLKDAFFQIPLPKQFQPYFAFTVPQQCNYGP GTRYAWKVLPQGFKNSPTLFQMQLASILQPIRQAFPQCVILQYMDDILLASPSPEDLQQLSEA TMASLISHGLPVSQDKTQQTPGTIKFLGQIISPNHITYDAVPTVPIRSRWALPELQALLGEIQWV SKGTPTLRQPLHSLYCALQGHTDPRDQIYLNPSQVQSLMQLQQALSQNCRSRLAQTLPLLGAI MLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTI HHNISIQTFNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALTPVFTL SPIIINTAPCLFSDGSTSQAAYILWDKHILSQRSFPLPPPHKSAQQAELLGLLHGLSSARSWHCL NIFLDSKYLYHYLRTLAWGTFQGKSSQAPFQALLPRLLAHKVIYLHHVRSHTNLPDPISKLNA LTDALLITPIL
HTL1L_P0C211_2mutB	GLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHD LRATNSLTVDLSSSSPGPPDLSSPPTTLAHLQTIDLKDAFFQIPLPKQFQPYFAFTVPQQCNYGP GTRYAWKVLPQGFKNSPTLFQMQLASILQPIRQAFPQCVILQYMDDILLASPSPEDLQQLSEA TMASLISHGLPVSQDKTQQTPGTIKFLGQIISPNHITYDAVPTVPIRSRWALPELQALLGEIQWV SKGTPTLRQPLHSLYCALQGHTDPRDQIYLNPSQVQSLMQLQQALSQNCRSRLAQTLPLLGAI MLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLLCQTI HHNISIQTFNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALTPVFTL SPIIINTAPCLFSDGSTSQAAYILWDKHILSQRSFPLPPPHKSAQQAELLGLLHGLSSARSWHCL NIFLDSKYLYHYLRTLAWGTFQGKSSQAPFQALLPRLLAHKVIYLHHVRSHTNLPDPISKLNA LTDALLITPIL
HTL32_Q0R5R2	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDL RATNSVTRDLASPSPGPPDLTSLPQGLPHLRTIDLTDAFFQIPLPTIFQPYFAFTLPQPNNYGPGT RYSWRVLPQGFKNSPTLFEQQLSHILTPVRKTFPNSLIIQYMDDILLASPAPGELAALTDKVTN ALTKEGLPLSPEKTQATPGPIHFLGQVISQDCITYETLPSINVKSTWSLAELQSMLGELQWVSK GTPVLRSSLHQLYLALRGHRDPRDTIKLTSIQVQALRTIQKALTLNCRSRLVNQLPILALIMLR PTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAVIILDKYSLQHYGQVCKSFHHNI SNQALTYYLHTSDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQMPQIFQNIDVLRPPFTISPVV INHAPCLFSDGSASKAAFIIWDRQVIHQQVLSLPSTCSAQAGELFGLLAGLQKSQPWVALNIFL DSKFLIGHLRRMALGAFPGPSTQCELHTQLLPLLQGKTVYVHHVRSHTLLQDPISRLNEATDA LMLAPLLPL

HTL32_Q0R5R2_2mut	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDL RATNSVTRDLASPSPGPPDLTSLPQGLPHLRTIDLTDAFFQIPLPTIFQPYFAFTLPQPNNYGPGT RYSWRVLPQGFKNSPTLFQQQLSHILTPVRKTFPNSLIIQYMDDILLASPAPGELAALTDKVTN ALTKEGLPLSPEKTQATPGPIHFLGQVISQDCITYETLPSINVKSTWSLAELQSMLGELQWVSK GTPVLRSSLHQLYLALRGHRDPRDTIKLTSIQVQALRTIQKALTLNCRSRLVNQLPILALIMLR PTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAVIILDKYSLQHYGQVCKSFHHNI SNQALTYYLHTSDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQMPQIFQNIDVLRPPFTISPVV INHAPCLFSDGSASKAAFIIWDRQVIHQQVLSLPSTCSAQAGELFGLLAGLQKSQPWVALNIFL DSKFLIGHLRRMAWGAFPGPSTQCELHTQLLPLLQGKTVYVHHVRSHTLLQDPISRLNEATD ALMLAPLLPL
HTL32_Q0R5R2_2mutB	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDL RATNSVTRDLASPSPGPPDLTSPPQGLPHLRTIDLTDAFFQIPLPTIFQPYFAFTLPQPNNYGPGT RYSWRVLPQGFKNSPTLFQQQLSHILTPVRKTFPNSLIIQYMDDILLASPAPGELAALTDKVTN ALTKEGLPLSPEKTQATPGPIHFLGQVISQDCITYETLPSINVKSTWSLAELQSMLGELQWVSK GTPVLRSSLHQLYLALRGHRDPRDTIKLTSIQVQALRTIQKALTLNCRSRLVNQLPILALIMLR PTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAVIILDKYSLQHYGQVCKSFHHNI SNQALTYYLHTSDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQMPQIFQNIDVLRPPFTISPVV INHAPCLFSDGASKAAFIIWDRQVIHQQVLSLPSTCSAQAGELFGLLAGLQKSQPWVALNIFL DSKFLIGHLRRMAWGAFPGPSTQCELHTQLLPLLQGKTVYVHHVRSHTLLQDPISRLNEATD ALMLAPLLPL
HTL3P_Q4U0X6	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDL RATNSLTRDLASPSPGPPDLTSLPQDLPHLRTIDLTDAFFQIPLPAVFQPYFAFTLPQPNNHGPG TRYSWRVLPQGFKNSPTLFEQQLSHILAPVRKAFPNSLIIQYMDDILLASPALRELTALTDKVT NALTKEGLPMSLEKTQATPGSIHFLGQVISPDCITYETLPSIHVKSIWSLAELQSMLGELQWVS KGTPVLRSSLHQLYLALRGHRDPRDTIELTSTQVQALKTIQKALALNCRSRLVSQLPILALIILR PTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAIITLDKYSLQHYGQICKSFHHNIS NQALTYYLHTSDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQVPQIFQNIDVLRPPFIISPVVID HAPCLFSDGATSKAAFILWDKQVIHQQVLPLPSTCSAQAGELFGLLAGLQKSKPWPALNIFLD SKFLIGHLRRMALGAFLGPSTQCDLHARLFPLLQGKTVYVHHVRSHTLLQDPISRLNEATDAL MLAPLLPL
HTL3P_Q4U0X6_2mut	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDL RATNSLTRDLASPSPGPPDLTSLPQDLPHLRTIDLTDAFFQIPLPAVFQPYFAFTLPQPNNHGPG TRYSWRVLPQGFKNSPTLFQQQLSHILAPVRKAFPNSLIIQYMDDILLASPALRELTALTDKVT NALTKEGLPMSLEKTQATPGSIHFLGQVISPDCITYETLPSIHVKSIWSLAELQSMLGELQWVS KGTPVLRSSLHQLYLALRGHRDPRDTIELTSTQVQALKTIQKALALNCRSRLVSQLPILALIILR PTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAIITLDKYSLQHYGQICKSFHHNIS NQALTYYLHTSDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQVPQIFQNIDVLRPPFIISPVVID HAPCLFSDGATSKAAFILWDKQVIHQQVLPLPSTCSAQAGELFGLLAGLQKSKPWPALNIFLD SKFLIGHLRRMAWGAFLGPSTQCDLHARLFPLLQGKTVYVHHVRSHTLLQDPISRLNEATDA LMLAPLLPL
HTL3P_Q4U0X6_2mutB	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDL RATNSLTRDLASPSPGPPDLTSPPQDLPHLRTIDLTDAFFQIPLPAVFQPYFAFTLPQPNNHGPG TRYSWRVLPQGFKNSPTLFQQQLSHILAPVRKAFPNSLIIQYMDDILLASPALRELTALTDKVT NALTKEGLPMSLEKTQATPGSIHFLGQVISPDCITYETLPSIHVKSIWSLAELQSMLGELQWVS KGTPVLRSSLHQLYLALRGHRDPRDTIELTSTQVQALKTIQKALALNCRSRLVSQLPILALIILR PTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAIITLDKYSLQHYGQICKSFHHNIS NQALTYYLHTSDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQVPQIFQNIDVLRPPFIISPVVID HAPCLFSDGATSKAAFILWDKQVIHQQVLPLPSTCSAQAGELFGLLAGLQKSKPWPALNIFLD SKFLIGHLRRMAWGAFLGPSTQCDLHARLFPLLQGKTVYVHHVRSHTLLQDPISRLNEATDA LMLAPLLPL
HTLV2_P03363_2mut	HLPPPPQVDQFPLNLPERLQALNDLVSKALEAGHIEPYSGPGNNPVFPVKKPNGKWRFIHDLR ATNAITTTLTSPSPGPPDLTSLPTALPHLQTIDLTDAFFQIPLPKQYQPYFAFTIPQPCNYGPGTR YAWTVLPQGFKNSPTLFQQQLAAVLNPMRKMFPTSTIVQYMDDILLASPTNEELQQLSQLTL QALTTHGLPISQEKTQQTPGQIRFLGQVISPNHITYESTPTIPIKSQWTLTELQVILGEIQWVSKG TPILRKHLQSLYSALHPYRDPRACITLTPQQLHALHAIQQALQHNCRGRLNPALPLLGLISLST SGTTSVIFQPKQNWPLAWLHTPHPPTSLCPWGHLLACTILTLDKYTLQHYGQLCQSFHHNMS

	KQALCDFLRNSPHPSVGILIHHMGRFHNLGSQPSGPWKTLLHLPTLLQEPRLLRPIFTLSPVVL DTAPCLFSDGSPQKAAYVLWDQTILQQDITPLPSHETHSAQKGELLALICGLRAAKPWPSLNIF LDSKYLIKYLHSLAIGAFLGTSAHQTLQAALPPLLQGKTIYLHHVRSHTNLPDPISTFNEYTDS LILAPLVPL
JSRV_P31623	PLGTSDSPVTHADPIDWKSEEPVWVDQWPLTQEKLSAAQQLVQEQLRLGHIEPSTSAWNSPIF VIKKKSGKWRLLQDLRKVNETMMHMGALQPGLPTPSAIPDKSYIIVIDLKDCFYTIPLAPQDC KRFAFSLPSVNFKEPMQRYQWRVLPQGMTNSPTLCQKFVATAIAPVRQRFPQLYLVHYMDDI LLAHTDEHLLYQAFSILKQHLSLNGLVIADEKIQTHFPYNYLGFSLYPRVYNTQLVKLQTDHL KTLNDFQKLLGDINWIRPYLKLPTYTLQPLFDILKGDSDPASPRTLSLEGRTALQSIEEAIRQQQ ITYCDYQRSWGLYILPTPRAPTGVLYQDKPLRWIYLSATPTKHLLPYYELVAKIIAKGRHEAIQ YFGMEPPFICVPYALEQQDWLFQFSDNWSIAFANYPGQITHHYPSDKLLQFASSHAFIFPKIVR RQPIPEATLIFTDGSSNGTAALIINHQTYYAQTSFSSAQVVELFAVHQALLTVPTSFNLFTDSSY VVGALQMIETVPIIGTTSPEVLNLFTLIQQVLHCRQHPCFFGHIRAHSTLPGALVQGNHTADVL TKQVFFQS
JSRV_P31623_2mutB	PLGTSDSPVTHADPIDWKSEEPVWVDQWPLTQEKLSAAQQLVQEQLRLGHIEPSTSAWNSPIF VIKKKSGKWRLLQDLRKVNETMMHMGALQPGLPTPSPIPDKSYIIVIDLKDCFYTIPLAPQDC KRFAFSLPSVNFKEPMQRYQWRVLPQGMTNSPTLCQKFVATAIAPVRQRFPQLYLVHYMDDI LLAHTDEHLLYQAFSILKQHLSLNGLVIADEKIQTHFPYNYLGFSLYPRVYNTQLVKLQTDHL KTLNDFQKLLGDINWIRPYLKLPTYTLQPLFDILKGDSDPASPRTLSLEGRTALQSIEEAIRQQQ ITYCDYQRSWGLYILPTPRAPTGVLYQDKPLRWIYLSATPTKHLLPYYELVAKIIAKGRHEAIQ YFGMEPPFICVPYALEQQDWLFQFSDNWSIAFANYPGQITHHYPSDKLLQFASSHAFIFPKIVR RQPIPEATLIFTDGSSNGTAALIINHQTYYAQTSFSSAQVVELFAVHQALLTVPTSFNLFTDSSY VVGALQMIETVPIIGTTSPEVLNLFTLIQQVLHCRQHPCFFGHIRAHSTLPGALVQGNHTADVL TKQVFFQS
KORV_Q9TTC1	TLGDQGSRGSDPLPEPRVTLTVEGIPTEFLVNTGAEHSVLTKPMGKMGSKRTVVAGATGSKV YPWTTKRLLKIGQKQVTHSFLVIPECPAPLLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLV LNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQY PMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHP TVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLP QGFKNSPTLFDEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKL GYRVSAKKAQLCREEVTYLGYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGTAGFCR LWIPGFASLAAPLYPLTREKVPFTWTEAHQEAFGRIKEALLSAPALALPDLTKPFALYVDEKE GVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVI APHNLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEIL AEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKA ELIALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGKDIKNKEEILALLEAIH LPKRVAIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKN
KORV_Q9TTC1_3mut	TLGDQGSRGSDPLPEPRVTLTVEGIPTEFLVNTGAEHSVLTKPMGKMGSKRTVVAGATGSKV YPWTTKRLLKIGQKQVTHSFLVIPECPAPLLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLV LNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQY PMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHP TVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLP QGFKNSPTLFNEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKL GYRVSAKKAQLCREEVTYLGYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGTAGFCR LWIPGFASLAAPLYPLTRPKVPFTWTEAHQEAFGRIKEALLSAPALALPDLTKPFALYVDEKE GVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVI APHNLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEIL AEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKA ELIALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGKDIKNKEEILALLEAIH LPKRVAIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKN
KORV_Q9TTC1_3mutA	TLGDQGSRGSDPLPEPRVTLTVEGIPTEFLVNTGAEHSVLTKPMGKMGSKRTVVAGATGSKV YPWTTKRLLKIGQKQVTHSFLVIPECPAPLLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLV LNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQY PMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHP TVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLP QGFKNSPTLFNEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKL

	GYRVSAKKAQLCREEVTYLGYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGKAGFCR LFIPGFASLAAPLYPLTRPKVPFTWTEAHQEAFGRIKEALLSAPALALPDLTKPFALYVDEKEG VARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIA PHNLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAE ETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELI ALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLP KRVAIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKN
KORV_ Q9TTC 1-Pro	LLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMV WAEKAGMGLANQVPPVVVELKSDASPVAVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFF CLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLASFRALNPQVV MLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKR WLTPARKATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTREKVPFTWTEAHQ EAFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVA SGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNARMTHYQSLL LNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIM DGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALRLAEGKSINIYTDSRYAFATAH VHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQRGTDPVATGNRKADEA AKQAAQSTRILTETTKN
KORV_ Q9TTC 1- Pro_3m ut	LLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMV WAEKAGMGLANQVPPVVVELKSDASPVAVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFF CLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLASFRALNPQVV MLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKR WLTPARKATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTRPKVPFTWTEAHQ EAFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVA SGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNARMTHYQSLL LNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIM DGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALRLAEGKSINIYTDSRYAFATAH VHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQRGTDPVATGNRKADEA AKQAAQSTRILTETTKN
KORV_ Q9TTC 1- Pro_3m utA	LLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMV WAEKAGMGLANQVPPVVVELKSDASPVAVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFF CLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLASFRALNPQVV MLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKR WLTPARKATVMKIPTPTTPRQVREFLGKAGFCRLFIPGFASLAAPLYPLTRPKVPFTWTEAHQ EAFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVA SGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNARMTHYQSLL LNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIM DGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALRLAEGKSINIYTDSRYAFATAH VHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQRGTDPVATGNRKADEA AKQAAQSTRILTETTKN
MLVA V_P033 56	TLNLEDEYRLYETSAEPEVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEAKLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHRWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLP QGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLLTLGNLGY RASAKKAQLCQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRL WIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQ GYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLRKDAGKLTMGQPLVI LAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEEGAPHDC LEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTS AQRAELIALTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGREIKNKSEILAL LKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVA V_P033	TLNLEDEYRLYETSAEPEVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEAKLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH

56_3mut	PTVPNPYNLLSGLPPSHRWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLP QGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLLTLGNLGY RASAKKAQLCQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRL WIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQ GYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLRKDAGKLTMGQPLVI LAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEEGAPHDC LEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTS AQRAELIALTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGREIKNKSEILAL LKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVAV_P03356_3mutA	TLNLEDEYRLYETSAEPEVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEAKLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHRWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLP QGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLLTLGNLGY RASAKKAQLCQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLF IPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQG YAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLRKDAGKLTMGQPLVIL APHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEEGAPHDCL EILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSA QRAELIALTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGREIKNKSEILALL KALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVBM_Q7SVK7	TLGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQY PMSHEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHP TVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLP QGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGY RASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMGQPVPKTPRQLREFLGTAGFCRLW IPGFAEMAAPLYPLTKTGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEEGAPHDCLE ILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQ RAELIALTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGREIKNKSEILALLK ALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVBM_Q7SVK7	TLGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQY PMSHEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHP TVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLP QGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGY RASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMGQPVPKTPRQLREFLGTAGFCRLW IPGFAEMAAPLYPLTKTGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEEGAPHDCLE ILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQ RAELIALTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGREIKNKSEILALLK ALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVBM_Q7SVK7_3mut	TLGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQY PMSHEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHP TVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLP QGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGY RASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMGQPVPKTPRQLREFLGTAGFCRLW IPGFAEMAAPLYPLTKPGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEEGAPHDCLE ILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQ RAELIALTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGREIKNKSEILALLK ALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVBM_Q7S	TLGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQY PMSHEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHP TVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLP

VK7_3 mut	QGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGY RASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMGQPVPKTPRQLREFLGTAGFCRLW IPGFAEMAAPLYPLTKPGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEEGAPHDCLE ILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQ RAELIALTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGREIKNKSEILALLK ALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVB M_Q7S VK7_3 mutA_ WS	LGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQYP MSHEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQ GFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYR ASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMGQPVPKTPRQLREFLGKAGFCRLFIP GFAEMAAPLYPLTKPGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYA KGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAP HAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEEGAPHDCLEI LAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQR AELIALTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGREIKNKSEILALLKA LFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLLI
MLVB M_Q7S VK7_3 mutA_ WS	LGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQYP MSHEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLPQ GFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYR ASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMGQPVPKTPRQLREFLGKAGFCRLFIP GFAEMAAPLYPLTKPGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYA KGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAP HAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEEGAPHDCLEI LAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQR AELIALTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGREIKNKSEILALLKA LFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLLI
MLVC B_P083 61	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLFDEALHRDLAGFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPIPKTPRQLREFLGTAGFCRLWI PGFAEMAAPLYPLTKTGTLFNWGPDQQKAFQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHDCLD ILAEAHGTRSDLMDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSA QRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAAREVATRETPETSTLL
MLVC B_P083 61_3mu t	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLFNEALHRDLAGFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPIPKTPRQLREFLGTAGFCRLWI PGFAEMAAPLYPLTKPGTLFNWGPDQQKAFQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHDCLD ILAEAHGTRSDLMDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSA QRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILAL LKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAAREVATRETPETSTLL
MLVC B_P083 61_3mu tA	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLFNEALHRDLAGFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGY

	RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPIPKTPRQLREFLGKAGFCRLFI PGFAEMAAPLYPLTKPGTLFNWGPDQQKAFQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHDCLD ILAEAHGTRSDLMDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSA QRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILAL LKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAAREVATRETPETSTLL
MLVF5_P26810	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIISLKATSTPVSIKQY PMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHP TVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQSLFAFEWKDPEMGISGQLTWTRLP QGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGLCRLW IPGFAEMAAPLYPLTKTGTLFKWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQG YAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDVGKLTMGQPLVIL APHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPIVALNPATLLPLPEEGLQHDCL DILAEAHGTRPDLTDQPLPDADHTWYTDGSSFLQEGQRRAGAAVTTETEVIWAKALPAGTSA QRAELIALTQALKMAAGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALL KALFLPKRLSIIHCPGHQKGNHAEARGNRMADQAAREVATRETPETSTLL
MLVF5_P26810_3mut	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIISLKATSTPVSIKQY PMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHP TVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQSLFAFEWKDPEMGISGQLTWTRLP QGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGLCRLW IPGFAEMAAPLYPLTKPGTLFKWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQG YAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDVGKLTMGQPLVIL APHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPIVALNPATLLPLPEEGLQHDCL DILAEAHGTRPDLTDQPLPDADHTWYTDGSSFLQEGQRRAGAAVTTETEVIWAKALPAGTSA QRAELIALTQALKMAAGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILAL LKALFLPKRLSIIHCPGHQKGNHAEARGNRMADQAAREVATRETPETSTLL
MLVF5_P26810_3mutA	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIISLKATSTPVSIKQY PMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHP TVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQSLFAFEWKDPEMGISGQLTWTRLP QGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGLCRLF IPGFAEMAAPLYPLTKPGTLFKWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQG YAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDVGKLTMGQPLVIL APHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPIVALNPATLLPLPEEGLQHDCL DILAEAHGTRPDLTDQPLPDADHTWYTDGSSFLQEGQRRAGAAVTTETEVIWAKALPAGTSA QRAELIALTQALKMAAGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILAL LKALFLPKRLSIIHCPGHQKGNHAEARGNRMADQAAREVATRETPETSTLL
MLVFF_P26809_3mut	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQSLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLW IPGFAEMAAPLYPLTKPGTLFEWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPIVALNPATLLPLPEEGLQHDCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVVWAKALPAGTSA QRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILAL LKALFLPKRLSIIHCPGHQKGNRAEARGNRMADQAAREVATRETPETSTLL
MLVFF_P26809_3mutA	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQSLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFI

	PGFAEMAAPLYPLTKPGTLFEWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPIVALNPATLLPLPEEGLQHDCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVVWAKALPAGTSA QRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILAL LKALFLPKRLSIIHCPGHQKGNRAEARGNRMADQAARKVATRETPETSTLL
MLVM S_P033 55	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLW IPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQG YAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVIL APHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCL DILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTS AQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILAL LKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL
MLVM S_P033 55	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLW IPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQG YAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVIL APHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCL DILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTS AQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILAL LKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL
MLVM S_P033 55_3mu t	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLW IPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQG YAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVIL APHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCL DILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTS AQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILA LLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL
MLVM S_P033 55_3mu t	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLW IPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQG YAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVIL APHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCL DILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTS AQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILA LLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL
MLVM S_P033 55_3mu tA_WS	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFI PGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY

	AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLD ILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSA QRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILAL LKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL
MLVM S_P033 55_3mu tA_WS	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFI PGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLD ILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSA QRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILAL LKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL
MLVM S_P033 55_PLV 919	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFI PGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLD ILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSA QRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILAL LKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRT ADGSEFE
MLVM S_P033 55_PLV 919	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFI PGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLD ILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSA QRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILAL LKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRT ADGSEFE
MLVR D_P112 27	TLNIEDEYRLHEISTEPDVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQY PMSQEAKLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQGLREVNKRVEDIHP TVPNPYNLLSGLPTSHRWYTVLDLKDAFFCLRLHPTSQPLFASEWRDPGMGISGQLTWTRLP QGFKNSPTLFDEALHRGLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLKTLGNLGY RASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLW IPRFAEMAAPLYPLTKTGTLFNWGPDQQKAYHEIKQALLTAPALGLPDLTKPFELFVDEKQG YAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVIL APHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEEGAPHDCL EILAETHGTEPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSA QRAELIALTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYKRRGLLTSEGREIKNKSEILALL KALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MLVR D_P112 27_3mu t	TLNIEDEYRLHEISTEPDVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQY PMSQEAKLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQGLREVNKRVEDIHP TVPNPYNLLSGLPTSHRWYTVLDLKDAFFCLRLHPTSQPLFASEWRDPGMGISGQLTWTRLP QGFKNSPTLFNEALHRGLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLKTLGNLGY RASAKKAQICQKQVKYLGYLLREGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLW

	IPRFAEMAAPLYPLTKPGTLFNWGPDQQKAYHEIKQALLTAPALGLPDLTKPFELFVDEKQG YAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVIL APHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEEGAPHDCL EILAETHGTEPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSA QRAELIALTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYKRRGWLTSEGREIKNKSEILALL KALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREAAIKTPPDTSTLL
MMTV B_P03365	WVQEISDSRPMLHIYLNGRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGV ARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLF ADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGK WRLLQDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVP SPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRS IVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQK LLGNINWIRPFLKLTTGELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLS QPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDY IVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVI FTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIE TATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTV B_P03365	WVQEISDSRPMLHIYLNGRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGV ARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLF ADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGK WRLLQDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVP SPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRS IVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQK LLGNINWIRPFLKLTTGELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLS QPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDY IVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVI FTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIE TATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTV B_P03365_2mut	WVQEISDSRPMLHIYLNGRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGV ARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLF ADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGK WRLLQDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVP SPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRS IVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQK LLGNINWIRPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLS QPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDY IVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVI FTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIE TATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTV B_P03365_2mut_WS	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVAR SSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFAD QISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRL LQDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNF KRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDE ILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGN INWIRPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWS LCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVP YTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTD GSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETAT LSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MMTV B_P03365_2mut_WS	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVAR SSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFAD QISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRL LQDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNF KRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDE ILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGN

	INWIRPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWS LCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVP YTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTD GSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETAT LSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MMTV B_P033 65_2mu tB	WVQEISDSRPMLHIYLNGRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGV ARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLF ADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGK WRLLQDLRAVNATMHDMGALQPGLPSPVAPPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVP SPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRS IVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQK LLGNINWIRPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLS QPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDY IVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVI FTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIE TATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTV B_P033 65_2mu tB	WVQEISDSRPMLHIYLNGRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGV ARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLF ADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGK WRLLQDLRAVNATMHDMGALQPGLPSPVAPPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVP SPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRS IVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQK LLGNINWIRPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLS QPWSLCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDY IVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVI FTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIE TATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTV B_P033 65_2mu tB_WS	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVAR SSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFAD QISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRL LQDLRAVNATMHDMGALQPGLPSPPAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNF KRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDE ILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGN INWIRPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWS LCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVP YTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTD GSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETAT LSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MMTV B_P033 65_WS	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVAR SSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFAD QISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRL LQDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNF KRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDE ILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGN INWIRPFLKLTTGELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWS

	LCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVP YTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTD GSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETAT LSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MMTV B_P03365_WS	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVAR SSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFAD QISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRL LQDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNF KRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDE ILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGN INWIRPFLKLTTGELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWS LCILKTEYTPTACLWQDGVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVP YTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTD GSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETAT LSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MMTV B_P03365-Pro	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQ QLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLRAVNATMHDMGALQPGLPSPVA VPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKF VDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNL KYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTGELKPLFEILNGDSNP ISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLP HISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGE VHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQ QAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYI GHIRGHTGLPGPLAQGNAYADSLTRILT
MMTV B_P03365-Pro	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQ QLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLRAVNATMHDMGALQPGLPSPVA VPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKF VDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNL KYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTGELKPLFEILNGDSNP ISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLP HISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGE VHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQ QAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYI GHIRGHTGLPGPLAQGNAYADSLTRILT
MMTV B_P03365-Pro_2mut	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQ QLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLRAVNATMHDMGALQPGLPSPVA VPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKF VDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNL KYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTGELKPLFEILNPDSNPI STRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPH ISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEV HFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQ AEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIG HIRGHTGLPGPLAQGNAYADSLTRILT
MMTV B_P03365-Pro_2mut	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQ QLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLRAVNATMHDMGALQPGLPSPVA VPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKF VDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNL KYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTGELKPLFEILNPDSNPI STRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPH ISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEV HFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQ AEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIG HIRGHTGLPGPLAQGNAYADSLTRILT

MMTV B_P03365-Pro_2mutB	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQ QLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLRAVNATMHDMGALQPGLPSPVA PPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKF VDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNL KYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTGELKPLFEILNPDSNPI STRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPH ISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEV HFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQ AEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIG HIRGHTGLPGPLAQGNAYADSLTRILT
MMTV B_P03365-Pro_2mutB	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQ QLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLRAVNATMHDMGALQPGLPSPVA PPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKF VDKAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNL KYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTGELKPLFEILNPDSNPI STRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPH ISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEV HFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQ AEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIG HIRGHTGLPGPLAQGNAYADSLTRILT
MPMV_P07572	LTAAIDILAPQQCAEPITWKSDEPVWVDQWPLTNDKLAAAQQLVQEQLEAGHITESSSPWNT PIFVIKKKSGKWRLLQDLRAVNATMVLMGALQPGLPSPVAIPQGYLKIIIDLKDCFFSIPLHPS DQKRFAFSLPSTNFKEPMQRFQWKVLPQGMANSPTLCQKYVATAIHKVRHAWKQMYIIHY MDDILIAGKDGQQVLQCFDQLKQELTAAGLHIAPEKVQLQDPYTYLGFELNGPKITNQKAVI RKDKLQTLNDFQKLLGDINWLRPYLKLTTGDLKPLFDTLKGDSDPNSHRSLSKEALASLEKV ETAIAEQFVTHINYSLPLIFLIFNTALTPTGLFWQDNPIMWIHLPASPKKVLLPYYDAIADLIILG RDHSKKYFGIEPSTIIQPYSKSQIDWLMQNTEMWPIACASFVGILDNHYPPNKLIQFCKLHTFV FPQIISKTPLNNALLVFTDGSSTGMAAYTLTDTTIKFQTNLNSAQLVELQALIAVLSAFPNQPL NIYTDSAYLAHSIPLLETVAQIKHISETAKLFLQCQQLIYNRSIPFYIGHVRAHSGLPGPIAQGN QRADLATKIVASNINT
MPMV_P07572_2mutB	LTAAIDILAPQQCAEPITWKSDEPVWVDQWPLTNDKLAAAQQLVQEQLEAGHITESSSPWNT PIFVIKKKSGKWRLLQDLRAVNATMVLMGALQPGLPSPVAPPQGYLKIIIDLKDCFFSIPLHPS DQKRFAFSLPSTNFKEPMQRFQWKVLPQGMANSPTLCQKYVATAIHKVRHAWKQMYIIHY MDDILIAGKDGQQVLQCFDQLKQELTAAGLHIAPEKVQLQDPYTYLGFELNGPKITNQKAVI RKDKLQTLNDFQKLLGDINWLRPYLKLTTGDLKPLFDTLKPDSDPNSHRSLSKEALASLEKVE TAIAEQFVTHINYSLPLIFLIFNTALTPTGLFWQDNPIMWIHLPASPKKVLLPYYDAIADLIILGR DHSKKYFGIEPSTIIQPYSKSQIDWLMQNTEMWPIACASFVGILDNHYPPNKLIQFCKLHTFVF PQIISKTPLNNALLVFTDGSSTGMAAYTLTDTTIKFQTNLNSAQLVELQALIAVLSAFPNQPLNI YTDSAYLAHSIPLLETVAQIKHISETAKLFLQCQQLIYNRSIPFYIGHVRAHSGLPGPIAQGNQR ADLATKIVASNINT
PERV_Q4VFZ2	TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVR QYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDI HPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTR LPQGFKNSPTIFDEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDL GYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRL WIPGFATLAAPLYPLTKEKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERK GVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIA PHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQL LIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQK AELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSLLEA LHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLL
PERV_Q4VFZ2	TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVR QYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDI HPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTR LPQGFKNSPTIFDEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDL GYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRL

	WIPGFATLAAPLYPLTKEKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERK GVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIA PHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQL LIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQK AELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSLLEA LHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLL
PERV_ Q4VFZ 2_3mut	TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVR QYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDI HPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTR LPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDL GYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRL WIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKG VARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAP HALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLL IEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKA ELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEAL HLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLL
PERV_ Q4VFZ 2_3mut	TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVR QYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDI HPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTR LPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDL GYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRL WIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKG VARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAP HALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLL IEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKA ELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEAL HLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLL
PERV_ Q4VFZ 2_3mut A_WS	LDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYP LSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPT VPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQ GFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYR ASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRLFIP GFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVA RGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHA LENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEE TGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAEL MALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHL PKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLP
PERV_ Q4VFZ 2_3mut A_WS	LDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYP LSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPT VPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQ GFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYR ASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRLFIP GFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVA RGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHA LENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEE TGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAEL MALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHL PKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLP
SFV1_P 23074	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPEAFLEDERPIQTMLIKTIHGEKQQDVYYLT FKVQGRKVEAEVLASPYDYILLNPSDVPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKE LLQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLKQ GVLIQQNSTMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTT LDLTNGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPALFTADVVDLLKEIPNVQA YVDDIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQ

	KLLNITPPKDLKQLQSILGLLNFARNFIPNYSELVKPLYTIVANANGKFISWTEDNSNQLQHIIS VLNQADNLEERNPETRLIIKVNSSPSAGYIRYYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLTT MHKGLIKAMDLAMGQEILVYSPIVSMTKIQRTPLPERKALPVRWITWMTYLEDPRIQFHYDK SLPELQQIPNVTEDVIAKTKHPSEFAMVFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVH QWSIPLGDHTAQLAEIAAVEFACKKALKISGPVLIVTDSFYVAESANKELPYWKSNGFLNNK KKPLRHVSKWKSIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGSYVVH
SFV1_P23074_2mut	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPEAFLEDERPIQTMLIKTIHGEKQQDVYYLT FKVQGRKVEAEVLASPYDYILLNPSDVPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKE LLQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLKQ GVLIQQNSTMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTT LDLTNGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPALFNADVVDLLKEIPNVQA YVDDIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQ KLLNITPPKDLKQLQSILGLLNFARNFIPNYSELVKPLYTIVAPANGKFISWTEDNSNQLQHIIS VLNQADNLEERNPETRLIIKVNSSPSAGYIRYYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLTT MHKGLIKAMDLAMGQEILVYSPIVSMTKIQRTPLPERKALPVRWITWMTYLEDPRIQFHYDK SLPELQQIPNVTEDVIAKTKHPSEFAMVFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVH QWSIPLGDHTAQLAEIAAVEFACKKALKISGPVLIVTDSFYVAESANKELPYWKSNGFLNNK KKPLRHVSKWKSIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGSYVVH
SFV1_P23074_2mutA	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPEAFLEDERPIQTMLIKTIHGEKQQDVYYLT FKVQGRKVEAEVLASPYDYILLNPSDVPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKE LLQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLKQ GVLIQQNSTMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTT LDLTNGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPALFNADVVDLLKEIPNVQA YVDDIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQ KLLNITPPKDLKQLQSILGKLNFARNFIPNYSELVKPLYTIVAPANGKFISWTEDNSNQLQHIIS VLNQADNLEERNPETRLIIKVNSSPSAGYIRYYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLTT MHKGLIKAMDLAMGQEILVYSPIVSMTKIQRTPLPERKALPVRWITWMTYLEDPRIQFHYDK SLPELQQIPNVTEDVIAKTKHPSEFAMVFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVH QWSIPLGDHTAQLAEIAAVEFACKKALKISGPVLIVTDSFYVAESANKELPYWKSNGFLNNK KKPLRHVSKWKSIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGSYVVH
SFV1_P23074-Pro	VPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRR IKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLKQGVLIQQNSTMNTPVYPVPKPDGKWR MVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQG KQYCWTRLPQGFLNSPALFTADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNA GYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLNITPPKDLKQLQSILGLLNFARNFIP NYSELVKPLYTIVANANGKFISWTEDNSNQLQHIISVLNQADNLEERNPETRLIIKVNSSPSAG YIRYYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLTTMHKGLIKAMDLAMGQEILVYSPIVSMT KIQRTPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSEFAMV FYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFACKKAL KISGPVLIVTDSFYVAESANKELPYWKSNGFLNNKKKPLRHVSKWKSIAECLQLKPDIIIMHE KGHQQPMTTLHTEGNNLADKLATQGSYVVH
SFV1_P23074-Pro_2mut	VPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRR IKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLKQGVLIQQNSTMNTPVYPVPKPDGKWR MVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQG KQYCWTRLPQGFLNSPALFNADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLN AGYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLNITPPKDLKQLQSILGLLNFARNF IPNYSELVKPLYTIVAPANGKFISWTEDNSNQLQHIISVLNQADNLEERNPETRLIIKVNSSPSA GYIRYYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLTTMHKGLIKAMDLAMGQEILVYSPIVSM TKIQRTPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSEFAM VFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFACKKA LKISGPVLIVTDSFYVAESANKELPYWKSNGFLNNKKKPLRHVSKWKSIAECLQLKPDIIIMHE KGHQQPMTTLHTEGNNLADKLATQGSYVVH
SFV1_P23074-Pro_2mutA	VPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRR IKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLKQGVLIQQNSTMNTPVYPVPKPDGKWR MVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQG KQYCWTRLPQGFLNSPALFNADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLN

	AGYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLNITPPKDLKQLQSILGKLNFARNF IPNYSELVKPLYTIVAPANGKFISWTEDNSNQLQHIISVLNQADNLEERNPETRLIIKVNSSPSA GYIRYYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLTTMHKGLIKAMDLAMGQEILVYSPIVSM TKIQRTPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSEFAM VFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFACKKA LKISGPVLIVTDSFYVAESANKELPYWKSNGFLNNKKKPLRHVSKWKSIAECLQLKPDIIIMHE KGHQQPMTTLHTEGNNLADKLATQGSYVVH
SFV3L_ P27401	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPQAFLEEEVPIKNIWIKTIHGEKEQPVYYLTF KIQGRKVEAEVISSPYDYILVSPSDIPWLMKKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKL QSLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQG VLIQQNSIMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLD LSNGFWAHSITPESYWLTAFTWLGQQYCWTRLPQGFLNSPALFTADVVDLLKEVPNVQVYV DDIYISHDDPREHLEQLEKVFSLLLNAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLTETFKQK LLNITPPRDLKQLQSILGLLNFARNFIPNFSELVKPLYNIIATANGKYITWTTDNSQQLQNIISML NSAENLEERNPEVRLIMKVNTSPSAGYIRFYNEFAKRPIMYLNYVYTKAEVKFTNTEKLLTTI HKGLIKALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDPRIQFHYDKTLP ELQQVPTVTDDIIAKIKHPSEFSMVFYTDGSAIKHPNVNKSHNAGMGIAQVQFKPEFTVINTW SIPLGDHTAQLAEVAAVEFACKKALKIDGPVLIVTDSFYVAESVNKELPYWQSNGFFNNKKK PLKHVSKWKSIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFV3L_ P27401 _2mut	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPQAFLEEEVPIKNIWIKTIHGEKEQPVYYLTF KIQGRKVEAEVISSPYDYILVSPSDIPWLMKKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKL QSLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQG VLIQQNSIMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLD LSNGFWAHSITPESYWLTAFTWLGQQYCWTRLPQGFLNSPALFNADVVDLLKEVPNVQVYV DDIYISHDDPREHLEQLEKVFSLLLNAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLTETFKQK LLNITPPRDLKQLQSILGLLNFARNFIPNFSELVKPLYNIIATAPGKYITWTTDNSQQLQNIISML NSAENLEERNPEVRLIMKVNTSPSAGYIRFYNEFAKRPIMYLNYVYTKAEVKFTNTEKLLTTI HKGLIKALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDPRIQFHYDKTLP ELQQVPTVTDDIIAKIKHPSEFSMVFYTDGSAIKHPNVNKSHNAGMGIAQVQFKPEFTVINTW SIPLGDHTAQLAEVAAVEFACKKALKIDGPVLIVTDSFYVAESVNKELPYWQSNGFFNNKKK PLKHVSKWKSIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFV3L_ P27401 _2mutA	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPQAFLEEEVPIKNIWIKTIHGEKEQPVYYLTF KIQGRKVEAEVISSPYDYILVSPSDIPWLMKKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKL QSLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQG VLIQQNSIMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLD LSNGFWAHSITPESYWLTAFTWLGQQYCWTRLPQGFLNSPALFNADVVDLLKEVPNVQVYV DDIYISHDDPREHLEQLEKVFSLLLNAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLTETFKQK LLNITPPRDLKQLQSILGKLNFARNFIPNFSELVKPLYNIIATAPGKYITWTTDNSQQLQNIISML NSAENLEERNPEVRLIMKVNTSPSAGYIRFYNEFAKRPIMYLNYVYTKAEVKFTNTEKLLTTI HKGLIKALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDPRIQFHYDKTLP ELQQVPTVTDDIIAKIKHPSEFSMVFYTDGSAIKHPNVNKSHNAGMGIAQVQFKPEFTVINTW SIPLGDHTAQLAEVAAVEFACKKALKIDGPVLIVTDSFYVAESVNKELPYWQSNGFFNNKKK PLKHVSKWKSIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFV3L_ P27401- Pro	IPWLMKKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRI KPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQGVLIQQNSIMNTPVYPVPKPDGKWR MVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLG QQYCWTRLPQGFLNSPALFTADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLL NAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLTETFKQKLLNITPPRDLKQLQSILGLLNFARN FIPNFSELVKPLYNIIATANGKYITWTTDNSQQLQNIISMLNSAENLEERNPEVRLIMKVNTSPS AGYIRFYNEFAKRPIMYLNYVYTKAEVKFTNTEKLLTTIHKGLIKALDLGMGQEILVYSPIVS MTKIQKTPLPERKALPIRWITWMSYLEDPRIQFHYDKTLPELQQVPTVTDDIIAKIKHPSEFSM VFYTDGSAIKHPNVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFACKK ALKIDGPVLIVTDSFYVAESVNKELPYWQSNGFFNNKKKPLKHVSKWKSIADCIQLKPDIIIIH EKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFV3L_ P27401-	IPWLMKKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRI KPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQGVLIQQNSIMNTPVYPVPKPDGKWR

Pro_2mut	MVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLG QQYCWTRLPQGFLNSPALFNADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLL NAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLTETFKQKLLNITPPRDLKQLQSILGLLNFARN FIPNFSELVKPLYNIIATAPGKYITWTTDNSQQLQNIISMLNSAENLEERNPEVRLIMKVNTSPS AGYIRFYNEFAKRPIMYLNYVVYTKAEVKFTNTEKLLTTIHKGLIKALDLGMGQEILVYSPIVS MTKIQKTPLPERKALPIRWITWMSYLEDPRIQFHYDKTLPELQQVPTVTDDIIAKIKHPSEFSM VFYTDGSAIKHPNVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFACKK ALKIDGPVLIVTDSFYVAESVNKELPYWQSNGFFNNKKKPLKHVSKWKSIADCIQLKPDIIIIH EKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFV3L_P27401-Pro_2mutA	IPWLMKKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRI KPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQGVLIQQNSIMNTPVYPVPKPDGKWR MVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLG QQYCWTRLPQGFLNSPALFNADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLL NAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLTETFKQKLLNITPPRDLKQLQSILGKLNFARN FIPNFSELVKPLYNIIATAPGKYITWTTDNSQQLQNIISMLNSAENLEERNPEVRLIMKVNTSPS AGYIRFYNEFAKRPIMYLNYVVYTKAEVKFTNTEKLLTTIHKGLIKALDLGMGQEILVYSPIVS MTKIQKTPLPERKALPIRWITWMSYLEDPRIQFHYDKTLPELQQVPTVTDDIIAKIKHPSEFSM VFYTDGSAIKHPNVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFACKK ALKIDGPVLIVTDSFYVAESVNKELPYWQSNGFFNNKKKPLKHVSKWKSIADCIQLKPDIIIIH EKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFVCP_Q87040	MNPLQLLQPLPAEVKGTKLLAHWNSGATITCIPESFLEDEQPIKQTLIKTIHGEKQQNVYYLTF KVKGRKVEAEVIASPYEYILLSPTDVPWLTQQPLQLTILVPLQEYQDRILNKTALPEEQKQQL KALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQG VLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTL DLANGFWAHPITPDSYWLTAFTWQGKQYCWTRLPQGFLNSPALFTADAVDLLKEVPNVQV YVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFK TKLLNVTPPKDLKQLQSILGLLNFARNFIPNFAELVQTLYNLIASSKGKYIEWTEDNTKQLNK VIEALNTASNLEERLPDQRLVIKVNTSPSAGYVRYYNESGKKPIMYLNYVFSKAELKFSMLEK LLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFH YDKTLPELKHIPDVYTSSIPPLKHPSQYEGVFCTDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKI LNQWSIPLGHHTAQMAEIAAVEFACKKALKVPGPVLVITDSFYVAESANKELPYWKSNGFVN NKKEPLKHISKWKSIAECLSIKPDITIQHEKGHQPINTSIHTEGNALADKLATQGSYVVN
SFVCP_Q87040_2mut	MNPLQLLQPLPAEVKGTKLLAHWNSGATITCIPESFLEDEQPIKQTLIKTIHGEKQQNVYYLTF KVKGRKVEAEVIASPYEYILLSPTDVPWLTQQPLQLTILVPLQEYQDRILNKTALPEEQKQQL KALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQG VLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTL DLANGFWAHPITPDSYWLTAFTWQGKQYCWTRLPQGFLNSPALFNADAVDLLKEVPNVQV YVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFK TKLLNVTPPKDLKQLQSILGLLNFARNFIPNFAELVQTLYNLIASSPGKYIEWTEDNTKQLNKV IEALNTASNLEERLPDQRLVIKVNTSPSAGYVRYYNESGKKPIMYLNYVFSKAELKFSMLEKL LTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHY DKTLPELKHIPDVYTSSIPPLKHPSQYEGVFCTDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKIL NQWSIPLGHHTAQMAEIAAVEFACKKALKVPGPVLVITDSFYVAESANKELPYWKSNGFVN NKKEPLKHISKWKSIAECLSIKPDITIQHEKGHQPINTSIHTEGNALADKLATQGSYVVN
SFVCP_Q87040_2mutA	MNPLQLLQPLPAEVKGTKLLAHWNSGATITCIPESFLEDEQPIKQTLIKTIHGEKQQNVYYLTF KVKGRKVEAEVIASPYEYILLSPTDVPWLTQQPLQLTILVPLQEYQDRILNKTALPEEQKQQL KALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQG VLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTL DLANGFWAHPITPDSYWLTAFTWQGKQYCWTRLPQGFLNSPALFNADAVDLLKEVPNVQV YVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFK TKLLNVTPPKDLKQLQSILGKLNFARNFIPNFAELVQTLYNLIASSPGKYIEWTEDNTKQLNK VIEALNTASNLEERLPDQRLVIKVNTSPSAGYVRYYNESGKKPIMYLNYVFSKAELKFSMLEK LLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFH YDKTLPELKHIPDVYTSSIPPLKHPSQYEGVFCTDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKI LNQWSIPLGHHTAQMAEIAAVEFACKKALKVPGPVLVITDSFYVAESANKELPYWKSNGFVN NKKEPLKHISKWKSIAECLSIKPDITIQHEKGHQPINTSIHTEGNALADKLATQGSYVVN

SFVCP_Q87040-Pro	VPWLTQQPLQLTILVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKI RPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLTPQNSTMNTPVYPVPKPDGRWR MVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLTAFTWQ GKQYCWTRLPQGFLNSPALFTADAVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQIL LQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFKTKLLNVTPPKDLKQLQSILGLLNFA RNFIPNFAELVQTLYNLIASSKGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNT SPSAGYVRYYNESGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVY SPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSIPPLKHPS QYEGVFCTDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFA CKKALKVPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKEPLKHISKWKSIAECLSIKPDI TIQHEKGHQPINTSIHTEGNALADKLATQGSYVVN
SFVCP_Q87040-Pro_2mut	VPWLTQQPLQLTILVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKI RPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLTPQNSTMNTPVYPVPKPDGRWR MVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLTAFTWQ GKQYCWTRLPQGFLNSPALFNADAVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQIL LQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFKTKLLNVTPPKDLKQLQSILGLLNFA RNFIPNFAELVQTLYNLIASSPGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNT SPSAGYVRYYNESGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVY SPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSIPPLKHPS QYEGVFCTDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFA CKKALKVPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKEPLKHISKWKSIAECLSIKPDI TIQHEKGHQPINTSIHTEGNALADKLATQGSYVVN
SFVCP_Q87040-Pro_2mutA	VPWLTQQPLQLTILVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKI RPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQGVLTPQNSTMNTPVYPVPKPDGRWR MVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLTAFTWQ GKQYCWTRLPQGFLNSPALFNADAVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQIL LQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFKTKLLNVTPPKDLKQLQSILGKLNFA RNFIPNFAELVQTLYNLIASSPGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNT SPSAGYVRYYNESGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVY SPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFHYDKTLPELKHIPDVYTSSIPPLKHPS QYEGVFCTDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFA CKKALKVPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKEPLKHISKWKSIAECLSIKPDI TIQHEKGHQPINTSIHTEGNALADKLATQGSYVVN
SMRVH_P03364	PRSRAIDIPVPHADKISWKITDPVWVDQWPLTYEKTLAAIALVQEQLAAGHIEPTNSPWNTPIF IIKKKSGSWRLLQDLRAVNKVMVPMGALQPGLPSPVAIPLNYHKIVIDLKDCFFTIPLHPEDRP YFAFSVPQINFQSPMPRYQWKVLPQGMANSPTLCQKFVAAAIAPVRSQWPEAYILHYMDDIL LACDSAEAAKACYAHIISCLTSYGLKIAPDKVQVSEPFSYLGFELHHQQVFTPRVCLKTDHLK TLNDFQKLLGDIQWLRPYLKLPTSALVPLNNILKGDPNPLSVRALTPEAKQSLALINKAIQNQS VQQISYNLPLVLLLLPTPHTPTAVFWQPNGTDPTKNGSPLLWLHLPASPSKVLLTYPSLLAMLI IKGRYTGRQLFGRDPHSIIIPYTQDQLTWLLQTSDEWAIALSSFTGDIDNHYPSDPVIQFAKLH QFIFPKITKCAPIPQATLVFTDGSSNGIAAYVIDNQPISIKSPYLSAQLVELYAILQVFTVLAHQP FNLYTDSAYIAQSVPLLETVPFIKSSTNATPLFSKLQQLILNRQHPFFIGHLRAHLNLPGPLAEG NALADAATQIFPIISD
SMRVH_P03364_2mut	PRSRAIDIPVPHADKISWKITDPVWVDQWPLTYEKTLAAIALVQEQLAAGHIEPTNSPWNTPIF IIKKKSGSWRLLQDLRAVNKVMVPMGALQPGLPSPVAIPLNYHKIVIDLKDCFFTIPLHPEDRP YFAFSVPQINFQSPMPRYQWKVLPQGMANSPTLCQKFVAAAIAPVRSQWPEAYILHYMDDIL LACDSAEAAKACYAHIISCLTSYGLKIAPDKVQVSEPFSYLGFELHHQQVFTPRVCLKTDHLK TLNDFQKLLGDIQWLRPYLKLPTSALVPLNNILKPDPNPLSVRALTPEAKQSLALINKAIQNQS VQQISYNLPLVLLLLPTPHTPTAVFWQPNGTDPTKNGSPLLWLHLPASPSKVLLTYPSLLAMLI IKGRYTGRQLFGRDPHSIIIPYTQDQLTWLLQTSDEWAIALSSFTGDIDNHYPSDPVIQFAKLH QFIFPKITKCAPIPQATLVFTDGSSNGIAAYVIDNQPISIKSPYLSAQLVELYAILQVFTVLAHQP FNLYTDSAYIAQSVPLLETVPFIKSSTNATPLFSKLQQLILNRQHPFFIGHLRAHLNLPGPLAEG NALADAATQIFPIISD
SMRVH_P033	PRSRAIDIPVPHADKISWKITDPVWVDQWPLTYEKTLAAIALVQEQLAAGHIEPTNSPWNTPIF IIKKKSGSWRLLQDLRAVNKVMVPMGALQPGLPSPVAPPLNYHKIVIDLKDCFFTIPLHPEDR PYFAFSVPQINFQSPMPRYQWKVLPQGMANSPTLCQKFVAAAIAPVRSQWPEAYILHYMDDI

64_2mutB	LLACDSAEAAKACYAHIISCLTSYGLKIAPDKVQVSEPFSYLGFELHHQQVFTPRVCLKTDHL KTLNDFQKLLGDIQWLRPYLKLPTSALVPLNNILKPDPNPLSVRALTPEAKQSLALINKAIQNQ SVQQISYNLPLVLLLLPTPHTPTAVFWQPNGTDPTKNGSPLLWLHLPASPSKVLLTYPSLLAM LIIKGRYTGRQLFGRDPHSIIIPYTQDQLTWLLQTSDEWAIALSSFTGDIDNHYPSDPVIQFAKL HQFIFPKITKCAPIPQATLVFTDGSSNGIAAYVIDNQPISIKSPYLSAQLVELYAILQVFTVLAHQ PFNLYTDSAYIAQSVPLLETVPFIKSSTNATPLFSKLQQLILNRQHPFFIGHLRAHLNLPGPLAE GNALADAATQIFPIISD
SRV2_P51517	LATAVDILAPQRYADPITWKSDEPVWVDQWPLTQEKLAAAQQLVQEQLQAGHIIESNSPWN TPIFVIKKKSGKWRLLQDLRAVNATMVLMGALQPGLPSPVAIPQGYFKIVIDLKDCFFTIPLQP VDQKRFAFSLPSTNFKQPMKRYQWKVLPQGMANSPTLCQKYVAAAIEPVRKSWAQMYIIHY MDDILIAGKLGEQVLQCFAQLKQALTTTGLQIAPEKVQLQDPYTYLGFQINGPKITNQKAVIR RDKLQTLNDFQKLLGDINWLRPYLHLTTGDLKPLFDILKGDSNPNSPRSLSEAALASLQKVET AIAEQFVTQIDYTQPLTFLIFNTTLTPTGLFWQNNPVMWVHLPASPKKVLLPYYDAIADLIILG RDNSKKYFGLEPSTIIQPYSKSQIHWLMQNTETWPIACASYAGNIDNHYPPNKLIQFCKLHAV VFPRIISKTPLDNALLVFTDGSSTGIAAYTFEKTTVRFKTSHTSAQLVELQALIAVLSAFPHRAL NVYTDSAYLAHSIPLLETVSHIKHISDTAKFFLQCQQLIYNRSIPFYLGHIRAHSGLPGPLSQGN HITDLATKVVATTLTT
SRV2_P51517_2mutB	LATAVDILAPQRYADPITWKSDEPVWVDQWPLTQEKLAAAQQLVQEQLQAGHIIESNSPWN TPIFVIKKKSGKWRLLQDLRAVNATMVLMGALQPGLPSPVAPPQGYFKIVIDLKDCFFTIPLQP VDQKRFAFSLPSTNFKQPMKRYQWKVLPQGMANSPTLCQKYVAAAIEPVRKSWAQMYIIHY MDDILIAGKLGEQVLQCFAQLKQALTTTGLQIAPEKVQLQDPYTYLGFQINGPKITNQKAVIR RDKLQTLNDFQKLLGDINWLRPYLHLTTGDLKPLFDILKGDSNPNSPRSLSEAALASLQKVET AIAEQFVTQIDYTQPLTFLIFNTTLTPTGLFWQNNPVMWVHLPASPKKVLLPYYDAIADLIILG RDNSKKYFGLEPSTIIQPYSKSQIHWLMQNTETWPIACASYAGNIDNHYPPNKLIQFCKLHAV VFPRIISKTPLDNALLVFTDGSSTGIAAYTFEKTTVRFKTSHTSAQLVELQALIAVLSAFPHRAL NVYTDSAYLAHSIPLLETVSHIKHISDTAKFFLQCQQLIYNRSIPFYLGHIRAHSGLPGPLSQGN HITDLATKVVATTLTT
WDSV_O92815	SCQTKNTLNIDEYLLQFPDQLWASLPTDIGRMLVPPITIKIKDNASLPSIRQYPLPKDKTEGLRP LISSLENQGILIKCHSPCNTPIFPIKKAGRDEYRMIHDLRAINNIVAPLTAVVASPTTVLSNLAPS LHWFTVIDLSNAFFSVPIHKDSQYLFAFTFEGHQYTWTVLPQGFIHSPTLFSQALYQSLHKIKF KISSEICIYMDDVLIASKDRDTNLKDTAVMLQHLASEGHKVSKKKLQLCQQEVVYLGQLLTP EGRKILPDRKVTVSQFQQPTTIRQIRAFLGLVGYCRHWIPEFSIHSKFLEKQLKKDTAEPFQLD DQQVEAFNKLKHAITTAPVLVVPDPAKPFQLYTSHSEHASIAVLTQKHAGRTRPIAFLSSKFD AIESGLPPCLKACASIHRSLTQADSFILGAPLIIYTTHAICTLLQRDRSQLVTASRFSKWEADLL RPELTFVACSAVSPAHLYMQSCENNIPPHDCVLLTHTISRPRPDLSDLPIPDPDMTLFSDGSYTT GRGGAAVVMHRPVTDDFIIIHQQPGGASAQTAELLALAAACHLATDKTVNIYTDSRYAYGV VHDFGHLWMHRGFVTSAGTPIKNHKEIEYLLKQIMKPKQVSVIKIEAHTKGVSMEVRGNAA ADEAAKNAVFLVQR
WDSV_O92815_2mut	SCQTKNTLNIDEYLLQFPDQLWASLPTDIGRMLVPPITIKIKDNASLPSIRQYPLPKDKTEGLRP LISSLENQGILIKCHSPCNTPIFPIKKAGRDEYRMIHDLRAINNIVAPLTAVVASPTTVLSNLAPS LHWFTVIDLSNAFFSVPIHKDSQYLFAFTFEGHQYTWTVLPQGFIHSPTLFNQALYQSLHKIKF KISSEICIYMDDVLIASKDRDTNLKDTAVMLQHLASEGHKVSKKKLQLCQQEVVYLGQLLTP EGRKILPDRKVTVSQFQQPTTIRQIRAFLGLVGYCRHWIPEFSIHSKFLEKQLKPDTAEPFQLD DQQVEAFNKLKHAITTAPVLVVPDPAKPFQLYTSHSEHASIAVLTQKHAGRTRPIAFLSSKFD AIESGLPPCLKACASIHRSLTQADSFILGAPLIIYTTHAICTLLQRDRSQLVTASRFSKWEADLL RPELTFVACSAVSPAHLYMQSCENNIPPHDCVLLTHTISRPRPDLSDLPIPDPDMTLFSDGSYTT GRGGAAVVMHRPVTDDFIIIHQQPGGASAQTAELLALAAACHLATDKTVNIYTDSRYAYGV VHDFGHLWMHRGFVTSAGTPIKNHKEIEYLLKQIMKPKQVSVIKIEAHTKGVSMEVRGNAA ADEAAKNAVFLVQR
WDSV_O92815_2mutA	SCQTKNTLNIDEYLLQFPDQLWASLPTDIGRMLVPPITIKIKDNASLPSIRQYPLPKDKTEGLRP LISSLENQGILIKCHSPCNTPIFPIKKAGRDEYRMIHDLRAINNIVAPLTAVVASPTTVLSNLAPS LHWFTVIDLSNAFFSVPIHKDSQYLFAFTFEGHQYTWTVLPQGFIHSPTLFNQALYQSLHKIKF KISSEICIYMDDVLIASKDRDTNLKDTAVMLQHLASEGHKVSKKKLQLCQQEVVYLGQLLTP EGRKILPDRKVTVSQFQQPTTIRQIRAFLGKVGYCRHFIPEFSIHSKFLEKQLKPDTAEPFQLDD QQVEAFNKLKHAITTAPVLVVPDPAKPFQLYTSHSEHASIAVLTQKHAGRTRPIAFLSSKFDAI ESGLPPCLKACASIHRSLTQADSFILGAPLIIYTTHAICTLLQRDRSQLVTASRFSKWEADLLRP

	ELTFVACSAVSPAHLYMQSCENNIPPHDCVLLTHTISRPRPDLSDLPIPDPDMTLFSDGSYTTG RGGAAVVMHRPVTDDFIIIHQQPGGASAQTAELLALAAACHLATDKTVNIYTDSRYAYGVV HDFGHLWMHRGFVTSAGTPIKNHKEIEYLLKQIMKPKQVSVIKIEAHTKGVSMEVRGNAAA DEAAKNAVFLVQR
WMSV_P03359	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQY PMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTPLLPVKKPGTNDYRPVQDLREINKRVQDIHP TVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLP QGFKNSPTLFDEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYRDCKEGTQKLLQELSKLG YRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKATVMKIPPPTTPRQVREFLGTAGFCRL WIPGFASLAAPLYPLTKESIPFIWTEEHQKAFDRIKEALLSAPALALPDLTKPFTLYVDERAGV ARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIAS HSLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILA EETGTRRDLKDQPLPGVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAEL VALTQALRLAEGKDINIYTDSRYAFATAHIHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLP KRVAIIHCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKP
WMSV_P03359_3mut	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQY PMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTPLLPVKKPGTNDYRPVQDLREINKRVQDIHP TVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLP QGFKNSPTLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYRDCKEGTQKLLQELSKLG YRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKATVMKIPPPTTPRQVREFLGTAGFCRL WIPGFASLAAPLYPLTKPSIPFIWTEEHQKAFDRIKEALLSAPALALPDLTKPFTLYVDERAGV ARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIAS HSLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILA EETGTRRDLKDQPLPGVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAEL VALTQALRLAEGKDINIYTDSRYAFATAHIHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLP KRVAIIHCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKP
WMSV_P03359_3mutA	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQY PMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTPLLPVKKPGTNDYRPVQDLREINKRVQDIHP TVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLP QGFKNSPTLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYRDCKEGTQKLLQELSKLG YRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKATVMKIPPPTTPRQVREFLGKAGFCRL FIPGFASLAAPLYPLTKPSIPFIWTEEHQKAFDRIKEALLSAPALALPDLTKPFTLYVDERAGVA RGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASH SLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAE ETGTRRDLKDQPLPGVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELV ALTQALRLAEGKDINIYTDSRYAFATAHIHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPK RVAIIHCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKP
XMRV6_A1Z651	TLNIEDEYRLHETSKEPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSEQDCQRGTRALLQTLGNLG YRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRL WIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQ GYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVI LAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEKEAPHDC LEILAETHGTRPDLTDQPIPDADYTWYTDGSSFLQEGQRRAGAAVTTETEVIWARALPAGTSA QRAELIALTQALKMAEGKKLNVYTDSRYAFATAHVHGEIYRRRGLLTSEGREIKNKNEILAL LKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAAREAAMKAVLETSTLL
XMRV6_A1Z651_3mut	TLNIEDEYRLHETSKEPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSEQDCQRGTRALLQTLGNLG YRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRL WIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQ GYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVI LAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEKEAPHDC

	LEILAETHGTRPDLTDQPIPDADYTWYTDGSSFLQEGQRRAGAAVTTETEVIWARALPAGTSA QRAELIALTQALKMAEGKKLNVYTDSRYAFATAHVHGEIYRRRGWLTSEGREIKNKNEILAL LKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAAREAAMKAVLETSTLL
XMRV 6_A1Z6 51_3mu tA	TLNIEDEYRLHETSKEPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQ YPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIH PTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP QGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSEQDCQRGTRALLQTLGNLG YRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRL FIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQG YAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVIL APHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEKEAPHDCL EILAETHGTRPDLTDQPIPDADYTWYTDGSSFLQEGQRRAGAAVTTETEVIWARALPAGTSA QRAELIALTQALKMAEGKKLNVYTDSRYAFATAHVHGEIYRRRGWLTSEGREIKNKNEILAL LKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAAREAAMKAVLETSTLL

In some embodiments, reverse transcriptase domains are modified, for example by site-specific mutation. In some embodiments, reverse transcriptase domains are engineered to have improved properties, e.g. SuperScript IV (SSIV) reverse transcriptase derived from the MMLV RT. In some embodiments, the reverse transcriptase domain may be engineered to have lower error rates, e.g., as described in WO2001068895, incorporated herein by reference. In some embodiments, the reverse transcriptase domain may be engineered to be more thermostable. In some embodiments, the reverse transcriptase domain may be engineered to be more processive. In some embodiments, the reverse transcriptase domain may be engineered to have tolerance to inhibitors. In some embodiments, the reverse transcriptase domain may be engineered to be faster. In some embodiments, the reverse transcriptase domain may be engineered to better tolerate modified nucleotides in the RNA template. In some embodiments, the reverse transcriptase domain may be engineered to insert modified DNA nucleotides. In some embodiments, the reverse transcriptase domain is engineered to bind a template RNA. In some embodiments, one or more mutations are chosen from D200N, L603W, T330P, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, W313F, L435G, N454K, H594Q, L671P, E69K, or D653N in the RT domain of murine leukemia virus reverse transcriptase or a corresponding mutation at a corresponding position of another RT domain.

In some embodiments, an RT domain (e.g., as listed in Table 6) comprises one or more mutations as listed in Table 2 below. In some embodiment, an RT domain as listed in Table 6 comprises one, two, three, four, five, or six of the mutations listed in the corresponding row of Table 2 below.

Table 2. Exemplary RT domain mutations (relative to corresponding wild-type sequences as listed in the corresponding row of Table 6)

RT Domain Name	Mutation(s)

AVIRE_P03360						
AVIRE_P03360_3mut	D200N	G330P	L605W			
AVIRE_P03360_3mutA	D200N	G330P	L605W	T306K	W313F	
BAEVM_P10272						
BAEVM_P10272_3mut	D198N	E328P	L602W			
BAEVM_P10272_3mutA	D198N	E328P	L602W	T304K	W311F	
BLVAU_P25059						
BLVAU_P25059_2mut	E159Q	G286P				
BLVJ_P03361						
BLVJ_P03361_2mut	E159Q	L524W				
BLVJ_P03361_2mutB	E159Q	L524W	I97P			
FFV_O93209	D21N					
FFV_O93209_2mut	D21N	T293N	T419P			
FFV_O93209_2mutA	D21N	T293N	T419P	L393K		
FFV_O93209-Pro						
FFV_O93209-Pro_2mut	T207N	T333P				
FFV_O93209-Pro_2mutA	T207N	T333P	L307K			
FLV_P10273						
FLV_P10273_3mut	D199N	L602W				
FLV_P10273_3mutA	D199N	L602W	T305K	W312F		
FOAMV_P14350	D24N					
FOAMV_P14350_2mut	D24N	T296N	S420P			
FOAMV_P14350_2mutA	D24N	T296N	S420P	L396K		
FOAMV_P14350-Pro						
FOAMV_P14350-Pro_2mut	T207N	S331P				
FOAMV_P14350-Pro_2mutA	T207N	S331P	L307K			
GALV_P21414						
GALV_P21414_3mut	D198N	E328P	L600W			
GALV_P21414_3mutA	D198N	E328P	L600W	T304K	W311F	
HTL1A_P03362						
HTL1A_P03362_2mut	E152Q	R279P				
HTL1A_P03362_2mutB	E152Q	R279P	L90P			
HTL1C_P14078						
HTL1C_P14078_2mut	E152Q	R279P				
HTL1L_P0C211						
HTL1L_P0C211_2mut	E149Q	L527W				
HTL1L_P0C211_2mutB	E149Q	L527W	L87P			
HTL32_Q0R5R2						
HTL32_Q0R5R2_2mut	E149Q	L526W				

HTL32_Q0R5R2_2mutB	E149Q	L526W	L87P			
HTL3P_Q4U0X6						
HTL3P_Q4U0X6_2mut	E149Q	L526W				
HTL3P_Q4U0X6_2mutB	E149Q	L526W	L87P			
HTLV2_P03363_2mut	E147Q	G274P				
JSRV_P31623						
JSRV_P31623_2mutB	A100P					
KORV_Q9TTC1	D32N					
KORV_Q9TTC1_3mut	D32N	D322N	E452P	L724W		
KORV_Q9TTC1_3mutA	D32N	D322N	E452P	L724W	T428K	W435F
KORV_Q9TTC1-Pro						
KORV_Q9TTC1-Pro_3mut	D231N	E361P	L633W			
KORV_Q9TTC1-Pro_3mutA	D231N	E361P	L633W	T337K	W344F	
MLVAV_P03356						
MLVAV_P03356_3mut	D200N	T330P	L603W			
MLVAV_P03356_3mutA	D200N	T330P	L603W	T306K	W313F	
MLVBM_Q7SVK7						
MLVBM_Q7SVK7						
MLVBM_Q7SVK7_3mut	D200N	T330P	L603W			
MLVBM_Q7SVK7_3mut	D200N	T330P	L603W			
MLVBM_Q7SVK7_3mutA_WS	D199N	T329P	L602W	T305K	W312F	
MLVBM_Q7SVK7_3mutA_WS	D199N	T329P	L602W	T305K	W312F	
MLVCB_P08361						
MLVCB_P08361_3mut	D200N	T330P	L603W			
MLVCB_P08361_3mutA	D200N	T330P	L603W	T306K	W313F	
MLVF5_P26810						
MLVF5_P26810_3mut	D200N	T330P	L603W			
MLVF5_P26810_3mutA	D200N	T330P	L603W	T306K	W313F	
MLVFF_P26809_3mut	D200N	T330P	L603W			
MLVFF_P26809_3mutA	D200N	T330P	L603W	T306K	W313F	
MLVMS_P03355						
MLVMS_P03355						
MLVMS_P03355_3mut	D200N	T330P	L603W			
MLVMS_P03355_3mut	D200N	T330P	L603W			
MLVMS_P03355_3mutA_WS	D200N	T330P	L603W	T306K	W313F	
MLVMS_P03355_3mutA_WS	D200N	T330P	L603W	T306K	W313F	
MLVMS_P03355_PLV919	D200N	T330P	L603W	T306K	W313F	H8Y
MLVMS_P03355_PLV919	D200N	T330P	L603W	T306K	W313F	H8Y
MLVRD_P11227						

MLVRD_P11227_3mut	D200N	T330P	L603W			
MMTVB_P03365	D26N					
MMTVB_P03365	D26N					
MMTVB_P03365_2mut	D26N	G401P				
MMTVB_P03365_2mut_WS	G400P					
MMTVB_P03365_2mut_WS	G400P					
MMTVB_P03365_2mutB	D26N	G401P	V215P			
MMTVB_P03365_2mutB	D26N	G401P	V215P			
MMTVB_P03365_2mutB_WS	G400P	V212P				
MMTVB_P03365_2mutB_WS	G400P	V212P				
MMTVB_P03365_WS						
MMTVB_P03365_WS						
MMTVB_P03365-Pro						
MMTVB_P03365-Pro						
MMTVB_P03365-Pro_2mut	G309P					
MMTVB_P03365-Pro_2mut	G309P					
MMTVB_P03365-Pro_2mutB	G309P	V123P				
MMTVB_P03365-Pro_2mutB	G309P	V123P				
MPMV_P07572						
MPMV_P07572_2mutB	G289P	I103P				
PERV_Q4VFZ2						
PERV_Q4VFZ2						
PERV_Q4VFZ2_3mut	D199N	E329P	L602W			
PERV_Q4VFZ2_3mut	D199N	E329P	L602W			
PERV_Q4VFZ2_3mutA_WS	D196N	E326P	L599W	T302K	W309F	
PERV_Q4VFZ2_3mutA_WS	D196N	E326P	L599W	T302K	W309F	
SFV1_P23074	D24N					
SFV1_P23074_2mut	D24N	T296N	N420P			
SFV1_P23074_2mutA	D24N	T296N	N420P	L396K		
SFV1_P23074-Pro						
SFV1_P23074-Pro_2mut	T207N	N331P				
SFV1_P23074-Pro_2mutA	T207N	N331P	L307K			
SFV3L_P27401	D24N					
SFV3L_P27401_2mut	D24N	T296N	N422P			
SFV3L_P27401_2mutA	D24N	T296N	N422P	L396K		
SFV3L_P27401-Pro						
SFV3L_P27401-Pro_2mut	T307N	N333P				
SFV3L_P27401-Pro_2mutA	T307N	N333P	L307K			
SFVCP_Q87040	D24N					

SFVCP_Q87040_2mut	D24N	T296N	K422P			
SFVCP_Q87040_2mutA	D24N	T296N	K422P	L396K		
SFVCP_Q87040-Pro						
SFVCP_Q87040-Pro_2mut	T207N	K333P				
SFVCP_Q87040-Pro_2mutA	T207N	K333P	L307K			
SMRVH_P03364						
SMRVH_P03364_2mut	G288P					
SMRVH_P03364_2mutB	G288P	I102P				
SRV2_P51517						
SRV2_P51517_2mutB	I103P					
WDSV_O92815						
WDSV_O92815_2mut	S183N	K312P				
WDSV_O92815_2mutA	S183N	K312P	L288K	W295F		
WMSV_P03359						
WMSV_P03359_3mut	D198N	E328P	L600W			
WMSV_P03359_3mutA	D198N	E328P	L600W	T304K	W311F	
XMRV6_A1Z651						
XMRV6_A1Z651_3mut	D200N	T330P	L603W			
XMRV6_A1Z651_3mutA	D200N	T330P	L603W	T306K	W313F	

In some embodiments, a gene modifying polypeptide comprises the RT domain from a retroviral reverse transcriptase, e.g., a wild-type M-MLV RT, e.g., comprising the following sequence:

M-MLV (WT):

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYP
MSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVP
NPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN
SPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKA
QICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAA
PLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLG
PWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPD
RWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQP
LPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGK
KLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGH
SAEARGNRMADQAARKAAITETPDTSTLLI (SEQ ID NO: 2)

In some embodiments, a gene modifying polypeptide comprises the RT domain from a retroviral reverse transcriptase, e.g., an M-MLV RT, e.g., comprising the following sequence:

TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYP
MSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVP
NPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN
SPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKA
QICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAA
PLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLG
PWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPD
RWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQP
LPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGK
KLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGH
SAEARGNRMADQAARKAAITETPDTSTLL (SEQ ID NO: 3)

In some embodiments, a gene modifying polypeptide comprises the RT domain from a retroviral reverse transcriptase comprising the sequence of amino acids 659-1329 of NP_057933. In embodiments, the gene modifying polypeptide further comprises one additional amino acid at the N-terminus of the sequence of amino acids 659-1329 of NP_057933, e.g., as shown below:

TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYP
MSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP**VKKPGTNDYRPVQDLREVNKRVEDIHPT**
VPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLP
QGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNL
GYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRL
WIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY
AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPH
AVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAE
AHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELI
ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKR
LSIIHCPGHQKGHSAEARGNRMADQAARKAA (SEQ ID NO: 4)

Core RT (bold), annotated per above
RNAseH (underlined), annotated per above

In embodiments, the gene modifying polypeptide further comprises one additional amino acid at the C-terminus of the sequence of amino acids 659-1329 of NP_057933. In embodiments, the gene modifying polypeptide comprises an RNaseH1 domain (e.g., amino acids 1178-1318 of NP_057933).

In some embodiments, a retroviral reverse transcriptase domain, e.g., M-MLV RT, may comprise one or more mutations from a wild-type sequence that may improve features of the RT, e.g.,

thermostability, processivity, and/or template binding. In some embodiments, an M-MLV RT domain comprises, relative to the M-MLV (WT) sequence above, one or more mutations, e.g., selected from D200N, L603W, T330P, T306K, W313F, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, L435G, N454K, H594Q, D653N, R110S, K103L, e.g., a combination of mutations, such as D200N, L603W, and T330P, optionally further including T306K and W313F. In some embodiments, an M-MLV RT used herein comprises the mutations D200N, L603W, T330P, T306K and W313F. In embodiments, the mutant M-MLV RT comprises the following amino acid sequence:

M-MLV (PE2):

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYP
MSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVP
NPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN
SPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKA
QICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAP
LYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGP
WRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDR
WLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPL
PDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKK
LNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHS
AEARGNRMADQAARKAAITETPDTSTLLI (SEQ ID NO: 5)

In some embodiments, a writing domain (e.g., RT domain) comprises an RNA-binding domain, e.g., that specifically binds to an RNA sequence. In some embodiments, a template RNA comprises an RNA sequence that is specifically bound by the RNA-binding domain of the writing domain.

In some embodiments, the reverse transcription domain only recognizes and reverse transcribes a specific template, e.g., a template RNA of the system. In some embodiments, the template comprises a sequence or structure that enables recognition and reverse transcription by a reverse transcription domain. In some embodiments, the template comprises a sequence or structure that enables association with an RNA-binding domain of a polypeptide component of a genome engineering system described herein. In some embodiments, the genome engineering system reverse preferably transcribes a template comprising an association sequence over a template lacking an association sequence.

The writing domain may also comprise DNA-dependent DNA polymerase activity, e.g., comprise enzymatic activity capable of writing DNA into the genome from a template DNA sequence. In some

embodiments, DNA-dependent DNA polymerization is employed to complete second-strand synthesis of a target site edit. In some embodiments, the DNA-dependent DNA polymerase activity is provided by a DNA polymerase domain in the polypeptide. In some embodiments, the DNA-dependent DNA polymerase activity is provided by a reverse transcriptase domain that is also capable of DNA-dependent DNA polymerization, e.g., second-strand synthesis. In some embodiments, the DNA-dependent DNA polymerase activity is provided by a second polypeptide of the system. In some embodiments, the DNA-dependent DNA polymerase activity is provided by an endogenous host cell polymerase that is optionally recruited to the target site by a component of the genome engineering system.

In some embodiments, the reverse transcriptase domain has a lower probability of premature termination rate (P_{off}) *in vitro* relative to a reference reverse transcriptase domain. In some embodiments, the reference reverse transcriptase domain is a viral reverse transcriptase domain, e.g., the RT domain from M-MLV.

In some embodiments, the reverse transcriptase domain has a lower probability of premature termination rate (P_{off}) *in vitro* of less than about 5 x 10^{-3}/nt, 5 x 10^{-4}/nt, or 5 x 10^{-6}/nt, e.g., as measured on a 1094 nt RNA. In embodiments, the *in vitro* premature termination rate is determined as described in Bibillo and Eickbush (2002) J Biol Chem 277(38):34836-34845 (incorporated by reference herein its entirety).

In some embodiments, the reverse transcriptase domain is able to complete at least about 30% or 50% of integrations in cells. The percent of complete integrations can be measured by dividing the number of substantially full-length integration events (e.g., genomic sites that comprise at least 98% of the expected integrated sequence) by the number of total (including substantially full-length and partial) integration events in a population of cells. In embodiments, the integrations in cells is determined (e.g., across the integration site) using long-read amplicon sequencing, e.g., as described in Karst et al. (2020) *bioRxiv* doi.org/10.1101/645903 (incorporated by reference herein in its entirety).

In embodiments, quantifying integrations in cells comprises counting the fraction of integrations that contain at least about 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of the DNA sequence corresponding to the template RNA (e.g., a template RNA having a length of at least 0.05, 0.1, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 3, 4, or 5 kb, e.g., a length between 0.5-0.6, 0.6-0.7, 0.7-0.8, 0.8-0.9, 1.0-1.2, 1.2-1.4, 1.4-1.6, 1.6-1.8, 1.8-2.0, 2-3, 3-4, or 4-5 kb).

In some embodiments, the reverse transcriptase domain is capable of polymerizing dNTPs *in vitro*. In embodiments, the reverse transcriptase domain is capable of polymerizing dNTPs *in vitro* at a rate between 0.1 – 50 nt/sec (e.g., between 0.1-1, 1-10, or 10-50 nt/sec). In embodiments, polymerization of dNTPs by the reverse transcriptase domain is measured by a single-molecule assay, e.g., as described in Schwartz and Quake (2009) *PNAS* 106(48):20294-20299 (incorporated by reference in its entirety).

In some embodiments, the reverse transcriptase domain has an *in vitro* error rate (e.g., misincorporation of nucleotides) of between 1 x 10^{-3} – 1 x 10^{-4} or 1 x 10^{-4} – 1 x 10^{-5} substitutions/nt , e.g., as described in Yasukawa et al. (2017) *Biochem Biophys Res Commun* 492(2):147-153 (incorporated herein by reference in its entirety). In some embodiments, the reverse transcriptase domain has an error rate (e.g., misincorporation of nucleotides) in cells (e.g., HEK293T cells) of between 1 x 10^{-3} – 1 x 10^{-4} or 1 x 10^{-4} – 1 x 10^{-5} substitutions/nt, e.g., by long-read amplicon sequencing, e.g., as described in Karst et al. (2020) *bioRxiv* doi.org/10.1101/645903 (incorporated by reference herein in its entirety).

In some embodiments, the reverse transcriptase domain is capable of performing reverse transcription of a target RNA *in vitro*. In some embodiments, the reverse transcriptase requires a primer of at least 3 nucleotides to initiate reverse transcription of a template. In some embodiments, reverse transcription of the target RNA is determined by detection of cDNA from the target RNA (e.g., when provided with a ssDNA primer, e.g., which anneals to the target with at least 3, 4, 5, 6, 7, 8, 9, or 10 nt at the 3′ end), e.g., as described in Bibillo and Eickbush (2002) *J Biol Chem* 277(38):34836-34845 (incorporated herein by reference in its entirety).

In some embodiments, the reverse transcriptase domain performs reverse transcription at least 5 or 10 times more efficiently (e.g., by cDNA production), e.g., when converting its RNA template to cDNA, for example, as compared to an RNA template lacking the protein binding motif (e.g., a 3′ UTR). In embodiments, efficiency of reverse transcription is measured as described in Yasukawa et al. (2017) *Biochem Biophys Res Commun* 492(2):147-153 (incorporated by reference herein in its entirety).

In some embodiments, the reverse transcriptase domain specifically binds a specific RNA template with higher frequency (e.g., about 5 or 10-fold higher frequency) than any endogenous cellular RNA, e.g., when expressed in cells (e.g., HEK293T cells). In embodiments, frequency of specific binding between the reverse transcriptase domain and the template RNA are measured by CLIP-seq, e.g., as described in Lin and Miles (2019) *Nucleic Acids Res* 47(11):5490-5501 (incorporated herein by reference in its entirety).

Template nucleic acid binding domain

The gene modifying polypeptide typically contains regions capable of associating with the template nucleic acid (e.g., template RNA). In some embodiments, the template nucleic acid binding domain is an RNA binding domain. In some embodiments, the RNA binding domain is a modular domain that can associate with RNA molecules containing specific signatures, e.g., structural motifs. In other embodiments, the template nucleic acid binding domain (e.g., RNA binding domain) is contained within the reverse transcription domain, e.g., the reverse transcriptase-derived component has a known signature for RNA preference.

In other embodiments, the template nucleic acid binding domain (e.g., RNA binding domain) is contained within the target DNA binding domain. For example, in some embodiments, the DNA binding domain is a CRISPR-associated protein that recognizes the structure of a template nucleic acid (e.g., template RNA) comprising a gRNA. In some embodiments, a gene modifying polypeptide comprises a DNA-binding domain comprising a CRISPR-associated protein that associates with a gRNA scaffold that allows the DNA-binding domain to bind a target genomic DNA sequence. In some embodiments, the gRNA scaffold and gRNA spacer is comprised within the template nucleic acid (e.g., template RNA), thus the DNA-binding domain is also the template nucleic acid binding domain. In some embodiments, the polypeptide possesses RNA binding function in multiple domains, e.g., can bind a gRNA structure in a CRISPR-associated DNA binding domain and an additional sequence or structure in a reverse transcriptase domain.

In some embodiments, the RNA binding domain is capable of binding to a template RNA with greater affinity than a reference RNA binding domain. In some embodiments, the reference RNA binding domain is an RNA binding domain from Cas9 of S. pyogenes. In some embodiments, the RNA binding domain is capable of binding to a template RNA with an affinity between 100 pM – 10 nM (e.g., between 100 pM-1 nM or 1 nM – 10 nM). In some embodiments, the affinity of a RNA binding domain for its template RNA is measured *in vitro*, e.g., by thermophoresis, e.g., as described in Asmari et al. Methods 146:107-119 (2018) (incorporated by reference herein in its entirety). In some embodiments, the affinity of a RNA binding domain for its template RNA is measured in cells (e.g., by FRET or CLIP-Seq).

In some embodiments, the RNA binding domain is associated with the template RNA *in vitro* at a frequency at least about 5-fold or 10-fold higher than with a scrambled RNA. In some embodiments, the frequency of association between the RNA binding domain and the template RNA or scrambled RNA is measured by CLIP-seq, e.g., as described in Lin and Miles (2019) *Nucleic Acids Res* 47(11):5490-5501 (incorporated by reference herein in its entirety). In some embodiments, the RNA binding domain is associated with the template RNA in cells (e.g., in HEK293T cells) at a frequency at least about 5-fold or 10-fold higher than with a scrambled RNA. In some embodiments, the frequency of association between the RNA binding domain and the template RNA or scrambled RNA is measured by CLIP-seq, e.g., as described in Lin and Miles (2019), *supra.*

RNA binding domains (RBDs)

In some embodiments, a gene modifying polypeptide as described herein comprises an RNA binding domain (RBD). In some embodiments, a gene modifying polypeptide as described herein comprises an RBD comprising the amino acid sequence of an RBD as listed in Table 31, or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some

embodiments, the RBD of a gene modifying polypeptide as described herein binds to an RNA binding partner, e.g., as listed in Table 31. In embodiments, the RBD comprises the amino acid sequence of an RBD as listed in any one row of Table 31, or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, and binds to the RNA binding partner listed in the same row of Table 31.

Table 31. Exemplary RNA binding domain sequences

Name	RNA binding partner	Amino Acid sequence
MCP_v1	MS2	MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSGRA
MCP_v2	MS2	MASNFTQFVLVDNGGTGDVTVAPSNFANGVAEWISSNSRSQAYKVTCSVRQSSAQKRKYTIKVEVPKVATQTVGGVELPVAAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGLY
PCP	PP7	MGSMSKTIVLSVGEATRTLTEIQSTADRQIFEEKVGPLVGRLRLTASLRQNGAKTAYRVNLKLDQADVVDSGLPKVRYTQVWSHDVTIVANSTEASRKSLYDLTKSLVATSQVEDLVVNLVPLGRGSGRA
Com	com	MGSMKSIRCKNCNKLLFKADSFDHIEIRCPRCKRHIIMLNACEHPTEKHCGKREKITHSDETVRYGSGRA
LS4	LS4-1, CS1	YVRFEVPEDMQNEALSLLEKVRESGKVKKGTNRTTHAVYRGLAKLVYIAEDVDPPEIVAHLPLLCEEKNVPYIYVKSKNDLGRAVGGWPIGASAAIINEGELRKELGSLVEKIKGLQKRSHMHLE
LS12	LS12-1, CS2	YVRFEVPEDMQNEALSLLEKVRESGKVKKGTNSTTLAVSRGLAKLVYIAEDVDPPEIVAHLPLLCEEKNVPYIYVKSKNDLGRAVGRVYPGASAAIINEGELRKELGSLVEKIKGLQKRSHMHLE
lambdaN(1-22)	BoxB	MDAQTRRRERRAEKQAQWKAAN
L7Ae	Kt	MYVRFEVPEDMQNEALSLLEKVRESGKVKKGTNETTKAVERGLAKLVYIAEDVDPPEIVAHLPLLCEEKNVPYIYVKSKNDLGRAVGIEVPCASAAIINEGELRKELGSLVEKIKGLQK
L7Ae	Kt	YVRFEVPEDMQNEALSLLEKVRESGKVKKGTNETTKAVERGLAKLVYIAEDVDPPEIVAHLPLLCEEKNVPYIYVKSKNDLGRAVGIEVPCASAAIINEGELRKELGSLVEKIKGLQKRSHMHLE

Endonuclease domains and DNA binding domains

In some embodiments, a gene modifying polypeptide possesses the function of DNA target site cleavage via an endonuclease domain. In some embodiments, a gene modifying polypeptide comprises a DNA binding domain, e.g., for binding to a target nucleic acid. In some embodiments, a domain (e.g., a Cas domain) of the gene modifying polypeptide comprises two or more smaller domains, e.g., a DNA binding domain and an endonuclease domain. It is understood that when a DNA binding domain (e.g., a

Cas domain) is said to bind to a target nucleic acid sequence, in some embodiments, the binding is mediated by a gRNA.

In some embodiments, a domain has two functions. For example, in some embodiments, the endonuclease domain is also a DNA-binding domain. In some embodiments, the endonuclease domain is also a template nucleic acid (e.g., template RNA) binding domain. For example, in some embodiments, a polypeptide comprises a CRISPR-associated endonuclease domain that binds a template RNA comprising a gRNA, binds a target DNA sequence (e.g., with complementarity to a portion of the gRNA), and cuts the target DNA sequence. In some embodiments, an endonuclease domain or endonuclease/DNA-binding domain from a heterologous source can be used or can be modified (e.g., by insertion, deletion, or substitution of one or more residues) in a gene modifying system described herein.

In some embodiments, a nucleic acid encoding the endonuclease domain or endonuclease/DNA binding domain is altered from its natural sequence to have altered codon usage, e.g. improved for human cells. In some embodiments, the endonuclease element is a heterologous endonuclease element, such as a Cas endonuclease (e.g., Cas9), a type-II restriction endonuclease (e.g., Fok1), a meganuclease (e.g., I-SceI), or other endonuclease domain.

In certain aspects, the DNA-binding domain of a gene modifying polypeptide described herein is selected, designed, or constructed for binding to a desired host DNA target sequence. In certain embodiments, the DNA-binding domain of the polypeptide is a heterologous DNA-binding element. In some embodiments the heterologous DNA binding element is a zinc-finger element or a TAL effector element, e.g., a zinc-finger or TAL polypeptide or functional fragment thereof. In some embodiments the heterologous DNA binding element is a sequence-guided DNA binding element, such as Cas9, Cpf1, or other CRISPR-related protein that has been altered to have no endonuclease activity. In some embodiments the heterologous DNA binding element retains endonuclease activity. In some embodiments, the heterologous DNA binding element retains partial endonuclease activity to cleave ssDNA, e.g., possesses nickase activity. In specific embodiments, the heterologous DNA-binding domain can be any one or more of Cas9, TAL domain, ZF domain, Myb domain, combinations thereof, or multiples thereof.

In some embodiments, DNA-binding domains are modified, for example by site-specific mutation, increasing or decreasing DNA-binding elements (for example, number and/or specificity of zinc fingers), etc., to alter DNA-binding specificity and affinity. In some embodiments a nucleic acid sequence encoding the DNA binding domain is altered from its natural sequence to have altered codon usage, e.g. improved for human cells. In embodiments, the DNA binding domain comprises one or more modifications relative to a wild-type DNA binding domain, e.g., a modification via directed evolution, e.g., phage-assisted continuous evolution (PACE).

In some embodiments, the DNA binding domain comprises a meganuclease domain (e.g., as described herein, e.g., in the endonuclease domain section), or a functional fragment thereof. In some embodiments, the meganuclease domain possesses endonuclease activity, e.g., double-strand cleavage and/or nickase activity. In other embodiments, the meganuclease domain has reduced activity, e.g., lacks endonuclease activity, e.g., the meganuclease is catalytically inactive. In some embodiments, a catalytically inactive meganuclease is used as a DNA binding domain, e.g., as described in Fonfara et al. Nucleic Acids Res 40(2):847-860 (2012), incorporated herein by reference in its entirety.

In some embodiments, a gene modifying polypeptide comprises a modification to a DNA-binding domain, e.g., relative to the wild-type polypeptide. In some embodiments, the DNA-binding domain comprises an addition, deletion, replacement, or modification to the amino acid sequence of the original DNA-binding domain. In some embodiments, the DNA-binding domain is modified to include a heterologous functional domain that binds specifically to a target nucleic acid (e.g., DNA) sequence of interest. In some embodiments, the functional domain replaces at least a portion (e.g., the entirety of) the prior DNA-binding domain of the polypeptide. In some embodiments, the functional domain comprises a zinc finger (e.g., a zinc finger that specifically binds to the target nucleic acid (e.g., DNA) sequence of interest. In some embodiments, the functional domain comprises a Cas domain (e.g., a Cas domain that specifically binds to the target nucleic acid (e.g., DNA) sequence of interest. In some embodiments, the Cas domain comprises a Cas9 or a mutant or variant thereof (e.g., as described herein). In embodiments, the Cas domain is associated with a guide RNA (gRNA), e.g., as described herein. In embodiments, the Cas domain is directed to a target nucleic acid (e.g., DNA) sequence of interest by the gRNA. In embodiments, the Cas domain is encoded in the same nucleic acid (e.g., RNA) molecule as the gRNA. In embodiments, the Cas domain is encoded in a different nucleic acid (e.g., RNA) molecule from the gRNA.

In some embodiments, the DNA binding domain is capable of binding to a target sequence (e.g., a dsDNA target sequence) with greater affinity than a reference DNA binding domain. In some embodiments, the reference DNA binding domain is a DNA binding domain from Cas9 of S. pyogenes. In some embodiments, the DNA binding domain is capable of binding to a target sequence (e.g., a dsDNA target sequence) with an affinity between 100 pM – 10 nM (e.g., between 100 pM-1 nM or 1 nM – 10 nM).

In some embodiments, the affinity of a DNA binding domain for its target sequence (e.g., dsDNA target sequence) is measured in vitro, e.g., by thermophoresis, e.g., as described in Asmari et al. Methods 146:107-119 (2018) (incorporated by reference herein in its entirety).

In embodiments, the DNA binding domain is capable of binding to its target sequence (e.g., dsDNA target sequence), e.g, with an affinity between 100 pM – 10 nM (e.g., between 100 pM-1 nM or 1

nM – 10 nM) in the presence of a molar excess of scrambled sequence competitor dsDNA, e.g., of about 100-fold molar excess.

In some embodiments, the DNA binding domain is found associated with its target sequence (e.g., dsDNA target sequence) more frequently than any other sequence in the genome of a target cell, e.g., human target cell, e.g., as measured by ChIP-seq (e.g., in HEK293T cells), e.g., as described in He and Pu (2010) *Curr. Protoc Mol Biol* Chapter 21 (incorporated herein by reference in its entirety). In some embodiments, the DNA binding domain is found associated with its target sequence (e.g., dsDNA target sequence) at least about 5-fold or 10-fold, more frequently than any other sequence in the genome of a target cell, e.g., as measured by ChIP-seq (e.g., in HEK293T cells), e.g., as described in He and Pu (2010), *supra*.

In some embodiments, the endonuclease domain has nickase activity and cleaves one strand of a target DNA. In some embodiments, nickase activity reduces the formation of double-stranded breaks at the target site. In some embodiments, the endonuclease domain creates a staggered nick structure in the first and second strands of a target DNA. In some embodiments, a staggered nick structure generates free 3' overhangs at the target site. In some embodiments, free 3' overhangs at the target site improve editing efficiency, e.g., by enhancing access and annealing of a 3' homology region of a template nucleic acid. In some embodiments, a staggered nick structure reduces the formation of double-stranded breaks at the target site.

In some embodiments, the endonuclease domain cleaves both strands of a target DNA, e.g., results in blunt-end cleavage of a target with no ssDNA overhangs on either side of the cut-site. The amino acid sequence of an endonuclease domain of a gene modifying system described herein may be at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99% identical to the amino acid sequence of an endonuclease domain described herein, e.g., an endonuclease domain as described herein.

In certain embodiments, the heterologous endonuclease is Fok1 or a functional fragment thereof. In certain embodiments, the heterologous endonuclease is a Holliday junction resolvase or homolog thereof, such as the Holliday junction resolving enzyme from Sulfolobus solfataricus—Ssol Hje (Govindaraju et al., *Nucleic Acids Research* 44:7, 2016). In certain embodiments, the heterologous endonuclease is the endonuclease of the large fragment of a spliceosomal protein, such as Prp8 (Mahbub et al., *Mobile DNA* 8:16, 2017). In certain embodiments, the heterologous endonuclease is derived from a CRISPR-associated protein, e.g., Cas9. In certain embodiments, the heterologous endonuclease is engineered to have only ssDNA cleavage activity, e.g., only nickase activity, e.g., be a Cas9 nickase, e.g.,

SpCas9 with D10A, H840A, or N863A mutations. Table 8 provides exemplary Cas proteins and mutations associated with nickase activity. In still other embodiments, homologous endonuclease domains are modified, for example by site-specific mutation, to alter DNA endonuclease activity. In still other embodiments, endonuclease domains are modified to reduce DNA-sequence specificity, e.g., by truncation to remove domains that confer DNA-sequence specificity or mutation to inactivate regions conferring DNA-sequence specificity.

In some embodiments, the endonuclease domain has nickase activity and does not form double-stranded breaks. In some embodiments, the endonuclease domain forms single-stranded breaks at a higher frequency than double-stranded breaks, e.g., at least 90%, 95%, 96%, 97%, 98%, or 99% of the breaks are single-stranded breaks, or less than 10%, 5%, 4%, 3%, 2%, or 1% of the breaks are double-stranded breaks. In some embodiments, the endonuclease forms substantially no double-stranded breaks. In some embodiments, the endonuclease does not form detectable levels of double-stranded breaks.

In some embodiments, the endonuclease domain has nickase activity that nicks the target site DNA of the first strand; e.g., in some embodiments, the endonuclease domain cuts the genomic DNA of the target site near to the site of alteration on the strand that will be extended by the writing domain. In some embodiments, the endonuclease domain has nickase activity that nicks the target site DNA of the first strand and does not nick the target site DNA of the second strand. For example, when a polypeptide comprises a CRISPR-associated endonuclease domain having nickase activity, in some embodiments, said CRISPR-associated endonuclease domain nicks the target site DNA strand containing the PAM site (e.g., and does not nick the target site DNA strand that does not contain the PAM site). As a further example, when a polypeptide comprises a CRISPR-associated endonuclease domain having nickase activity, in some embodiments, said CRISPR-associated endonuclease domain nicks the target site DNA strand not containing the PAM site (e.g., and does not nick the target site DNA strand that contains the PAM site).

In some other embodiments, the endonuclease domain has nickase activity that nicks the target site DNA of the first strand and the second strand. Without wishing to be bound by theory, after a writing domain (e.g., RT domain) of a polypeptide described herein polymerizes (e.g., reverse transcribes) from the heterologous object sequence of a template nucleic acid (e.g., template RNA), the cellular DNA repair machinery must repair the nick on the first DNA strand. The target site DNA now contains two different sequences for the first DNA strand: one corresponding to the original genomic DNA (e.g., having a free 5′ end) and a second corresponding to that polymerized from the heterologous object sequence (e.g., having a free 3′ end). It is thought that the two different sequences equilibrate with one another, first one hybridizing the second strand, then the other, and which sequence the cellular DNA repair apparatus incorporates into its repaired target site may be a stochastic process. Without wishing to be bound by

theory, it is thought that introducing an additional nick to the second-strand may bias the cellular DNA repair machinery to adopt the heterologous object sequence-based sequence more frequently than the original genomic sequence (Anzalone et al. Nature 576:149-157 (2019)). In some embodiments, the additional nick is positioned at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, or 150 nucleotides 5′ or 3′ of the target site modification (e.g., the insertion, deletion, or substitution) or to the nick on the first strand.

Alternatively or additionally, without wishing to be bound by theory, it is thought that an additional nick to the second strand may promote second-strand synthesis. In some embodiments, where the gene modifying system has inserted or substituted a portion of the first strand, synthesis of a new sequence corresponding to the insertion/substitution in the second strand is necessary.

In some embodiments, the polypeptide comprises a single domain having endonuclease activity (e.g., a single endonuclease domain) and said domain nicks both the first strand and the second strand. For example, in such an embodiment the endonuclease domain may be a CRISPR-associated endonuclease domain, and the template nucleic acid (e.g., template RNA) comprises a gRNA spacer that directs nicking of the first strand and an additional gRNA spacer that directs nicking of the second strand. In some embodiments, the polypeptide comprises a plurality of domains having endonuclease activity, and a first endonuclease domain nicks the first strand and a second endonuclease domain nicks the second strand (optionally, the first endonuclease domain does not (e.g., cannot) nick the second strand and the second endonuclease domain does not (e.g., cannot) nick the first strand).

In some embodiments, the endonuclease domain is capable of nicking a first strand and a second strand. In some embodiments, the first and second strand nicks occur at the same position in the target site but on opposite strands. In some embodiments, the second strand nick occurs in a staggered location, e.g., upstream or downstream, from the first nick. In some embodiments, the endonuclease domain generates a target site deletion if the second strand nick is upstream of the first strand nick. In some embodiments, the endonuclease domain generates a target site duplication if the second strand nick is downstream of the first strand nick. In some embodiments, the endonuclease domain generates no duplication and/or deletion if the first and second strand nicks occur in the same position of the target site. In some embodiments, the endonuclease domain has altered activity depending on protein conformation or RNA-binding status, e.g., which promotes the nicking of the first or second strand (e.g., as described in Christensen et al. PNAS 2006; incorporated by reference herein in its entirety).

In some embodiments, the endonuclease domain comprises a meganuclease, or a functional fragment thereof. In some embodiments, the endonuclease domain comprises a homing endonuclease, or a functional fragment thereof. In some embodiments, the endonuclease domain comprises a meganuclease from the LAGLIDADG, GIY-YIG, HNH, His-Cys Box, or PD-(D/E) XK families, or a

functional fragment or variant thereof, e.g., which possess conserved amino acid motifs, e.g., as indicated in the family names. In some embodiments, the endonuclease domain comprises a meganuclease, or fragment thereof, chosen from, e.g., I-SmaMI (Uniprot F7WD42), I-SceI (Uniprot P03882), I-AniI (Uniprot P03880), I-DmoI (Uniprot P21505), I-CreI (Uniprot P05725), I-TevI (Uniprot P13299), I-OnuI (Uniprot Q4VWW5), or I-BmoI (Uniprot Q9ANR6). In some embodiments, the meganuclease is naturally monomeric, e.g., I-SceI, I-TevI, or dimeric, e.g., I-CreI, in its functional form. For example, the LAGLIDADG meganucleases with a single copy of the LAGLIDADG motif generally form homodimers, whereas members with two copies of the LAGLIDADG motif are generally found as monomers. In some embodiments, a meganuclease that normally forms as a dimer is expressed as a fusion, e.g., the two subunits are expressed as a single ORF and, optionally, connected by a linker, e.g., an I-CreI dimer fusion (Rodriguez-Fornes et al. Gene Therapy 2020; incorporated by reference herein in its entirety). In some embodiments, a meganuclease, or a functional fragment thereof, is altered to favor nickase activity for one strand of a double-stranded DNA molecule, e.g., I-SceI (K122I and/or K223I) (Niu et al. J Mol Biol 2008), I-AniI (K227M) (McConnell Smith et al. PNAS 2009), I-DmoI (Q42A and/or K120M) (Molina et al. J Biol Chem 2015). In some embodiments, a meganuclease or functional fragment thereof possessing this preference for single-strand cleavage is used as an endonuclease domain, e.g., with nickase activity. In some embodiments, an endonuclease domain comprises a meganuclease, or a functional fragment thereof, which naturally targets or is engineered to target a safe harbor site, e.g., an I-CreI targeting SH6 site (Rodriguez-Fornes et al., *supra*). In some embodiments, an endonuclease domain comprises a meganuclease, or a functional fragment thereof, with a sequence tolerant catalytic domain, e.g., I-TevI recognizing the minimal motif CNNNG (Kleinstiver et al. PNAS 2012). In some embodiments, a target sequence tolerant catalytic domain is fused to a DNA binding domain, e.g., to direct activity, e.g., by fusing I-TevI to: (i) zinc fingers to create Tev-ZFEs (Kleinstiver et al. PNAS 2012), (ii) other meganucleases to create MegaTevs (Wolfs et al. Nucleic Acids Res 2014), and/or (iii) Cas9 to create TevCas9 (Wolfs et al. PNAS 2016).

In some embodiments, the endonuclease domain comprises a restriction enzyme, e.g., a Type IIS or Type IIP restriction enzyme. In some embodiments, the endonuclease domain comprises a Type IIS restriction enzyme, e.g., FokI, or a fragment or variant thereof. In some embodiments, the endonuclease domain comprises a Type IIP restriction enzyme, e.g., PvuII, or a fragment or variant thereof. In some embodiments, a dimeric restriction enzyme is expressed as a fusion such that it functions as a single chain, e.g., a FokI dimer fusion (Minczuk et al. Nucleic Acids Res 36(12):3926-3938 (2008)).

The use of additional endonuclease domains is described, for example, in Guha and Edgell Int J Mol Sci 18(22):2565 (2017), which is incorporated herein by reference in its entirety.

In some embodiments, a gene modifying polypeptide comprises a modification to an endonuclease domain, e.g., relative to a wild-type Cas protein. In some embodiments, the endonuclease domain comprises an addition, deletion, replacement, or modification to the amino acid sequence of the wild-type Cas protein. In some embodiments, the endonuclease domain is modified to include a heterologous functional domain that binds specifically to and/or induces endonuclease cleavage of a target nucleic acid (e.g., DNA) sequence of interest. In some embodiments, the endonuclease domain comprises a zinc finger. In embodiments, the endonuclease domain comprising the Cas domain is associated with a guide RNA (gRNA), e.g., as described herein. In some embodiments, the endonuclease domain is modified to include a functional domain that does not target a specific target nucleic acid (e.g., DNA) sequence. In embodiments, the endonuclease domain comprises a Fok1 domain.

In some embodiments, the endonuclease domain is associated with the target dsDNA *in vitro* at a frequency at least about 5-fold or 10-fold higher than with a scrambled dsDNA. In some embodiments, the endonuclease domain is associated with the target dsDNA *in vitro* at a frequency at least about 5-fold or 10-fold higher than with a scrambled dsDNA, e.g., in a cell (e.g., a HEK293T cell). In some embodiments, the frequency of association between the endonuclease domain and the target DNA or scrambled DNA is measured by ChIP-seq, e.g., as described in He and Pu (2010) *Curr. Protoc Mol Biol* Chapter 21 (incorporated by reference herein in its entirety).

In some embodiments, the endonuclease domain can catalyze the formation of a nick at a target sequence, e.g., to an increase of at least about 5-fold or 10-fold relative to a non-target sequence (e.g., relative to any other genomic sequence in the genome of the target cell). In some embodiments, the level of nick formation is determined using NickSeq, e.g., as described in Elacqua et al. (2019) *bioRxiv* doi.org/10.1101/867937 (incorporated herein by reference in its entirety).

In some embodiments, the endonuclease domain is capable of nicking DNA *in vitro*. In embodiments, the nick results in an exposed base. In embodiments, the exposed base can be detected using a nuclease sensitivity assay, e.g., as described in Chaudhry and Weinfeld (1995) *Nucleic Acids Res* 23(19):3805-3809 (incorporated by reference herein in its entirety). In embodiments, the level of exposed bases (e.g., detected by the nuclease sensitivity assay) is increased by at least 10%, 50%, or more relative to a reference endonuclease domain. In some embodiments, the reference endonuclease domain is an endonuclease domain from Cas9 of S. pyogenes.

In some embodiments, the endonuclease domain is capable of nicking DNA in a cell. In embodiments, the endonuclease domain is capable of nicking DNA in a HEK293T cell. In embodiments, an unrepaired nick that undergoes replication in the absence of Rad51 results in increased NHEJ rates at the site of the nick, which can be detected, e.g., by using a Rad51 inhibition assay, e.g., as described in Bothmer et al. (2017) *Nat Commun* 8:13905 (incorporated by reference herein in its entirety). In

embodiments, NHEJ rates are increased above 0-5%. In embodiments, NHEJ rates are increased to 20-70% (e.g., between 30%-60% or 40-50%), e.g., upon Rad51 inhibition.

In some embodiments, the endonuclease domain releases the target after cleavage. In some embodiments, release of the target is indicated indirectly by assessing for multiple turnovers by the enzyme, e.g., as described in Yourik at al. RNA 25(1):35-44 (2019) (incorporated herein by reference in its entirety) and shown in FIG. 2. In some embodiments, the k_{exp} of an endonuclease domain is 1 x 10^{-3} – 1 x 10-5 min-1 as measured by such methods.

In some embodiments, the endonuclease domain has a catalytic efficiency (k_{cat}/K_m) greater than about 1 x 10^8 s^{-1} M^{-1} *in vitro*. In embodiments, the endonuclease domain has a catalytic efficiency greater than about 1 x 10^5, 1 x 10^6, 1 x 10^7, or 1 x 10^8, s^{-1} M^{-1} *in vitro*. In embodiments, catalytic efficiency is determined as described in Chen et al. (2018) *Science* 360(6387):436-439 (incorporated herein by reference in its entirety). In some embodiments, the endonuclease domain has a catalytic efficiency (k_{cat}/K_m) greater than about 1 x 10^8 s^{-1} M^{-1} in cells. In embodiments, the endonuclease domain has a catalytic efficiency greater than about 1 x 10^5, 1 x 10^6, 1 x 10^7, or 1 x 10^8 s^{-1} M^{-1} in cells.

Gene modifying polypeptides comprising Cas domains

In some embodiments, a gene modifying polypeptide described herein comprises a Cas domain. In some embodiments, the Cas domain can direct the gene modifying polypeptide to a target site specified by a gRNA spacer, thereby modifying a target nucleic acid sequence in "cis". In some embodiments, a gene modifying polypeptide is fused to a Cas domain. In some embodiments, a gene modifying polypeptide comprises a CRISPR/Cas domain (also referred to herein as a CRISPR-associated protein). In some embodiments, a CRISPR/Cas domain comprises a protein involved in the clustered regulatory interspaced short palindromic repeat (CRISPR) system, e.g., a Cas protein, and optionally binds a guide RNA, e.g., single guide RNA (sgRNA).

CRISPR systems are adaptive defense systems originally discovered in bacteria and archaea. CRISPR systems use RNA-guided nucleases termed CRISPR-associated or "Cas" endonucleases (e. g., Cas9 or Cpf1) to cleave foreign DNA. For example, in a typical CRISPR-Cas system, an endonuclease is directed to a target nucleotide sequence (e. g., a site in the genome that is to be sequence-edited) by sequence-specific, non-coding "guide RNAs" that target single- or double-stranded DNA sequences. Three classes (I-III) of CRISPR systems have been identified. The class II CRISPR systems use a single Cas endonuclease (rather than multiple Cas proteins). One class II CRISPR system includes a type II Cas endonuclease such as Cas9, a CRISPR RNA ("crRNA"), and a trans-activating crRNA ("tracrRNA"). The crRNA contains a "spacer" sequence, a typically about 20-nucleotide RNA sequence that corresponds to a target DNA sequence ("protospacer"). In the wild-type system, and in some engineered systems, crRNA also contains a region that binds to the tracrRNA to form a partially double-stranded

structure that is cleaved by RNase III, resulting in a crRNA/tracrRNA hybrid molecule. A crRNA/tracrRNA hybrid then directs the Cas endonuclease to recognize and cleave a target DNA sequence. A target DNA sequence is generally adjacent to a "protospacer adjacent motif" ("PAM") that is specific for a given Cas endonuclease and required for cleavage activity at a target site matching the spacer of the crRNA. CRISPR endonucleases identified from various prokaryotic species have unique PAM sequence requirements, e.g., as listed for exemplary Cas enzymes in Table 7; examples of PAM sequences include 5′-NGG (Streptococcus pyogenes), 5′-NNAGAA (Streptococcus thermophilus CRISPR1), 5′-NGGNG (Streptococcus thermophilus CRISPR3), and 5′-NNNGATT (Neisseria meningiditis). Some endonucleases, e.g., Cas9 endonucleases, are associated with G-rich PAM sites, e. g., 5′-NGG, and perform blunt-end cleaving of the target DNA at a location 3 nucleotides upstream from (5′ from) the PAM site. Another class II CRISPR system includes the type V endonuclease Cpf1, which is smaller than Cas9; examples include AsCpf1 (from Acidaminococcus sp.) and LbCpf1 (from Lachnospiraceae sp.). Cpf1-associated CRISPR arrays are processed into mature crRNAs without the requirement of a tracrRNA; in other words, a Cpf1 system, in some embodiments, comprises only Cpf1 nuclease and a crRNA to cleave a target DNA sequence. Cpf1 endonucleases, are typically associated with T-rich PAM sites, e. g., 5′-TTN. Cpf1 can also recognize a 5′-CTA PAM motif. Cpf1 typically cleaves a target DNA by introducing an offset or staggered double-strand break with a 4- or 5-nucleotide 5′ overhang, for example, cleaving a target DNA with a 5-nucleotide offset or staggered cut located 18 nucleotides downstream from (3′ from) from a PAM site on the coding strand and 23 nucleotides downstream from the PAM site on the complimentary strand; the 5-nucleotide overhang that results from such offset cleavage allows more precise genome editing by DNA insertion by homologous recombination than by insertion at blunt-end cleaved DNA. See, e.g., Zetsche et al. (2015) Cell, 163:759 – 771.

A variety of CRISPR associated (Cas) genes or proteins can be used in the technologies provided by the present disclosure and the choice of Cas protein will depend upon the particular conditions of the method. Specific examples of Cas proteins include class II systems including Cas1, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas10, Cpf1, C2C1, or C2C3. In some embodiments, a Cas protein, e.g., a Cas9 protein, may be from any of a variety of prokaryotic species. In some embodiments a particular Cas protein, e.g., a particular Cas9 protein, is selected to recognize a particular protospacer-adjacent motif (PAM) sequence. In some embodiments, a DNA-binding domain or endonuclease domain includes a sequence targeting polypeptide, such as a Cas protein, e.g., Cas9. In certain embodiments a Cas protein, e.g., a Cas9 protein, may be obtained from a bacteria or archaea or synthesized using known methods. In certain embodiments, a Cas protein may be from a gram-positive bacteria or a gram-negative bacteria. In certain embodiments, a Cas protein may be from a Streptococcus (e.g., a S. pyogenes, or a S.

thermophilus), a Francisella (e.g., an F. novicida), a Staphylococcus (e.g., an S. aureus), an Acidaminococcus (e.g., an Acidaminococcus sp. BV3L6), a Neisseria (e.g., an N. meningitidis), a Cryptococcus, a Corynebacterium, a Haemophilus, a Eubacterium, a Pasteurella, a Prevotella, a Veillonella, or a Marinobacter.

In some embodiments, a gene modifying polypeptide may comprise the amino acid sequence of SEQ ID NO: 4000 below, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity thereto. In embodiments, the amino acid sequence of SEQ ID NO: 4000 below, or the sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity thereto, is positioned at the N-terminal end of the gene modifying polypeptide. In embodiments, the amino acid sequence of SEQ ID NO: 4000 below, or the sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity thereto, is positioned within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, or 30 amino acids of the N-terminal end of the gene modifying polypeptide.

Exemplary N-terminal NLS-Cas9 domain

MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHP
IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVN
TEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNR
EKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDR
EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPE
NIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD
MYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY
KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY
SVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR
KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGG (SEQ ID NO: 4000)

In some embodiments, a gene modifying polypeptide may comprise the amino acid sequence of SEQ ID NO: 4001 below, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%,

98%, 99% identity thereto. In embodiments, the amino acid sequence of SEQ ID NO: 4001 below, or the sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity thereto, is positioned at the C-terminal end of the gene modifying polypeptide. In embodiments, the amino acid sequence of SEQ ID NO: 4001 below, or the sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity thereto, is positioned within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, or 30 amino acids of the C-terminal end of the gene modifying polypeptide.

Exemplary C-terminal sequence comprising an NLS

AGKRTADGSEFEKRTADGSEFESPKKKAKVE **(SEQ ID NO: 4001)**

Exemplary benchmarking sequence

MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHP
IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVN
TEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY
KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNR
EKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDR
EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPE
NIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD
MYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY
KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY
SVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR

KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGSGGSSGGSSGSETPGTSESATPESSGG
SSGGSSGGTLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATS
TPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNK
RVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLT
WTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLG
NLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRL
FIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY
AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV
EALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHG
TRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQ
ALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSI
IHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSEFEAGKRT
ADGSEFEKRTADGSEFESPKKKAKVE (SEQ ID NO: 4002)

In some embodiments, a gene modifying polypeptide may comprise a Cas domain as listed in Table 7 or 8, or a functional fragment thereof, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity thereto.

Table 7. CRISPR/Cas Proteins, Species, and Mutations

Name	*Enzyme*	*Species*	*# of AAs*	*PAM*	*Mutations to alter PAM recognition*	*Mutations to make catalytically dead*
FnCas9	Cas9	*Francisella novicida*	1629	5'-NGG-3'	Wt	D11A/H969A/N995A
FnCas9 RHA	Cas9	*Francisella novicida*	1629	5'-YG-3'	E1369R/E1449H/R1556A	D11A/H969A/N995A
SaCas9	Cas9	*Staphylococcus aureus*	1053	5'-NNGRRT-3'	Wt	D10A/H557A

SaCas9 KKH	Cas9	*Staphylococcus aureus*	1053	5'-NNNRRT-3'	E782K/N968K/R1015H	D10A/H557A
SpCas9	Cas9	*Streptococcus pyogenes*	1368	5'-NGG-3'	Wt	D10A/D839A/H840A/N863A
SpCas9 VQR	Cas9	*Streptococcus pyogenes*	1368	5'-NGA-3'	D1135V/R1335Q/T1337R	D10A/D839A/H840A/N863A
AsCpf1 RR	Cpf1	*Acidaminococcus sp. BV3L6*	1307	5'-TYCV-3'	S542R/K607R	E993A
AsCpf1 RVR	Cpf1	*Acidaminococcus sp. BV3L6*	1307	5'-TATV-3'	S542R/K548V/N552R	E993A
FnCpf1	Cpf1	*Francisella novicida*	1300	5'-NTTN-3'	Wt	D917A/E1006A/D1255A
NmCas9	Cas9	*Neisseria meningitidis*	1082	5'-NNNGATT-3'	Wt	D16A/D587A/H588A/N611A

Table 8 Amino Acid Sequences of CRISPR/Cas Proteins, Species, and Mutations

Variant	Parental Host(s)	Protein Sequence	Nickase (HNH)	Nickase (HNH)	Nickase (RuvC)
Nme2Cas9	Neisseria meningitidis	MAAFKPNPINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKTGDSLAMARRLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDENGLIKSLPNTPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVANNAHALQTGDFRTPAELALNKFEKESGHIRNQRGDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSSELQDEIGTAFSLFKTDEDITGRLKDRVQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLVRLNEKGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSR	N611A	H588A	D16A

		EWQEFKARVETSRFPRSKKQRILLQKFDEDGFKECNLNDTRYVNRFLCQFVA DHILLTGKGKRRVFASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVACS TVAMQQKITRFVRYKEMNAFDGKTIDKETGKVLHQKTHFPQPWEFFAQEV MIRVFGKPDGKPEFEEADTPEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNR KMSGAHKDTLRSAKRFVKHNEKISVKRVWLTEIKLADLENMVNYKNGREIEL YEALKARLEAYGGNAKQAFDPKDNPFYKKGGQLVKAVRVEKTQESGVLLNK KNAYTIADNGDMVRVDVFCKVDKKGKNQYFIVPIYAWQVAENILPDIDCKG YRIDDSYTFCFSLHKYDLIAFQKDEKSKVEFAYYINCDSSNGRFYLAWHDKGS KEQQFRISTQNLVLIQKYQVNELGKEIRPCRLKKRPPVR			
PpnCas9	Pasteurella pneumotropica	MQNNPLNYILGLDLGIASIGWAVVEIDEESSPIRLIDVGVRTFERAEVAKTGE SLALSRRLARSSRRLIKRRAERLKKAKRLLKAEKILHSIDEKLPINVWQLRVKGL KEKLERQEWAAVLLHLSKHRGYLSQRKNEGKSDNKELGALLSGIASNHQML QSSEYRTPAEIAVKKFQVEEGHIRNQRGSYTHTFSRLDLLAEMELLFQRQAEL GNSYTSTTLLENLTALLMWQKPALAGDAILKMLGKCTFEPSEYKAAKNSYSA ERFVWLTKLNNLRILENGTERALNDNERFALLEQPYEKSKLTYAQVRAMLAL SDNAIFKGVRYLGEDKKTVESKTTLIEMKFYHQIRKTLGSAELKKEWNELKGN SDLLDEIGTAFSLYKTDDDICRYLEGKLPERVLNALLENLNFDKFIQLSLKALHQ ILPLMLQGQRYDEAVSAIYGDHYGKKSTETTRLLPTIPADEIRNPVVLRTLTQA RKVINAVVRLYGSPARIHIETAREVGKSYQDRKKLEKQQEDNRKQRESAVKK FKEMFPHFVGEPKGKDILKMRLYELQQAKCLYSGKSLELHRLLEKGYVEVDH ALPFSRTWDDSFNNKVLVLANENQNKGNLTPYEWLDGKNNSERWQHFVV RVQTSGFSYAKKQRILNHKLDEKGFIERNLNDTRYVARFLCNFIADNMLLVG KGKRNVFASNGQITALLRHRWGLQKVREQNDRHHALDAVVVACSTVAMQ QKITRFVRYNEGNVFSGERIDRETGEIIPLHFPSPWAFFKENVEIRIFSENPKLE LENRLPDYPQYNHEWVQPLFVSRMPTRKMTGQGHMETVKSAKRLNEGLS VLKVPLTQLKLSDLERMVNRDREIALYESLKARLEQFGNDPAKAFAEPFYKKG GALVKAVRLEQTQKSGVLVRDGNGVADNASMVRVDVFTKGGKYFLVPIYT WQVAKGILPNRAATQGKDENDWDIMDEMATFQFSLCQNDLIKLVTKKKTI FGYFNGLNRATSNINIKEHDLDKSKGKLGIYLEVGVKLAISLEKYQVDELGKNI RPCRPTKRQHVR	N605A	H582A	D13A
SauCas9	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKK DGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYE GPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLN NLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTST GKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSE LTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKV DLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSK DAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYS LEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQ YLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNL VDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG YKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQ EYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVN NLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPL YKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKL SLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQA EFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPP RIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A	H557A	D10A
SauCas9-KKH	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKK DGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYE GPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLN NLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTST GKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSE LTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKV DLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSK	N580A	H557A	D10A

		DAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYS LEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQ YLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNL VDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG YKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQ EYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIV NNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNP LYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVK LSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQ AEFIASFYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRP PHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG			
SauriCas9	Staphylococcus auricularis	MQENQQKQNYILGLDIGITSVGYGLIDSKTREVIDAGVRLFPEADSENNSNR RSKRGARRLKRRRIHRLNRVKDLLADYQMIDLNNVPKSTDPYTIRVKGLREPL TKEEFAIALLHIAKRRGLHNISVSMGDEEQDNELSTKQQLQKNAQQLQDKY VCELQLERLTNINKVRGEKNRFKTEDFVKEVKQLCETQRQYHNIDDQFIQQY IDLVSTRREYFEGPGNGSPYGWDGDLLKWYEKLMGRCTYFPEELRSVKYAYS ADLFNALNDLNNLVVTRDDNPKLEYYEKYHIIENVFKQKKNPTLKQIAKEIGV QDYDIRGYRITKSGKPQFTSFKLYHDLKNIFEQAKYLEDVEMLDEIAKILTIYQ DEISIKKALDQLPELLTESEKSQIAQLTGYTGTHRLSLKCIHIVIDELWESPENQ MEIFTRLNLKPKKVEMSEIDSIPTTLVDEFILSPVVKRAFIQSIKVINAVINRFGL PEDIIIELAREKNSKDRRKFINKLQKQNEATRKKIEQLLAKYGNTNAKYMIEKI KLHDMQEGKCLYSLEAIPLEDLLSNPTHYEVDHIIPRSVSFDNSLNNKVLVKQ SENSKKGNRTPYQYLSSNESKISYNQFKQHILNLSKAKDRISKKKRDMLLEER DINKFEVQKEFINRNLVDTRYATRELSNLLKTYFSTHDYAVKVKTINGGFTNH LRKVWDFKKHRNHGYKHHAEDALVIANADFLFKTHKALRRTDKILEQPGLE VNDTTVKVDTEEKYQELFETPKQVKNIKQFRDFKYSHRVDKKPNRQLINDTL YSTREIDGETYVVQTLKDLYAKDNEKVKKLFTERPQKILMYQHDPKTFEKLM TILNQYAEAKNPLAAYYEDKGEYVTKYAKKGNGPAIHKIKYIDKKLGSYLDVS NKYPETQNKLVKLSLKSFRFDIYKCEQGYKMVSIGYLDVLKKDNYYYIPKDKYE AEKQKKKIKESDLFVGSFYYNDLIMYEDELFRVIGVNSDINNLVELNMVDITY KDFCEVNNVTGEKRIKKTIGKRVVLIEKYTTDILGNLYKTPLPKKPQLIFKRGEL	N588A	H565A	D15A
SauriCas9-KKH	Staphylococcus auricularis	MQENQQKQNYILGLDIGITSVGYGLIDSKTREVIDAGVRLFPEADSENNSNR RSKRGARRLKRRRIHRLNRVKDLLADYQMIDLNNVPKSTDPYTIRVKGLREPL TKEEFAIALLHIAKRRGLHNISVSMGDEEQDNELSTKQQLQKNAQQLQDKY VCELQLERLTNINKVRGEKNRFKTEDFVKEVKQLCETQRQYHNIDDQFIQQY IDLVSTRREYFEGPGNGSPYGWDGDLLKWYEKLMGRCTYFPEELRSVKYAYS ADLFNALNDLNNLVVTRDDNPKLEYYEKYHIIENVFKQKKNPTLKQIAKEIGV QDYDIRGYRITKSGKPQFTSFKLYHDLKNIFEQAKYLEDVEMLDEIAKILTIYQ DEISIKKALDQLPELLTESEKSQIAQLTGYTGTHRLSLKCIHIVIDELWESPENQ MEIFTRLNLKPKKVEMSEIDSIPTTLVDEFILSPVVKRAFIQSIKVINAVINRFGL PEDIIIELAREKNSKDRRKFINKLQKQNEATRKKIEQLLAKYGNTNAKYMIEKI KLHDMQEGKCLYSLEAIPLEDLLSNPTHYEVDHIIPRSVSFDNSLNNKVLVKQ SENSKKGNRTPYQYLSSNESKISYNQFKQHILNLSKAKDRISKKKRDMLLEER DINKFEVQKEFINRNLVDTRYATRELSNLLKTYFSTHDYAVKVKTINGGFTNH LRKVWDFKKHRNHGYKHHAEDALVIANADFLFKTHKALRRTDKILEQPGLE VNDTTVKVDTEEKYQELFETPKQVKNIKQFRDFKYSHRVDKKPNRKLINDTL YSTREIDGETYVVQTLKDLYAKDNEKVKKLFTERPQKILMYQHDPKTFEKLM TILNQYAEAKNPLAAYYEDKGEYVTKYAKKGNGPAIHKIKYIDKKLGSYLDVS NKYPETQNKLVKLSLKSFRFDIYKCEQGYKMVSIGYLDVLKKDNYYYIPKDKYE AEKQKKKIKESDLFVGSFYKNDLIMYEDELFRVIGVNSDINNLVELNMVDITY KDFCEVNNVTGEKHIKKTIGKRVVLIEKYTTDILGNLYKTPLPKKPQLIFKRGEL	N588A	H565A	D15A
ScaCas9-Sc++	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALL FDSGETAEATRLKRTARRRYTRRKNRIRYLQEIFANEMAKLDDSFFQRLEESF LVEEDKKNERHPIFGNLADEVAYHRNYPTIYHLRKKLADSPEKADLRLIYLALA HIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEESPLDEIEVDAKGILSA RLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTPNFKSNFDLTEDAKLQLSKD TYDDDLDELLGQIGDQYADLFSAAKNLSDAILLSDILRSNSEVTKAPLSASMV KRYDEHHQDLALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVGADKKLRKRS GKLATEEEFYKFIKPILEKMDGAEELLAKLNREDLLRKQRTFDNGSIPHQIHLK	N872A	H849A	D10A

		ELHAILRRQEEFYPFLKENREKIEKILTFRIPYYVGPLARGNSRFAWLTRKSEEA ITPWNFEEVVDKGASAQSFIERMTNFDEQLPNKKVLPKHSLLYEYFTVYNEL TKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS VEIIGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIE ERLKTYAHLFDDKVMKQLKRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKS DGFSNRNFMQLIHDDSLTFKEEIEKAQVSGQGDSLHEQIADLAGSPAIKKGIL QTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKRIEEGIKELE SQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSEADKAGFIKRQLVETRQITKHVARILDSRMNTKRDKN DKPIREVKVITLKSKLVSDFRKDFQLYKVRDINNYHHAHDAYLNAVVGTALIK KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNIMNFFKTEVKL ANGEIRKRPLIETNGETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTG GFSKESILSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKVEKGKAKKL KSVKVLVGITIMEKGSYEKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRR MLASAKELQKANELVLPQHLVRLLYYTQNISATTGSNNLGYIEQHREEFKEIF EKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLLKYTSFGASGGFT FLDLDVKQGRLRYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD			
SpyCas9	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
SpyCas9-NG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ	N863A	H840A	D10A

		ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES IRPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA RFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAF KYFDTTIDRKVYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD			
SpyCas9-SpRY	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAERTRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES IRPKRNSDKLIARKKDWDPKKYGGFLWPTVAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AKQLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTRLGAPRAF KYFDTTIDPKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
St1Cas9	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQG RRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFI ALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLER YQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEF INRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEF RAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAK LFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETL DKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGW HNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIY NPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKAN KDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYT GKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ ALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLV DTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYH HHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFK APYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADE TYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPN KQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDIT PKDSNNKVVLQSVSPWRADVYFNKTTGKYEILGLKYADLQFEKGTGTYKISQ EKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPKQKH YVELKPYDKQKFEGGEALIKVLGNVANSGQCKKGLGKSNISIYKVRTDVLGN QHIIKNEGDKPKLDF	N622A	H599A	D9A

BlatCas9	Brevibacillus laterosporus	MAYTMGIDVGIASCGWAIVDLERQRIIDIGVRTFEKAENPKNGEALAVPRRE ARSSRRRLRRKKHRIERLKHMFVRNGLAVDIQHLEQTLRSQNEIDVWQLRV DGLDRMLTQKEWLRVLIHLAQRRGFQSNRKTDGSSEDGQVLVNVTENDRL MEEKDYRTVAEMMVKDEKFSDHKRNKNGNYHGVVSRSSLLVEIHTLFETQ RQHHNSLASKDFELEYVNIWSAQRPVATKDQIEKMIGTCTFLPKEKRAPKAS WHFQYFMLLQTINHIRITNVQGTRSLNKEEIEQVVNMALTKSKVSYHDTRKI LDLSEEYQFVGLDYGKEDEKKKVESKETIIKLDDYHKLNKIFNEVELAKGETWE ADDYDTVAYALTFFKDDEDIRDYLQNKYKDSKNRLVKNLANKEYTNELIGKV STLSFRKVGHLSLKALRKIIPFLEQGMTYDKACQAAGFDFQGISKKKRSVVLP VIDQISNPVVNRALTQTRKVINALIKKYGSPETIHIETARELSKTFDERKNITKD YKENRDKNEHAKKHLSELGIINPTGLDIVKYKLWCEQQGRCMYSNQPISFER LKESGYTEVDHIIPYSRSMNDSYNNRVLVMTRENREKGNQTPFEYMGNDT QRWYEFEQRVTTNPQIKKEKRQNLLLKGFTNRRELEMLERNLNDTRYITKYL SHFISTNLEFSPSDKKKKVVNTSGRITSHLRSRWGLEKNRGQNDLHHAMDAI VIAVTSDSFIQQVTNYYKRKERRELNGDDKFPLPWKFFREEVIARLSPNPKEQ IEALPNHFYSEDELADLQPIFVSRMPKRSITGEAHQAQFRRVVGKTKEGKNIT AKKTALVDISYDKNGDFNMYGRETDPATYEAIKERYLEFGGNVKKAFSTDLH KPKKDGTKGPLIKSVRIMENKTLVHPVNKGKGVVYNSSIVRTDVFQRKEKYY LLPVYVTDVTKGKLPNKVIVAKKGYHDWIEVDDSFTFLFSLYPNDLIFIRQNPK KKISLKKRIESHSISDSKEVQEIHAYYKGVDSSTAAIEFIIHDGSYYAKGVGVQN LDCFEKYQVDILGNYFKVKGEKRLELETSDSNHKGKDVNSIKSTSR	N607A	H584A	D8A
cCas9-v16	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKK DGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYE GPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLN NLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTST GKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSE LTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKV DLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSK DAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYS LEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQ YLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNL VDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG YKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQ EYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIV NNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNP LYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVK LSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQ AEFIASFYKNDLIKINGELYRVIGVNSDKNNLIEVNMIDITYREYLENMNDKRP PHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A	H557A	D10A
cCas9-v17	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKK DGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYE GPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLN NLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTST GKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSE LTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKV DLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSK DAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYS LEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQ YLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNL VDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG YKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQ EYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIV NNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNP LYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVK LSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQ	N580A	H557A	D10A

		AEFIASFYKNDLIKINGELYRVIGVNNSTRNIVELNMIDITYREYLENMNDKRP PHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG			
cCas9-v21	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKK DGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYE GPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLN NLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTST GKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSE LTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKV DLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSK DAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYS LEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQ YLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNL VDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG YKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQ EYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIV NNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNP LYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVK LSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQ AEFIASFYKNDLIKINGELYRVIGVNSDDRNIIELNMIDITYREYLENMNDKRP PHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A	H557A	D10A
cCas9-v42	Staphylococcus aureus	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKK DGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYE GPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLN NLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTST GKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSE LTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKV DLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSK DAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYS LEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQ YLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNL VDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKG YKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQ EYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIV NNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNP LYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVK LSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQ AEFIASFYKNDLIKINGELYRVIGVNNNRLNKIELNMIDITYREYLENMNDKRP PHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A	H557A	D10A
CdiCas9	Corynebacterium diphtheriae	MKYHVGIDVGTFSVGLAAIEVDDAGMPIKTLSLVSHIHDSGLDPDEIKSAVT RLASSGIARRTRRLYRRKRRRLQQLDKFIQRQGWPVIELEDYSDPLYPWKVR AELAASYIADEKERGEKLSVALRHIARHRGWRNPYAKVSSLYLPDGPSDAFK AIREEIKRASGQPVPETATVGQMVTLCELGTLKLRGEGGVLSARLQQSDYAR EIQEICRMQEIGQELYRKIIDVVFAAESPKGSASSRVGKDPLQPGKNRALKAS DAFQRYRIAALIGNLRVRVDGEKRILSVEEKNLVFDHLVNLTPKKEPEWVTIA EILGIDRGQLIGTATMTDDGERAGARPPTHDTNRSIVNSRIAPLVDWWKTA SALEQHAMVKALSNAEVDDFDSPEGAKVQAFFADLDDDVHAKLDSLHLPV GRAAYSEDTLVRLTRRMLSDGVDLYTARLQEFGIEPSWTPPTPRIGEPVGNP AVDRVLKTVSRWLESATKTWGAPERVIIEHVREGFVTEKRAREMDGDMRR RAARNAKLFQEMQEKLNVQGKPSRADLWRYQSVQRQNCQCAYCGSPITF SNSEMDHIVPRAGQGSTNTRENLVAVCHRCNQSKGNTPFAIWAKNTSIEG VSVKEAVERTRHWVTDTGMRSTDFKKFTKAVVERFQRATMDEEIDARSME SVAWMANELRSRVAQHFASHGTTVRVYRGSLTAEARRASGISGKLKFFDGV GKSRLDRRHHAIDAAVIAFTSDYVAETLAVRSNLKQSQAHRQEAPQWREFT GKDAEHRAAWRVWCQKMEKLSALLTEDLRDDRVVVMSNVRLRLGNGSA HKETIGKLSKVKLSSQLSVSDIDKASSEALWCALTREPGFDPKEGLPANPERHI	N597A	H573A	D8A

		RVNGTHVYAGDNIGLFPVSAGSIALRGGYAELGSSFHHARVYKITSGKKPAF AMLRVYTIDLLPYRNQDLFSVELKPQTMSMRQAEKKLRDALATGNAEYLG WLVVDDELVVDTSKIATDQVKAVEAELGTIRRWRVDGFFSPSKLRLRPLQM SKEGIKKESAPELSKIIDRPGWLPAVNKLFSDGNVTVVRRDSLGRVRLESTAH LPVTWKVQ			
CjeCas9	Campylobacter jejuni	MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSA RKRLARRKARLNHLKHLIANEFKLNYEDYQSFDESLAKAYKGSLISPYELRFRA LNELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEKLANYQS VGEYLYKEYFQKFKENSKEFTNVRNKKESYERCIAQSFLKDELKLIFKKQREFG FSFSKKFEEEVLSVAFYKRALKDFSHLVGNCSFFTDEKRAPKNSPLAFMFVAL TRIINLLNNLKNTEGILYTKDDLNALLNEVLKNGTLTYKQTKKLLGLSDDYEFK GEKGTYFIEFKKYKEFIKALGEHNLSQDDLNEIAKDITLIKDEIKLKKALAKYDLN QNQIDSLSKLEFKDHLNISFKALKLVTPLMLEGKKYDEACNELNLKVAINEDK KDFLPAFNETYYKDEVTNPVVLRAIKEYRKVLNALLKKYGKVHKINIELAREVG KNHSQRAKIEKEQNENYKAKKDAELECEKLGLKINSKNILKLRLFKEQKEFCAY SGEKIKISDLQDEKMLEIDHIYPYSRSFDDSYMNKVLVFTKQNQEKLNQTPFE AFGNDSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNLNDTRYI ARLVLNYTKDYLDFLPLSDDENTKLNDTQKGSKVHVEAKSGMLTSALRHTW GFSAKDRNNHLHHAIDAVIIAYANNSIVKAFSDFKKEQESNSAELYAKKISELD YKNKRKFFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALHEETFRKEEEFYQSY GGKEGVLKALELGKIRKVNGKIVKNGDMFRVDIFKHKKTNKFYAVPIYTMDF ALKVLPNKAVARSKKGEIKDWILMDENYEFCFSLYKDSLILIQTKDMQEPEFV YYNAFTSSTVSLIVSKHDNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVFEK YIVSALGEVTKAEFRQREDFKK	N582A	H559A	D8A
GeoCas9	Geobacillus stearothermophilus	MRYKIGLDIGITSVGWAVMNLDIPRIEDLGVRIFDRAENPQTGESLALPRRLA RSARRRLRRRKHRLERIRRLVIREGILTKEELDKLFEEKHEIDVWQLRVEALDR KLNNDELARVLLHLAKRRGFKSNRKSERSNKENSTMLKHIEENRAILSSYRTV GEMIVKDPKFALHKRNKGENYTNTIARDDLEREIRLIFSKQREFGNMSCTEEF ENEYITIWASQRPVASKDDIEKKVGFCTFEPKEKRAPKATYTFQSFIAWEHIN KLRLISPSGARGLTDEERRLLYEQAFQKNKITYHDIRTLLHLPDDTYFKGIVYDR GESRKQNENIRFLELDAYHQIRKAVDKVYGKGKSSSFLPIDFDTFGYALTLFKD DADIHSYLRNEYEQNGKRMPNLANKVYDNELIEELLNLSFTKFGHLSLKALRS ILPYMEQGEVYSSACERAGYTFTGPKKKQKTMLLPNIPPIANPVVMRALTQA RKVVNAIIKKYGSPVSIHIELARDLSQTFDERRKTKKEQDENRKKNETAIRQL MEYGLTLNPTGHDIVKFKLWSEQNGRCAYSLQPIEIERLLEPGYVEVDHVIPY SRSLDDSYTNKVLVLTRENREKGNRIPAEYLGVGTERWQQFETFVLTNKQFS KKKRDRLLRLHYDENEETEFKNRNLNDTRYISRFFANFIREHLKFAESDDKQK VYTVNGRVTAHLRSRWEFNKNREESDLHHAVDAVIVACTTPSDIAKVTAFY QRREQNKELAKKTEPHFPQPWPHFADELRARLSKHPKESIKALNLGNYDDQ KLESLQPVFVSRMPKRSVTGAAHQETLRRYVGIDERSGKIQTVVKTKLSEIKL DASGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGEPGP VIRTVKIIDTKNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPVYTMDIM KGILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIELPREKTVKTAAGEE INVKDVFVYYKTIDSANGGLELISHDHRFSLRGVGSRTLKRFEKYQVDVLGNI YKVRGEKRVGLASSAHSKPGKTIRPLQSTRD	N605A	H582A	D8A
iSpyMacCas9	Streptococcus spp.	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRKLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLKREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV	N863A	H840A	D10A

		VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEIQTVGQNGG LFDDNPKSPLEVTPSKLVPLKKELNPKKYGGYQKPTTAYPVLLITDTKQLIPISV MNKKQFEQNPVKFLRDRGYQQVGKNDFIKLPKYTLVDIGDGIKRLWASSKEI HKGNQLVVSKKSQILLYHAHHLDSDLSNDYLQNHNQQFDVLFNEIISFSKKC KLGKEHIQKIENVYSNKKNSASIEELAESFIKLLGFTQLGATSPFNFLGVKLNQ KQYKGKKDYILPCTEGTLIRQSITGLYETRVDLSKIGEDSGGSGGSKRTADGSE FES			
NmeCas9	Neisseria meningitidis	MAAFKPNSINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPK TGDSLAMARRLARSVRRLTRRRAHRLLRTRRLLKREGVLQAANFDENGLIKS LPNTPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELG ALLKGVAGNAHALQTGDFRTPAELALNKFEKESGHIRNQRSDYSHTFSRKDL QAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTF EPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKS KLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGL KDKKSPLNLSPELQDEIGTAFSLFKTDEDITGRLKDRIQPEILEALLKHISFDKFV QISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNP VVLRALSQARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRK DREKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLGRLNEK GYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSRE WQEFKARVETSRFPRSKKQRILLQKFDEDGFKERNLNDTRYVNRFLCQFVA DRMRLTGKGKKRVFASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVA CSTVAMQQKITRFVRYKEMNAFDGKTIDKETGEVLHQKTHFPQPWEFFAQ EVMIRVFGKPDGKPEFEEADTLEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAP NRKMSGQGHMETVKSAKRLDEGVSVLRVPLTQLKLKDLEKMVNREREPKL YEALKARLEAHKDDPAKAFAEPFYKYDKAGNRTQQVKAVRVEQVQKTGVW VRNHNGIADNATMVRVDVFEKGDKYYLVPIYSWQVAKGILPDRAVVQGKD EEDWQLIDDSFNFKFSLHPNDLVEVITKKARMFGYFASCHRGTGNINIRIHD LDHKIGKNGILEGIGVKTALSFQKYQIDELGKEIRPCRLKKRPPVR	N611A	H588A	D16A
ScaCas9	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALL FDSGETAEATRLKRTARRRYTRRKNRIRYLQEIFANEMAKLDDSFFQRLEESF LVEEDKKNERHPIFGNLADEVAYHRNYPTIYHLRKKLADSPEKADLRLIYLALA HIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEESPLDEIEVDAKGILSA RLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTPNFKSNFDLTEDAKLQLSKD TYDDDLDELLGQIGDQYADLFSAAKNLSDAILLSDILRSNSEVTKAPLSASMV KRYDEHHQDLALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVDGIGIKHRKRTT KLATQEEFYKFIKPILEKMDGAEELLAKLNRDDLLRKQRTFDNGSIPHQIHLKE LHAILRRQEEFYPFLKENREKIEKILTFRIPYYVGPLARGNSRFAWLTRKSEEAI TPWNFEEVVDKGASAQSFIERMTNFDEQLPNKKVLPKHSLLYEYFTVYNELT KVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSV EIIGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEE RLKTYAHLFDDKVMKQLKRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKS DGFSNRNFMQLIHDDSLTFKEEIEKAQVSGQGDSLHEQIADLAGSPAIKKGIL QTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKRIEEGIKELE SQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSEADKAGFIKRQLVETRQITKHVARILDSRMNTKRDKN DKPIREVKVITLKSKLVSDFRKDFQLYKVRDINNYHHAHDAYLNAVVGTALIK KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNIMNFFKTEVKL ANGEIRKRPLIETNGETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTG GFSKESILSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKVEKGKAKKL KSVKVLVGITIMEKGSYEKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRR MLASATELQKANELVLPQHLVRLLYYTQNISATTGSNNLGYIEQHREEFKEIF	N872A	H849A	D10A

		EKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLLKYTSFGASGGFT FLDLDVKQGRLRYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD			
ScaCas9-HiFi-Sc++	Streptococcus canis	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALL FDSGETAEATRLKRTARRRYTRRKNRIRYLQEIFANEMAKLDDSFFQRLEESF LVEEDKKNERHPIFGNLADEVAYHRNYPTIYHLRKKLADSPEKADLRLIYLALA HIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEESPLDEIEVDAKGILSA RLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTPNFKSNFDLTEDAKLQLSKD TYDDDLDELLGQIGDQYADLFSAAKNLSDAILLSDILRSNSEVTKAPLSASMV KRYDEHHQDLALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVGADKKLRKRS GKLATEEEFYKFIKPILEKMDGAEELLAKLNRDDLLRKQRTFDNGSIPHQIHLK ELHAILRRQEEFYPFLKENREKIEKILTFRIPYYVGPLARGNSRFAWLTRKSEEA ITPWNFEEVVDKGASAQSFIERMTNFDEQLPNKKVLPKHSLLYEYFTVYNEL TKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS VEIIGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIE ERLKTYAHLFDDKVMKQLKRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKS DGFSNANFMQLIHDDSLTFKEEIEKAQVSGQGDSLHEQIADLAGSPAIKKGIL QTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKRIEEGIKELE SQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSEADKAGFIKRQLVETRQITKHVARILDSRMNTKRDKN DKPIREVKVITLKSKLVSDFRKDFQLYKVRDINNYHHAHDAYLNAVVGTALIK KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNIMNFFKTEVKL ANGEIRKRPLIETNGETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTG GFSKESILSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKVEKGKAKKL KSVKVLVGITIMEKGSYEKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRR MLASAKELQKANELVLPQHLVRLLYYTQNISATTGSNNLGYIEQHREEFKEIF EKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLLKYTSFGASGGFT FLDLDVKQGRLRYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD	N872A	H849A	D10A
SpyCas9-3var-NRRH	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MVKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQ GDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKGNSDKLIARKKDWDPKKYGGFNSPTAAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIGFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGVLHKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGVPAA FKYFDTTIDKKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
SpyCas9-3var-NRTH	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS			

		KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MVKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQ GDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKGNSDKLIARKKDWDPKKYGGFNSPTVAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIGFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS ASVLHKGNELALPSKYVNFLYLASHYEKLKGSSEDNKQKQLFVEQHKHYLDEI IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGASAAF KYFDTTIGRKLYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD			
SpyCas9-3var-NRCH	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MVKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQ GDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRN FMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVV DELVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKGNSDKLIARKKDWDPKKYGGFNSPTVAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGVLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAA FKYFDTTINRKQYNTTKEVLDATLIRQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
SpyCas9-HF1	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA	N863A	H840A	D10A

		HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD			
SpyCas9-QQR1	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA RELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADAQLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTFKQKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
SpyCas9-SpG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK	N863A	H840A	D10A

		LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFLWPTVAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AKQLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAA FKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD			
SpyCas9-VQR	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
SpyCas9-VRER	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAH MIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLS KDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQ EDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED RFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYA HLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANR NFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKV VDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQ ILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSF LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPK LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKES ILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKE LLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA RELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEII EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A

SpyCas9-xCas	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKLYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
SpyCas9-xCas-NG	Streptococcus pyogenes	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKLYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIDRKVYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
St1Cas9-CNRZ1066	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETL	N622A	H599A	D9A

		DKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGW HNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIY NPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKAN KDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYT GKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ ALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLV DTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYH HHAVDALIIAASSQLNLWKKQKNTLVSYSEEQLLDIETGELISDDEYKESVFKA PYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKKDET YVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNK QMNEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLLGNPIDI TPENSKNKVVLQSLKPWRTDVYFNKATGKYEILGLKYADLQFEKGTGTYKIS QEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTLPKQK HYVELKPYDKQKFEGGEALIKVLGNVANGGQCIKGLAKSNISIYKVRTDVLG NQHIIKNEGDKPKLDF			
St1Cas9-LMG1831	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQG RRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFI ALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLER YQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEF INRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEF RAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAK LFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETL DKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGW HNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIY NPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKAN KDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYT GKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ ALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLV DTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYH HHAVDALIIAASSQLNLWKKQKNTLVSYSEEQLLDIETGELISDDEYKESVFKA PYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKKDET YVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNK QMNEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLLGNPIDI TPENSKNKVVLQSLKPWRTDVYFNKNTGKYEILGLKYADLQFEKKTGTYKISQ EKYNGIMKEEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPNVK YYVELKPYSKDKFEKNESLIEILGSADKSGRCIKGLGKSNISIYKVRTDVLGNQH IIKNEGDKPKLDF	N622A	H599A	D9A
St1Cas9-MTH17CL396	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQG RRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFI ALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLER YQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEF INRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEF RAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAK LFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETL DKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGW HNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIY NPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKAN KDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYT GKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQ ALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLV DTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYH HHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFK APYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADE TYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPN KQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDIT PKDSNNKVVLQSLKPWRTDVYFNKNTGKYEILGLKYSDMQFEKGTGKYSISK EQYENIKVREGVDENSEFKFTLYKNDLLLLKDSENGEQILLRFTSRNDTSKHYV ELKPYNRQKFEGSEYLIKSLGTVAKGGQCIKGLGKSNISIYKVRTDVLGNQHII KNEGDKPKLDF	N622A	H599A	D9A

St1Cas9-TH1477	Streptococcus thermophilus	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSLKPWRTDVYFNKNTGKYEILGLKYSDMQFEKGTGKYSISKEQYENIKVREGVDENSEFKFTLYKNDLLLLKDSENGEQILLRFTSRNDTSKHYVELKPYNRQKFEGSEYLIKSLGTVVKGGRCIKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF	N622A	H599A	D9A
sRGN3.1	Staphylococcus spp.	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGSRRLKRRRIHRLERVKLLLTEYDLINKEQIPTSNNPYQIRVKGLSEILSKDELAIALLHLAKRRGIHNVDVAADKEETASDSLSTKDQINKNAKFLESRYVCELQKERLENEGHVRGVENRFLTKDIVREAKKIIDTQMQYYPEIDETFKEKYISLVETRREYFEGPGQGSPFGWNGDLKKWYEMLMGHCTYFPQELRSVKYAYSADLFNALNDLNNLIIQRDNSEKLEYHEKYHIIENVFKQKKKPTLKQIAKEIGVNPEDIKGYRITKSGTPEFTSFKLFHDLKKVVKDHAILDDIDLLNQIAEILTIYQDKDSIVAELGQLEYLMSEADKQSISELTGYTGTHSLSLKCMNMIIDELWHSSMNQMEVFTYLNMRPKKYELKGYQRIPTDMIDDAILSPVVKRTFIQSINVINKVIEKYGIPEDIIIELARENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHDQQEGKCLYSLESIPLEDLLNNPNHYEVDHIIPRSVSFDNSYHNKVLVKQSENSKKSNLTPYQYFNSGKSKLSYNQFKQHILNLSKSQDRISKKKKEYLLEERDINKFEVQKEFINRNLVDTRYATRELTNYLKAYFSANNMNVKVKTINGSFTDYLRKVWKFKKERNHGYKHHAEDALIIANADFLFKENKKLKAVNSVLEKPEIETKQLDIQVDSEDNYSEMFIIPKQVQDIKDFRNFKYSHRVDKKPNRQLINDTLYSTRKKDNSTYIVQTIKDIYAKDNTTLKKQFDKSPEKFLMYQHDPRTFEKLEVIMKQYANEKNPLAKYHEETGEYLTKYSKKNNGPIVKKIKLLGNKVGNHLDVTNKYENKKLVKLSIKNYRFDVYLTEKGYKFVTIAYLNVFKKDNYYYIPKDKYQELKEKKKIKDTDQFIASFYKNDLIKLNGDLYKIIGVNSDDRNIIELDYYDIKYKDYCEINNIKGEPRIKKTIGKKTESIEKFTTDVLGNLYLHSTEKAPQLIFKRGL	N585A	H562A	D10A
sRGN3.3	Staphylococcus spp.	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGSRRLKRRRIHRLERVKLLLTEYDLINKEQIPTSNNPYQIRVKGLSEILSKDELAIALLHLAKRRGIHNVDVAADKEETASDSLSTKDQINKNAKFLESRYVCELQKERLENEGHVRGVENRFLTKDIVREAKKIIDTQMQYYPEIDETFKEKYISLVETRREYFEGPGQGSPFGWNGDLKKWYEMLMGHCTYFPQELRSVKYAYSADLFNALNDLNNLIIQRDNSEKLEYHEKYHIIENVFKQKKKPTLKQIAKEIGVNPEDIKGYRITKSGTPEFTSFKLFHDLKKVVKDHAILDDIDLLNQIAEILTIYQDKDSIVAELGQLEYLMSEADKQSISELTGYTGTHSLSLKCMNMIIDELWHSSMNQMEVFTYLNMRPKKYELKGYQRIPTDMIDDAILSPVVKRTFIQSINVINKVIEKYGIPEDIIIELARENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHDQQEGKCLYSLESIPLEDLLNNPNHYEVDHIIPRSVSFDNSYHNKVLVKQSENSKKSNLTPYQYFNSGKSKLSYNQFKQHILNLSKSQDRISKKKKEYLLEERDINKFEVQKEFINRNLVDTRYATRELTSYLKAYFSANNMDVKVKTINGSFTNHLRKVWRFDKYRNHGYKHHAEDALIIANADFLFKENKKLQNTNKILEKPTIENNTKKVTVEKEEDYNNVFETPKLVEDIKQYRDYKFSHRVDKKPNRQLINDTLYSTRMKDEHDYIVQTITDIYGKDNTNLKKQFNKNPEKFLMYQNDPKTFEKLSIIMKQYSDEKNPLAKYYEETGEYLTKYSKKNNGPIVKKIKLLGNKVGNHLDVTNKYEN	N585A	H562A	D10A

		STKKLVKLSIKNYRFDVYLTEKGYKFVTIAYLNVFKKDNYYYIPKDKYQELKEKK KIKDTDQFIASFYKNDLIKLNGDLYKIIGVNSDDRNIIELDYYDIKYKDYCEINNI KGEPRIKKTIGKKTESIEKFTTDVLGNLYLHSTEKAPQLIFKRGL			

In some embodiments, a Cas protein requires a protospacer adjacent motif (PAM) to be present in or adjacent to a target DNA sequence for the Cas protein to bind and/or function. In some embodiments, the PAM is or comprises, from 5′ to 3′, NGG, YG, NNGRRT, NNNRRT, NGA, TYCV, TATV, NTTN, or NNNGATT, where N stands for any nucleotide, Y stands for C or T, R stands for A or G, and V stands for A or C or G. In some embodiments, a Cas protein is a protein listed in Table 7 or 8. In some embodiments, a Cas protein comprises one or more mutations altering its PAM. In some embodiments, a Cas protein comprises E1369R, E1449H, and R1556A mutations or analogous substitutions to the amino acids corresponding to said positions. In some embodiments, a Cas protein comprises E782K, N968K, and R1015H mutations or analogous substitutions to the amino acids corresponding to said positions. In some embodiments, a Cas protein comprises D1135V, R1335Q, and T1337R mutations or analogous substitutions to the amino acids corresponding to said positions. In some embodiments, a Cas protein comprises S542R and K607R mutations or analogous substitutions to the amino acids corresponding to said positions. In some embodiments, a Cas protein comprises S542R, K548V, and N552R mutations or analogous substitutions to the amino acids corresponding to said positions. Exemplary advances in the engineering of Cas enzymes to recognize altered PAM sequences are reviewed in Collias et al Nature Communications 12:555 (2021), incorporated herein by reference in its entirety.

In some embodiments, the Cas protein is catalytically active and cuts one or both strands of the target DNA site. In some embodiments, cutting the target DNA site is followed by formation of an alteration, e.g., an insertion or deletion, e.g., by the cellular repair machinery.

In some embodiments, the Cas protein is modified to deactivate or partially deactivate the nuclease, e.g., nuclease-deficient Cas9. Whereas wild-type Cas9 generates double-strand breaks (DSBs) at specific DNA sequences targeted by a gRNA, a number of CRISPR endonucleases having modified functionalities are available, for example: a “nickase” version of Cas9 that has been partially deactivated generates only a single-strand break; a catalytically inactive Cas9 (“dCas9”) does not cut target DNA. In some embodiments, dCas9 binding to a DNA sequence may interfere with transcription at that site by steric hindrance. In some embodiments, dCas9 binding to an anchor sequence may interfere with (e.g., decrease or prevent) genomic complex (e.g., ASMC) formation and/or maintenance. In some embodiments, a DNA-binding domain comprises a catalytically inactive Cas9, e.g., dCas9. Many catalytically inactive Cas9 proteins are known in the art. In some embodiments, dCas9 comprises

mutations in each endonuclease domain of the Cas protein, e.g., D10A and H840A or N863A mutations. In some embodiments, a catalytically inactive or partially inactive CRISPR/Cas domain comprises a Cas protein comprising one or more mutations, e.g., one or more of the mutations listed in Table 7. In some embodiments, a Cas protein described on a given row of Table 7 comprises one, two, three, or all of the mutations listed in the same row of Table 7. In some embodiments, a Cas protein, e.g., not described in Table 7, comprises one, two, three, or all of the mutations listed in a row of Table 7 or a corresponding mutation at a corresponding site in that Cas protein.

In some embodiments, a Cas9 derivative with enhanced activity may be used in the gene modification polypeptide. In some embodiments, a Cas9 derivative may comprise mutations that improve activity of the HNH endonuclease domain, e.g., SpyCas9 R221K, N394K, or mutations that improve R-loop formation, e.g., SpyCas9 L1245V, or comprise a combination of such mutations, e.g., SpyCas9 R221K/N394K, SpyCas9 N394K/L1245V, SpyCas9 R221K/L1245V, or SpyCas9 R221K/N394K/L1245V (see, e.g., Spencer and Zhang Sci Rep 7:16836 (2017), the Cas9 derivatives and comprising mutations of which are incorporated herein by reference). In some embodiments, a Cas9 derivative may comprise one or more types of mutations described herein, e.g., PAM-modifying mutations, protein stabilizing mutations, activity enhancing mutations, and/or mutations partially or fully inactivating one or two endonuclease domains relative to the parental enzyme (e.g., one or more mutations to abolish endonuclease activity towards one or both strands of a target DNA, e.g., a nickase or catalytically dead enzyme). In some embodiments, a Cas9 enzyme used in a system described herein may comprise mutations that confer nickase activity toward the enzyme (e.g., SpyCas9 N863A or H840A) in addition to mutations improving catalytic efficiency (e.g., SpyCas9 R221K, N394K, and/or L1245V). In some embodiments, a Cas9 enzyme used in a system described herein is a SpyCas9 enzyme or derivative that further comprises an N863A mutation to confer nickase activity in addition to R221K and N394K mutations to improve catalytic efficiency.

In some embodiments, a catalytically inactive, e.g., dCas9, or partially deactivated Cas9 protein comprises a D11 mutation (e.g., D11A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a H969 mutation (e.g., H969A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a N995 mutation (e.g., N995A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, comprises mutations at one, two, or three of positions D11, H969, and N995 (e.g., D11A, H969A, and N995A mutations) or analogous substitutions to the amino acids corresponding to said positions.

In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a D10 mutation (e.g., a D10A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a H557 mutation (e.g., a H557A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, comprises a D10 mutation (e.g., a D10A mutation) and a H557 mutation (e.g., a H557A mutation) or analogous substitutions to the amino acids corresponding to said positions.

In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a D839 mutation (e.g., a D839A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a H840 mutation (e.g., a H840A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a N863 mutation (e.g., a N863A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, comprises a D10 mutation (e.g., D10A), a D839 mutation (e.g., D839A), a H840 mutation (e.g., H840A), and a N863 mutation (e.g., N863A) or analogous substitutions to the amino acids corresponding to said positions.

In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a E993 mutation (e.g., a E993A mutation) or an analogous substitution to the amino acid corresponding to said position.

In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a D917 mutation (e.g., a D917A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a a E1006 mutation (e.g., a E1006A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a D1255 mutation (e.g., a D1255A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, comprises a D917 mutation (e.g., D917A), a E1006 mutation (e.g., E1006A), and a D1255 mutation (e.g., D1255A) or analogous substitutions to the amino acids corresponding to said positions.

In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a D16 mutation (e.g., a D16A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g.,

dCas9, or partially deactivated Cas9 protein comprises a D587 mutation (e.g., a D587A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a partially deactivated Cas domain has nickase activity. In some embodiments, a partially deactivated Cas9 domain is a Cas9 nickase domain. In some embodiments, the catalytically inactive Cas domain or dead Cas domain produces no detectable double strand break formation. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a H588 mutation (e.g., a H588A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, or partially deactivated Cas9 protein comprises a N611 mutation (e.g., a N611A mutation) or an analogous substitution to the amino acid corresponding to said position. In some embodiments, a catalytically inactive Cas9 protein, e.g., dCas9, comprises a D16 mutation (e.g., D16A), a D587 mutation (e.g., D587A), a H588 mutation (e.g., H588A), and a N611 mutation (e.g., N611A) or analogous substitutions to the amino acids corresponding to said positions.

In some embodiments, a DNA-binding domain or endonuclease domain may comprise a Cas molecule comprising or linked (e.g., covalently) to a gRNA (e.g., a template nucleic acid, e.g., template RNA, comprising a gRNA).

In some embodiments, an endonuclease domain or DNA binding domain comprises a *Streptococcus pyogenes* Cas9 (SpCas9) or a functional fragment or variant thereof. In some embodiments, the endonuclease domain or DNA binding domain comprises a modified SpCas9. In embodiments, the modified SpCas9 comprises a modification that alters protospacer-adjacent motif (PAM) specificity. In embodiments, the PAM has specificity for the nucleic acid sequence 5′-NGT-3′. In embodiments, the modified SpCas9 comprises one or more amino acid substitutions, e.g., at one or more of positions L1111, D1135, G1218, E1219, A1322, of R1335, e.g., selected from L1111R, D1135V, G1218R, E1219F, A1322R, R1335V. In embodiments, the modified SpCas9 comprises the amino acid substitution T1337R and one or more additional amino acid substitutions, e.g., selected from L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337H, T1337Q, and T1337M, or corresponding amino acid substitutions thereto. In embodiments, the modified SpCas9 comprises: (i) one or more amino acid substitutions selected from D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, and T1337; and (ii) one or more amino acid substitutions selected from L1111R, G1218R, E1219F, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, T1337L, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, and T1337M, or corresponding amino acid substitutions thereto.

In some embodiments, the endonuclease domain or DNA binding domain comprises a Cas domain, e.g., a Cas9 domain. In embodiments, the endonuclease domain or DNA binding domain comprises a nuclease-active Cas domain, a Cas nickase (nCas) domain, or a nuclease-inactive Cas (dCas) domain. In embodiments, the endonuclease domain or DNA binding domain comprises a nuclease-active Cas9 domain, a Cas9 nickase (nCas9) domain, or a nuclease-inactive Cas9 (dCas9) domain. In some embodiments, the endonuclease domain or DNA binding domain comprises a Cas9 domain of Cas9 (e.g., dCas9 and nCas9), Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, or Cas12i. In some embodiments, the endonuclease domain or DNA binding domain comprises a Cas9 (e.g., dCas9 and nCas9), Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, or Cas12i. In some embodiments, the endonuclease domain or DNA binding domain comprises an S. pyogenes or an S. thermophilus Cas9, or a functional fragment thereof. In some embodiments, the endonuclease domain or DNA binding domain comprises a Cas9 sequence, e.g., as described in Chylinski, Rhun, and Charpentier (2013) RNA Biology 10:5, 726-737; incorporated herein by reference. In some embodiments, the endonuclease domain or DNA binding domain comprises the HNH nuclease subdomain and/or the RuvC1 subdomain of a Cas, e.g., Cas9, e.g., as described herein, or a variant thereof. In some embodiments, the endonuclease domain or DNA binding domain comprises Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, or Cas12i. In some embodiments, the endonuclease domain or DNA binding domain comprises a Cas polypeptide (e.g., enzyme), or a functional fragment thereof. In embodiments, the Cas polypeptide (e.g., enzyme) is selected from Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas8a, Cas8b, Cas8c, Cas9 (e.g., Csn1 or Csx12), Cas10, Cas10d, Cas12a/Cpfl, Cas12b/C2cl, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, Csy1 , Csy2, Csy3, Csy4, Cse1, Cse2, Cse3, Cse4, Cse5e, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csx11, Csf1, Csf2, CsO, Csf4, Csd1, Csd2, Cst1, Cst2, Csh1, Csh2, Csa1, Csa2, Csa3, Csa4, Csa5, Type II Cas effector proteins, Type V Cas effector proteins, Type VI Cas effector proteins, CARF, DinG, Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12b/C2c1, Cas12c/C2c3, SpCas9(K855A), eSpCas9(1.1), SpCas9-HF1, hyper accurate Cas9 variant (HypaCas9), homologues thereof, modified or engineered versions thereof, and/or functional fragments thereof. In embodiments, the Cas9 comprises one or more substitutions, e.g., selected from H840A, D10A, P475A, W476A, N477A, D1125A, W1126A, and D1127A. In embodiments, the Cas9 comprises one or more mutations at positions selected from: D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or A987, e.g., one or more substitutions selected from D10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A, H983A, A984A, and/or D986A. In some embodiments, the endonuclease domain or DNA binding domain

comprises a Cas (e.g., Cas9) sequence from Corynebacterium ulcerans, Corynebacterium diphtheria, Spiroplasma syrphidicola, Prevotella intermedia, Spiroplasma taiwanense, Streptococcus iniae, Belliella baltica, Psychroflexus torquis, Streptococcus thermophilus, Listeria innocua, Campylobacter jejuni, Neisseria meningitidis, Streptococcus pyogenes, or Staphylococcus aureus, or a fragment or variant thereof.

In some embodiments, the endonuclease domain or DNA binding domain comprises a Cpf1 domain, e.g., comprising one or more substitutions, e.g., at position D917, E1006A, D1255 or any combination thereof, e.g., selected from D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A, and D917A/E1006A/D1255A.

In some embodiments, the endonuclease domain or DNA binding domain comprises spCas9, spCas9-VRQR(SEQ ID NO: 19), spCas9- VRER(SEQ ID NO: 20), xCas9 (sp), saCas9, saCas9-KKH, spCas9-MQKSER(SEQ ID NO: 21), spCas9-LRKIQK(SEQ ID NO: 22), or spCas9- LRVSQL(SEQ ID NO: 23).

In some embodiments, a gene modifying polypeptide has an endonuclease domain comprising a Cas9 nickase, e.g., Cas9 H840A. In embodiments, the Cas9 H840A has the following amino acid sequence:

Cas9 nickase (H840A):

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR
TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEK
YPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLF
EENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDA
KLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDE
HHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK
LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN
SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNE
LTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRF
NASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK
RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERM
KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSF
LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS
ELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF

FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD

In some embodiments, a gene modifying polypeptide comprises a dCas9 sequence comprising a D10A and/or H840A mutation, e.g., the following sequence:

SMDKKYSIGLAIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL
KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYH
EKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQ
LFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED
AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYD
EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV
KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARG
NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN
ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRF
NASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLK
RRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERM
KRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSF
LKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS
ELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYK
VREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF
FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKY
VNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH
RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQL
GGD (SEQ ID NO: 7)

TAL Effectors and Zinc Finger Nucleases

In some embodiments, an endonuclease domain or DNA-binding domain comprises a TAL effector molecule. A TAL effector molecule, e.g., a TAL effector molecule that specifically binds a DNA sequence, typically comprises a plurality of TAL effector domains or fragments thereof, and optionally one or more additional portions of naturally occurring TAL effectors (e.g., N- and/or C-terminal of the plurality of TAL effector domains). Many TAL effectors are known to those of skill in the art and are commercially available, e.g., from Thermo Fisher Scientific.

Naturally occurring TALEs are natural effector proteins secreted by numerous species of bacterial pathogens including the plant pathogen *Xanthomonas* which modulates gene expression in host plants and facilitates bacterial colonization and survival. The specific binding of TAL effectors is based on a central repeat domain of tandemly arranged nearly identical repeats of typically 33 or 34 amino acids (the repeat-variable di-residues, RVD domain).

Members of the TAL effectors family differ mainly in the number and order of their repeats. The number of repeats typically ranges from 1.5 to 33.5 repeats and the C-terminal repeat is usually shorter in length (e.g., about 20 amino acids) and is generally referred to as a "half-repeat." Each repeat of the TAL effector generally features a one-repeat-to-one-base-pair correlation with different repeat types exhibiting different base-pair specificity (one repeat recognizes one base-pair on the target gene sequence). Generally, the smaller the number of repeats, the weaker the protein-DNA interactions. A number of 6.5 repeats has been shown to be sufficient to activate transcription of a reporter gene (Scholze et al., 2010).

Repeat to repeat variations occur predominantly at amino acid positions 12 and 13, which have therefore been termed "hypervariable" and which are responsible for the specificity of the interaction with the target DNA promoter sequence, as shown in Table 9 listing exemplary repeat variable diresidues (RVD) and their correspondence to nucleic acid base targets.

Table 9 – RVDs and Nucleic Acid Base Specificity

Target	Possible RVD Amino Acid Combinations												
A	NI	NN	CI	HI	KI								
G	NN	GN	SN	VN	LN	DN	QN	EN	HN	RH	NK	AN	FN
C	HD	RD	KD	ND	AD								
T	NG	HG	VG	IG	EG	MG	YG	AA	EP	VA	QG	KG	RG

Accordingly, it is possible to modify the repeats of a TAL effector to target specific DNA sequences. Further studies have shown that the RVD NK can target G. Target sites of TAL effectors also tend to include a T flanking the 5′ base targeted by the first repeat, but the exact mechanism of this recognition is not known. More than 113 TAL effector sequences are known to date. Non-limiting examples of TAL effectors from *Xanthomonas* include, Hax2, Hax3, Hax4, AvrXa7, AvrXa10 and AvrBs3.

Accordingly, the TAL effector domain of a TAL effector molecule described herein may be derived from a TAL effector from any bacterial species (e.g., *Xanthomonas* species such as the African strain of *Xanthomonas oryzae* pv. *Oryzae* (Yu et al. 2011), *Xanthomonas campestris* pv. *raphani* strain

756C and *Xanthomonas oryzae* pv. *oryzicola*strain BLS256 (Bogdanove et al. 2011). In some embodiments, the TAL effector domain comprises an RVD domain as well as flanking sequence(s) (sequences on the N-terminal and/or C-terminal side of the RVD domain) also from the naturally occurring TAL effector. It may comprise more or fewer repeats than the RVD of the naturally occurring TAL effector. The TAL effector molecule can be designed to target a given DNA sequence based on the above code and others known in the art. The number of TAL effector domains (e.g., repeats (monomers or modules)) and their specific sequence can beselected based on the desired DNA target sequence. For example, TAL effector domains, e.g., repeats, may be removed or added in order to suit a specific target sequence. In an embodiment, the TAL effector molecule of the present invention comprises between 6.5 and 33.5 TAL effector domains, e.g., repeats. In an embodiment, TAL effector molecule of the present invention comprises between 8 and 33.5 TAL effector domains, e.g., repeats, e.g., between 10 and 25 TAL effector domains, e.g., repeats, e.g., between 10 and 14 TAL effector domains, e.g., repeats.

In some embodiments, the TAL effector molecule comprises TAL effector domains that correspond to a perfect match to the DNA target sequence. In some embodiments, a mismatch between a repeat and a target base-pair on the DNA target sequence is permitted as along as it allows for the function of the polypeptide comprising the TAL effector molecule. In general, TALE binding is inversely correlated with the number of mismatches. In some embodiments, the TAL effector molecule of a polypeptide of the present invention comprises no more than 7 mismatches, 6 mismatches, 5 mismatches, 4 mismatches, 3 mismatches, 2 mismatches, or 1 mismatch, and optionally no mismatch, with the target DNA sequence. Without wishing to be bound by theory, in general the smaller the number of TAL effector domains in the TAL effector molecule, the smaller the number of mismatches will be tolerated and still allow for the function of the polypeptide comprising the TAL effector molecule. The binding affinity is thought to depend on the sum of matching repeat-DNA combinations. For example, TAL effector molecules having 25 TAL effector domains or more may be able to tolerate up to 7 mismatches.

In addition to the TAL effector domains, the TAL effector molecule of the present invention may comprise additional sequences derived from a naturally occurring TAL effector. The length of the C-terminal and/or N-terminal sequence(s) included on each side of the TAL effector domain portion of the TAL effector molecule can vary and be selected by one skilled in the art, for example based on the studies of Zhang et al. (2011). Zhang et al., have characterized a number of C-terminal and N-terminal truncation mutants in Hax3 derived TAL-effector based proteins and have identified key elements, which contribute to optimal binding to the target sequence and thus activation of transcription. Generally, it was found that transcriptional activity is inversely correlated with the length of N-terminus. Regarding the C-terminus, an important element for DNA binding residues within the first 68 amino acids of the Hax 3 sequence was identified. Accordingly, in some embodiments, the first 68 amino acids on the C-terminal side of the

TAL effector domains of the naturally occurring TAL effector is included in the TAL effector molecule. Accordingly, in an embodiment, a TAL effector molecule comprises 1) one or more TAL effector domains derived from a naturally occurring TAL effector; 2) at least 70, 80, 90, 100, 110, 120, 130, 140, 150, 170, 180, 190, 200, 220, 230, 240, 250, 260, 270, 280 or more amino acids from the naturally occurring TAL effector on the N-terminal side of the TAL effector domains; and/or 3) at least 68, 80, 90, 100, 110, 120, 130, 140, 150, 170, 180, 190, 200, 220, 230, 240, 250, 260 or more amino acids from the naturally occurring TAL effector on the C-terminal side of the TAL effector domains.

In some embodiments, an endonuclease domain or DNA-binding domain is or comprises a Zn finger molecule. A Zn finger molecule comprises a Zn finger protein, e.g., a naturally occurring Zn finger protein or engineered Zn finger protein, or fragment thereof. Many Zn finger proteins are known to those of skill in the art and are commercially available, e.g., from Sigma-Aldrich.

In some embodiments, a Zn finger molecule comprises a non-naturally occurring Zn finger protein that is engineered to bind to a target DNA sequence of choice. See, for example, Beerli, et al. (2002) Nature Biotechnol. 20:135-141; Pabo, et al. (2001) Ann. Rev. Biochem. 70:313-340; Isalan, et al. (2001) Nature Biotechnol. 19:656-660; Segal, et al. (2001) Curr. Opin. Biotechnol. 12:632-637; Choo, et al. (2000) Curr. Opin. Struct. Biol. 10:411-416; U.S. Pat. Nos. 6,453,242; 6,534,261; 6,599,692; 6,503,717; 6,689,558; 7,030,215; 6,794,136; 7,067,317; 7,262,054; 7,070,934; 7,361,635; 7,253,273; and U.S. Patent Publication Nos. 2005/0064474; 2007/0218528; 2005/0267061, all incorporated herein by reference in their entireties.

An engineered Zn finger protein may have a novel binding specificity, compared to a naturally-occurring Zn finger protein. Engineering methods include, but are not limited to, rational design and various types of selection. Rational design includes, for example, using databases comprising triplet (or quadruplet) nucleotide sequences and individual Zn finger amino acid sequences, in which each triplet or quadruplet nucleotide sequence is associated with one or more amino acid sequences of zinc fingers which bind the particular triplet or quadruplet sequence. See, for example, U.S. Pat. Nos. 6,453,242 and 6,534,261, incorporated by reference herein in their entireties.

Exemplary selection methods, including phage display and two-hybrid systems, are disclosed in U.S. Pat. Nos. 5,789,538; 5,925,523; 6,007,988; 6,013,453; 6,410,248; 6,140,466; 6,200,759; and 6,242,568; as well as International Patent Publication Nos. WO 98/37186; WO 98/53057; WO 00/27878; and WO 01/88197 and GB 2,338,237. In addition, enhancement of binding specificity for zinc finger proteins has been described, for example, in International Patent Publication No. WO 02/077227.

In addition, as disclosed in these and other references, zinc finger domains and/or multi-fingered zinc finger proteins may be linked together using any suitable linker sequences, including for example,

linkers of 5 or more amino acids in length. See, also, U.S. Pat. Nos. 6,479,626; 6,903,185; and 7,153,949 for exemplary linker sequences 6 or more amino acids in length. The proteins described herein may include any combination of suitable linkers between the individual zinc fingers of the protein. In addition, enhancement of binding specificity for zinc finger binding domains has been described, for example, in co-owned International Patent Publication No. WO 02/077227.

Zn finger proteins and methods for design and construction of fusion proteins (and polynucleotides encoding same) are known to those of skill in the art and described in detail in U.S. Pat. Nos. 6,140,0815; 789,538; 6,453,242; 6,534,261; 5,925,523; 6,007,988; 6,013,453; and 6,200,759; International Patent Publication Nos. WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970; WO 01/88197; WO 02/099084; WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536; and WO 03/016496.

In addition, as disclosed in these and other references, Zn finger proteins and/or multi-fingered Zn finger proteins may be linked together, e.g., as a fusion protein, using any suitable linker sequences, including for example, linkers of 5 or more amino acids in length. See, also, U.S. Pat. Nos. 6,479,626; 6,903,185; and 7,153,949 for exemplary linker sequences 6 or more amino acids in length. The Zn finger molecules described herein may include any combination of suitable linkers between the individual zinc finger proteins and/or multi-fingered Zn finger proteins of the Zn finger molecule.

In certain embodiments, the DNA-binding domain or endonuclease domain comprises a Zn finger molecule comprising an engineered zinc finger protein that binds (in a sequence-specific manner) to a target DNA sequence. In some embodiments, the Zn finger molecule comprises one Zn finger protein or fragment thereof. In other embodiments, the Zn finger molecule comprises a plurality of Zn finger proteins (or fragments thereof), e.g., 2, 3, 4, 5, 6 or more Zn finger proteins (and optionally no more than 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, or 2 Zn finger proteins). In some embodiments, the Zn finger molecule comprises at least three Zn finger proteins. In some embodiments, the Zn finger molecule comprises four, five or six fingers. In some embodiments, the Zn finger molecule comprises 8, 9, 10, 11 or 12 fingers. In some embodiments, a Zn finger molecule comprising three Zn finger proteins recognizes a target DNA sequence comprising 9 or 10 nucleotides. In some embodiments, a Zn finger molecule comprising four Zn finger proteins recognizes a target DNA sequence comprising 12 to 14 nucleotides. In some embodiments, a Zn finger molecule comprising six Zn finger proteins recognizes a target DNA sequence comprising 18 to 21 nucleotides.

In some embodiments, a Zn finger molecule comprises a two-handed Zn finger protein. Two handed zinc finger proteins are those proteins in which two clusters of zinc finger proteins are separated by intervening amino acids so that the two zinc finger domains bind to two discontinuous target DNA sequences. An example of a two handed type of zinc finger binding protein is SIP1, where a cluster of

four zinc finger proteins is located at the amino terminus of the protein and a cluster of three Zn finger proteins is located at the carboxyl terminus (see Remade, et al. (1999) EMBO Journal 18(18):5073-5084). Each cluster of zinc fingers in these proteins is able to bind to a unique target sequence and the spacing between the two target sequences can comprise many nucleotides.

Linkers

In some embodiments, a gene modifying polypeptide may comprise a linker, e.g., a peptide linker, e.g., a linker as described in Table 1 or Table 10. In some embodiments, a gene modifying polypeptide comprises, in an N-terminal to C-terminal direction, a Cas domain (e.g., a Cas domain of Table 8), a linker of Table 10 (or a sequence having at least 70%, 80%, 85%, 90%, 95%, or 99% identity thereto), and an RT domain (e.g., an RT domain of Table 6). In some embodiments, a gene modifying polypeptide comprises a flexible linker between the endonuclease and the RT domain, e.g., a linker comprising the amino acid sequence SGGSSGGSSGSETPGTSESATPESSGGSSGGSS. In some embodiments, an RT domain of a gene modifying polypeptide may be located C-terminal to the endonuclease domain. In some embodiments, an RT domain of a gene modifying polypeptide may be located N-terminal to the endonuclease domain.

Table 10. Exemplary linker sequences

Amino Acid Sequence	SEQ ID NO
GGS	101
GGSGGS	102
GGSGGSGGS	103
GGSGGSGGSGGS	104
GGSGGSGGSGGSGGS	105
GGSGGSGGSGGSGGSGGS	106
GGGGS	107
GGGGSGGGGS	108
GGGGSGGGGSGGGGS	109
GGGGSGGGGSGGGGSGGGGS	110

GGGGSGGGGSGGGGSGGGGSGGGGS	111
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	112
GGG	113
GGGG	114
GGGGG	115
GGGGGG	116
GGGGGGG	117
GGGGGGGG	118
GSS	119
GSSGSS	120
GSSGSSGSS	121
GSSGSSGSSGSS	122
GSSGSSGSSGSSGSS	123
GSSGSSGSSGSSGSSGSS	124
EAAAK	125
EAAAKEAAAK	126
EAAAKEAAAKEAAAK	127
EAAAKEAAAKEAAAKEAAAK	128
EAAAKEAAAKEAAAKEAAAKEAAAK	129
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	130
PAP	131
PAPAP	132
PAPAPAP	133
PAPAPAPAP	134
PAPAPAPAPAP	135
PAPAPAPAPAPAP	136
GGSGGG	137
GGGGGS	138
GGSGSS	139

GSSGGS	140
GGSEAAAK	141
EAAAKGGS	142
GGSPAP	143
PAPGGS	144
GGGGSS	145
GSSGGG	146
GGGEAAAK	147
EAAAKGGG	148
GGGPAP	149
PAPGGG	150
GSSEAAAK	151
EAAAKGSS	152
GSSPAP	153
PAPGSS	154
EAAAKPAP	155
PAPEAAAK	156
GGSGGGGSS	157
GGSGSSGGG	158
GGGGGSGSS	159
GGGGSSGGS	160
GSSGGSGGG	161
GSSGGGGGS	162
GGSGGGEAAAK	163
GGSEAAAKGGG	164
GGGGGSEAAAK	165
GGGEAAAKGGS	166
EAAAKGGSGGG	167
EAAAKGGGGGS	168

GGSGGGPAP	169
GGSPAPGGG	170
GGGGGSPAP	171
GGGPAPGGS	172
PAPGGSGGG	173
PAPGGGGGS	174
GGSGSSEAAAK	175
GGSEAAAKGSS	176
GSSGGSEAAAK	177
GSSEAAAKGGS	178
EAAAKGGSGSS	179
EAAAKGSSGGS	180
GGSGSSPAP	181
GGSPAPGSS	182
GSSGGSPAP	183
GSSPAPGGS	184
PAPGGSGSS	185
PAPGSSGGS	186
GGSEAAAKPAP	187
GGSPAPEAAAK	188
EAAAKGGSPAP	189
EAAAKPAPGGS	190
PAPGGSEAAAK	191
PAPEAAAKGGS	192
GGGGSSEAAAK	193
GGGEAAAKGSS	194
GSSGGGEAAAK	195
GSSEAAAKGGG	196
EAAAKGGGGSS	197

EAAAKGSSGGG	198
GGGGSSPAP	199
GGGPAPGSS	200
GSSGGGPAP	201
GSSPAPGGG	202
PAPGGGGSS	203
PAPGSSGGG	204
GGGEAAAKPAP	205
GGGPAPEAAAK	206
EAAAKGGGPAP	207
EAAAKPAPGGG	208
PAPGGGEAAAK	209
PAPEAAAKGGG	210
GSSEAAAKPAP	211
GSSPAPEAAAK	212
EAAAKGSSPAP	213
EAAAKPAPGSS	214
PAPGSSEAAAK	215
PAPEAAAKGSS	216
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	217
GGGGSEAAAKGGGGS	218
EAAAKGGGGSEAAAK	219
SGSETPGTSESATPES	220
GSAGSAAGSGEF	221
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	222

In some embodiments, a linker of a gene modifying polypeptide comprises a motif chosen from: $(SGGS)_n$ (SEQ ID NO: 25), $(GGGS)_n$ (SEQ ID NO: 26), $(GGGGS)_n$ (SEQ ID NO: 27), $(G)_n$, $(EAAAK)_n$ (SEQ ID NO: 28), $(GGS)_n$, or $(XP)_n$.

Gene modifying polypeptide selection by pooled screening

Candidate gene modifying polypeptides may be screened to evaluate a candidate's gene editing ability. For example, an RNA gene modifying system designed for the targeted editing of a coding sequence in the human genome may be used. In certain embodiments, such a gene modifying system may be used in conjunction with a pooled screening approach.

For example, a library of gene modifying polypeptide candidates and a template guide RNA (tgRNA) may be introduced into mammalian cells to test the candidates' gene editing abilities by a pooled screening approach. In specific embodiments, a library of gene modifying polypeptide candidates is introduced into mammalian cells followed by introduction of the tgRNA into the cells.

Representative, non-limiting examples of mammalian cells that may be used in screening include HEK293T cells, U2OS cells, HeLa cells, HepG2 cells, Huh7 cells, K562 cells, or iPS cells.

A gene modifying polypeptide candidate may comprise 1) a Cas-nuclease, for example a wild-type Cas nuclease, e.g., a wild-type Cas9 nuclease, a mutant Cas nuclease, e.g., a Cas nickase, for example, a Cas9 nickase such as a Cas9 N863A nickase, or a Cas nuclease selected from **Table 7 or 8**, 2) a peptide linker, e.g., a sequence from **Table 1 or 10,** that may exhibit varying degrees of length, flexibility, hydrophobicity, and/or secondary structure; and 3) a reverse transcriptase (RT), e.g. an RT domain from **Table 1 or 6**. A gene modifying polypeptide candidate library comprises: a plurality of different gene modifying polypeptide candidates that differ from each other with respect to one, two or all three of the Cas nuclease, peptide linker or RT domain components, or a plurality of nucleic acid expression vectors that encode such gene modifying polypeptide candidates.

For screening of gene modifying polypeptide candidates, a two-component system may be used that comprises a gene modifying polypeptide component and a tgRNA component. A gene modifying component may comprise, for example, an expression vector, e.g., an expression plasmid or lentiviral vector, that encodes a gene modifying polypeptide candidate, for example, comprises a human codon-optimized nucleic acid that encodes a gene modifying polypeptide candidate, e.g., a Cas-linker-RT fusion as described above. In a particular embodiment, a lentiviral cassette is utilized that comprises: (i) a promoter for expression in mammalian cells, e.g., a CMV promoter; (ii) a gene modifying library candidate, e.g. a Cas-linker-RT fusion comprising a Cas nuclease of Table CC, a peptide linker of Table AA and an RT of Table BB, for example a Cas-linker-RT fusion as in **Table 1**; (iii) a self-cleaving polypeptide, e.g.,

a T2A peptide; (iv) a marker enabling selection in mammalian cells, e.g., a puromycin resistance gene; and (v) a termination signal, e.g., a poly A tail.

The tgRNA component may comprise a tgRNA or expression vector, e.g., an expression plasmid, that produces the tgRNA, for example, utilizes a U6 promoter to drive expression of the tgRNA, wherein the tgRNA is a non-coding RNA sequence that is recognized by Cas and localizes it to the genomic locus of interest, and that also templates reverse transcription of the desired edit into the genome by the RT domain.

To prepare a pool of cells expressing gene modifying polypeptide library candidates, mammalian cells, e.g., HEK293T or U2OS cells, may be transduced with pooled gene modifying polypeptide candidate expression vector preparations, e.g., lentiviral preparations, of the gene modifying candidate polypeptide library. In a particular embodiment, lentiviral plasmids are utilized, and HEK293 Lenti-X cells are seeded in 15 cm plates (~$12x10^6$ cells) prior to lentiviral plasmid transfection. In such an embodiment, lentiviral plasmid transfection may be performed using the Lentiviral Packaging Mix (Biosettia) and transfection of the plasmid DNA for the gene modifying candidate library is performed the following day using Lipofectamine 2000 and Opti-MEM media according to the manufacturer's protocol. In such an embodiment, extracellular DNA may be removed by a full media change the next day and virus-containing media may be harvested 48 hours after. Lentiviral media may be concentrated using Lenti-X Concentrator (TaKaRa Biosciences) and 5 mL lentiviral aliquots may be made and stored at -80°C. Lentiviral titering is performed by enumerating colony forming units post-selection, e.g., post Puromycin selection.

For monitoring gene editing of a target DNA, mammalian cells, e.g., HEK293T or U2OS cells, carrying a target DNA may be utilized. In other embodiments for monitoring gene editing of a target DNA, mammalian cells, e.g., HEK293T or U2OS cells, carrying a target DNA genomic landing pad may be utilized. In particular embodiments, the target DNA genomic landing pad may comprise a gene to be edited for treatment of a disease or disorder of interest. In other particular embodiments, the target DNA is a gene sequence that expresses a protein that exhibits detectable characteristics that may be monitored to determine whether gene editing has occurred. For example, in certain embodiments, a blue fluorescence protein (BFP)- or green fluorescence protein (GFP)-expressing genomic landing pad is utilized. In certain embodiments, mammalian cells, e.g., HEK293T or U2OS cells, comprising a target DNA, e.g., a target DNA genomic landing pad, are seeded in culture plates at 500x-3000x cells per gene modifying library candidate and transduced at a 0.2-0.3 multiplicity of infection (MOI) to minimize multiple infections per cell. Puromycin (2.5 ug/mL) may be added 48 hours post infection to allow for selection of infected cells. In such an embodiment, cells may be kept under puromycin selection for at least 7 days and then scaled up for tgRNA introduction, e.g., tgRNA electroporation.

To ascertain whether gene editing occurs, mammalian cells containing a target DNA to be edited may be infected with gene modifying polypeptide library candidates then transfected with tgRNA designed for use in editing of the target DNA. Subsequently, the cells may be analyzed to determine whether editing of the target locus has occurred according to the designed outcome, or whether no editing or imperfect editing has occurred, e.g., by using cell sorting and sequence analysis.

In a particular embodiment, to ascertain whether genome editing occurs, BFP- or GFP-expressing mammalian cells, e.g., HEK293T or U2OS cells, may be infected with gene modifying library candidates and then transfected or electroporated with tgRNA plasmid or RNA, e.g., by electroporation of 250,000 cells/well with 200 ng of a tgRNA plasmid designed to convert BFP-to-GFP or GFP-to-BFP, at a cell count ensuring >250x-1000x coverage per library candidate. In such an embodiment, the genome-editing capacity of the various constructs in this assay may be assessed by sorting the cells by Fluorescence-Activated Cell Sorting (FACS) for expression of the color-converted fluorescent protein (FP) at 4-10 days post-electroporation. Cells are sorted and harvested as distinct populations of unedited cells (exhibiting original florescence protein signal), edited cells (exhibiting converted fluorescence protein signal), and imperfect edit (exhibiting no florescence protein signal) cells. A sample of unsorted cells may also be harvested as the input population to determine candidate enrichment during analysis.

To determine which gene modifying library candidates exhibit genome-editing capacity in an assay, genomic DNA (gDNA) is harvested from the sorted cell populations, and analyzed by sequencing the gene modifying library candidates in each population. Briefly, gene modifying candidates may be amplified from the genome using primers specific to the gene modifying polypeptide expression vector, e.g., the lentiviral cassette, amplified in a second round of PCR to dilute genomic DNA, and then sequenced, for example, sequenced by a next-generation sequencing platform. After quality control of sequencing reads, reads of at least about 1500 nucleotides and generally no more than about 3200 nucleotides are mapped to the gene modifying polypeptide library sequences and those containing a minimum of about an 80% match to a library sequence are considered to be successfully aligned to a given candidate for purposes of this pooled screen. In order to identify candidates capable of performing gene editing in the assay, e.g., the BFP-to-GFP or GFP-to-BFP edit, the read count of each library candidate in the edited population is compared to its read count in the initial, unsorted population.

For purposes of pooled screening, gene modifying candidates with genome-editing capacity are identified based on enrichment in the edited (converted FP) population relative to unsorted (input) cells. In some embodiments, an enrichment of at least 1.0, 1.5, 2.0, 2.5, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, or at least 100-fold over the input indicates potentially useful gene editing activity, e.g., at least 2-fold enrichment. In some embodiments, the enrichment is converted to a log-value

by taking the log base 2 of the enrichment ratio. In some embodiments, a log2 enrichment score of at least 0, 1, 2, 3, 4, 5, 5.5, 6.0, 6.2, 6.3, 6.4, 6.5, or at least 6.6 indicates potentially useful gene editing activity, e.g., a log2 enrichment score of at least 1.0. In particular embodiments, enrichment values observed for gene modifying candidates may be compared to enrichment values observed under similar conditions utilizing a reference, e.g., Element ID No: 17380.

In some embodiments, multiple tgRNAs may be used to screen the gene modifying candidate library. In particular embodiments, a plurality of tgRNAs may be utilized to optimize template/Cas-linker-RT fusion pairs, e.g., for gene editing of particular target genes, for example, gene targets for the treatment of disease. In specific embodiments, a pooled approach to screening gene modifying candidates may be performed using a multiplicity of different tgRNAs in an arrayed format.

In some embodiments, multiple types of edits, e.g., insertions, substitutions, and/or deletions of different lengths, may be used to screen the gene modifying candidate library.

In some embodiments, multiple target sequences, e.g., different fluorescent proteins, may be used to screen the gene modifying candidate library. In some embodiments, multiple target sequences, e.g., different fluorescent proteins, may be used to screen the gene modifying candidate library. In some embodiments, multiple cell types, e.g., HEK293T or U2OS, may be used to screen the gene modifying candidate library. The person of ordinary skill in the art will appreciate that a given candidate may exhibit altered editing capacity or even the gain or loss of any observable or useful activity across different conditions, including tgRNA sequence (e.g., nucleotide modifications, PBS length, RT template length), target sequence, target location, type of edit, location of mutation relative to the first-strand nick of the gene modifying polypeptide, or cell type. Thus, in some embodiments, gene modifying library candidates are screened across multiple parameters, e.g., with at least two distinct tgRNAs in at least two cell types, and gene editing activity is identified by enrichment in any single condition. In other embodiments, a candidate with more robust activity across different tgRNA and cell types is identified by enrichment in at least two conditions, e.g., in all conditions screened. For clarity, candidates found to exhibit little to no enrichment under any given condition are not assumed to be inactive across all conditions and may be screened with different parameters or reconfigured at the polypeptide level, e.g., by swapping, shuffling, or evolving domains (e.g., RT domain), linkers, or other signals (e.g., NLS).

Sequences of exemplary Cas9-linker-RT fusions

In some embodiments, a gene modifying polypeptide comprises a linker sequence and an RT sequence. In some embodiments, a gene modifying polypeptide comprises a linker sequence as listed in

Table 1, or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide comprises the amino acid sequence of an RT domain as listed in Table 1, or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide comprises a linker sequence as listed in Table 1, or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and the amino acid sequence of an RT domain as listed in Table 1, or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a gene modifying polypeptide comprises: (i) a linker sequence as listed in a row of Table 1, or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto; and (ii) the amino acid sequence of an RT domain as listed in the same row of Table 1, or an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. For each RT domain named in Table 1, the corresponding amino acid sequence can be found in Table 6 herein.

Dimerization domains

In some embodiments, a gene modifying system as described herein comprises a DNA binding domain (DBD), e.g., comprising a Cas domain (e.g., a Cas9 domain, e.g., an nCas9 or dCas9 domain); an RNA binding domain (RBD); and a retroviral reverse transcriptase (RT) domain. In some embodiments, the DBD is attached to the RBD via binding between two dimerization domains. In some embodiments, the DBD is attached to the RT domain via binding between two dimerization domains. In some embodiments, the RT domain is attached to the RBD via binding between two dimerization domains.

In some embodiments, a pair of dimerization domains comprised in a gene modifying polypeptide or complex as described herein can be induced to dimerize by a compound (e.g., a small molecule). In some embodiments, a pair of dimerization domains comprised in a gene modifying polypeptide or complex as described herein can be induced to dimerize by exposure to light (e.g., of a specific color and/or wavelength). In some embodiments, a pair of dimerization domains comprised in a gene modifying polypeptide or complex as described herein comprise a Chain A sequence (or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto) and a Chain B sequence (or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto), as listed in a single row of Table 34. In embodiments, the pair of dimerization domains can be induced by the inducer listed in the same row of Table 34.

ble 34. Exemplary chemical- or light-induced dimerization domains

ducer(s)	chain A name	chain A sequence	Exemplary chain A source	chain B name	chain B sequence	Exemplary chain B source
rapamycin /rapalog	FKBP	GVQVETISPGDGRTFPKRGQTCVVHY TGMLEDGKKFDSSRDRNKPFKFMLGK QEVIRGWEEGVAQMSVGQRAKLTISP DYAYGATGHPGIIPPHATLVFDVELL KLE	snapgene	FRB	ILWHEMWHEGLEEASRLYFGERNV KGMFEVLEPLHAMMERGPQTLKET SFNQAYGRDLMEAQEWCRKYMKSG NVKDLTQAWDLYYHVFRRISK	snapgene
rapamycin /rapalog	FKBP	GVQVETISPGDGRTFPKRGQTCVVHY TGMLEDGKKFDSSRDRNKPFKFMLGK QEVIRGWEEGVAQMSVGQRAKLTISP DYAYGATGHPGIIPPHATLVFDVELL KLE	snapgene	FRB	EMWHEGLEEASRLYFGERNVKGMF EVLEPLHAMMERGPQTLKETSFNQ AYGRDLMEAQEWCRKYMKSGNVKD LTQAWDLYYHVFRRISKQL	snapgene
rapamycin /rapalog	FKBP	GVQVETISPGDGRTFPKRGQTCVVHY TGMLEDGKKFDSSRDRNKPFKFMLGK QEVIRGWEEGVAQMSVGQRAKLTISP DYAYGATGHPGIIPPHATLVFDVELL KLE	snapgene	FRB*	ILWHEMWHEGLEEASRLYFGERNV KGMFEVLEPLHAMMERGPQTLKET SFNQAYGRDLMEAQEWCRKYMKSG NVKDLLQAWDLYYHVFRRIS	addgene 108836 (pBW1308)
rapamycin /rapalog	FKBP	GVQVETISPGDGRTFPKRGQTCVVHY TGMLEDGKKFDSSRDRNKPFKFMLGK QEVIRGWEEGVAQMSVGQRAKLTISP DYAYGATGHPGIIPPHATLVFDVELL KLE	snapgene	FRB*	ILWHEMWHEGLEEASRLYFGERNV KGMFEVLEPLHAMMERGPQTLKET SFNQAYGRDLMEAQEWCRKYMKSG NVKDLLQAWDLYYHVFRRISK	snapgene
rapamycin /rapalog	FKBP	SRGVQVETISPGDGRTFPKRGQTCVV HYTGMLEDGKKFDSSRDRNKPFKFML GKQEVIRGWEEGVAQMSVGQRAKLTI SPDYAYGATGHPGIIPPHATLVFDVE LLKLE	addgene 108837 (pBHW1309)	FRB	ILWHEMWHEGLEEASRLYFGERNV KGMFEVLEPLHAMMERGPQTLKET SFNQAYGRDLMEAQEWCRKYMKSG NVKDLTQAWDLYYHVFRRISK	snapgene
rapamycin /rapalog	FKBP	SRGVQVETISPGDGRTFPKRGQTCVV HYTGMLEDGKKFDSSRDRNKPFKFML GKQEVIRGWEEGVAQMSVGQRAKLTI	addgene 108837	FRB	EMWHEGLEEASRLYFGERNVKGMF EVLEPLHAMMERGPQTLKETSFNQ	snapgene

		SPDYAYGATGHPGIIPPHATLVFDVE LLKLE	(pBHW130 9)		AYGRDLMEAQEWCRKYMKSGNVKD LTQAWDLYYHVFRRISKQL	
ıpamycin apalog	FKBP	SRGVQVETISPGDGRTFPKRGQTCVV HYTGMLEDGKKFDSSRDRNKPFKFML GKQEVIRGWEEGVAQMSVGQRAKLTI SPDYAYGATGHPGIIPPHATLVFDVE LLKLE	addgene 108837 (pBHW130 9)	FRB*	ILWHEMWHEGLEEASRLYFGERNV KGMFEVLEPLHAMMERGPQTLKET SFNQAYGRDLMEAQEWCRKYMKSG NVKDLLQAWDLYYHVFRRIS	addgene 108836 (pBW1308)
rapamycin /rapalog	FKBP	SRGVQVETISPGDGRTFPKRGQTCVV HYTGMLEDGKKFDSSRDRNKPFKFML GKQEVIRGWEEGVAQMSVGQRAKLTI SPDYAYGATGHPGIIPPHATLVFDVE LLKLE	addgene 108837 (pBHW130 9)	FRB*	ILWHEMWHEGLEEASRLYFGERNV KGMFEVLEPLHAMMERGPQTLKET SFNQAYGRDLMEAQEWCRKYMKSG NVKDLLQAWDLYYHVFRRISK	snapgene
abscisic acid	ABI	PLYGFTSICGRRPEMEAAVSTIPRFL QSSSGSMLDGRFDPQSAAHFFGVYDG HGGSQVANYCRERMHLALAEEIAKEK PMLCDGDTWLEKWKKALFNSFLRVDS EIESVAPETVGSTSVVAVVFPSHIFV ANCGDSRAVLCRGKTALPLSVDHKPD REDEAARIEAAGGKVIQWNGARVFGV LAMSRSIGDRYLKPSIIPDPEVTAVK RVKEDDCLILASDGVWDVMTDEEACE MARKRILLWHKKNAVAGDASLLADER RKEGKDPAAMSAAEYLSKLAIQRGSK DNISVVVVDLK	addgene 135985 (TL)	PYL	APTQDEFTQLSQSIAEFHTYQLGN GRCSSLLAQRIHAPPETVWSVVRR FDRPQIYKHFIKSCNVSEDFEMRV GCTRDVNVISGLPANTSRERLDLL DDDRRVTGFSITGGEHRLRNYKSV TTVHRFEKEEEEERIWTVVLESYV VDVPEGNSEEDTRLFADTVIRLNL QKLASITEAMN	addgene 135988 (TL)
abscisic acid	ABI	PLYGFTSICGRRPEMEAAVSTIPRFL QSSSGSMLDGRFDPQSAAHFFGVYDG HGGSQVANYCRERMHLALAEEIAKEK PMLCDGDTWLEKWKKALFNSFLRVDS EIESVAPETVGSTSVVAVVFPSHIFV ANCGDSRAVLCRGKTALPLSVDHKPD REDEAARIEAAGGKVIQWNGARVFGV LAMSRSIGDRYLKPSIIPDPEVTAVK RVKEDDCLILASDGVWDVMTDEEACE MARKRILLWHKKNAVAGDASLLADER	addgene 135985 (TL)	PYL	APTQDEFTQLSQSIAEFHTYQLGN GRCSSLLAQRIHAPPETVWSVVRR FDRPQIYKHFIKSCNVSEDFEMRV GCTRDVNVISGLPANTSRERLDLL DDDRRVTGFSITGGEHRLRNYKSV TTVHRFEKEEEEERIWTVVLESYV VDVPEGNSEEDTRLFADTVIRLNL QKLASITEAMNYPYDVPDYA	addgene 108841 (pBW1313

WO 2023/039441 PCT/US2022/076064

		RKEGKDPAAMSAAEYLSKLAIQRGSK DNISVVVVDLK				
ɔscisic :id	ABI	PLYGFTSICGRRPEMEDAVSTIPRFL QSSSGSMLDGRFDPQSAAHFFGVYDG HGGSQVANYCRERMHLALAEEIAKEK PMLCDGDTWLEKWKKALFNSFLRVDS EIGSVAPETVGSTSVVAVVFPSHIFV ANCGDSRAVLCRGKTALPLSVDHKPD REDEAARIEAAGGKVIQWNGARVFGV LAMSRSIGDRYLKPSIIPDPEVTAVK RVKEDDCLILASDGVWDVMTDEEACE MARKRILLWHKKNAVAGDASLLADER RKEGKDPAAMSAAEYLSKLAIQRGSK DNISVVVVDLK	addgene 108839 (pBW1311)	PYL	APTQDEFTQLSQSIAEFHTYQLGN GRCSSLLAQRIHAPPETVWSVVRR FDRPQIYKHFIKSCNVSEDFEMRV GCTRDVNVISGLPANTSRERLDLL DDDRRVTGFSITGGEHRLRNYKSV TTVHRFEKEEEEERIWTVVLESYV VDVPEGNSEEDTRLFADTVIRLNL QKLASITEAMN	addgene 135988 (TL)
abscisic acid	ABI	PLYGFTSICGRRPEMEDAVSTIPRFL QSSSGSMLDGRFDPQSAAHFFGVYDG HGGSQVANYCRERMHLALAEEIAKEK PMLCDGDTWLEKWKKALFNSFLRVDS EIGSVAPETVGSTSVVAVVFPSHIFV ANCGDSRAVLCRGKTALPLSVDHKPD REDEAARIEAAGGKVIQWNGARVFGV LAMSRSIGDRYLKPSIIPDPEVTAVK RVKEDDCLILASDGVWDVMTDEEACE MARKRILLWHKKNAVAGDASLLADER RKEGKDPAAMSAAEYLSKLAIQRGSK DNISVVVVDLK	addgene 108839 (pBW1311)	PYL	APTQDEFTQLSQSIAEFHTYQLGN GRCSSLLAQRIHAPPETVWSVVRR FDRPQIYKHFIKSCNVSEDFEMRV GCTRDVNVISGLPANTSRERLDLL DDDRRVTGFSITGGEHRLRNYKSV TTVHRFEKEEEEERIWTVVLESYV VDVPEGNSEEDTRLFADTVIRLNL QKLASITEAMNYPYDVPDYA	addgene 108841 (pBW1313)
gibberellin /GA3-AM (gibberelli c ester)	GAI	KRDHHHHHHQDKKTMMMNEEDDGNGM DELLAVLGYKVRSSEMADVAQKLEQL EVMMSNVQEDDLSQLATETVHYNPAE LYTWLDSMLTDLN	addgene 108845 (pBW2067)	GID1	AASDEVNLIESRTVVPLNTWVLIS NFKVAYNILRRPDGTFNRHLAEYL DRKVTANANPVDGVFSFDVLIDRR INLLSRVYRPAYADQEQPPSILDL EKPVDGDIVPVILFFHGGSFAHSS ANSAIYDTLCRRLVGLCKCVVVSV NYRRAPENPYPCAYDDGWIALNWV NSRSWLKSKKDSKVHIFLAGDSSG GNIAHNVALRAGESGIDVLGNILL	addgene 108843 (pBW2065

WO 2023/039441 PCT/US2022/076064

					NPMFGGNERTESEKSLDGKYFVTV RDRDWYWKAFLPEGEDREHPACNP FSPRGKSLEGVSFPKSLVVVAGLD LIRDWQLAYAEGLKKAGQEVKLMH LEKATVGFYLLPNNNHFHNVMDEI SAFVNAEC	
blue light	CRY2	KMDKKTIVWFRRDLRIEDNPALAAAA HEGSVFPVFIWCPEEEGQFYPGRASR WWMKQSLAHLSQSLKALGSDLTLIKT HNTISAILDCIRVTGATKVVFNHLYD PVSLVRDHTVKEKLVERGISVQSYNG DLLYEPWEIYCEKGKPFTSFNSYWKK CLDMSIESVMLPPPWRLMPITAAAEA IWACSIEELGLENEAEKPSNALLTRA WSPGWSNADKLLNEFIEKQLIDYAKN SKKVVGNSTSLLSPYLHFGEISVRHV FQCARMKQIIWARDKNSEGEESADLF LRGIGLREYSRYICFNFPFTHEQSLL SHLRFFPWDADVDKFKAWRQGRTGYP LVDAGMRELWATGWMHNRIRVIVSSF AVKFLLLPWKWGMKYFWDTLLDADLE CDILGWQYISGSIPDGHELDRLDNPA LQGAKYDPEGEYIRQWLPELARLPTE WIHHPWDAPLTVLKASGVELGTNYAK PIVDIDTARELLAKAISRTREAQIMI GAA	addgene 135989 (TL)	CIBN	NGAIGGDLLLNFPDMSVLERQRAH LKYLNPTFDSPLAGFFADSSMITG GEMDSYLSTAGLNLPMMYGETTVE GDSRLSISPETTLGTGNFKAAKFD TETKDCNEAAKKMTMNRDDLVEEG EEEKSKITEQNNGSTKSIKKMKHK AKKEENNFSNDSSKVTKELEKTDY I	addgene 135986 (TL)
blue light	CRY2	KMDKKTIVWFRRDLRIEDNPALAAAA HEGSVFPVFIWCPEEEGQFYPGRASR WWMKQSLAHLSQSLKALGSDLTLIKT HNTISAILDCIRVTGATKVVFNHLYD PVSLVRDHTVKEKLVERGISVQSYNG DLLYEPWEIYCEKGKPFTSFNSYWKK CLDMSIESVMLPPPWRLMPITAAAEA IWACSIEELGLENEAEKPSNALLTRA WSPGWSNADKLLNEFIEKQLIDYAKN	addgene 135989 (TL)	CIBN	NGAIGGDLLLNFPDMSVLERQRAH LKYLNPTFDSPLAGFFADSSMITG GEMDSYLSTAGLNLPMMYGETTVE GDSRLSISPETTLGTGNFKKRKFD TETKDCNEKKKKMTMNRDDLVEEG EEEKSKITEQNNGSTKSIKKMKHK AKKEENNFSNDSSKVTKELEKTDY I	snapgene

		SKKVVGNSTSLLSPYLHFGEISVRHV FQCARMKQIIWARDKNSEGEESADLF LRGIGLREYSRYICFNFPFTHEQSLL SHLRFFPWDADVDKFKAWRQGRTGYP LVDAGMRELWATGWMHNRIRVIVSSF AVKFLLLPWKWGMKYFWDTLLDADLE CDILGWQYISGSIPDGHELDRLDNPA LQGAKYDPEGEYIRQWLPELARLPTE WIHHPWDAPLTVLKASGVELGTNYAK PIVDIDTARELLAKAISRTREAQIMI GAA				
blue light	pMag	HTLYAPGGYDIMGYLRQIRNRPNPQV ELGPVDTSCALILCDLKQKDTPIVYA SEAFLYMTGYSNAEVLGRNCRFLQSP DGMVKPKSTRKYVDSNTINTMRKAID RNAEVQVEVVNFKKNGQRFVNFLTMI PVRDETGEYRYSMGFQCETE	snapgene , addgene 108848 (pBW2655)	nMagHigh1	HTLYAPGGYDIMGYLDQIGNRPNP QVELGPVDTSCALILCDLKQKDTP IVYASEAFLYMTGYSNAEVLGRNC RFLQSPDGMVKPKSTRKYVDSNTI NTIRKAIDRNAEVQVEVVNFKKNG QRFVNFLTIIPVRDETGEYRYSMG FQCETE	
blue light	pMag	HTLYAPGGYDIMGYLRQIRNRPNPQV ELGPVDTSCALILCDLKQKDTPIVYA SEAFLYMTGYSNAEVLGRNCRFLQSP DGMVKPKSTRKYVDSNTINTMRKAID RNAEVQVEVVNFKKNGQRFVNFLTMI PVRDETGEYRYSMGFQCETE	snapgene , addgene 108848 (pBW2655)	nMag	HTLYAPGGYDIMGYLDQIGNRPNP QVELGPVDTSCALILCDLKQKDTP IVYASEAFLYMTGYSNAEVLGRNC RFLQSPDGMVKPKSTRKYVDSNTI NTMRKAIDRNAEVQVEVVNFKKNG QRFVNFLTMIPVRDETGEYRYSMG FQCETE	
blue light	pMagFast2	HTLYAPGGYDIMGYLRQIRNRPNPQV ELGPVDTSCALVLCDLKQKDTPVVYA SEAFLYMTGYSNAEVLGRNCRFLQSP DGMVKPKSTRKYVDSNTINTMRKAID RNAEVQVEVVNFKKNGQRFVNFLTMI PVRDETGEYRYSMGFQCETE		nMagHigh1	HTLYAPGGYDIMGYLDQIGNRPNP QVELGPVDTSCALILCDLKQKDTP IVYASEAFLYMTGYSNAEVLGRNC RFLQSPDGMVKPKSTRKYVDSNTI NTIRKAIDRNAEVQVEVVNFKKNG QRFVNFLTIIPVRDETGEYRYSMG FQCETE	
blue light	pMagFast2	HTLYAPGGYDIMGYLRQIRNRPNPQV ELGPVDTSCALVLCDLKQKDTPVVYA SEAFLYMTGYSNAEVLGRNCRFLQSP		nMag	HTLYAPGGYDIMGYLDQIGNRPNP QVELGPVDTSCALILCDLKQKDTP IVYASEAFLYMTGYSNAEVLGRNC	

		DGMVKPKSTRKYVDSNTINTMRKAID RNAEVQVEVVNFKKNGQRFVNFLTMI PVRDETGEYRYSMGFQCETE			RFLQSPDGMVKPKSTRKYVDSNTI NTMRKAIDRNAEVQVEVVNFKKNG QRFVNFLTMIPVRDETGEYRYSMG FQCETE	
.ɔd light	PhyB	VSGVGGSGGGRGGGRGGEEEPSSSHT PNNRRGGEQAQSSGTKSLRPRSNTES MSKAIQQYTVDARLHAVFEQSGESGK SFDYSQSLKTTTYGSSVPEQQITAYL SRIQRGGYIQPFGCMIAVDESSFRII GYSENAREMLGIMPQSVPTLEKPEIL AMGTDVRSLFTSSSSILLERAFVARE ITLLNPVWIHSKNTGKPFYAILHRID VGVVIDLEPARTEDPALSIAGAVQSQ KLAVRAISQLQALPGGDIKLLCDTVV ESVRDLTGYDRVMVYKFHEDEHGEVV AESKRDDLEPYIGLHYPATDIPQASR FLFKQNRVRMIVDCNATPVLVVQDDR LTQSMCLVGSTLRAPHGCHSQYMANM GSIASLAMAVIINGNEDDGSNVASGR SSMRLWGLVVCHHTSSRCIPFPLRYA CEFLMQAFGLQLNMELQLALQMSEKR VLRTQTLLCDMLLRDSPAGIVTQSPS IMDLVKCDGAAFLYHGKYYPLGVAPS EVQIKDVVEWLLANHADSTGLSTDSL GDAGYPGAAALGDAVCGMAVAYITKR DFLFWFRSHTAKEIKWGGAKHHPEDK DDGQRMHPRSSFQAFLEVVKSRSQPW ETAEMDAIHSLQLILRDSFKESEAAM NSKVVDGVVQPCRDMAGEQGIDELGA VAREMVRLIETATVPIFAVDAGGCIN GWNAKIAELTGLSVEEAMGKSLVSDL IYKENEATVNKLLSRALRGDEEKNVE VKLKTFSPELQGKAVFVVVNACSSKD YLNNIVGVCFVGQDVTSQKIVMDKFI NIQGDYKAIVHSPNPLIPPIFAADEN	pBW2682	PIF6	MFLPTDYCCRLSDQEYMELVFENG QILAKGQRSNVSLHNQRTKSIMDL YEAEYNEDFMKSIIHGGGGAITNL GDTQVVPQSHVAAAHETNMLESNK HVD	pBW2684

		TCCLEWNMAMEKLTGWSRSEVIGKMI VGEVFGSCCMLKGPDALTKFMIVLHN AIGGQDTDKFPFPFFDRNGKFVQALL TANKRVSLEGKVIGAFCFLQIPS				
.ɔd light	PhyB	VSGVGGSGGGRGGGRGGEEEPSSSHT PNNRRGGEQAQSSGTKSLRPRSNTES MSKAIQQYTVDARLHAVFEQSGESGK SFDYSQSLKTTTYGSSVPEQQITAYL SRIQRGGYIQPFGCMIAVDESSFRII GYSENAREMLGIMPQSVPTLEKPEIL AMGTDVRSLFTSSSSILLERAFVARE ITLLNPVWIHSKNTGKPFYAILHRID VGVVIDLEPARTEDPALSIAGAVQSQ KLAVRAISQLQALPGGDIKLLCDTVV ESVRDLTGYDRVMVYKFHEDEHGEVV AESKRDDLEPYIGLHYPATDIPQASR FLFKQNRVRMIVDCNATPVLVVQDDR LTQSMCLVGSTLRAPHGCHSQYMANM GSIASLAMAVIINGNEDDGSNVASGR SSMRLWGLVVCHHTSSRCIPFPLRYA CEFLMQAFGLQLNMELQLALQMSEKR VLRTQTLLCDMLLRDSPAGIVTQSPS IMDLVKCDGAAFLYHGKYYPLGVAPS EVQIKDVVEWLLANHADSTGLSTDSL GDAGYPGAAALGDAVCGMAVAYITKR DFLFWFRSHTAKEIKWGGAKHHPEDK DDGQRMHPRSSFQAFLEVVKSRSQPW ETAEMDAIHSLQLILRDSFKESEAAM NSKVVDGVVQPCRDMAGEQGIDELGA VAREMVRLIETATVPIFAVDAGGCIN GWNAKIAELTGLSVEEAMGKSLVSDL IYKENEATVNKLLSRALRGDEEKNVE VKLKTFSPELQGKAVFVVVNACSSKD YLNNIVGVCFVGQDVTSQKIVMDKFI NIQGDYKAIVHSPNPLIPPIFAADEN	pBW2682	PIF6	MFLPTDYCCRLSDQEYMELVFENG QILAKGQRSNVSLHNQRTKSIMDL YEAEYNEDFMKSIIHGGGGAITNL GDTQVVPQSHVAAAHETNMLESNK HVD	pBW2684

		TCCLEWNMAMEKLTGWSRSEVIGKMI VGEVFGSCCMLKGPDALTKFMIVLHN AIGGQDTDKFPFPFFDRNGKFVQALL TANKRVSLEGKVIGAFCFLQIPS				
.ed light	PhyBNT	VSGVGGSGGGRGGGRGGEEEPSSSHT PNNRRGGEQAQSSGTKSLRPRSNTES MSKAIQQYTVDARLHAVFEQSGESGK SFDYSQSLKTTTYGSSVPEQQITAYL SRIQRGGYIQPFGCMIAVDESSFRII GYSENAREMLGIMPQSVPTLEKPEIL AMGTDVRSLFTSSSSILLERAFVARE ITLLNPVWIHSKNTGKPFYAILHRID VGVVIDLEPARTEDPALSIAGAVQSQ KLAVRAISQLQALPGGDIKLLCDTVV ESVRDLTGYDRVMVYKFHEDEHGEVV AESKRDDLEPYIGLHYPATDIPQASR FLFKQNRVRMIVDCNATPVLVVQDDR LTQSMCLVGSTLRAPHGCHSQYMANM GSIASLAMAVIINGNEDDGSNVASGR SSMRLWGLVVCHHTSSRCIPFPLRYA CEFLMQAFGLQLNMELQLALQMSEKR VLRTQTLLCDMLLRDSPAGIVTQSPS IMDLVKCDGAAFLYHGKYYPLGVAPS EVQIKDVVEWLLANHADSTGLSTDSL GDAGYPGAAALGDAVCGMAVAYITKR DFLFWFRSHTAKEIKWGGAKHHPEDK DDGQRMHPRSSFQAFLEVVKSRSQPW ETAEMDAIHSLQLILRDSFKES	pBW2682	PIF6	MFLPTDYCCRLSDQEYMELVFENG QILAKGQRSNVSLHNQRTKSIMDL YEAEYNEDFMKSIIHGGGGAITNL GDTQVVPQSHVAAAHETNMLESNK HVD	pBW2684
red light	PhyBNT	VSGVGGSGGGRGGGRGGEEEPSSSHT PNNRRGGEQAQSSGTKSLRPRSNTES MSKAIQQYTVDARLHAVFEQSGESGK SFDYSQSLKTTTYGSSVPEQQITAYL SRIQRGGYIQPFGCMIAVDESSFRII GYSENAREMLGIMPQSVPTLEKPEIL AMGTDVRSLFTSSSSILLERAFVARE	pBW2682	PIF6	MFLPTDYCCRLSDQEYMELVFENG QILAKGQRSNVSLHNQRTKSIMDL YEAEYNEDFMKSIIHGGGGAITNL GDTQVVPQSHVAAAHETNMLESNK HVD	pBW2684

		ITLLNPVWIHSKNTGKPFYAILHRID VGVVIDLEPARTEDPALSIAGAVQSQ KLAVRAISQLQALPGGDIKLLCDTVV ESVRDLTGYDRVMVYKFHEDEHGEVV AESKRDDLEPYIGLHYPATDIPQASR FLFKQNRVRMIVDCNATPVLVVQDDR LTQSMCLVGSTLRAPHGCHSQYMANM GSIASLAMAVIINGNEDDGSNVASGR SSMRLWGLVVCHHTSSRCIPFPLRYA CEFLMQAFGLQLNMELQLALQMSEKR VLRTQTLLCDMLLRDSPAGIVTQSPS IMDLVKCDGAAFLYHGKYYPLGVAPS EVQIKDVVEWLLANHADSTGLSTDSL GDAGYPGAAALGDAVCGMAVAYITKR DFLFWFRSHTAKEIKWGGAKHHPEDK DDGQRMHPRSSFQAFLEVVKSRSQPW ETAEMDAIHSLQLILRDSFKES				
near infrared light	PpsR2	ASKSVHADITLLLDMEGVIREATLSP TMAAESVDGWLGRRWSDIAGAEGGDK VRRMVEDARRSGISAFRQINQPFPSG VEIPIEFTTMLLGDRTGMIAVGKNMQ AVTELHSRLIAAQQAMERDYWRLREL ETRYRLVFDAAADAVMIVSAGDMRIV EANRAAVNAISRVERGNDDLAGRDFL AEVAAADRDAVRDMLAQVRQRGTALS VLVHLGRYDRAWMLRGSLMSSERRQV FLLHFTPVTTTPAIDDVDDDAVLRGL IDRIPDGFVALDSEGVVRHANQAFLD LVQIGSKPAAVGRSLGVWMGRPGADL SSLLTLLRRYKTVRLFQTTIRGELGT ETEVEVSAVDGEDDQYIGVLMRNVAR RLDAADDHDALRQALGPISKQLGRSS LRKLVKNAVSIVEQHYVKEALLRSKG NRTATAELLGLSRQSLYAKLNSYGFD DKGVVASAADGAEGASDDAED	pBW2780	BphP1	VAGHASGSPAFGTADLSNCEREEI HLAGSIQPHGALLVVSEPDHRIIQ ASANAAEFLNLGSVLGVPLAEIDG DLLIKILPHLDPTAEGMPVAVRCR IGNPSTEYDGLMHRPPEGGLIIEL ERAGPPIDLSGTLAPALERIRTAG SLRALCDDTALLFQQCTGYDRVMV YRFDEQGHGEVFSERHVPGLESYF GNRYPSSDIPQMARRLYERQRVRV LVDVSYQPVPLEPRLSPLTGRDLD MSGCFLRSMSPIHLQYLKNMGVRA TLVVSLVVGGKLWGLVACHHYLPR FIHFELRAICELLAEAIATRITAL ESFAQSQSELFVQRLEQRMIEAIT REGDWRAAIFDTSQSILQPLHADG CALVYEDQIRTIGDVPSTQDVREI AGWLDRQPRAAVTSTASLGLDVPE LAHLTRMASGVVAAPISDHRGEFL	pBW2779

					MWFRPERVHTVTWGGDPKKPFTMG DTPADLSPRRSFAKWHQVVEGTSD PWTAADLAAARTIGQTVADIVLQF RAVRTLIAREQYEQFSSQVHASMQ PVLITDAEGRILLMNDSFRDMLPA GSPSAVHLDDLAGFFVESNDFLRN VAELIDHGRGWRGEVLLRGAGNRP LPLAVRADPVTRTEDQSLGFVLIF SDATDRRTADAARTRFQEGILASA RPGVRLDSKSDLLHEKLLSALVEN AQLAALEITYGVETGRIAELLEGV RQSMLRTAEVLGHLVQHAARTAGS DSSSNGSQNKK	

WO 2023/039441 PCT/US2022/076064

In some embodiments, a pair of dimerization domains comprised in a gene modifying polypeptide or complex as described herein comprise an antibody, or a functional fragment thereof, and a peptide recognized by the antibody or fragment thereof. In some embodiments, a pair of dimerization domains comprised in a gene modifying polypeptide or complex as described herein comprise a Chain A sequence (or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto) and a Chain B sequence (or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto), as listed in a single row of Table 35.

Table 35. Exemplary antibody-peptide dimerization domains

system	chain A name	chain A sequence	Exemplary chain A source	chain B name	chain B sequence	Exemplary chain B source
SunTag	GCN4_v4	EELLSKNYHLENEVARLKK	snapgene	GCN4_scFv	GPDIVMTQSPSSLSASVGDRVTITCRSSTGAVTTSNYASWVQEKPGKLFKGLIGGTNNRAPGVPSRFSGSLIGDKATLTISSLQPEDFATYFCALWYSNHWVFGQGTKVELKRGGGGSGGGGSGGGGSSGGGSEVKLLESGGGLVQPGGSLKLSCAVSGFSLTDYGVNWVRQAPGRGLEWIGVIWGDGITDYNSALKDRFIISKDNGKNTVYLQMSKVRSDDTALYYCVTGLFDYWGQGTLVTVSS	addgene 60904
SunTag	GCN4_v4	EELLSKNYHLENEVARLKK	snapgene	GCN4_scFv	GPDIVMTQSPSSLSASVGDRVTITCRSSTGAVTTSNYASWVQEKPGKLFKGLIGGTNNRAPGVPSRFSGSLIGDKATLTISSLQPEDFATYFCALWYSNHWVFGQGTKVELKRGGGGSGGGGSGGGGSSGGGSEVKLLESGGGLVQPGGSLKLSCAVSGFSLTDYGVNWVRQAPGRGLEWIGVIWGDGITDYNSALKDRFIISKDNGKNTVYLQMSKVRSDDTALYYCVTGLFDYWGQGTLVTVSS	addgene 60904
SunTag	GCN4_v1	LLPKNYHLENEVARLKKLVGER	snapgene	GCN4_scFv	GPDIVMTQSPSSLSASVGDRVTITCRSSTGAVTTSNYASWVQEKPGKLFKGLIGGTNNRAPGVPSRFSGSLIGDKATLTISSLQPEDFATYFCALWYSNHWVFGQGTKVELKRGGGGSGGGGSGGGGSSGGGSEVKLLESGGGLVQPGGSLKLSCAVSGFSLTDYGVNWVRQAPGRGLEWIGVIWGDGITDYNSALKDRFIISKDNGKNTVYLQMSKVRSDDTALYYCVTGLFDYWGQGTLVTVSS	addgene 60904
SunTag	GCN4_v1	LLPKNYHLENEVARLKKLVGER	snapgene	GCN4_scFv	GPDIVMTQSPSSLSASVGDRVTITCRSSTGAVTTSNYASWVQEKPGKLFKGLIGGTNNRAPGVPSRFSGSLIGDKATLTISSLQPEDFATYFCALWYSNHWVFGQGTKVELKRGGGGSGGGGSGGGGSSGGGSEVKLLESGGGLVQPGGSLKLSCAVSGFSLTDYGVNWVRQAPGRGLEWIGVIWGDGITDYNSALKDRFIISKDNGKNTVYLQMSKVRSDDTALYYCVTGLFDYWGQGTLVTVSS	addgene 60904

MoonTag	gp41_peptide	KNEQELLELDKWASL	addgene 128605	moontag_nanobody	EVQLVESGGGLVQPGGSLRLSCAASGSISSVDVMSWYRQAPGKQRELVAFITDRGRTNYKVSVKGRFTISRDNSKNMVYLQMNSLKPEDTADYLCRAESRTSWSSPSPLDVWGRGTQVTVSS	addgene 128602

In some embodiments, a dimerization domain comprised in a gene modifying polypeptide or complex as described herein comprises a coiled-coil dimerization domain. In some embodiments, a dimerization domain comprised in a gene modifying polypeptide or complex as described herein comprises a sequence as listed in a single row of Table 36, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a pair of dimerization domains comprised in a gene modifying polypeptide or complex as described herein comprise copies of the same coiled-coil dimerization domain (or coiled-coil dimerization domains having at least 90%, 95%, 96%, 97%, 98%, or 99% identity relative to each other).

Table 36. Exemplary coiled coil dimerization domains

Name	Sequence
P1	SPEDEIQALEEENAQLEQENAALEEEIAQLEYG
P2	SPEDKIAQLKEKNAALKEKNQQLKEKIQALKYG
P3	SPEDEIQQLEEEIAQLEQKNAALKEKNQALKYG
P4	SPEDKIAQLKQKIQALKQENQQLEEENAALEYG
P3S	SPEDEIQQLEEEISQLEQKNSQLKEKNQQLKYG
P4S	SPEDKISQLKQKIQQLKQENQQLEEENSQLEYG
P5	SPEDENAALEEKIAQLKQKNAALKEEIQALEYG
P6	SPEDKNAALKEEIQALEEENQALEEKIAQLKYG
P7	SPEDEIQALEEKNAQLKQEIAALEEKNQALKYG
P8	SPEDKIAQLKEENQQLEQKIQALKEENAALEYG
P9	SPEDENQALEQKNAQLKQEIAALEQEIAQLEYG
P10	SPEDKNAQLKEENAALEEKIQQLKEKIQALKYG
P11	SPEDENQALEQEIAQLEQEIAALEQKNAQLKYG
P12	SPEDKNAQLKEKIAALKEKIQQLKEENQALEYG
DHD9	GSPKEEARELIRKQKELIKEQKKLIKEAKQKSDSRDAERIWKRSREINRESKKINKRIKELIKS
DHD9	PKKEAEELAEESEELHDRSEKLHERAEQSSNSEEARKILEDIERISERIEEISDRIERLLRS
DHD13_XAAA	GTKEDILERQRKIIERAQEIHRRQQEILEELERIIRKPGSSEEAMKRMLKLLEESLRLLKELLELSEESAQLLYEQR
DHD13_XAAA	GTEKRLLEEAERAHREQKEIIKKAQELHRRLEEIVRQSGSSEEAKKEAKKILEEIRELSKRSLELLREILYLSQEQK GSLVPR
DHD13_XAXA	TKEDILERQRKIIERAQEIIRRQQEILEELERIIRKPGSSEEAMKRMLKLLEESLRLLKELLELLEESAQLLYEQR
DHD13_XAXA	GSTEKRLLEEAERAHREAKEIIKKAQELHRRLEEIVRQSGSSEEAKKEAKKILEEIRELSKRLLELLREILYLSQEQ K
DHD13_XAAX	TKEDILERARKIIERAQEIHRRQQEILEELERIIRKPGSSEEAMKRMLKLLEESLRLLKELLELSEELAQLLYEQR
DHD13_XAAX	GSTEKRLLEEAERAIREQKEIIKKAQELHRRLEEIVRQSGSSEEAKKEAKKILEEIRELSKRSLELLREILYLLQEQ K
DHD13_2:341	TKEDILERQRKIIERAQEIHRRQQEILEELEYIIR

DHD13_2:341	MSEEAMKRMLKLLEESLRLLKELLELSEESAQLLYEQRKANNGSETEKRLLEEAERAHREQKEIIKKAQELHRRLEE IVRQSGSSEEAKKEAKKILEEIRELSKRSLELLREILYLSQEQK
DHD13_AAAA	MTKEDILERQRKIIERAQEIHRRQQEILKEQEKIIRKPGSSEEAMKRSLKLIEESLRLLKELLELSEESAQLLYEQR
DHD13_AAAA	GTEKRLLEEAERAHREQKEIIKKAQELHKELTKIHQQSGSSEEAKKRALKISQEIRELSKRSLELLREILYLSQEQK
DHD13_BAAA	TKEDILERQRKIIERAQEIHRRQQEILKRSEEIIRKPGSSEEALETLRELQEESLRLLKELLELSEESAQLLYEQR
DHD13_BAAA	GSTEKRLLEEAERAHREQKEIIKKAQELHRRTEEIIRQSGSSEEAKDELRRIQEEIRELSKRSLELLREILYLSQEQ K
DHD13_4:123	TTKRYLEEAERAHREQKEIIKKAQELHRRLEEIVRQ
DHD13_4:123	GSSEEAKKEAKKILEEIRELSKRSLELLREILYLSQQVNDVDEKALERQRKIIERAQEIHRRQQEILEELERIIRKP GSSEEAMKRMLKLLEESLRLLKELLELSEESAQLLYEAR
DHD13_1:234	EAMKRMLKLLEESLRLLKELLELSEESAQLLYEAR
DHD13_1:234	TTKRYLEEAERAHREQKEIIKKAQELHRRLEEIVRQSGSSEEAKKEAKKILEEIRELSKRSLELLREILYLSQQVND VDEKALERQRKIIERAQEIHRRQQEILEELERIIRKPGS
DHD15	TREELLRENIELAKEHIEIMREILELLQKMEELLEKARGADEDVAKTIKELLRRLKEIIERNQRIAKEHEYIARERS
DHD15	GTERKLLERSRRLQEESKRLLDEMAEIMRRIKKLLKKARGADEKVLDELRKIIERIRELLDRSRKIHERSEEIAYKE E
DHD20	GDRQELIRRNIELLKEHIKILEEISQLIEELSELLDKSSSEEVVKRYKKILERYKQLLRKSQEIHKESSEIAKKES
DHD20	GDEQKLIERSQRMQKESLELLKEIIKILDTIEKLLDKPDSEELLDTIKKLHDTLKKIHDRNKKLLKEHEEILRQRSG SLVPR
DHD21	DKEEEYKRLLDEIKEILKESKEVLKDSKRVLEDIKRKVPDDDLVKLLEKHVRLLEEHVKLLEQLIREAEKSSK
DHD21	QGSSAEELLKKIKESEKKIRDSLRKIKEIIKKSRKEGVDDKQLDLIRKVVESHRDLLRLHRDLLRLLREETS
DHD25	DIDESIKEVEKLLEEVEQSLQKLDDSLKKLLEKVNQDPDVDDSVRKIVKRHVEILKRHEEVLKRLIEVVKEHTKTVK
DHD25	GSDREEVHKEIVKLIREIIKIHKKILKIHEKIKNGEIDPSEILKLSEEIKKLTDTIIKIIEDLEQLTRDLRR
DHD27	DRKEIVKRHQKVVELLKESSKLLRESSKLLQRLLDKTGDENLQKAVDDQDKAIKRQETAIRKSQEASKKLD
DHD27	DNSEEIKKVAKTSREVAEYSERVAKENDKVVKTLEEGKIDESELLRLLEESIKIFDTALKLHEEAYKLHQDLVRKVS
DHD30	DESEAASVAIESVQILVESVKLLEESVRILLDAVKKNGVEDLLRVAQRWEKLVDEWLKVVKRWLDNVRDIQR
DHD30	GSDKAEEVEKSVRKIEESIKKIRKSIKKAEDAVQLLKEGKIDAKDFLRIVREDLEVVKEDVEIVKEDVENVREFSS
DHD33	SDKEVSDKLLKASKKLLKVSEELLEVVRRLLKALKDDELIKKIADLLRKIIDKDKKFIRTSEEIVKESR
DHD33	GSDLKEVLKTVEEAVKEIIKSSEELLQISRKILEISRVGVDEHEYISAIREYLKALEKHIQILKKFIEILKELIRAV S

DHD34_XAAXA	SKEEIDKIVKKHKKKIEEHKKKVDELKKLVEEHDKRVSQDKDDKVKKLSEEVKKIIKRLEEVSKRLEEVSKKLLKVISDKR
DHD34_XAAXA	GSNDEELKKILETLDRILKKLDKILTRLIEVLKKSEDPNLDDKDYTELVKQFIELIKKYEEVVKEYEEVVRQLIRLFS
DHD34_XAXXA	SKEEIDKIVKKHKKKIEELKKLVDELKKLVEEHDKRVSQDKDDKVKKLSEEVKKIIKRVEEVAKRLEEVSKKLLKVISDKR
DHD34_XAXXA	GSNDEELKKILETLDRILKKLEKILTRLIEVLKKSEDPNLDDKDYTELVKQFIELIKKFEEVIKEYEEVVRQLIRLFS
DHD34_XAAAA	SKEEIDKIVKKHKKKIEEHKKKVDEHKKLVEEHDKRVSQDKDDKVKKLSEELKKISKRLEEVSKRLEEVSKKLLKVISDKR
DHD34_XAAAA	GSNDEELKKILETLDRILKKLDKILTRLDEVLKKSEDPNLDDKDYTELVKQYIELVKKYEEVVKEYEEVVRQLIRLFS
DHD36	DHSRKLKEILDRLRKHVKRLKEHLDELRDLVRQVPEDKLLEHVVKLSDKILQISERAVREFTKSVDKDS
DHD36	GSDKKDELERILDEIRRLIERLDEILSRLNKLLELLKHGVPNAKEVVKDYIRLLKEYLELVKEFLKLVKRHADLVS
DHD37_ABXB	DSDEHLKKLKTFLENLRRHLDRLDKHIKQLRDILSENPEDERVKDVIDLSERSVRIVKTVIKIFEDSVRKKE
DHD37_ABXB	GSDDKELDKLLDTLEKILQTATKIIDDANKLLEKLRRSERKDPKVVETYVELLKRHEKAVKELLEIAKTHAKKVE
DHD37_BBBB	MDEEDHLKKLKTHLEKLERHLKLLEDHAKKLEDILKERPEDSAVKESIDELRRSIELVRESIEIFRQSVEEEE
DHD37_BBBB	GDVKELTKILDTLTKILETATKVIKDATKLLEEHRKSDKPDPRLIETHKKLVEEHETLVRQHKELAEEHLKRTR
DHD37_XBXB	DSDEHLKKLKTFLENLRRHLDRLDKLLKELRDILSENPEDERVKDVIDELERVIRIVKTVIKIFEDSVRKKE
DHD37_XBXB	GSDDKELDKLLDTLEKILQTATKIIDDLNKVLEKLRRSERKDPKVIETVVELLKRHEKAVKELLEIAKTHAKKVE
DHD37_AXXB	DSDEHLKKLKTFLENLRRLEDLLDKHIKQLRDILSENPEDERVKDVIDLSERVVRTVKTVIKIFEDSVRKKE
DHD37_AXXB	GSDDKELDKLLDTLEKILQTATKVVDDANKLLEKLRRSERKDPKVVETYVELLKRLEKLIKELLEIAKTHAKKVE
DHD37_3:124	DSDEHLKKLKTFLENLRRHLDRLDKHIKQLRDILSEN
DHD37_3:124	EDERVKDVIDLSERSVRIVKTVIKIFEDSVRKLEKTKPDSKTAKELDKLLDTLEKILQTATKIIDDANKLLEKLRRSERKDPKVVETYVELLKRHEKAVKELLEIAKTHAKKVE
DHD37_1:234	DSDEHLYKLKTFLENLRRHLDRLDKHIKQLRDILSENPEDERVKDAIDLSERSVRIVKTVIKIFEDSVRKKEKRPIDKRDDKELDKLLDTLEKILQTATKIIDDANKLLEYLRR
DHD37_1:234	GDPKVVETYVELLKRHEKAVKELLEIAKTHAKKVE
DHD37_AXBB	DSDEHLDRLDKHLKKLKTFLENLRRHIKQLRDILSENPEDERVKDVIDLSKTVIKIFEDSVRKKERSVRIVE
DHD37_AXBB	GSDDKEATKIIDDLDKLLDTLEKILQTANKLLEKLRRSERKDPKVVETYVKAVKELLEIAKTHAELLKRHEKKVE

DHD37_XBBA	DSDEHIKQLRDHLDRLDKHLKKLKTFLENLRRILSENPEDERVKTVIKIFEDSVRKKERSVRIVKDVIDLSE
DHD37_XBBA	GSDDKEANKLLEKATKIIDDLDKLLDTLEKILQTLRRSERKDPKAVKELLEIAKTHAELLKRHEKVVETYVKKVE
DHD39	DHSRKLEEILDRLRKHVKRLLEHLRELLSLVKENPEDKDLVEVLELSLAILRRSLEAVEAFLKSVTKKDPDDEDLRR KADEIRKEVEEIKKSLAEVEKEIYKLK
DHD39	GSSADDVLEDILKIIRELIEILDQILSLLNQLLKLLRHGVPNAKKVVEKYKEILELYLQLVSLFLKIVKTHADAVSG KIDKKAEEEIKKEEEKIKEKLRQAKDILKKLQEEIDKTR
DHD40	DRDAHLYKLLTFLEQLVRHLDRLVKHITQLRDIVKKDPEDERAVDVIRQSVRSLEIVITVLKIFVDSVSDAARSKEA EKIVRKIRKEIDEIRQKLREIDKEVKKTTS
DHD40	GSNDKVLDKILDILDRILRLATRVIDLANKLLQVKKKSTHKDPRIVETYKELLKIHETAVRLLLELADLHRRLKSKD EEANKRVETELDRIRKKVKDIEDKVRKLEDKVRKTAS
DHD43	NDLSKEVLKKLEKSVEELLRRVQKSVKEAQKRGLLSDELVDRHLKILNQLVKRHLELLQEVIKRSDKK
DHD43	GSDEAVKRVVEKSLKILDEVIKKSLDILRELIELQIRHAKDDESVIRASKSALKDAIEALKKSLDEIKKALKRSADE G
DHD65	SSEEVVKVHEKVVKLHKEILELLKKIIKIHETAARDPDDKDSIKKLSDEIKKIVKRIEDISDQAKRESSDAQRKQS
DHD65	DKEEESKELLKKLKEILKRSEELLEESKELLKLAKNGEIDESELADADRKLNKKHEKLVQDIQDLLREHERQDR
DHD70	DEKKKIDKIVKETEDLLQKSEKLLQQSKEAVKRIRSQVKENEIVDRLLRISEELLKISRRLVEISRRIASTLS
DHD70	GSSKEEVIRLLKENVRLIKENLELLTRNLKLITDLVRGSNGSEEKIKTLKELLKEYRELLKRYRKLVEDYKRLVDKH D
DHD88	EIQELIKSSRRIIEESKELIKESEEVLRRIKEILDRIRNGVDNQEDLLREILKLLTKNLKIIQRNLKLLQDNAEILK RLVS
DHD88	GSYIEDVIKKILDVSRELIKLSRTIIKISEEINKQLQQGRDTKDLVKKYDEIIKKYTRIVQHYTELIKELQKLLS
DHD89	SPTEEAIQLSQRVIELSKRVIELSKEILKLLKRVLDLLPDLDKNEEKRLDDYDKELKEYDKELKKYEKRLKDLAS
DHD89	GSEEEEILKIQKELLRIQSEILDKQKKILDTLRSNGAVTEEVRSILEKVERLSEEAKELSKEAKELTKEVSKLIS
DHD90	SPLKELNNQLLRLLRELVKVSKKIVDLSKTIIEVLKHTDLDPRLLDSLEKSQQELDKSQKELDKVVKELTKVNKKLQ
DHD90	GSPLEDLVRKYDELVKTYEKLVEEFKKAVDKYDKAVKKAPVSKEATDSLDLIRKVLELLDRNLKLIKENAKLIKELL K
DHD91	SPTRENEKVIKENEKVISDNERVLEEVVKVVETATDRKEIQDAVDEVRKSVDKLRDSVRKLEESVRTLD
DHD91	GSPIKDISKRLLEISKRLVEISDRIVELLQRIADSKDPNKDLQKEVKDVLEEYKRLVREYREVVKEYEKVVS
DHD92	DEDEHVKQLIKNADLLRKHAELLKELVKLFQEIASQIPDDRVAKKVTDVVDRIDKILKQTEKLVRRTKQILDYSR
DHD92	GSNLEELVKLLKEVLEMHERLLRIHEDLVEAHKSNASDKESERKLKKSDKDIKESLKKIKSIIDQVRYIQS
DHD93	PVEDIIEESLRLLEESLKLLNRILKLLEDSLRKLPRSEEWRQRLDEFRKKLEDWKEELERWIEDVRYKKT

DHD93	GSDEDYESREIIDEIRKLLDRSKKIVHRSQRLVERVKSTPLSEDQEDLIRRHEETINRHRELVKELEKVLEDHERHIR
DHD94	PEEDSRRVLERFVRVSREVLKVLEEFLRVSEELLREADRDRDRRLEEYERQVDELREEIRRYKEEVDKFDKEVKYYKK
DHD94	GSPEKDENRKLLDKVRKLVEKSRRLVEELRKLVDQSTKNGLIDEKALRKQQEVLRKVEEVLEKQERVLRELEEISYRVI
DHD94_3:214	GSPERDENRKLLDKVRKLVEKSRRLVEELRKLVDQSTKN
DHD94_3:214	GSDEKALRKQQEVLRKVEEVLEKQERVLRELEEISYRVITRGEDHKAEEDSRRVLERFVRVSREVLKVLEEFLRVSEELLREADRDRDRRLEEYERQVDELREEIRRYKEEVDKFDKEVKYYKK
DHD94_2:143	GSDRRLEEYERQVDELREEIRRYKEEVDKFDKEVKYYKK
DHD94_2:143	GSPERDENRKLLDKVRKLVEKSRRLVEELRKLVDQSTKNGLIDEKALRKQQEVLRKVEEVLEKQERVLRELEEISYRVITRGEDHKAEEDSRRVLERFVRVSREVLKVLEEFLRVSEELLREADR
DHD95	DLSEESKKFVEKVKKLEKESRELEKQVKKIEEDSRSVENDVQKEFLELLKRLLDIQKKVVEVLREVVKVQQYVDS
DHD95	GSDSEYESRQVLRELDTVLKDSHTVLEALRQVIRDSQDVVSKSDEESRRVIDDLEKVIQDSKKVLDDIKRLIDKSKSIKS
DHD96	NEDELLKLLTENLKLLDENLKLLRENLSLLRQANNITDKNRIREIVKQSKEIVKQSREILKQSKEIVERIKYIVS
DHD96	GSSLYELTQRYEKLVQQYEELVKDYRRLVKKLEKLKRDNKPDKRLLKEIVDVIKKSVEIIDRSLKLLEESIKILEETD
DHD97	SQERSLEILKRILDVLKESLEILKESLSILRQLASRIKNPNRKIEEILKESDKIIKESDKVLKEIEEVIRYSS
DHD97	GSDIEYESKEILELIKELLKLSRELLKESRRALELVRKSRDDSIVEEVIQVHKKVLDIHKEVLKIVRKVVEVHRRVKS
DHD98	SKKDESTKLERLAEKIDEITKRIEELVKDVKRKSSEGVDKDQQQKIDEVFQKLLDLQREILEILDRILKVQQYILD
DHD98	GSDLEYLNRRLLQLIKTLIDLNRHLLKLIDKLKKLNSREGDEEKIKEESKQIQEQFKEIVERSKEIIKQIKEIIKRSQ
DHD99	DFERSSRRLEKVVEDLRRSSDRLREVIDELRKSADEKDEDEDLRRARKEHRDLIEELKRALEKQEEIIKHLQELVYRQL
DHD99	GSEESEEVRKVVERIKKISRELEEVVKELDRVSKEFDRHGETDEIVREHERIVEKLEEIVKKHTKIVEELAEIVYKQQ
DHD100	SDDDSVRVLDEIVKILDESVKLLKESLKLLDDFLRTKPDDHLKEVVKESKKVVEQSKKVLDRIKKIIYESK
DHD100	GSDLLYLSKELLKLVRELLKLSRELVELSRRLVNSTHKSPELVKKYDKLVKKYQDLLKKLADVADEYLRQRS
DHD101	DEKDYHRRLIEHLEDLVRRHEELIKRQKKVVEELERRGLDERLRRVVDRFRRSSERWEEVIERFRQVVDKLRKSVE

DHD101	GSDAYDLDRIVKEHRRLVEEQRELVEELEKLVRRQEDHRVDKKESHEILERLERIIRRSTRILTELEKLTDEFERRT R
DHD102	DERYRAREHIRRVEEHTKRLRHILKRLREHEEKLRRELKPGDEITESVDRFKKIVDQFEESIKKFETVSEELRKSDS
DHD102	GSDRQRILDRLDKILEKLDDILKKLKDILETLSKDDVSDRRHKDLVEKFRELVDTHHKLVERYRELVYQNR
DHD102_1:243	GSDEITESVDRFKKIVDQFEESIKKFETVSEELRKSIS
DHD102_1:243	GSDPQRAADRLDKILEKLDDILKKLKDILETLSKDDVKDRRAKDLVEKFRELVDTHHKLVERYRELVYTATAGSDLA RELIRRVEEHTKRLRHILKRLREHEEKLRR
DHD103	NADDQLATSIKKLEDSIDQLIKIVRKFEESVKKLQKHGVDQHHVEILRKIVEIFRQHIEKLKKHLEKLRYTSS
DHD103	GSDKEYLVTEHEKLVREHEKIVSEIEKLVKKHEAGVDESELEEILKKVEKLLRKLDEILEQLTQLLRKTE
DHD103_1:423	GSDQHVVEILRKIVEIFRQHIEKLKKHLEKLRYTSS
DHD103_1:423	GSDAEYLVTEHEKLVREHEKIVSEIEKLVKKHEKGVDESELEEILKKVEKLLRKLDEILEQLTQLLRKAEKHIDKHS KAADQLATSIKKLEDSIDQLIKIVRKFEESVKKLQKH
DHD104	DEDDDIRRVLDESRRVLEHSRRVLKRSEEVLEKASRKKEKDTEEIEKHLKRLREHAKKLEKHRRELDDFLYKEI
DHD104	GSRDKYLLERLNDILKKLDEIVDKLSDILKRLKDVRHDDRLQELVERYKEIVKEYKRIVEEYEKLVREFEEQQR
DHD105	DRDYEDKEFKKIIKELEDVQEELKKLQEKIKRFSSELEEPNELLKEQLKVNEEQLEVNKKILKILRDQLKQNE
DHD105	GSDAEYKVRESVKRSKESVKHSEDVVDKLNKSVKLSESGHSDAEKASRELVKLVREVVELSREVIKLSEKVLRVIS
DHD106	DLQYKQEKLIRHFDRVVREWDKLVRKFSKVLEKQKHESKDKELEEASRRVDELIKRLREQLKRSKEILRRLKELSRK SS
DHD106	GSDWEELLRRLEKVLQEYEEIVKELIDLIERLIKVSEDKSKDASEYKKLVTELEKLISKLEEISKKLEELVKEYEYK TE
DHD107	DAKDELEKSLQEIEESLKELKKLLEELDKSLRELTSQGRNKKLEEHIKKVQKFIELVKKYIKAVQDYLKEVRYDNS
DHD107	GSDKERAARATEEMVKLTKKLLKAVEDLVRDVRRLLKEGLISEKHARIAETILEVFKKHAKIIKKHVDIVKYDES
DHD108	GSPLKERLLEIQRDLDRVLEEVVERLLRIQERLDSVVERKPPDVHEEYKYIVDEIREIVERVVREYEEIVKRIDEEV R
DHD108	GSEEDERIRYDLDRIRKDVRRKLEEIRQRVRELEKKLRDAGHRRDEKELLRELIETSKDILRLVEELLKKIIDKSED LLRKTE
DHD109	GSDEEDYINENVEKDVRDIEDDVRRINERIRELLEKIRTEEVLQRVLEEHHELVERVLRKLVEILRKHEEENR
DHD109	GSDEEEYYKEKLHKLLREIEELLKHYRELVRRLEELVKRGELDKDTAAHILERLSELLERIIRRVAHTLRRLSEERR

DHD110	GSDEDEISYDSKRRVEEIVRQAREKSEKSRKDIEDVAEVLRKGDVSEKEVVDELVKVLEEQVKVLREAVERLREVLK KQVDDVR
DHD110	GSDIVELVDHLLKRSLKLLEELAELVRRLLEKSTELLKRRTEEHKEEVVEESEYMVRELEERLRRVVDESEKLVRDA DKHIR
DHD111	GSKEKDIVKTLVDLLRENLETLERLIEEVVRLLKENVDVRDEGRDDKDSERILRDIKRRIDEAAKESREIIERIEKE VEYRSR
DHD111	GSPEVDVLRRIVREILKASEELLRLLRKLIDEALKLSERKRDSQEYREVVDRVKKELERLLDEYRKLVEELKEKLRY DTR
DHD112	GSDKRYESEKLKRRLDEAVEKVREVVERVERESDRVLEEVRRRRESKEVVDKVIEDNDKALEDVLRVVDEVAKVVRD VVRENTR
DHD112	GSPREYHSKDILRKVDEILERIRRHADRVKKKSERLKRENVDVNEHSKDVKRVIRELLELVKELLRLAKKHSDDQQE
DHD113	GSDEDEILYHSERLLQKLKKELDDLKEKSRELLEELKKEDPDDRLIERIIRLHDEVLKDLDEVLKNILEVHREVLER LR
DHD113	DKLDRLLKIHEEALRRAEELIKRLLDIHRRALDLARRGELDDYLLKESERELREIIRRAREELKESRDRLEEISR
DHD114	GSPKEELIRRVLEEVKRLNEKLLEIIRRAAELVKRANDELPETEKLREIDRELEKKLKEIEDELRRIDKELDDALYE IED
DHD114	GSPKLDKLRELLERNLEKLREILEEVLKILRTNLERVREDIRDEDVLQEYERLIRKAEEDLRRVLKEYDDLLKKLVY ELR
DHD115	GSKEDESVKRAEEIVRTLLKLLEDSLREAERSLRDIKNGEDEHNLRRISEKLEELSKRITETIERLLRELQYTSR
DHD115	GSPNQELLDRVRKILEDLLRLNEELVRLNKELLKRALEMRRKNRDSEEVLERLAEEYRKRLEEYRRELEKLLEELEE TIYRYKR
DHD116	GSDESEEAQHEVEKVLDDIRRLSEHLQKRLEEVLEEVYELRREGSDRTEVVELLKEVIREIVRVNREALERLLRVVE EAVKRNE
DHD116	GSDEEELVETVKRIQKEILDRLTELAKLLVEIQREIKKLKDEGEDDKELKRLSDELEEKVRQVVEEIKRLSDELEET VEYVSR
DHD117	GSDEEEEVVRRAEELVKEHEELIERVIRTHEELVYKLEDQGADKKLVDVLKRVVEESERVAREIVKVSRELIRLLEE ASR
DHD117	GSSKEEILKELEDLQRRLIEELKKLQERVVELLEELIKRLRDRGRDDKHLKRLVKEVRRLSEEVLRSIKEVSDRVRY QLR
DHD118	GSDKEEESEYLLRDLVRLLEKVKEKIEEVNREVEKLLKKVKDGRLDRREVLREILRLNRELAEIIKEVVDRIRHVVE RSER
DHD118	GSDLHEVVYETKELLKRIEEVVEELRKKSEDIIRKAERGEISEDELKRLQEEIAREAKKLLDEIKRVLERHLEQTL

DHD119	GSPVEEIIKEVVKRVIEVQEKVLRIISHAVKRVVEVQKKYDPGSEESNRVVEEVKKTIEDAIRESDEVVDEVVKRIQ YTVR
DHD119	GSPEQEIADRILTEIRESQKELERLARKILKLLDESQEKAKRGRLSEEESDELLERIKKELDELLERSKELLKKIEY ELR
DHD120	GSDEDKEANRVLDEVLKTVRDLLETANEVLKEVLYRLKRTDDQEKVVRTLTEVLKEHLKLVEEIVRILDKVLKEHLE TEK
DHD120	GSPEDDVLRRLEEVSEKILRVAEDVARQLREVSEKITQGKVDRKEWEEDIKRLKRELEELLREWKEEIERLTYELR
DHD121	GSRREEVVKRIRELLKRNKELIDRIRELLEENEYLDKDARDKDVLRRSVELLEELVRILEESVELAKEIIKLLREVV E
DHD121	GSDEKEDNRRLQHKIERILEKNEDLQRKLEEILELLERGEADEEKIDRLRKAVEDYRRVVEEIKEDVKRHKYTVR
DHD122	GSDEKEEAKKASEESVRTVERILEELLKASEESVELLRRGEDAKDVVERSKEALKRVKELLDEVVKRSDEILKYIHN
DHD122	GSDEKKLINEVVETQKRLIKEAAKRLSEVVRHQTELIRELREKNVDDKDVEKLLKESLDLAEEIVRRIKELLDESKK LVEYVSN
DHD123	GSPDMDEVKRVLDELIEIQEEILREIKRVLEKLIKIQEDNGSEYESREVVREIVEIARKLVERSRRVVKKITETLQ
DHD123	GSDERYATREIVERIERIAREILKRTEEIVREVREVLSRDVDQEEVVRRLADLLRESVELVQHLVRRVEELLQESVE RKK
DHD124	GSPEREALREVLEDLKRVTDRLRELVERVLEELKKVTDHVDSERILRESRRVLKELKDIIEEILRESEKVLEKLKYT ED
DHD124	GSPAREILEEVVKKHLEVVEDAARILEEIIREHEKAVREDRDKKELEEISRDLLRKAREALKKVKDISDDLSREIEY VAS
DHD125	GSPVEEAIKKVIDDLRDVQRKIRELVEELIRLLEEVQRDNDKRESEYVVERVEEILRRITETSREVVRKAVEDLS
DHD125	GSDSDEKAEYLLKEMERVVRESDEVVKKILRDLEEVLERLRRGEISEDDVTEILKELAERHIRAIEELVRRLRELLE RHKR
DHD126	GSPVEEVLKELSEVNERVRDIAREIIERLSEVNEEVKETDDEDELKKISKKVVDEVEDLLRKILEVSEEVVRRVEYH DR
DHD126	GSPKEDILREVLRRHKEIVREIVRLVREAVETHLELVKRNSDDRDAQDVIRKLEEDLERLVRHAQEVIEEIFYRLH
DHD127	GSPRSYLLKELADLSQHLVRLLERLVRESERVVEVLERGEVDEEELKRLEDLHRELEKAVREVRETHREIRERSR
DHD127	GSDREYIIKDILDSQEHLLRLIEELLETQKELLEILKRRPDSVERVRELVRRSKEIADEIRRQSDRNVRLLEEVSK
DHD128	GSDEKDEIRHVIESVERLIEDIKRLLKTLRELAHDDSDKKTVKEVLDRVKEMIERHRRELEEHRKELERAEYEVR
DHD128	GSESEDRIKELLKRHIELVERHEELLHEIKKLIDLEEKDDKDREEAVKRIDDAIKESEEMLEESKEILEEIEYLNR
DHD129	GSSLEDSVRLNDEVVKVVERVVRLNQEVVRLIKHATDVEDEETVKYVLERVREVLDESREVLKRVHELLEESERRLE

DHD129	GSHEKDIVYKVEDLVRKSDRIAERAREIVKRSRDIMREIRKDKDNKKLSDDLLKVTRDLQRVVDELEELSRELLRVAEESRK
DHD130	GSPELDEVKKLIDELKKSVERLEESIREVKESIKKLRKGDIDAEENIKLLKENIKIVRENIKIIKEIIDVVQYVLR
DHD130	GSDEEEIEELLRELEKLLKKSEEALEESKKLIDESEELLRRDRLDKEKHVRASEEHVKLSEEHLRISREIVKILEKAVYSTR
DHD131	GSDESDRIRKIVEESDEIVKESRKLAERARELIKESEDKRVSEERNERLLEELLRILDENAELLKRNLELLKEVLYRTR
DHD131	GSDEDDELERLLREYHRVLREYEKLLEELRRLYEEYKRGEVSEEESDRILREIKEILDKSERLWDLSEEVWRTLLYQAE
DHD132	GSDKKDASRRAIRVLHEFVRVSEEVLEVLRKSVESLKRLDVDEKIKRTHDRIEEELRRWKRELEELIERLREWEYHQD
DHD132	GSDDEEEDKRLLEEVKRSLDTDERILEKLRHSLERQLEDVDKDEDSRRVLRELDEITKRSREVVKRLRKLAYESK
DHD133	GSDKEYKLDRILRRLDELIKQLSRILEEIERLVDELEREPLDDKEVQDVIERIVELIDEHLELLKEYIKLLEEYIKTTK
DHD133	GSPSKEYQEKSAERQKELLHEYEKLVRHLRELVEKLQRRELDKEEVLRRLVEILERLKDLHKKIEDAHRKNEEAHKENK
DHD134	GSRDRKISEELIKALEDHIRMLEELIRAIEEHIKLAERGVDEKELRESLEELKKIVDELEKSLEELRKLAERYKYETR
DHD134	GSPKEESVEELKRVIDKHEEILRELKRVLEEHERVSHDEDENELRRSLERLKHILDRLHESLKELHELLKKNEYTER
DHD135	GSDHEYWVKIVERILRVMEKHAEIVKKHLEIVERVVREGPSEDLRRKLKESLREIEESLRELKELLDELDELSEKTR
DHD135	GSDEEYVTRSQRRLKRLLEEYIKVVEEHARLVERNERDDKELKRSIDELDKLTKELLELVKRYKELVDKTET
DHD136	GSDKEEIVKLQDEVIKTLERHLDILRKHIDLLEKLKDHLSEELKERVDRSIKKLEESIKRLERIIEELQELAEYSL
DHD136	GSREEELKESAEELERSVRELKKEADKYKEEVDRLHYRGKVDKDWVRVVEKLIKLVEEHLELIREHLELLKEERR
DHD137	GSDMEYELKKSAEELRKSLEELKRILDELHKSLRELRRHGDDEEYVQTVEELRKELEEHAKKLEEHLKELERVAT
DHD137	PEYELKKSVDDLKRDVDRLVEEVEEVFELSKERLREDRKHLELVEEMVRLIEKHLELIKEHLKLADDHVR
DHD138	GSREKDESKELNDEYKKLLEEYERLLRRSEELVKRAKGPRDEKELKRILEENEDILRRTKEILERTKEISEEQKYRRR
DHD138	GSDKDERQERLNEESDKSNEESERSNRESEELNRRARGPNDEKELQEILDRHLELLERNQRLLDENKEILRESQYLND
DHD139	GSENKYILKEILKLLRENLKLLHDILRLLDENLEELEKHGAKDLDDYRRKIEEIRKKVEDYREKIEEIEKKVERDR
DHD139	GSESEYTQEEILELLKESIKLLREILRLLEESEELWRRENTKSERSEEIKERAKEAIKRSEEILERVKRLSDHSR

DHD140	GSDEEEANYVSDKAVKIAEDVQELLKELLELSEVVRRGEVDEDEYDRVLRKLQEVMKEYEEVLKEYEEVSRKHE
DHD140	GSPEKYLIKTQEELLRRHAEILEDLIRKVERQVDLRRKVDERDEDLKRELERSLRELERLVRESSRLVEEIRELSKE IKR
DHD141	GSDEEYELERISRESKELLERYKRLLREYQELLKELRHVKDLDRAVKIIHELMRVSKELVEISHRLLELHERLVRRR K
DHD141	GSEKEYIEKLSRKIEEDIRRSEERAKDSERLVRRLEELAKRKRLDLDDVLRVAEENLEILEDNLRILEEILKEQDKS NR
DHD142	GSPHEEVVELHERVMEISERAVELIQRIIDIIRRIREDDKDIEKLVKTIRDLVREYEELHRELEEIDEEIYKKSE
DHD142	GSDHEDVVRLHEDLVRKQEDARRVLEEIVRLAEEIVEVIKKDEKDKDRVTRLVEEIEKLVEEYKKKVDEMRKISDEI KYRSR
DHD143	GSRAREVVKRAKRIIEEWQKILEEWRRILEEWRRLLEDERVDDRDNERIIRENERVIRENEKIIRDVIRLLEELLYE RR
DHD143	GSREDEELEEEIDRIRQMVEEYEELVKEYEELTEKYKQGKVDKEESKKIIEKSERLLDLSQDAVRKVKEIIRRILYT NR
DHD144	GSPKEEIVKLHDESAELHRRSVEVADEILKMHERSKDVDDERESRELSKEIERLIREVEEVSKRIKRLSEEVEYLVR
DHD144	GSPLEEILKIQRRINKIQDDINKILHEILRMQEKLNRSSDKDEVEESLRRIRELIKRIKDLSKEIEDLSREVKYRTT
DHD145	GSPEDEHVYVVREIYEVLREHAEVLEENREVIERLLEAKKRGDKSEELVKELKKSIDKLKEISRKLEEIVKELEKVS EKLK
DHD145	GSDEDETSYRILELLREIVRASRELIRLSEELLEVARRDDKDETVLETLIREYKELLDRYRRLIEELTRLVEEYEER SR
DHD146	GSTQEEINRIQHEVLRIQEEIDEILRDIVEKLKAISRGELDHEVVKDVEDKVREALEKSEELLDKSRKVEYKSE
DHD146	GSDEEELNRELLEKSKRLVDINRDIIRTAQELIEMLKDSKDGRVDEDTKRELRDKLRKLEEKLERVREELRKYEELL RYVQR
DHD147	GSDEKDRVYEILKEVQRLVKEYRDISKEIEDLVKHYEHITDDEAQEVSKELIDKSLRASEIVRELIRLIKELLDELE
DHD147	GSDEEDVLYHLRELLEELKRVSDDYERLVREIKETSERKDRDTKENKDMLDELVKAHREQEKLLERLVRLLEELFER KR
DHD1	PREQAIRISEEIIRISKKIIEILERTRSSTAREAMKWAKDSIRLAEESKYLLDK
DHD1	IEDDVKKIQDSTKKAQKETIEALERSTSSTARKQMEEQKEQIRLQKEAMYLLKK
DHD2	SREEIAKLQEEVIKLQRRVIELQKEVIELQRRAKELTSSYTKEILEIQRRIEEIQREIEEIQKRIEEIQEEIQRRT
DHD2	SDEEIKRLSEEVIQLSRRVIKMSREAIKLSREVQKLTPSYQKRIKEIADRSIELARESIEIAKRSEKIAEESQRRT
DHD3	PAKDEALKMANESLELAKKSARLIQESSSKEILERIEKIQRRIAELQDRIAYLIKK

DHD3	PAKDEALRMIDESRELIKKSNELIQRSSSKEILERILEIQRKIAELQKRIQYLLKS
DHD4	TDEARYRSERIVKEAKRLLDEARRRSEKIVREAKQRSNSEDAKRIMEENLRESEEAARRLREIIRRNLEESRETG
DHD4	TREALEYQRKMAEEIEDLLREALRRQEEMVREAKQRSLSEEFKRIMERILEEQERVMRLAKEALERILEEQKRTG
DHD5	SERTKREAKRSQEEILREAKEAMRRAKESQDHRQNRDGSNSEDLERLSQEQKRELEEVERRLKELAREQKYKLEDS
DHD5	SEDLKRILKEITERELKLMQDLMEILKKITEDENNLDSNNSEDLKRSIEKARRILDEALRKLEESARRAKYIQEDN
DHD6	TEDEIRESLKWLDEVLQELREIARESNEVLERNRQKSRSDKLREDIERYKKRMEEARKKLDDQLNKYKKRMDENRS
DHD6	TEEELKESKKFAEDLARSARRALKESKRVLEEISQASRSKKLEEIVRRYKEQVKRWQDEWDERAREYRKRMKENRS
DHD7	TKTEEIERLAREIKKLSEKVERLAQEIEELSRRVKEENSTDRELKEANREIERAIREIEKANKRMEEALRRMKYNG
DHD7	TKTEEHERLAREISKLADEHRKLAKIIEELARRIKEENLTDDELREAIRKIEDALRKNKEALKIMKEAAERNRYNT
DHD8	TKKEESRELARESEELARESEKLARKSLELARRAESSGSEEEKRRIIDENRKIIERNREIIERNKEIIEYNKELIS
DHD8	TKDEESLELNRESEELNRKSEELNRKSKELNDRAESSNSEEEEKEILREHKEILREHLEILRRHKEILRRHKYLTS
DHD16	TREELLRENIELAKEHIEIMREILELLQKMEELLERQSSEDILEELRKIIERIRELLDRSRKIHERSEEIAYKEE
DHD16	SEDIAREIKELLRRLKEIIERNQRIAKEHEYIARERKKLDPSNEKERKLLERSRRLQEESKRLLDEMAEIMRRIKKL LD
DHD18	DRQKLIEENIKLLDKHIKILEEILRLLKKDIDLLKKSSSEEVLEELKKIHRRIDKLLDESKKIHKRSSEIVKKRS
DHD18	DEQKLIETSQRLQEKSERLLEKFEQILREASDLYRKPDSEELLRRVEKLLRELEKLIRENQDLARKHEKILRDQS
DHD19	DRQELIRENIELLKKHIKIVKEIQKLIETFIELLKKSSSEEILRRLKKILKRIEKLYRESQEIHKRSEEIAKKRQ
DHD19	DEERLIDKSRELQKESEELLKELLKIFKRIEELLEKPDSEELIREIKKLLETLSEIHKRNEKLARTHEEILRQQS
DHD22	STRDVQREIAKAFKKMADVQKKLAEEIKRHVKNVEKKNKDNDEYRKIATELLKKATESQKKLKELLDRIRKSDS
DHD22	DKDDRSTSLLKRVEKLIDESDRIIDKFTTLIELSRNGKIDDDQYKKELKEILELLKKYDKHVKEVEELLKRLNS
DHD23	SKRKALEVSERVVRISEKVVRVLDESSDLLKKSYDDSDKFAELIDRHEEKIKKWKKLIKEWLEIIQRHKS
DHD23	SAEEFVKLSEEAVKRSKEILDIVRKQVKLVKAGVDKHEITDSLRKSEKLIEEHKELIKTHRDLLRREN
DHD24	SSTEILKRFKRALRESEKIVKHSRRVLKIIREVLKQKPTQAVHDLVRIIETQVKALEEQLKVLKRIVEALERQS
DHD24	DKQKEIKDILEKTRRIAEESRKIAEKFDEIIKRSTEGKIDESLTKELEELVKEVIKLSEDDARTSDDLVRKES
DHD26	DEDESIKLTRKSIEETRKSLKIIKEVVELIREVLKHIKDLDKEIFERIDKILDKYKKQVDTYDEILKEYEKKQR
DHD26	SELDEQKELIKKQEKLIEEQQRLLSKIRRMFKERVKDQELLREIQKVLKRSQEIVETSKKILDRSDKTTE
DHD28	DQKEINTRIVEKLERIFKKSKEIVRQSERVISTIEKKTEDERELDLLRRHVKIVREHLKLLEELLKIIKEVQKESE
DHD28	DTEELVKRLNELLKELSKLVKEFIKILETYRKDQTKDTSKISERVDRILKTYEDLLQKYKEILEKIEKQLS
DHD29	DYARLIDQAVEVTRKVVEVNVTVARVNDKFAKHLGDEELRRVSEHLKEVSKDLQEVAKKSKDAARQVK

DHD29	DVSKVAEEYLQISKTLVDISRTLLEISERLVRLVRTVADDRSEVKKAIEDSIEVLKTSEEVVRQIKRASDKLVKAIS
DHD31	DAKEIQRRVVEIQTEVVKLQKKAVDIIRKIIEAFNNSNIDQSLLEAAKEIVKEIDKLEKLTESLLEESKKLLKRSS
DHD31	SAEEVVKLAKIFLELLRESIKLLKRSVDLLRKSSDPSLDKSEAEKVSREIEKVSDTSLKLSKKALDVVKRALKVAS
DHD32	DEKDAARKARKVSEEAKEASKKIEKALEESKRILNTLKQKKDEQEVKVIKEHEDVLRQIEKIQKQVLEIQKEVAKLL ESLD
DHD32	SADDVARASEKVLRVARESAKAADKSLEVFKEVVKRGDKEAFLQVVKINEEVVKINITVIRILIEVSKTAT
DHD38	DEYVKETLKQLREALASLREADKRITELVKEARKKPLSEAARKFAEAIVTHVKVVVEHVEVVLRHVEVLVEAKKNGV IDKSILDNALRIIENVIRLLSNVIRVVDEVLQDLD
DHD38	DASDVIRRIHELFEEVHRLIEAVHRAIEDVAKAAQKKGLDESAVEILAELSKELAKLSRRLAEISREIQKVVTDPDD KEAVERLKEIIKEIKKQLDELRDRLRKLQDLLYKLK
DHD60	SEDKAHHDIVRVLEELIKIHDELMKISEEILKATSDSTATDETKEELKRRSKEAQKKSDTLVKIVKELEKESRKAQS
DHD60	DDEEKYRQIIREAQEISKTAKRILRDAQEISKRIRHQGVDRSEHQRLVDLLRELIKEHHKLLRRQQEADTRND
DHD63	DRKDKARKASEKLEEVIQRWKTVADKWKKMVDLVSNGKLSQEEVARVTEELLKIQTELAKLLEEHAKVLQESAS
DHD63	SDEESIKTQSELIKTSEELLKDVKRIDEELQKLRDDPTLDESELKKRVKEWSDRVRKAKEISRKIQEIVKESKKRSS
DHD66	DKDEELRKVIEKYREMVKEYRKVIREYEEVIKSSKTIDKSSLISLSRKMVELSQRVIDVSDEVAKVLSRKQS
DHD66	TDEERLKKQTKELKEQTKQLEKQKDLLEKISNGEISKDEIQEIIKESKKIAKESQKALDSSRKALEEVS
DHD67	DEKEVSKEIIKVLKDIAKVQQKVIEVSQRLASVLRADDDNVVKRALEEYEKILEELRELNKEIEKLTDKYRKVTS
DHD67	DSDEQTKELEKLTELHKRHVEKLKKQTKESREVDSNKLWKSKDVKDKLSESEKELQKLSDQDKKAKDALESSRRKND
DHD69	DAEEQLKLLTKLLRHQQRLLQLIKESLKLIEKIDQSSQENQDEIRKWREVTKKLRELIKTSEKLVRELEKSYKKSS
DHD69	SLRDVVRRYQELVRRYDELIKTLTEILKKYQKKGAEDKDASTELVKAVRTSLKLSKELLKLNSELLKEDS
DHD71	SKEELKRKLDELKKRSDTLKELSKKLKEISERNPDDKSVHRTIIRIHREFVKNHKEIVRVIEEIVSDKS
DHD71	SKQDEHDRLLKIHDKLVKQHDELLKLLTKLSRAGDSVTKKKLEEILRKLQEVSKQLEESLKDADKVSKDIN
DHD72	TVQSLLEQHVKIVKRSIEILERHTQILQDIARSQGVSKELEDVERQVKEYRKEVKKLEEDLRQLSRNSK
DHD72	SDSDRIEKLIRESTELLKEQQKLAKRSRELAETVESLPLTEEYLKQQREHQKKIEKLLKDSEKHLEELKRLVKSEK
DHD73	DSEKRIEDILRTDLELAKRDAELVKEHIKLVKRIDLSEELKKQVEDVEKESKKLEDSSEKLVQKVRKRSS
DHD73	DEEERAKDLRKYLEEQTQYYRTVTEHLRNLEKVVEELERRGKPSSELQQILERSQRIYKETTEIYDTSKKLIEELDK HHR
DHD148	PLEDILKRHLDKVRELVRLSEEVNKLAKEVLDILKDKRVDEKELDKVLKELEKVVEEYERAVKESRDLLRELRETTR
DHD148	DKERLLEIHERIQKLLDRNLEIIERLLRLLREARDIKDDDKLDKVIKRLKELSEESKDILDKIKELLKESEKELT

DHD149	PEDEVIRVIEELLRIAAEVDEVHRRNVEVQEEASRVTDRERLERLNRESEELIKRSRELIEEQRKLIERLERLAT
DHD149	DLEELIKEYAEVVRRHHKAVRDLERLVRELANAKHASEEELKRIATEILRIVKELIRVQERLIKLSEDSNEESR
DHD150	PTDEVIEVLKELLRIHRENLRVNEEIVEVNERASRVTDREELERLLRRSNELIKRSRELNEESKKLIEKLERLAT
DHD150	DNEEIIKEARRVVEEYKKAVDRLEELVRRAENAKHASEKELKDIVREILRISKELNKVSERLIELWERSQERAR
DHD151	PKEDIDRVSRELVRVHKELLEVLRKSTEIVEAVARNEKDERTIEEVLEEQERAVRKLEEVSKKHKEAVKRLK
DHD151	ELERLSEEIQKLSDRLIELIRRHSKVLEEIVRLLKHKDNDEREVRRLLKLLRDLTRRYEEVLRKVEEIVKRQEDESR
DHD152	PEEDILRLLRKLVEVDKELLEVVRESTEVVRLVARNEKDVETVERVLRKQEEVVRKYERVSRELEEAVRRLK
DHD152	ELKDLVEEIVKLSKENLKLWEDHSRVLEEIVRLLKHKDNDEREVRRLLKLLEDLTRRAEETSRRIEEIVKEAEDRAR
DHD153	DEERELREVLRKHHRVVREWTKVVEELKRVVELLKRGETSEEDLLRVLKKLLEMDKRILEVNREVLRVLEKRLT
DHD153	SLEEIIEELVELVRRSVEIAKESDEVARRIVESEDKKKELIDTLRDLHREWQEVTKRAEELVREAEKEVR
DHD154	TAEELLEVHKKSDRVTKEHLRVSEEILKVVEVLTRGEVSSEVLKRVLRKLEELTDKLRRVTEEQRRVVEKLN
DHD154	DLEDLLRRLRRLVDEQRRLVEELERVSRRLEKAVRDNEDERELARLSREHSDIQDKHDKLAREILEVLKRLLERTE
DHD155	PEDDVVRIIKEDLESNREVLREQKEIHRILELVTRGEVSEEAIDRVLKRQEDLLKKQKESTDKARKVVEERR
DHD155	DEVRLITEWLKLSEESTRLLKELVELTRLLRNNVPNVEEILREHERISRELERLSRRLKDLADKLERTRR
DHD156	DEDEVVKVHEEHVKSHEEIHRSHEEVVRAAEEDKRDSRELRTLMEEHRKLLEENEKSIEEVKKIHERVKR
DHD156	KKEELIDISKEVLDLDDEINKISKEILELIKKLLRLKEEGREDKDKAREVKRRIRELHRRIQELNKRLRELHKRVQE TKR
DHD157	PEEDIARRVEDLLRKSEELIKESEKILKESKRLLDRNDSDKRVLETNLRLIDKHTKLLERNLELLEELLKLAEDVAK
DHD157	RFKDLSREYIEVVKRLLELSREALEVLREIKDTDKTDKKRIKELIDRLRKLIEEYKRIIDRLRKLSKDLEEEHR
DHD158	DEEELVKILKELQRLSEESLEINKRLVEILRLLRRGEVPKEEVEKKLREIKKEQEKLDREHEKIKKRIEEITK
DHD158	SLKEKILEIIERNMKLVELSNRSVEIVARILKGEKDDEETLERLLREWDKITRDYEEIIKESRKLVKELEEEAK
DHD159	SKTEILRKALEIHKEQIDIVRKLIELSEEVLKLVEESKEKNLEKLKRIDEETDRLLERLDELHKRLTELAERLK
DHD159	SDDEARKQLEEMKRRLREVEKKSKRVEERVRELERLVRENREDEDRVLKTLEDLLRENEKLVRTIERHVREQRELSK EVK
DHD160	SEEELEKKADELRKLSEEWRKLQEEDKRLSEMVEKGELDLQEVDEHSLRVLERATEVHRTVDKVIEEILRTTN
DHD160	SEKERHRESQETQEEIRRTHEEIIRKLEEILRRAKAGELPEETLDRLRRIMERLKELSERLDDLVRKLRDDHRREQK
DHD161	SEKEILEELKRILKRVKDISDRLEELDKRTEEIARREPTKELVDELVKIHRDWLRLHEEILKLVDDALKKVEDATK
DHD161	DLRELLELQREASRLHRELVKLLTELVKKLELIAKGEDIREEDLKRIKERLEEIKKRSKRIKEESDEIDKKTK
DHD162	SERELQRELNKIVRRILEIHREVSELHQRAVKLIRENDNSEELEEISRRIEELSKELEKLVREHDEIVKTIE

DHD162	SEREKLDRNDEELKEINKRVEEIKERSDRITEAIEKNERSEEEIRRLSREQNEALQRLLELHKKLVKLHRELLEDTR
DHD163	DKEDVIRVHDEQHKLIEEQLELTRRIAELVREIAKNTASEEEIKEMLKEIKRLDDRSREIQDRLQKLLEEIRRKTK
DHD163	TEEEIVELNKDIQRKSKEHIDLQNELVKKIERAIRENNITEELLEELERLLRESEKIVEEIRRITDKIRKDAK
DHD164	SEKEILERLLRLSKEQNEISEEIHRLTERLVELKRRKDDDERLKRILDRQKRLVERAREISKEYEDLLRKLE
DHD164	SMEELLRKNARLSRKQLKIIDEHLELSTKLTRGEAGDETLEEIERRSREMLEEQRRVDEESKRIREKLK
DHD165	SEEEIRDIVEKLLRTHEEVLKEIKKLLDDSERVRRRELDKKDLDRIQKEQRDIQEENKEKAKRFDELVKELKKAAK
DHD165	SEEEHRRTMEKVEKEVRDIKRRSEEVKKKVKANTLSEEDLVRLLERLVEDHKRLQDLSQEIIERDEKATK
DHD166	DEDELAKEIEDVQRRNKESQEEHDKSVKKLEAAERGEIDEDSLLRVLEEDIKVLEKDIEVLERSIEVIEKAE
DHD166	SEKELIRRLLEQQRQHLRLSERLIELSRRLVEVVRKGKDNRDLLRELKKLSEEHKKHSKDDHEKVREIREREK
SYNZIP1	NLVAQLENEVASLENENETLKKKNLHKKDLIAYLEKEIANLRKKIEE
SYNZIP2	ARNAYLRKKIARLKKDNLQLERDEQNLEKIIANLRDEIARLENEVASHEQ
SYNZIP3	NEVTTLENDAAFIENENAYLEKEIARLRKEKAALRNRLAHKK
SYNZIP4	QKVAELKNRVAVKLNRNEQLKNKVEELKNRNAYLKNELATLENEVARLENDVAE
SYNZIP5	NTVKELKNYIQELEERNAELKNLKEHLKFAKAELEFELAAHKFE
SYNZIP6	QKVAQLKNRVAYKLKENAKLENIVARLENDNANLEKDIANLEKDIANLERDVAR
SYNZIP7	KEIEYLEKEIERLKDLREHLKQDNAAHRQELNALRLEEAKLEFILAHLLST
SYNZIP8	KEIANLEKEIASLEKKVAVLKQRNAAHKQEVAALRKEIAYVEDEIQYVEDE
SYNZIP9	QKVESLKQKIEELKQRKAQLKNDIANLEKEIAYAET
SYNZIP10	NLLATLRSTAAVLENENHVLEKEKEKLRKEKEQLLNKLEAYK
SYNZIP11	ELTDELKNKKEALRKDNAALLNELASLENEIANLEKEIAYFK
SYNZIP12	NEDLVLENRLAALRNENAALENDLARLEKEIAYLEKEIEREK
SYNZIP13	QKVEELKNKIAELENRNAVKKNRVAHLKQEIAYLKDELAAHEFE
SYNZIP14	NDLDAYEREAEKLEKKNEVLRNRLAALENELATLRQEVASMKQELQS
SYNZIP15	FENVTHEFILATLENENAKLRRLEAKLERELARLRNEVAWL
SYNZIP16	NILASLENKKEELKKLNAHLLKEIENLEKEIANLEKEIAYFK
SYNZIP17	NEKEELKSKKAELRNRIEQLKQKREQLKQKIANLRKEIEAYK
SYNZIP18	SIAATLENDLARLENENARLEKDIANLERDLAKLEREEAYF
SYNZIP19	NELESLENKKEELKNRNEELKQKREQLKQKLAALRNKLDAYKNRL

SYNZIP20	STVEELLRAIQELEKRNAELKNRKEELKNLVAHLRQELAAHKYE
SYNZIP21	NEVAQLENDVAVIENENAYLEKEIARLRKEIAALRDRLAHKK
SYNZIP22	KRIAYLRKKIAALKKDNANLEKDIANLENEIERLIKEIKTLENEVASHEQ

In some embodiments, a pair of dimerization domains as described herein bind noncovalently to each other.

In some embodiments, a pair of dimerization domains as described herein bind covalently, e.g., to form a fusion (e.g., an intein mediated fusion, e.g., as described herein). In embodiments, a pair of intein dimerization domains comprise a Chain A sequence (or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto) and a Chain B sequence (or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto), as listed in a single row of Table 33.

Localization sequences for gene modifying systems

In certain embodiments, a gene editor system RNA further comprises an intracellular localization sequence, e.g., a nuclear localization sequence (NLS). In some embodiments, a gene modifying polypeptide comprises an NLS as comprised in SEQ ID NO: 4000 and/or SEQ ID NO: 4001, or an NLS having an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

The nuclear localization sequence may be an RNA sequence that promotes the import of the RNA into the nucleus. In certain embodiments the nuclear localization signal is located on the template RNA. In certain embodiments, the gene modifying polypeptide is encoded on a first RNA, and the template RNA is a second, separate, RNA, and the nuclear localization signal is located on the template RNA and not on an RNA encoding the gene modifying polypeptide. While not wishing to be bound by theory, in some embodiments, the RNA encoding the gene modifying polypeptide is targeted primarily to the cytoplasm to promote its translation, while the template RNA is targeted primarily to the nucleus to promote insertion into the genome. In some embodiments the nuclear localization signal is at the 3′ end, 5′ end, or in an internal region of the template RNA. In some embodiments the nuclear localization signal is 3′ of the heterologous sequence (e.g., is directly 3′ of the heterologous sequence) or is 5′ of the heterologous sequence (e.g., is directly 5′ of the heterologous sequence). In some embodiments the nuclear localization signal is placed outside of the 5′ UTR or outside of the 3′ UTR of the template RNA. In some embodiments the nuclear localization signal is placed between the 5′ UTR and the 3′ UTR, wherein optionally the nuclear localization signal is not transcribed with the transgene (e.g., the nuclear localization signal is an anti-sense orientation or is downstream of a transcriptional termination signal or polyadenylation signal). In some embodiments the nuclear localization sequence is situated inside of an

intron. In some embodiments a plurality of the same or different nuclear localization signals are in the RNA, e.g., in the template RNA. In some embodiments the nuclear localization signal is less than 5, 10, 25, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900 or 1000 bp in length. Various RNA nuclear localization sequences can be used. For example, Lubelsky and Ulitsky, *Nature* 555 (107-111), 2018 describe RNA sequences which drive RNA localization into the nucleus. In some embodiments, the nuclear localization signal is a SINE-derived nuclear RNA localization (SIRLOIN) signal. In some embodiments the nuclear localization signal binds a nuclear-enriched protein. In some embodiments the nuclear localization signal binds the HNRNPK protein. In some embodiments the nuclear localization signal is rich in pyrimidines, e.g., is a C/T rich, C/U rich, C rich, T rich, or U rich region. In some embodiments the nuclear localization signal is derived from a long non-coding RNA. In some embodiments the nuclear localization signal is derived from MALAT1 long non-coding RNA or is the 600 nucleotide M region of MALAT1 (described in Miyagawa et al., *RNA* 18, (738-751), 2012). In some embodiments the nuclear localization signal is derived from BORG long non-coding RNA or is a AGCCC motif (described in Zhang et al., *Molecular and Cellular Biology* 34, 2318-2329 (2014). In some embodiments the nuclear localization sequence is described in Shukla et al., *The EMBO Journal* e98452 (2018). In some embodiments the nuclear localization signal is derived from a retrovirus.

In some embodiments, a polypeptide described herein comprises one or more (e.g., 2, 3, 4, 5) nuclear targeting sequences, for example a nuclear localization sequence (NLS). In some embodiments, the NLS is a bipartite NLS. In some embodiments, an NLS facilitates the import of a protein comprising an NLS into the cell nucleus. In some embodiments, the NLS is fused to the N-terminus of a gene modifying polypeptide as described herein. In some embodiments, the NLS is fused to the C-terminus of the gene modifying polypeptide. In some embodiments, the NLS is fused to the N-terminus or the C-terminus of a Cas domain. In some embodiments, a linker sequence is disposed between the NLS and the neighboring domain of the gene modifying polypeptide.

In some embodiments, an NLS comprises the amino acid sequence MDSLLMNRRKFLYQFKNVRWAKGRRETYLC (SEQ ID NO: 9), PKKRKVEGADKRTADGSEFESPKKKRKV(SEQ ID NO: 10), RKSGKIAAIWKRPRKPKKKRKV (SEQ ID NO: 11) KRTADGSEFESPKKKRKV(SEQ ID NO: 12), KKTELQTTNAENKTKKL (SEQ ID NO: 13), or KRGINDRNFWRGENGRKTR (SEQ ID NO: 14), KRPAATKKAGQAKKKK (SEQ ID NO: 15), or a functional fragment or variant thereof. Exemplary NLS sequences are also described in PCT/EP2000/011690, the contents of which are incorporated herein by reference for their disclosure of exemplary nuclear localization sequences. In some embodiments, an NLS comprises an amino acid sequence as disclosed in Table 11. An NLS of this table may be utilized with one or more copies in a

polypeptide in one or more locations in a polypeptide, e.g., 1, 2, 3 or more copies of an NLS in an N-terminal domain, between peptide domains, in a C-terminal domain, or in a combination of locations, in order to improve subcellular localization to the nucleus. Multiple unique sequences may be used within a single polypeptide. Sequences may be naturally monopartite or bipartite, e.g., having one or two stretches of basic amino acids, or may be used as chimeric bipartite sequences. Sequence references correspond to UniProt accession numbers, except where indicated as SeqNLS for sequences mined using a subcellular localization prediction algorithm (Lin et al BMC Bioinformat 13:157 (2012), incorporated herein by reference in its entirety).

Table 11 Exemplary nuclear localization signals for use in gene modifying systems

Sequence	Sequence References	SEQ ID No.
AHFKISGEKRPSTDPGKKAK NPKKKKKKDP	Q76IQ7	223
AHRAKKMSKTHA	P21827	224
ASPEYVNLPINGNG	SeqNLS	225
CTKRPRW	O88622, Q86W56, Q9QYM2, O02776	226
DKAKRVSRNKSEKKRR	O15516, Q5RAK8, Q91YB2, Q91YB0, Q8QGQ6, O08785, Q9WVS9, Q6YGZ4	227
EELRLKEELLKGIYA	Q9QY16, Q9UHL0, Q2TBP1, Q9QY15	228
EEQLRRRKNSRLNNTG	G5EFF5	229
EVLKVIRTGKRKKKAWKR MVTKVC	SeqNLS	230
HHHHHHHHHHHHQPH	Q63934, G3V7L5, Q12837	231
HKKKHPDASVNFSEFSK	P10103, Q4R844, P12682, B0CM99, A9RA84, Q6YKA4, P09429, P63159, Q08IE6, P63158, Q9YH06, B1MTB0	232
HKRTKK	Q2R2D5	233
IINGRKLKLKKSRRRSSQTS NNSFTSRRS	SeqNLS	234
KAEQERRK	Q8LH59	235
KEKRKRREELFIEQKKRK	SeqNLS	236
KKGKDEWFSRGKKP	P30999	237
KKGPSVQKRKKT	Q6ZN17	238

KKKTVINDLLHYKKEK	SeqNLS, P32354	239
KKNGGKGKNKPSAKIKK	SeqNLS	240
KKPKWDDFKKKKK	Q15397, Q8BKS9, Q562C7	241
KKRKKD	SeqNLS, Q91Z62, Q1A730, Q969P5, Q2KHT6, Q9CPU7	242
KKRRKRRRK	SeqNLS	243
KKRRRRARK	Q9UMS6, D4A702, Q91YE8	244
KKSKRGR	Q9UBS0	245
KKSRKRGS	B4FG96	246
KKSTALSRELGKIMRRR	SeqNLS, P32354	247
KKSYQDPEIIAHSRPRK	Q9U7C9	248
KKTGKNRKLKSKRVKTR	Q9Z301, O54943, Q8K3T2	249
KKVSIAGQSGKLWRWKR	Q6YUL8	250
KKYENVVIKRSPRKRGRPRK	SeqNLS	251
KNKKRK	SeqNLS	252
KPKKKR	SeqNLS	253
KRAMKDDSHGNSTSPKRRK	Q0E671	254
KRANSNLVAAYEKAKKK	P23508	255
KRASEDTTSGSPPKKSSAGPKR	Q9BZZ5, Q5R644	256
KRFKRRWMVRKMKTKK	SeqNLS	257
KRGLNSSFETSPKKVK	Q8IV63	258
KRGNSSIGPNDLSKRKQRKK	SeqNLS	259
KRIHSVSLSQSQIDPSKKVKRAK	SeqNLS	260
KRKGKLKNKGSKRKK	O15381	261
KRRRRRRREKRKR	Q96GM8	262
KRSNDRTYSPEEEKQRRA	Q91ZF2	263
KRTVATNGDASGAHRAKKMSK	SeqNLS	264

KRVYNKGEDEQEHLPKGKKR	SeqNLS	265
KSGKAPRRRAVSMDNSNK	Q9WVH4, O43524	266
KVNFLDMSLDDIIIYKELE	Q9P127	267
KVQHRIAKKTTRRRR	Q9DXE6	268
LSPSLSPL	Q9Y261, P32182, P35583	269
MDSLLMNRRKFLYQFKNVRWAKGRRETYLC	Q9GZX7	270
MPQNEYIELHRKRYGYRLDYHEKKRKKESREAHERSKKAKKMIGLKAKLYHK	SeqNLS	271
MVQLRPRASR	SeqNLS	272
NNKLLAKRRKGGASPKDDPMDDIK	Q965G5	273
NYKRPMDGTYGPPAKRHEGE	O14497, A2BH40	274
PDTKRAKLDSSETTMVKKK	SeqNLS	275
PEKRTKI	SeqNLS	276
PGGRGKKK	Q719N1, Q9UBP0, A2VDN5	277
PGKMDKGEHRQERRDRPY	Q01844, Q61545	278
PKKGDKYDKTD	Q45FA5	279
PKKKSRK	O35914, Q01954	280
PKKNKPE	Q22663	281
PKKRAKV	P04295, P89438	282
PKPKKLKVE	P55263, P55262, P55264, Q64640	283
PKRGRGR	Q9FYS5, Q43386	284
PKRRLVDDA	P0C797	285
PKRRRTY	SeqNLS	286
PLFKRR	A8X6H4, Q9TXJ0	287
PLRKAKR	Q86WB0, Q5R8V9	288
PPAKRKCIF	Q6AZ28, O75928, Q8C5D8	289
PPARRRRL	Q8NAG6	290

PPKKKRKV	Q3L6L5, P03070, P14999, P03071	291
PPNKRMKVKH	Q8BN78	292
PPRIYPQLPSAPT	P0C799	293
PQRSPFPKSSVKR	SeqNLS	294
PRPRKVPR	P0C799	295
PRRRVQRKR	SeqNLS, Q5R448, Q5TAQ9	296
PRRVRLK	Q58DJ0, P56477, Q13568	297
PSRKRPR	Q62315, Q5F363, Q92833	298
PSSKKRKV	SeqNLS	299
PTKKRVK	P07664	300
QRPGPYDRP	SeqNLS	301
RGKGGKGLGKGGAKRHRK	SeqNLS	302
RKAGKGGGGHKTTKKRSA KDEKVP	B4FG96	303
RKIKLKRAK	A1L3G9	304
RKIKRKRAK	B9X187	305
RKKEAPGPREELRSRGR	O35126, P54258, Q5IS70, P54259	306
RKKRKGK	SeqNLS, Q29243, Q62165, Q28685, O18738, Q9TSZ6, Q14118	307
RKKRRQRRR	P04326, P69697, P69698, P05907, P20879, P04613, P19553, P0C1J9, P20893, P12506, P04612, Q73370, P0C1K0, P05906, P35965, P04609, P04610, P04614, P04608, P05905	308
RKKSIPLSIKNLKRKHKRKK NKITR	Q9C0C9	309
RKLVKPKNTKMKTKLRTNP Y	Q14190	310
RKRLILSDKGQLDWKK	SeqNLS, Q91Z62, Q1A730, Q2KHT6, Q9CPU7	311
RKRLKSK	Q13309	312
RKRRVRDNM	Q8QPH4, Q809M7, A8C8X1, Q2VNC5, Q38SQ0, O89749, Q6DNQ9, Q809L9, Q0A429, Q20NV3, P16509, P16505, Q6DNQ5, P16506, Q6XT06, P26118,	313

	Q2ICQ2, Q2RCG8, Q0A2D0, Q0A2H9, Q9IQ46, Q809M3, Q6J847, Q6J856, B4URE4, A4GCM7, Q0A440, P26120, P16511,	
RKRSPKDKKEKDLDGAGKRRKT	Q7RTP6	314
RKRTPRVDGQTGENDMNKRRRK	O94851	315
RLPVRRRRRR	P04499, P12541, P03269, P48313, P03270	316
RLRFRKPKSK	P69469	317
RQQRKR	Q14980	318
RRDLNSSFETSPKKVK	Q8K3G5	319
RRDRAKLR	Q9SLB8	320
RRGDGRRR	Q80WE1, Q5R9B4, Q06787, P35922	321
RRGRKRKAEKQ	Q812D1, Q5XXA9, Q99JF8, Q8MJG1, Q66T72, O75475	322
RRKKRR	Q0VD86, Q58DS6, Q5R6G2, Q9ERI5, Q6AYK2, Q6NYC1	323
RRKRSKSEDMDSVESKRRR	Q7TT18	324
RRKRSR	Q99PU7, D3ZHS6, Q92560, A2VDM8	325
RRPKGKTLQKRKPK	Q6ZN17	326
RRRGFERFGPDNMGRKRK	Q63014, Q9DBR0	327
RRRGKNKVAAQNCRK	SeqNLS	328
RRRKRR	Q5FVH8, Q6MZT1, Q08DH5, Q8BQP9	329
RRRQKQKGGASRRR	SeqNLS	330
RRRREGPRARRRR	P08313, P10231	331
RRTIRLKLVYDKCDRSCKIQKKNRNKCQYCRFHKCLSVGMSHNAIRFGRMPRSEKAKLKAE	SeqNLS	332
RRVPQRKEVSRCRKCRK	Q5RJN4, Q32L09, Q8CAK3, Q9NUL5	333
RVGGRRQAVECIEDLLNEPGQPLDLSCKRPRP	P03255	334

RVVKLRIAP	P52639, Q8JMN0	335
RVVRRR	P70278	336
SKRKTKISRKTR	Q5RAY1, O00443	337
SYVKTVPNRTRTYIKL	P21935	338
TGKNEAKKRKIA	P52739, Q8K3J5, Q5RAU9	339
TLSPASSPSSVSCPVIPASTD ESPGSALNI	SeqNLS	340
VSKKQRTGKKIH	P52739, Q8K3J5, Q5RAU9	341
SPKKKRKVE		342
KRTAD GSEFE SPKKKRKVE		343
PAAKRVKLD		344
PKKKRKV		345
MDSLLMNRRKFLYQFKNVR WAKGRRETYLC		346
SPKKKRKVEAS		347
MAPKKKRKVGIHRGVP		348

In some embodiments, the NLS is a bipartite NLS. A bipartite NLS typically comprises two basic amino acid clusters separated by a spacer sequence (which may be, e.g., about 10 amino acids in length). A monopartite NLS typically lacks a spacer. An example of a bipartite NLS is the nucleoplasmin NLS, having the sequence KR[PAATKKAGQA]KKKK (SEQ ID NO: 15), wherein the spacer is bracketed. Another exemplary bipartite NLS has the sequence PKKKRKVEGADKRTADGSEFESPKKKRKV (SEQ ID NO: 16). Exemplary NLSs are described in International Application WO2020051561, which is herein incorporated by reference in its entirety, including for its disclosures regarding nuclear localization sequences.

In certain embodiments, a gene editor system polypeptide (e.g., a gene modifying polypeptide as described herein) further comprises an intracellular localization sequence, e.g., a nuclear localization sequence and/or a nucleolar localization sequence. The nuclear localization sequence and/or nucleolar localization sequence may be amino acid sequences that promote the import of the protein into the nucleus and/or nucleolus, where it can promote integration of heterologous sequence into the genome. In certain embodiments, a gene editor system polypeptide (e.g., (e.g., a gene modifying polypeptide as described herein) further comprises a nucleolar localization sequence. In certain embodiments, the gene modifying polypeptide is encoded on a first RNA, and the template RNA is a second, separate, RNA, and

the nucleolar localization signal is encoded on the RNA encoding the gene modifying polypeptide and not on the template RNA. In some embodiments, the nucleolar localization signal is located at the N-terminus, C-terminus, or in an internal region of the polypeptide. In some embodiments, a plurality of the same or different nucleolar localization signals are used. In some embodiments, the nuclear localization signal is less than 5, 10, 25, 50, 75, or 100 amino acids in length. Various polypeptide nucleolar localization signals can be used. For example, Yang et al., *Journal of Biomedical Science* 22, 33 (2015), describe a nuclear localization signal that also functions as a nucleolar localization signal. In some embodiments, the nucleolar localization signal may also be a nuclear localization signal. In some embodiments, the nucleolar localization signal may overlap with a nuclear localization signal. In some embodiments, the nucleolar localization signal may comprise a stretch of basic residues. In some embodiments, the nucleolar localization signal may be rich in arginine and lysine residues. In some embodiments, the nucleolar localization signal may be derived from a protein that is enriched in the nucleolus. In some embodiments, the nucleolar localization signal may be derived from a protein enriched at ribosomal RNA loci. In some embodiments, the nucleolar localization signal may be derived from a protein that binds rRNA. In some embodiments, the nucleolar localization signal may be derived from MSP58. In some embodiments, the nucleolar localization signal may be a monopartite motif. In some embodiments, the nucleolar localization signal may be a bipartite motif. In some embodiments, the nucleolar localization signal may consist of a multiple monopartite or bipartite motifs. In some embodiments, the nucleolar localization signal may consist of a mix of monopartite and bipartite motifs. In some embodiments, the nucleolar localization signal may be a dual bipartite motif. In some embodiments, the nucleolar localization motif may be a KRASSQALGTIPKRRSSSRFIKRKK (SEQ ID NO: 17). In some embodiments, the nucleolar localization signal may be derived from nuclear factor-κB-inducing kinase. In some embodiments, the nucleolar localization signal may be an RKKRKKK motif (SEQ ID NO: 18) (described in Birbach et al., Journal of Cell Science, 117 (3615-3624), 2004).

Evolved Variants of Gene Modifying Polypeptides and Systems

In some embodiments, the invention provides evolved variants of gene modifying polypeptides as described herein. Evolved variants can, in some embodiments, be produced by mutagenizing a reference gene modifying polypeptide, or one of the fragments or domains comprised therein. In some embodiments, one or more of the domains (e.g., the reverse transcriptase domain) is evolved. One or more of such evolved variant domains can, in some embodiments, be evolved alone or together with other domains. An evolved variant domain or domains may, in some embodiments, be combined with

unevolved cognate component(s) or evolved variants of the cognate component(s), e.g., which may have been evolved in either a parallel or serial manner.

In some embodiments, the process of mutagenizing a reference gene modifying polypeptide, or fragment or domain thereof, comprises mutagenizing the reference gene modifying polypeptide or fragment or domain thereof. In embodiments, the mutagenesis comprises a continuous evolution method (e.g., PACE) or non-continuous evolution method (e.g., PANCE), e.g., as described herein. In some embodiments, the evolved gene modifying polypeptide, or a fragment or domain thereof, comprises one or more amino acid variations introduced into its amino acid sequence relative to the amino acid sequence of the reference gene modifying polypeptide, or fragment or domain thereof. In embodiments, amino acid sequence variations may include one or more mutated residues (e.g., conservative substitutions, non-conservative substitutions, or a combination thereof) within the amino acid sequence of a reference gene modifying polypeptide, e.g., as a result of a change in the nucleotide sequence encoding the gene modifying polypeptide that results in, e.g., a change in the codon at any particular position in the coding sequence, the deletion of one or more amino acids (e.g., a truncated protein), the insertion of one or more amino acids, or any combination of the foregoing. The evolved variant gene modifying polypeptide may include variants in one or more components or domains of the gene modifying polypeptide (e.g., variants introduced into a reverse transcriptase domain).

In some aspects, the disclosure provides gene modifying polypeptides, systems, kits, and methods using or comprising an evolved variant of a gene modifying polypeptide, e.g., employs an evolved variant of a gene modifying polypeptide or a gene modifying polypeptide produced or producible by PACE or PANCE. In embodiments, the unevolved reference gene modifying polypeptide is a gene modifying polypeptide as disclosed herein.

The term “phage-assisted continuous evolution (PACE),”as used herein, generally refers to continuous evolution that employs phage as viral vectors. Examples of PACE technology have been described, for example, in International PCT Application No. PCT/US 2009/056194, filed September 8, 2009, published as WO 2010/028347 on March 11, 2010; International PCT Application, PCT/US2011/066747, filed December 22, 2011, published as WO 2012/088381 on June 28, 2012; U.S. Patent No. 9,023,594, issued May 5, 2015; U.S. Patent No. 9,771,574, issued September 26, 2017; U.S. Patent No. 9,394,537, issued July 19, 2016; International PCT Application, PCT/US2015/012022, filed January 20, 2015, published as WO 2015/134121 on September 11, 2015; U.S. Patent No. 10,179,911, issued January 15, 2019; and International PCT Application, PCT/US2016/027795, filed April 15, 2016, published as WO 2016/168631 on October 20, 2016, the entire contents of each of which are incorporated herein by reference.

The term "phage-assisted non-continuous evolution (PANCE)," as used herein, generally refers to non-continuous evolution that employs phage as viral vectors. Examples of PANCE technology have been described, for example, in Suzuki T. et al, Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase, Nat Chem Biol. 13(12): 1261-1266 (2017), incorporated herein by reference in its entirety. Briefly, PANCE is a technique for rapid *in vivo* directed evolution using serial flask transfers of evolving selection phage (SP), which contain a gene of interest to be evolved, across fresh host cells (e.g., E. coli cells). Genes inside the host cell may be held constant while genes contained in the SP continuously evolve. Following phage growth, an aliquot of infected cells may be used to transfect a subsequent flask containing host E. coli. This process can be repeated and/or continued until the desired phenotype is evolved, e.g., for as many transfers as desired.

Methods of applying PACE and PANCE to gene modifying polypeptides may be readily appreciated by the skilled artisan by reference to, *inter alia*, the foregoing references. Additional exemplary methods for directing continuous evolution of genome-modifying proteins or systems, e.g., in a population of host cells, e.g., using phage particles, can be applied to generate evolved variants of gene modifying polypeptides, or fragments or subdomains thereof. Non-limiting examples of such methods are described in International PCT Application, PCT/US2009/056194, filed September 8, 2009, published as WO 2010/028347 on March 11, 2010; International PCT Application, PCT/US2011/066747, filed December 22, 2011, published as WO 2012/088381 on June 28, 2012; U.S. Patent No. 9,023,594, issued May 5, 2015; U.S. Patent No. 9,771,574, issued September 26, 2017; U.S. Patent No. 9,394,537, issued July 19, 2016; International PCT Application, PCT/US2015/012022, filed January 20, 2015, published as WO 2015/134121 on September 11, 2015; U.S. Patent No. 10,179,911, issued January 15, 2019; International Application No. PCT/US2019/37216, filed June 14, 2019, International Patent Publication WO 2019/023680, published January 31, 2019, International PCT Application, PCT/US2016/027795, filed April 15, 2016, published as WO 2016/168631 on October 20, 2016, and International Patent Publication No. PCT/US2019/47996, filed August 23, 2019, each of which is incorporated herein by reference in its entirety.

In some non-limiting illustrative embodiments, a method of evolution of a evolved variant gene modifying polypeptide, of a fragment or domain thereof, comprises: (a) contacting a population of host cells with a population of viral vectors comprising the gene of interest (the starting gene modifying polypeptide or fragment or domain thereof), wherein: (1) the host cell is amenable to infection by the viral vector; (2) the host cell expresses viral genes required for the generation of viral particles; (3) the expression of at least one viral gene required for the production of an infectious viral particle is dependent on a function of the gene of interest; and/or (4) the viral vector allows for expression of the protein in the host cell, and can be replicated and packaged into a viral particle by the host cell. In some embodiments,

the method comprises (b) contacting the host cells with a mutagen, using host cells with mutations that elevate mutation rate (e.g., either by carrying a mutation plasmid or some genome modification—e.g., proofing-impaired DNA polymerase, SOS genes, such as UmuC, UmuD', and/or RecA, which mutations, if plasmid-bound, may be under control of an inducible promoter), or a combination thereof. In some embodiments, the method comprises (c) incubating the population of host cells under conditions allowing for viral replication and the production of viral particles, wherein host cells are removed from the host cell population, and fresh, uninfected host cells are introduced into the population of host cells, thus replenishing the population of host cells and creating a flow of host cells. In some embodiments, the cells are incubated under conditions allowing for the gene of interest to acquire a mutation. In some embodiments, the method further comprises (d) isolating a mutated version of the viral vector, encoding an evolved gene product (e.g., an evolved variant gene modifying polypeptide, or fragment or domain thereof), from the population of host cells.

The skilled artisan will appreciate a variety of features employable within the above-described framework. For example, in some embodiments, the viral vector or the phage is a filamentous phage, for example, an M13 phage, e.g., an M13 selection phage. In certain embodiments, the gene required for the production of infectious viral particles is the M13 gene III (gIII). In embodiments, the phage may lack a functional gIII, but otherwise comprise gI, gII, gIV, gV, gVI, gVII, gVIII, gIX, and a gX. In some embodiments, the generation of infectious VSV particles involves the envelope protein VSV-G. Various embodiments can use different retroviral vectors, for example, Murine Leukemia Virus vectors, or Lentiviral vectors. In embodiments, the retroviral vectors can efficiently be packaged with VSV-G envelope protein, e.g., as a substitute for the native envelope protein of the virus.

In some embodiments, host cells are incubated according to a suitable number of viral life cycles, e.g., at least 10, at least 20, at least 30, at least 40, at least 50, at least 100, at least 200, at least 300, at least 400, at least, 500, at least 600, at least 700, at least 800, at least 900, at least 1000, at least 1250, at least 1500, at least 1750, at least 2000, at least 2500, at least 3000, at least 4000, at least 5000, at least 7500, at least 10000, or more consecutive viral life cycles, which in on illustrative and non-limiting examples of M13 phage is 10-20 minutes per virus life cycle. Similarly, conditions can be modulated to adjust the time a host cell remains in a population of host cells, e.g., about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19, about 20, about 21, about 22, about 23, about 24, about 25, about 30, about 35, about 40, about 45, about 50, about 55, about 60, about 70, about 80, about 90, about 100, about 120, about 150, or about 180 minutes. Host cell populations can be controlled in part by density of the host cells, or, in some embodiments, the host cell density in an inflow, e.g., 10^3 cells/ml, about 10^4 cells/ml, about 10^5 cells/ml, about 5- 10^5 cells/ml, about 10^6 cells/ml, about 5-

10^6 cells/ml, about 10^7 cells/ml, about 5- 10^7 cells/ml, about 10^8 cells/ml, about 5- 10^8 cells/ml, about 10^9 cells/ml, about $5\cdot 10^9$ cells/ml, about 10^{10} cells/ml, or about $5\cdot 10^{10}$ cells/ml.

Inteins

In some embodiments, as described in more detail below, an intein-N (intN) domain may be fused to the N-terminal portion of a first domain of a gene modifying polypeptide described herein, and an intein-C (intC) domain may be fused to the C-terminal portion of a second domain of a gene modifying polypeptide described herein for the joining of the N-terminal portion to the C-terminal portion, thereby joining the first and second domains. In some embodiments, the first and second domains are each independently chosen from a DNA binding domain, an RNA binding domain, an RT domain, and an endonuclease domain.

Inteins can occur as self-splicing protein intron (e.g., peptide), e.g., which ligates flanking N-terminal and C-terminal exteins (e.g., fragments to be joined). An intein may, in some instances, comprise a fragment of a protein that is able to excise itself and join the remaining fragments (the exteins) with a peptide bond in a process known as protein splicing. Inteins are also referred to as "protein introns." The process of an intein excising itself and joining the remaining portions of the protein is herein termed "protein splicing" or "intein-mediated protein splicing."

In some embodiments, an intein of a precursor protein (an intein containing protein prior to intein-mediated protein splicing) comes from two genes. Such intein is referred to herein as a split intein (e.g., split intein-N and split intein-C). Accordingly, an intein-based approach may be used to join a first polypeptide sequence and a second polypeptide sequence together. For example, in cyanobacteria, DnaE, the catalytic subunit a of DNA polymerase III, is encoded by two separate genes, dnaE-n and dnaE-c. An intein-N domain, such as that encoded by the dnaE-n gene, when situated as part of a first polypeptide sequence, may join the first polypeptide sequence with a second polypeptide sequence, wherein the second polypeptide sequence comprises an intein-C domain, such as that encoded by the dnaE-c gene. Accordingly, in some embodiments, a protein can be made by providing nucleic acid encoding the first and second polypeptide sequences (e.g., wherein a first nucleic acid molecule encodes the first polypeptide sequence and a second nucleic acid molecule encodes the second polypeptide sequence), and the nucleic acid is introduced into the cell under conditions that allow for production of the first and second polypeptide sequences, and for joining of the first to the second polypeptide sequence via an intein-based mechanism.

Use of inteins for joining heterologous protein fragments is described, for example, in Wood et al., J. Biol. Chem.289(21); 14512-9 (2014) (incorporated herein by reference in its entirety). For example, when fused to separate protein fragments, the inteins IntN and IntC may recognize each other, splice themselves out, and/or simultaneously ligate the flanking N- and C-terminal exteins of the protein fragments to which they were fused, thereby reconstituting a full-length protein from the two protein fragments.

In some embodiments, a synthetic intein based on the dnaE intein, the Cfa-N (e.g., split intein-N) and Cfa-C (e.g., split intein-C) intein pair, is used. Examples of such inteins have been described, e.g., in Stevens et al., J Am Chem Soc. 2016 Feb. 24; 138(7):2162-5 (incorporated herein by reference in its entirety). Non-limiting examples of intein pairs that may be used in accordance with the present disclosure include: Cfa DnaE intein, Ssp GyrB intein, Ssp DnaX intein, Ter DnaE3 intein, Ter ThyX intein, Rma DnaB intein and Cne Prp8 intein (e.g., as described in U.S. Pat. No. 8,394,604, incorporated herein by reference.

In some embodiments involving a split Cas9, an intein-N domain and an intein-C domain may be fused to the N-terminal portion of the split Cas9 and the C-terminal portion of a split Cas9, respectively, for the joining of the N-terminal portion of the split Cas9 and the C-terminal portion of the split Cas9. For example, in some embodiments, an intein-N is fused to the C-terminus of the N-terminal portion of the split Cas9, i.e., to form a structure of N— [N-terminal portion of the split Cas9]-[intein-N]~ C. In some embodiments, an intein-C is fused to the N-terminus of the C-terminal portion of the split Cas9, i.e., to form a structure of N-[intein-C]~ [C-terminal portion of the split Cas9]-C. The mechanism of intein-mediated protein splicing for joining the proteins the inteins are fused to (e.g., split Cas9) is described in Shah et al., Chem Sci. 2014; 5(l):446-46l, incorporated herein by reference. Methods for designing and using inteins are known in the art and described, for example by WO2020051561, W02014004336, WO2017132580, US20150344549, and US20180127780, each of which is incorporated herein by reference in their entirety.

In some embodiments, a split refers to a division into two or more fragments. In some embodiments, a split Cas9 protein or split Cas9 comprises a Cas9 protein that is provided as an N-terminal fragment and a C-terminal fragment encoded by two separate nucleotide sequences. The polypeptides corresponding to the N-terminal portion and the C-terminal portion of the Cas9 protein may be spliced to form a reconstituted Cas9 protein. In embodiments, the Cas9 protein is divided into two fragments within a disordered region of the protein, e.g., as described in Nishimasu et al., Cell, Volume 156, Issue 5, pp. 935-949, 2014, or as described in Jiang et al. (2016) Science 351: 867-871 and PDB file: 5F9R (each of which is incorporated herein by reference in its entirety). A disordered region may be determined by one or more protein structure determination techniques known in the art, including,

without limitation, X-ray crystallography, NMR spectroscopy, electron microscopy (e.g., cryoEM), and/or *in silico* protein modeling. In some embodiments, the protein is divided into two fragments at any C, T, A, or S, e.g., within a region of SpCas9 between amino acids A292- G364, F445-K483, or E565-T637, or at corresponding positions in any other Cas9, Cas9 variant (e.g., nCas9, dCas9), or other napDNAbp. In some embodiments, protein is divided into two fragments at SpCas9 T310, T313, A456, S469, or C574. In some embodiments, the process of dividing the protein into two fragments is referred to as splitting the protein.

In some embodiments, a protein fragment ranges from about 2-1000 amino acids (e.g., between 2-10, 10-50, 50-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, or 900-1000 amino acids) in length. In some embodiments, a protein fragment ranges from about 5-500 amino acids (e.g., between 5-10, 10-50, 50-100, 100-200, 200-300, 300-400, or 400-500 amino acids) in length. In some embodiments, a protein fragment ranges from about 20-200 amino acids (e.g., between 20-30, 30-40, 40-50, 50-100, or 100-200 amino acids) in length.

In some embodiments, a portion or fragment of a gene modifying polypeptide is fused to an intein. The nuclease can be fused to the N-terminus or the C-terminus of the intein. In some embodiments, a portion or fragment of a fusion protein is fused to an intein and fused to an AAV capsid protein. The intein, nuclease and capsid protein can be fused together in any arrangement (e.g., nuclease-intein-capsid, intein-nuclease-capsid, capsid-intein-nuclease, etc.). In some embodiments, the N-terminus of an intein is fused to the C-terminus of a fusion protein and the C-terminus of the intein is fused to the N-terminus of an AAV capsid protein.

In some embodiments, an endonuclease domain (e.g., a nickase Cas9 domain) is fused to intein-N and a polypeptide comprising an RT domain is fused to an intein-C.

Exemplary nucleotide and amino acid sequences of intein-N domains and compatible intein-C domains are provided below:

DnaE Intein-N DNA:
TGCCTGTCATACGAAACCGAGATACTGACAGTAGAATATGGCCTTCTGCCAATCGGGAAGAT
TGTGGAGAAACGGATAGAATGCACAGTTTACTCTGTCGATAACAATGGTAACATTTATACTC
AGCCAGTTGCCCAGTGGCACGACCGGGGAGAGCAGGAAGTATTCGAATACTGTCTGGAGGA
TGGAAGTCTCATTAGGGCCACTAAGGACCACAAATTTATGACAGTCGATGGCCAGATGCTGC
CTATAGACGAAATCTTTGAGCGAGAGTTGGACCTCATGCGAGTTGACAACCTTCCTAAT
(SEQ ID NO: 29)

DnaE Intein-N Protein:
CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLI
RATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN (SEQ ID NO: 30)

DnaE Intein-C DNA:

ATGATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGATATTGGAGTCG
AAAGAGATCACAACTTTGCTCTGAAGAACGGATTCATAGCTTCTAAT (SEQ ID NO: 31)

DnaE Intein-C Protein:
MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN (SEQ ID NO: 32)

Cfa-N DNA:
TGCCTGTCTTATGATACCGAGATACTTACCGTTGAATATGGCTTCTTGCCTATTGGAAAGATT
GTCGAAGAGAGAATTGAATGCACAGTATATACTGTAGACAAGAATGGTTTCGTTTACACACA
GCCCATTGCTCAATGGCACAATCGCGGCGAACAAGAAGTATTTGAGTACTGTCTCGAGGATG
GAAGCATCATACGAGCAACTAAAGATCATAAATTCATGACCACTGACGGGCAGATGTTGCC
AATAGATGAGATATTCGAGCGGGGCTTGGATCTCAAACAAGTGGATGGATTG CCA (SEQ ID
NO: 33)

Cfa-N Protein:
CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLEDGSIIR
ATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP (SEQ ID NO: 34)

Cfa-C DNA:
ATGAAGAGGACTGCCGATGGATCAGAGTTTGAATCTCCCAAGAAGAAGAGGAAAGTAAAGA
TAATATCTCGAAAAAGTCTTGGTACCCAAAATGTCTATGATATTGGAGTGGAGAAAGATCAC
AACTTCCTTCTCAAGAACGGTCTCGTAGCCAGCAAC (SEQ ID NO: 35)
Cfa-C Protein:
MKRTADGSEFESPKKKRKVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN (SEQ ID NO: 36)

In some embodiments, an RBD of a gene modifying polypeptide as described herein is attached to an RT domain via an intein-based fusion, e.g., via an intein dimerization sequence as listed in Table 33 below (or an intein dimerization sequence comprising an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto). In some embodiments, an RBD of a gene modifying polypeptide as described herein is attached to a DBD (e.g., a Cas domain, e.g., a Cas9 domain, e.g., an nCas9 or dCas9 domain) via an intein-based fusion, e.g., via an intein dimerization sequence as listed in Table 33 below (or an intein dimerization sequence comprising an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto). In some embodiments, an RT domain of a gene modifying polypeptide as described herein is attached to a DBD (e.g., a Cas domain, e.g., a Cas9 domain, e.g., an nCas9 or dCas9 domain) via an intein-based fusion, e.g., via an intein dimerization sequence as listed in Table 33 below (or an intein dimerization sequence comprising an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto). In some embodiments, a DBD (e.g., a Cas domain, e.g., a Cas9 domain, e.g., an nCas9 or dCas9 domain) of a gene modifying polypeptide as described herein is attached to an RBD and to an RT domain via intein-based fusions. In embodiments, the DBD is attached to the RBD and the RT domain via different intein dimerization sequences, e.g., intein dimerization sequences as listed in Table 33 below (or sequences comprising an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%,

98%, or 99% identity thereto). In embodiments, the DBD is attached to the RBD and the RT domain via the same intein dimerization sequence, e.g., an intein dimerization sequence as listed in Table 33 below (or a sequence comprising an amino acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto). In some embodiments, the intein dimerization sequences of an RBD and a DBD to be bound to each other comprise a Chain A sequence and a Chain B sequence, respectively, or a Chain B sequence and a Chain A sequence, respectively, as listed in a single row of Table 33 below (or sequences having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto). In some embodiments, the intein dimerization sequences of an RBD and an RT domain to be bound to each other comprise a Chain A sequence and a Chain B sequence, respectively, or a Chain B sequence and a Chain A sequence, respectively, as listed in a single row of Table 33 below (or sequences having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto). In some embodiments, the intein dimerization sequences of an RT domain and a DBD to be bound to each other comprise a Chain A sequence and a Chain B sequence, respectively, or a Chain B sequence and a Chain A sequence, respectively, as listed in a single row of Table 33 below (or sequences having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto).

Table 33. Exemplary intein dimerization sequences

System	Chain A name	Chain A sequence	Exemplary Chain A source	Chain B name	Chain B sequence	Exemplary Chain B source
Sce-VMA	Sce-VMA-5'-v1	CFAKGTNVLMADGSIECIENIEVGN KVMGKDGRPREVIKLPRGRETMYS VVQKSQHRAHKSDSSREVPELLKF TCNATHELVVRTPRSVRRLSRTIKG VEYFEVITFEMGQKKAPDGRIVELV KEVSKSYPISEGPERANELVESYR KASNKAYFEWTIEARDLSLLGSHV RKATYQTYAPILYENDHFFDYMQK SKFHLTIEGPKVLAYLLGLWIGDGL SDRATFSVDSRDTSLMERVTEYAE KLNLCAEYKDRKEPQVAKTVNLYS KVVRG	snapgene	Sce-VMA-3'-v1	GHGGIRNNLNTENPLWD AIVGLGFLKDGVKNIPSF LSTDNIGTRETFLAGLIDS DGYVTDEHGIKATIKTIHT SVRDGLVSLARSLGLVV SVNAEPAKVDMNVTKHK ISYAIYMSGGDVLLNVLS KCAGSKKFRPAPAAAFA RECRGFYFELQELKEDD YYGITLSDDSDHQFLLGS QVVVQN	snapgene common features
Sce-VMA	Sce-VMA-5'-v1	CFAKGTNVLMADGSIECIENIEVGN KVMGKDGRPREVIKLPRGRETMYS VVQKSQHRAHKSDSSREVPELLKF TCNATHELVVRTPRSVRRLSRTIKG VEYFEVITFEMGQKKAPDGRIVELV KEVSKSYPISEGPERANELVESYR KASNKAYFEWTIEARDLSLLGSHV RKATYQTYAPILYENDHFFDYMQK SKFHLTIEGPKVLAYLLGLWIGDGL SDRATFSVDSRDTSLMERVTEYAE KLNLCAEYKDRKEPQVAKTVNLYS KVVRG	snapgene	Sce-VMA-3'-v2	GHGGIRNNLNTENPLWD AIVGLGFLKDGVKNIPSF LSTDNIGTRETFLAGLIDS DGYVTDEHGIKATIKTIHT SVRDGLVSLARSLGLVV SVNAEPAKVDMNVTKHK ISYAIYMSGGDVLLNVLS KCAGSKKFRPAPAAAFA RECRGFYFELQELKEDD YYGITLSDDSDHQFLLGS QVVVQNCGERGNGSG	snapgene common features

Sce-VMA	Sce-VMA-5'-v1	CFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPILYENDHFFDYMQKSKFHLTIEGPKVLAYLLGLWIGDGLSDRATFSVDSRDTSLMERVTEYAEKLNLCAEYKDRKEPQVAKTVNLYSKVVRG	snapgene	Sce-VMA-3'-v3	GHGGIRNNLNTENPLWDAIVGLGFLKDGVKNIPSFLSTDNIGTRETFLAGLIDSDGYVTDEHGIKATIKTIHTSVRDGLVSLARSLGLVVSVNAEPAKVDMNVTKHKISYAIYMSGGDVLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLGSQVVVQNCTMTEKGSG	snapgene common features
Sce-VMA	Sce-VMA-5'-v1	CFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPILYENDHFFDYMQKSKFHLTIEGPKVLAYLLGLWIGDGLSDRATFSVDSRDTSLMERVTEYAEKLNLCAEYKDRKEPQVAKTVNLYSKVVRG	snapgene	Sce-VMA-3'-v4	GHGGIRNNLNTENPLWDAIVGLGFLKDGVKNIPSFLSTDNIGTRETFLAGLIDSDGYVTDEHGIKATIKTIHTSVRDGLVSLARSLGLVVSVNAEPAKVDMNVTKHKISYAIYMSGGDVLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLGSQVVVQNCGEKSMGSG	snapgene common features
Sce-VMA	Sce-VMA-5'-v1	CFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPILYENDHFFDYMQKSKFHLTIEGPKVLAYLLGLWIGDGL	snapgene	Sce-VMA-3'-v5	VLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLANQVVVHN	https://www.nature.com/articles/nmeth.3585

		SDRATFSVDSRDTSLMERVTEYAE KLNLCAEYKDRKEPQVAKTVNLYS KVVRG				
Sce-VMA	Sce-VMA-5'-v1	CFAKGTNVLMADGSIECIENIEVGN KVMGKDGRPREVIKLPRGRETMYS VVQKSQHRAHKSDSSREVPELLKF TCNATHELVVRTPRSVRRLSRTIKG VEYFEVITFEMGQKKAPDGRIVELV KEVSKSYPISEGPERANELVESYR KASNKAYFEWTIEARDLSLLGSHV RKATYQTYAPILYENDHFFDYMQK SKFHLTIEGPKVLAYLLGLWIGDGL SDRATFSVDSRDTSLMERVTEYAE KLNLCAEYKDRKEPQVAKTVNLYS KVVRG	snapgene	Sce-VMA-3'-v6	VLLNVLSKCAGSKKFRP APAAAFARECRGFYFEL QELKEDDYYGITLSDDS DHQFLLANQVVVHNCGE RGNGSG	https://www.nature.com/articles/nmeth.3585
Sce-VMA	Sce-VMA-5'-v1	CFAKGTNVLMADGSIECIENIEVGN KVMGKDGRPREVIKLPRGRETMYS VVQKSQHRAHKSDSSREVPELLKF TCNATHELVVRTPRSVRRLSRTIKG VEYFEVITFEMGQKKAPDGRIVELV KEVSKSYPISEGPERANELVESYR KASNKAYFEWTIEARDLSLLGSHV RKATYQTYAPILYENDHFFDYMQK SKFHLTIEGPKVLAYLLGLWIGDGL SDRATFSVDSRDTSLMERVTEYAE KLNLCAEYKDRKEPQVAKTVNLYS KVVRG	snapgene	Sce-VMA-3'-v7	VLLNVLSKCAGSKKFRP APAAAFARECRGFYFEL QELKEDDYYGITLSDDS DHQFLLANQVVVHNCTM TEKGSG	https://www.nature.com/articles/nmeth.3585
Sce-VMA	Sce-VMA-5'-v1	CFAKGTNVLMADGSIECIENIEVGN KVMGKDGRPREVIKLPRGRETMYS VVQKSQHRAHKSDSSREVPELLKF TCNATHELVVRTPRSVRRLSRTIKG VEYFEVITFEMGQKKAPDGRIVELV KEVSKSYPISEGPERANELVESYR	snapgene	Sce-VMA-3'-v8	VLLNVLSKCAGSKKFRP APAAAFARECRGFYFEL QELKEDDYYGITLSDDS DHQFLLANQVVVHNCGE KSMGSG	https://www.nature.com/articles/nmeth.3585

		KASNKAYFEWTIEARDLSLLGSHV RKATYQTYAPILYENDHFFDYMQK SKFHLTIEGPKVLAYLLGLWIGDGL SDRATFSVDSRDTSLMERVTEYAE KLNLCAEYKDRKEPQVAKTVNLYS KVVRG				
Sce-VMA	Sce-VMA-5'-v2	GGIIYVGCFAKGTNVLMADGSIECI ENIEVGNKVMGKDGRPREVIKLPR GRETMYSVVQKSQHRAHKSDSSR EVPELLKFTCNATHELVVRTPRSV RRLSRTIKGVEYFEVITFEMGQKKA PDGRIVELVKEVSKSYPISEGPERA NELVESYRKASNKAYFEWTIEARD LSLLGSHVRKATYQTYAPILYENDH FFDYMQKSKFHLTIEGPKVLAYLLG LWIGDGLSDRATFSVDSRDTSLME RVTEYAEKLNLCAEYKDRKEPQVA KTVNLYSKVVRG snapgene common features	snapgene	Sce-VMA-3'-v1	GHGGIRNNLNTENPLWD AIVGLGFLKDGVKNIPSF LSTDNIGTRETFLAGLIDS DGYVTDEHGIKATIKTIHT SVRDGLVSLARSLGLVV SVNAEPAKVDMNVTKHK ISYAIYMSGGDVLLNVLS KCAGSKKFRPAPAAAFA RECRGFYFELQELKEDD YYGITLSDDSDHQFLLGS QVVVQN	snapgene common features
Sce-VMA	Sce-VMA-5'-v2	GGIIYVGCFAKGTNVLMADGSIECI ENIEVGNKVMGKDGRPREVIKLPR GRETMYSVVQKSQHRAHKSDSSR EVPELLKFTCNATHELVVRTPRSV RRLSRTIKGVEYFEVITFEMGQKKA PDGRIVELVKEVSKSYPISEGPERA NELVESYRKASNKAYFEWTIEARD LSLLGSHVRKATYQTYAPILYENDH FFDYMQKSKFHLTIEGPKVLAYLLG LWIGDGLSDRATFSVDSRDTSLME RVTEYAEKLNLCAEYKDRKEPQVA KTVNLYSKVVRG snapgene common features	snapgene	Sce-VMA-3'-v2	GHGGIRNNLNTENPLWD AIVGLGFLKDGVKNIPSF LSTDNIGTRETFLAGLIDS DGYVTDEHGIKATIKTIHT SVRDGLVSLARSLGLVV SVNAEPAKVDMNVTKHK ISYAIYMSGGDVLLNVLS KCAGSKKFRPAPAAAFA RECRGFYFELQELKEDD YYGITLSDDSDHQFLLGS QVVVQNCGERGNGSG	snapgene common features

Sce-VMA	Sce-VMA-5'-v2	GGIIYVGCFAKGTNVLMADGSIECI ENIEVGNKVMGKDGRPREVIKLPR GRETMYSVVQKSQHRAHKSDSSR EVPELLKFTCNATHELVVRTPRSV RRLSRTIKGVEYFEVITFEMGQKKA PDGRIVELVKEVSKSYPISEGPERA NELVESYRKASNKAYFEWTIEARD LSLLGSHVRKATYQTYAPILYENDH FFDYMQKSKFHLTIEGPKVLAYLLG LWIGDGLSDRATFSVDSRDTSLME RVTEYAEKLNLCAEYKDRKEPQVA KTVNLYSKVVRG snapgene common features	snapgene	Sce-VMA-3'-v3	GHGGIRNNLNTENPLWD AIVGLGFLKDGVKNIPSF LSTDNIGTRETFLAGLIDS DGYVTDEHGIKATIKTIHT SVRDGLVSLARSLGLVV SVNAEPAKVDMNVTKHK ISYAIYMSGGDVLLNVLS KCAGSKKFRPAPAAAFA RECRGFYFELQELKEDD YYGITLSDDSDHQFLLGS QVVVQNCTMTEKGSG	snapgene common features
Sce-VMA	Sce-VMA-5'-v2	GGIIYVGCFAKGTNVLMADGSIECI ENIEVGNKVMGKDGRPREVIKLPR GRETMYSVVQKSQHRAHKSDSSR EVPELLKFTCNATHELVVRTPRSV RRLSRTIKGVEYFEVITFEMGQKKA PDGRIVELVKEVSKSYPISEGPERA NELVESYRKASNKAYFEWTIEARD LSLLGSHVRKATYQTYAPILYENDH FFDYMQKSKFHLTIEGPKVLAYLLG LWIGDGLSDRATFSVDSRDTSLME RVTEYAEKLNLCAEYKDRKEPQVA KTVNLYSKVVRG snapgene common features	snapgene	Sce-VMA-3'-v4	GHGGIRNNLNTENPLWD AIVGLGFLKDGVKNIPSF LSTDNIGTRETFLAGLIDS DGYVTDEHGIKATIKTIHT SVRDGLVSLARSLGLVV SVNAEPAKVDMNVTKHK ISYAIYMSGGDVLLNVLS KCAGSKKFRPAPAAAFA RECRGFYFELQELKEDD YYGITLSDDSDHQFLLGS QVVVQNCGEKSMGSG	snapgene common features
Sce-VMA	Sce-VMA-5'-v2	GGIIYVGCFAKGTNVLMADGSIECI ENIEVGNKVMGKDGRPREVIKLPR GRETMYSVVQKSQHRAHKSDSSR EVPELLKFTCNATHELVVRTPRSV RRLSRTIKGVEYFEVITFEMGQKKA PDGRIVELVKEVSKSYPISEGPERA NELVESYRKASNKAYFEWTIEARD	snapgene	Sce-VMA-3'-v5	VLLNVLSKCAGSKKFRP APAAAFARECRGFYFEL QELKEDDYYGITLSDDS DHQFLLANQVVVHN	https://www.nature.com/articles/nmeth.3585

		LSLLGSHVRKATYQTYAPILYENDH FFDYMQKSKFHLTIEGPKVLAYLLG LWIGDGLSDRATFSVDSRDTSLME RVTEYAEKLNLCAEYKDRKEPQVA KTVNLYSKVVRG snapgene common features				
Sce-VMA	Sce-VMA-5'-v2	GGIIYVGCFAKGTNVLMADGSIECI ENIEVGNKVMGKDGRPREVIKLPR GRETMYSVVQKSQHRAHKSDSSR EVPELLKFTCNATHELVVRTPRSV RRLSRTIKGVEYFEVITFEMGQKKA PDGRIVELVKEVSKSYPISEGPERA NELVESYRKASNKAYFEWTIEARD LSLLGSHVRKATYQTYAPILYENDH FFDYMQKSKFHLTIEGPKVLAYLLG LWIGDGLSDRATFSVDSRDTSLME RVTEYAEKLNLCAEYKDRKEPQVA KTVNLYSKVVRG snapgene common features	snapgene	Sce-VMA-3'-v6	VLLNVLSKCAGSKKFRP APAAAFARECRGFYFEL QELKEDDYYGITLSDDS DHQFLLANQVVVHNCGE RGNGSG	https://www.nature.com/articles/nmeth.3585
Sce-VMA	Sce-VMA-5'-v2	GGIIYVGCFAKGTNVLMADGSIECI ENIEVGNKVMGKDGRPREVIKLPR GRETMYSVVQKSQHRAHKSDSSR EVPELLKFTCNATHELVVRTPRSV RRLSRTIKGVEYFEVITFEMGQKKA PDGRIVELVKEVSKSYPISEGPERA NELVESYRKASNKAYFEWTIEARD LSLLGSHVRKATYQTYAPILYENDH FFDYMQKSKFHLTIEGPKVLAYLLG LWIGDGLSDRATFSVDSRDTSLME RVTEYAEKLNLCAEYKDRKEPQVA KTVNLYSKVVRG snapgene common features	snapgene	Sce-VMA-3'-v7	VLLNVLSKCAGSKKFRP APAAAFARECRGFYFEL QELKEDDYYGITLSDDS DHQFLLANQVVVHNCTM TEKGSG	https://www.nature.com/articles/nmeth.3585

Sce-VMA	Sce-VMA-5'-v2	GGIIYVGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPILYENDHFFDYMQKSKFHLTIEGPKVLAYLLGLWIGDGLSDRATFSVDSRDTSLMERVTEYAEKLNLCAEYKDRKEPQVAKTVNLYSKVVRG snapgene common features	snapgene	Sce-VMA-3'-v8	VLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLANQVVVHNCGEKSMGSG	https://www.nature.com/articles/nmeth.3585
Sce-VMA	Sce-VMA-5'-v3	GGVVLEKGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPILYENDHFFDYMQKSKFHLTIEGPKVLAYLLGLWIGDGLSDRATFSVDSRDTSLMERVTEYAEKLNLCAEYKDRKEPQVAKTVNLYSKVVRG	snapgene	Sce-VMA-3'-v1	GHGGIRNNLNTENPLWDAIVGLGFLKDGVKNIPSFLSTDNIGTRETFLAGLIDSDGYVTDEHGIKATIKTIHTSVRDGLVSLARSLGLVVSVNAEPAKVDMNVTKHKISYAIYMSGGDVLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLGSQVVVQN	snapgene common features
Sce-VMA	Sce-VMA-5'-v3	GGVVLEKGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPILYEND	snapgene	Sce-VMA-3'-v2	GHGGIRNNLNTENPLWDAIVGLGFLKDGVKNIPSFLSTDNIGTRETFLAGLIDSDGYVTDEHGIKATIKTIHTSVRDGLVSLARSLGLVVSVNAEPAKVDMNVTKHKISYAIYMSGGDVLLNVLSKCAGSKKFRPAPAAAFA	snapgene common features

		HFFDYMQKSKFHLTIEGPKVLAYLL GLWIGDGLSDRATFSVDSRDTSLM ERVTEYAEKLNLCAEYKDRKEPQV AKTVNLYSKVVRG			RECRGFYFELQELKEDD YYGITLSDDSDHQFLLGS QVVVQNCGERGNGSG	
Sce-VMA	Sce-VMA-5'-v3	GGVVLEKGCFAKGTNVLMADGSIE CIENIEVGNKVMGKDGRPREVIKLP RGRETMYSVVQKSQHRAHKSDSS REVPELLKFTCNATHELVVRTPRS VRRLSRTIKGVEYFEVITFEMGQKK APDGRIVELVKEVSKSYPISEGPER ANELVESYRKASNKAYFEWTIEAR DLSLLGSHVRKATYQTYAPILYEND HFFDYMQKSKFHLTIEGPKVLAYLL GLWIGDGLSDRATFSVDSRDTSLM ERVTEYAEKLNLCAEYKDRKEPQV AKTVNLYSKVVRG	snapgene	Sce-VMA-3'-v3	GHGGIRNNLNTENPLWD AIVGLGFLKDGVKNIPSF LSTDNIGTRETFLAGLIDS DGYVTDEHGIKATIKTIHT SVRDGLVSLARSLGLVV SVNAEPAKVDMNVTKHK ISYAIYMSGGDVLLNVLS KCAGSKKFRPAPAAAFA RECRGFYFELQELKEDD YYGITLSDDSDHQFLLGS QVVVQNCTMTEKGSG	snapgene common features
Sce-VMA	Sce-VMA-5'-v3	GGVVLEKGCFAKGTNVLMADGSIE CIENIEVGNKVMGKDGRPREVIKLP RGRETMYSVVQKSQHRAHKSDSS REVPELLKFTCNATHELVVRTPRS VRRLSRTIKGVEYFEVITFEMGQKK APDGRIVELVKEVSKSYPISEGPER ANELVESYRKASNKAYFEWTIEAR DLSLLGSHVRKATYQTYAPILYEND HFFDYMQKSKFHLTIEGPKVLAYLL GLWIGDGLSDRATFSVDSRDTSLM ERVTEYAEKLNLCAEYKDRKEPQV AKTVNLYSKVVRG	snapgene	Sce-VMA-3'-v4	GHGGIRNNLNTENPLWD AIVGLGFLKDGVKNIPSF LSTDNIGTRETFLAGLIDS DGYVTDEHGIKATIKTIHT SVRDGLVSLARSLGLVV SVNAEPAKVDMNVTKHK ISYAIYMSGGDVLLNVLS KCAGSKKFRPAPAAAFA RECRGFYFELQELKEDD YYGITLSDDSDHQFLLGS QVVVQNCGEKSMGSG	snapgene common features
Sce-VMA	Sce-VMA-5'-v3	GGVVLEKGCFAKGTNVLMADGSIE CIENIEVGNKVMGKDGRPREVIKLP RGRETMYSVVQKSQHRAHKSDSS REVPELLKFTCNATHELVVRTPRS VRRLSRTIKGVEYFEVITFEMGQKK	snapgene	Sce-VMA-3'-v5	VLLNVLSKCAGSKKFRP APAAAFARECRGFYFEL QELKEDDYYGITLSDDS DHQFLLANQVVVHN	https://www.nature.com/articles/nmeth.3585

		APDGRIVELVKEVSKSYPISEGPER ANELVESYRKASNKAYFEWTIEAR DLSLLGSHVRKATYQTYAPILYEND HFFDYMQKSKFHLTIEGPKVLAYLL GLWIGDGLSDRATFSVDSRDTSLM ERVTEYAEKLNLCAEYKDRKEPQV AKTVNLYSKVVRG				
Sce-VMA	Sce-VMA-5'-v3	GGVVLEKGCFAKGTNVLMADGSIE CIENIEVGNKVMGKDGRPREVIKLP RGRETMYSVVQKSQHRAHKSDSS REVPELLKFTCNATHELVVRTPRS VRRLSRTIKGVEYFEVITFEMGQKK APDGRIVELVKEVSKSYPISEGPER ANELVESYRKASNKAYFEWTIEAR DLSLLGSHVRKATYQTYAPILYEND HFFDYMQKSKFHLTIEGPKVLAYLL GLWIGDGLSDRATFSVDSRDTSLM ERVTEYAEKLNLCAEYKDRKEPQV AKTVNLYSKVVRG	snapgene	Sce-VMA-3'-v6	VLLNVLSKCAGSKKFRP APAAAFARECRGFYFEL QELKEDDYYGITLSDDS DHQFLLANQVVVHNCGE RGNGSG	https://www.nature.com/articles/nmeth.3585
Sce-VMA	Sce-VMA-5'-v3	GGVVLEKGCFAKGTNVLMADGSIE CIENIEVGNKVMGKDGRPREVIKLP RGRETMYSVVQKSQHRAHKSDSS REVPELLKFTCNATHELVVRTPRS VRRLSRTIKGVEYFEVITFEMGQKK APDGRIVELVKEVSKSYPISEGPER ANELVESYRKASNKAYFEWTIEAR DLSLLGSHVRKATYQTYAPILYEND HFFDYMQKSKFHLTIEGPKVLAYLL GLWIGDGLSDRATFSVDSRDTSLM ERVTEYAEKLNLCAEYKDRKEPQV AKTVNLYSKVVRG	snapgene	Sce-VMA-3'-v7	VLLNVLSKCAGSKKFRP APAAAFARECRGFYFEL QELKEDDYYGITLSDDS DHQFLLANQVVVHNCTM TEKGSG	https://www.nature.com/articles/nmeth.3585

Sce-VMA	Sce-VMA-5'-v3	GGVVLEKGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPILYENDHFFDYMQKSKFHLTIEGPKVLAYLLGLWIGDGLSDRATFSVDSRDTSLMERVTEYAEKLNLCAEYKDRKEPQVAKTVNLYSKVVRG	snapgene	Sce-VMA-3'-v8	VLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLANQVVVHNCGEKSMGSG	https://www.nature.com/articles/nmeth.3585
Sce-VMA	Sce-VMA-5'-v4	GGFQTVGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPILYENDHFFDYMQKSKFHLTIEGPKVLAYLLGLWIGDGLSDRATFSVDSRDTSLMERVTEYAEKLNLCAEYKDRKEPQVAKTVNLYSKVVRG	snapgene	Sce-VMA-3'-v1	GHGGIRNNLNTENPLWDAIVGLGFLKDGVKNIPSFLSTDNIGTRETFLAGLIDSDGYVTDEHGIKATIKTIHTSVRDGLVSLARSLGLVVSVNAEPAKVDMNVTKHKISYAIYMSGGDVLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLGSQVVVQN	snapgene common features
Sce-VMA	Sce-VMA-5'-v4	GGFQTVGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPILYENDHFFDYMQKSKFHLTIEGPKVLAYLLG	snapgene	Sce-VMA-3'-v2	GHGGIRNNLNTENPLWDAIVGLGFLKDGVKNIPSFLSTDNIGTRETFLAGLIDSDGYVTDEHGIKATIKTIHTSVRDGLVSLARSLGLVVSVNAEPAKVDMNVTKHKISYAIYMSGGDVLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDD	snapgene common features

		LWIGDGLSDRATFSVDSRDTSLME RVTEYAEKLNLCAEYKDRKEPQVA KTVNLYSKVVRG			YYGITLSDDSDHQFLLGS QVVVQNCGERGNGSG	
Sce-VMA	Sce-VMA-5'-v4	GGFQTVGCFAKGTNVLMADGSIEC IENIEVGNKVMGKDGRPREVIKLPR GRETMYSVVQKSQHRAHKSDSSR EVPELLKFTCNATHELVVRTPRSV RRLSRTIKGVEYFEVITFEMGQKKA PDGRIVELVKEVSKSYPISEGPERA NELVESYRKASNKAYFEWTIEARD LSLLGSHVRKATYQTYAPILYENDH FFDYMQKSKFHLTIEGPKVLAYLLG LWIGDGLSDRATFSVDSRDTSLME RVTEYAEKLNLCAEYKDRKEPQVA KTVNLYSKVVRG	snapgene	Sce-VMA-3'-v3	GHGGIRNNLNTENPLWD AIVGLGFLKDGVKNIPSF LSTDNIGTRETFLAGLIDS DGYVTDEHGIKATIKTIHT SVRDGLVSLARSLGLVV SVNAEPAKVDMNVTKHK ISYAIYMSGGDVLLNVLS KCAGSKKFRPAPAAAFA RECRGFYFELQELKEDD YYGITLSDDSDHQFLLGS QVVVQNCTMTEKGSG	snapgene common features
Sce-VMA	Sce-VMA-5'-v4	GGFQTVGCFAKGTNVLMADGSIEC IENIEVGNKVMGKDGRPREVIKLPR GRETMYSVVQKSQHRAHKSDSSR EVPELLKFTCNATHELVVRTPRSV RRLSRTIKGVEYFEVITFEMGQKKA PDGRIVELVKEVSKSYPISEGPERA NELVESYRKASNKAYFEWTIEARD LSLLGSHVRKATYQTYAPILYENDH FFDYMQKSKFHLTIEGPKVLAYLLG LWIGDGLSDRATFSVDSRDTSLME RVTEYAEKLNLCAEYKDRKEPQVA KTVNLYSKVVRG	snapgene	Sce-VMA-3'-v4	GHGGIRNNLNTENPLWD AIVGLGFLKDGVKNIPSF LSTDNIGTRETFLAGLIDS DGYVTDEHGIKATIKTIHT SVRDGLVSLARSLGLVV SVNAEPAKVDMNVTKHK ISYAIYMSGGDVLLNVLS KCAGSKKFRPAPAAAFA RECRGFYFELQELKEDD YYGITLSDDSDHQFLLGS QVVVQNCGEKSMGSG	snapgene common features
Sce-VMA	Sce-VMA-5'-v4	GGFQTVGCFAKGTNVLMADGSIEC IENIEVGNKVMGKDGRPREVIKLPR GRETMYSVVQKSQHRAHKSDSSR EVPELLKFTCNATHELVVRTPRSV RRLSRTIKGVEYFEVITFEMGQKKA PDGRIVELVKEVSKSYPISEGPERA	snapgene	Sce-VMA-3'-v5	VLLNVLSKCAGSKKFRP APAAAFARECRGFYFEL QELKEDDYYGITLSDDS DHQFLLANQVVVHN	https://www.nature.com/articles/nmeth.3585

		NELVESYRKASNKAYFEWTIEARD LSLLGSHVRKATYQTYAPILYENDH FFDYMQKSKFHLTIEGPKVLAYLLG LWIGDGLSDRATFSVDSRDTSLME RVTEYAEKLNLCAEYKDRKEPQVA KTVNLYSKVVRG				
Sce-VMA	Sce-VMA-5'-v4	GGFQTVGCFAKGTNVLMADGSIEC IENIEVGNKVMGKDGRPREVIKLPR GRETMYSVVQKSQHRAHKSDSSR EVPELLKFTCNATHELVVRTPRSV RRLSRTIKGVEYFEVITFEMGQKKA PDGRIVELVKEVSKSYPISEGPERA NELVESYRKASNKAYFEWTIEARD LSLLGSHVRKATYQTYAPILYENDH FFDYMQKSKFHLTIEGPKVLAYLLG LWIGDGLSDRATFSVDSRDTSLME RVTEYAEKLNLCAEYKDRKEPQVA KTVNLYSKVVRG	snapgene	Sce-VMA-3'-v6	VLLNVLSKCAGSKKFRP APAAAFARECRGFYFEL QELKEDDYYGITLSDDS DHQFLLANQVVVHNCGE RGNGSG	https://www.nature.com/articles/nmeth.3585
Sce-VMA	Sce-VMA-5'-v4	GGFQTVGCFAKGTNVLMADGSIEC IENIEVGNKVMGKDGRPREVIKLPR GRETMYSVVQKSQHRAHKSDSSR EVPELLKFTCNATHELVVRTPRSV RRLSRTIKGVEYFEVITFEMGQKKA PDGRIVELVKEVSKSYPISEGPERA NELVESYRKASNKAYFEWTIEARD LSLLGSHVRKATYQTYAPILYENDH FFDYMQKSKFHLTIEGPKVLAYLLG LWIGDGLSDRATFSVDSRDTSLME RVTEYAEKLNLCAEYKDRKEPQVA KTVNLYSKVVRG	snapgene	Sce-VMA-3'-v7	VLLNVLSKCAGSKKFRP APAAAFARECRGFYFEL QELKEDDYYGITLSDDS DHQFLLANQVVVHNCTM TEKGSG	https://www.nature.com/articles/nmeth.3585
Sce-VMA	Sce-VMA-5'-v4	GGFQTVGCFAKGTNVLMADGSIEC IENIEVGNKVMGKDGRPREVIKLPR GRETMYSVVQKSQHRAHKSDSSR	snapgene	Sce-VMA-3'-v8	VLLNVLSKCAGSKKFRP APAAAFARECRGFYFEL QELKEDDYYGITLSDDS	https://www.nature.com/articl

		EVPELLKFTCNATHELVVRTPRSV RRLSRTIKGVEYFEVITFEMGQKKA PDGRIVELVKEVSKSYPISEGPERA NELVESYRKASNKAYFEWTIEARD LSLLGSHVRKATYQTYAPILYENDH FFDYMQKSKFHLTIEGPKVLAYLLG LWIGDGLSDRATFSVDSRDTSLME RVTEYAEKLNLCAEYKDRKEPQVA KTVNLYSKVVRG			DHQFLLANQVVVHNCGE KSMGSG	es/nmeth. 3585
Sce-VMA	Sce-VMA-5'-v5	CFAKGTNVLMADGSIECIENIEVGN KVMGKDGRPREVIKLPRGRETMYS VVQKSQHRAHKSDSSREVPELLKF TCNATHELVVRTPRSVRRLSRTIKG VEYFEVITFEMGQKKAPDGRIVELV KEVSKSYPISEGPERANELVESYR KASNKAYFEWTIEARDLSLLGSHV RKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v1	GHGGIRNNLNTENPLWD AIVGLGFLKDGVKNIPSF LSTDNIGTRETFLAGLIDS DGYVTDEHGIKATIKTIHT SVRDGLVSLARSLGLVV SVNAEPAKVDMNVTKHK ISYAIYMSGGDVLLNVLS KCAGSKKFRPAPAAAFA RECRGFYFELQELKEDD YYGITLSDDSDHQFLLGS QVVVQN	snapgene common features
Sce-VMA	Sce-VMA-5'-v5	CFAKGTNVLMADGSIECIENIEVGN KVMGKDGRPREVIKLPRGRETMYS VVQKSQHRAHKSDSSREVPELLKF TCNATHELVVRTPRSVRRLSRTIKG VEYFEVITFEMGQKKAPDGRIVELV KEVSKSYPISEGPERANELVESYR KASNKAYFEWTIEARDLSLLGSHV RKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v2	GHGGIRNNLNTENPLWD AIVGLGFLKDGVKNIPSF LSTDNIGTRETFLAGLIDS DGYVTDEHGIKATIKTIHT SVRDGLVSLARSLGLVV SVNAEPAKVDMNVTKHK ISYAIYMSGGDVLLNVLS KCAGSKKFRPAPAAAFA RECRGFYFELQELKEDD YYGITLSDDSDHQFLLGS QVVVQNCGERGNGSG	snapgene common features

Sce-VMA	Sce-VMA-5'-v5	CFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v3	GHGGIRNNLNTENPLWDAIVGLGFLKDGVKNIPSFLSTDNIGTRETFLAGLIDSDGYVTDEHGIKATIKTIHTSVRDGLVSLARSLGLVVSVNAEPAKVDMNVTKHKISYAIYMSGGDVLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLGSQVVVQNCTMTEKGSG	snapgene common features
Sce-VMA	Sce-VMA-5'-v5	CFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v4	GHGGIRNNLNTENPLWDAIVGLGFLKDGVKNIPSFLSTDNIGTRETFLAGLIDSDGYVTDEHGIKATIKTIHTSVRDGLVSLARSLGLVVSVNAEPAKVDMNVTKHKISYAIYMSGGDVLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLGSQVVVQNCGEKSMGSG	snapgene common features
Sce-VMA	Sce-VMA-5'-v5	CFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v5	VLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLANQVVVHN	https://www.nature.com/articles/nmeth.3585
Sce-VMA	Sce-VMA-5'-v5	CFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKF	https://www.nature.com/articl	Sce-VMA-3'-v6	VLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDS	https://www.nature.com/articl

		TCNATHELVVRTPRSVRRLSRTIKG VEYFEVITFEMGQKKAPDGRIVELV KEVSKSYPISEGPERANELVESYR KASNKAYFEWTIEARDLSLLGSHV RKATYQTYAPI	es/nmeth. 3585		DHQFLLANQVVVHNCGE RGNGSG	es/nmeth. 3585
Sce-VMA	Sce-VMA-5'-v5	CFAKGTNVLMADGSIECIENIEVGN KVMGKDGRPREVIKLPRGRETMYS VVQKSQHRAHKSDSSREVPELLKF TCNATHELVVRTPRSVRRLSRTIKG VEYFEVITFEMGQKKAPDGRIVELV KEVSKSYPISEGPERANELVESYR KASNKAYFEWTIEARDLSLLGSHV RKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v7	VLLNVLSKCAGSKKFRP APAAAFARECRGFYFEL QELKEDDYYGITLSDDS DHQFLLANQVVVHNCTM TEKGSG	https://www.nature.com/articles/nmeth.3585
Sce-VMA	Sce-VMA-5'-v5	CFAKGTNVLMADGSIECIENIEVGN KVMGKDGRPREVIKLPRGRETMYS VVQKSQHRAHKSDSSREVPELLKF TCNATHELVVRTPRSVRRLSRTIKG VEYFEVITFEMGQKKAPDGRIVELV KEVSKSYPISEGPERANELVESYR KASNKAYFEWTIEARDLSLLGSHV RKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v8	VLLNVLSKCAGSKKFRP APAAAFARECRGFYFEL QELKEDDYYGITLSDDS DHQFLLANQVVVHNCGE KSMGSG	https://www.nature.com/articles/nmeth.3585
Sce-VMA	Sce-VMA-5'-v6	GGIIYVGCFAKGTNVLMADGSIECI ENIEVGNKVMGKDGRPREVIKLPR GRETMYSVVQKSQHRAHKSDSSR EVPELLKFTCNATHELVVRTPRSV RRLSRTIKGVEYFEVITFEMGQKKA PDGRIVELVKEVSKSYPISEGPERA NELVESYRKASNKAYFEWTIEARD LSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v1	GHGGIRNNLNTENPLWD AIVGLGFLKDGVKNIPSF LSTDNIGTRETFLAGLIDS DGYVTDEHGIKATIKTIHT SVRDGLVSLARSLGLVV SVNAEPAKVDMNVTKHK ISYAIYMSGGDVLLNVLS KCAGSKKFRPAPAAAFA RECRGFYFELQELKEDD YYGITLSDDSDHQFLLGS QVVVQN	snapgene common features

Sce-VMA	Sce-VMA-5'-v6	GGIIYVGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v2	GHGGIRNNLNTENPLWDAIVGLGFLKDGVKNIPSFLSTDNIGTRETFLAGLIDSDGYVTDEHGIKATIKTIHTSVRDGLVSLARSLGLVVSVNAEPAKVDMNVTKHKISYAIYMSGGDVLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLGSQVVVQNCGERGNGSG	snapgene common features
Sce-VMA	Sce-VMA-5'-v6	GGIIYVGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v3	GHGGIRNNLNTENPLWDAIVGLGFLKDGVKNIPSFLSTDNIGTRETFLAGLIDSDGYVTDEHGIKATIKTIHTSVRDGLVSLARSLGLVVSVNAEPAKVDMNVTKHKISYAIYMSGGDVLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLGSQVVVQNCTMTEKGSG	snapgene common features
Sce-VMA	Sce-VMA-5'-v6	GGIIYVGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v4	GHGGIRNNLNTENPLWDAIVGLGFLKDGVKNIPSFLSTDNIGTRETFLAGLIDSDGYVTDEHGIKATIKTIHTSVRDGLVSLARSLGLVVSVNAEPAKVDMNVTKHKISYAIYMSGGDVLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLGSQVVVQNCGEKSMGSG	snapgene common features

Sce-VMA	Sce-VMA-5'-v6	GGIIYVGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v5	VLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLANQVVVHN	https://www.nature.com/articles/nmeth.3585
Sce-VMA	Sce-VMA-5'-v6	GGIIYVGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v6	VLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLANQVVVHNCGERGNGSG	https://www.nature.com/articles/nmeth.3585
Sce-VMA	Sce-VMA-5'-v6	GGIIYVGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v7	VLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLANQVVVHNCTMTEKGSG	https://www.nature.com/articles/nmeth.3585
Sce-VMA	Sce-VMA-5'-v6	GGIIYVGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v8	VLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLANQVVVHNCGEKSMGSG	https://www.nature.com/articles/nmeth.3585

Sce-VMA	Sce-VMA-5'-v7	GGVVLEKGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v1	GHGGIRNNLNTENPLWDAIVGLGFLKDGVKNIPSFLSTDNIGTRETFLAGLIDSDGYVTDEHGIKATIKTIHTSVRDGLVSLARSLGLVVSVNAEPAKVDMNVTKHKISYAIYMSGGDVLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLGSQVVVQN	snapgene common features
Sce-VMA	Sce-VMA-5'-v7	GGVVLEKGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v2	GHGGIRNNLNTENPLWDAIVGLGFLKDGVKNIPSFLSTDNIGTRETFLAGLIDSDGYVTDEHGIKATIKTIHTSVRDGLVSLARSLGLVVSVNAEPAKVDMNVTKHKISYAIYMSGGDVLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLGSQVVVQNCGERGNGSG	snapgene common features
Sce-VMA	Sce-VMA-5'-v7	GGVVLEKGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v3	GHGGIRNNLNTENPLWDAIVGLGFLKDGVKNIPSFLSTDNIGTRETFLAGLIDSDGYVTDEHGIKATIKTIHTSVRDGLVSLARSLGLVVSVNAEPAKVDMNVTKHKISYAIYMSGGDVLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLGSQVVVQNCTMTEKGSG	snapgene common features

Sce-VMA	Sce-VMA-5'-v7	GGVVLEKGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v4	GHGGIRNNLNTENPLWDAIVGLGFLKDGVKNIPSFLSTDNIGTRETFLAGLIDSDGYVTDEHGIKATIKTIHTSVRDGLVSLARSLGLVVSVNAEPAKVDMNVTKHKISYAIYMSGGDVLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLGSQVVVQNCGEKSMGSG	snapgene common features
Sce-VMA	Sce-VMA-5'-v7	GGVVLEKGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v5	VLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLANQVVVHN	https://www.nature.com/articles/nmeth.3585
Sce-VMA	Sce-VMA-5'-v7	GGVVLEKGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v6	VLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLANQVVVHNCGERGNGSG	https://www.nature.com/articles/nmeth.3585
Sce-VMA	Sce-VMA-5'-v7	GGVVLEKGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPER	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v7	VLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLANQVVVHNCTMTEKGSG	https://www.nature.com/articles/nmeth.3585

		ANELVESYRKASNKAYFEWTIEAR DLSLLGSHVRKATYQTYAPI				
Sce-VMA	Sce-VMA-5'-v7	GGVVLEKGCFAKGTNVLMADGSIE CIENIEVGNKVMGKDGRPREVIKLP RGRETMYSVVQKSQHRAHKSDSS REVPELLKFTCNATHELVVRTPRS VRRLSRTIKGVEYFEVITFEMGQKK APDGRIVELVKEVSKSYPISEGPER ANELVESYRKASNKAYFEWTIEAR DLSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v8	VLLNVLSKCAGSKKFRP APAAAFARECRGFYFEL QELKEDDYYGITLSDDS DHQFLLANQVVVHNCGE KSMGSG	https://www.nature.com/articles/nmeth.3585
Sce-VMA	Sce-VMA-5'-v8	GGFQTVGCFAKGTNVLMADGSIEC IENIEVGNKVMGKDGRPREVIKLPR GRETMYSVVQKSQHRAHKSDSSR EVPELLKFTCNATHELVVRTPRSV RRLSRTIKGVEYFEVITFEMGQKKA PDGRIVELVKEVSKSYPISEGPERA NELVESYRKASNKAYFEWTIEARD LSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v1	GHGGIRNNLNTENPLWD AIVGLGFLKDGVKNIPSF LSTDNIGTRETFLAGLIDS DGYVTDEHGIKATIKTIHT SVRDGLVSLARSLGLVV SVNAEPAKVDMNVTKHK ISYAIYMSGGDVLLNVLS KCAGSKKFRPAPAAAFA RECRGFYFELQELKEDD YYGITLSDDSDHQFLLGS QVVVQN	snapgene common features
Sce-VMA	Sce-VMA-5'-v8	GGFQTVGCFAKGTNVLMADGSIEC IENIEVGNKVMGKDGRPREVIKLPR GRETMYSVVQKSQHRAHKSDSSR EVPELLKFTCNATHELVVRTPRSV RRLSRTIKGVEYFEVITFEMGQKKA PDGRIVELVKEVSKSYPISEGPERA NELVESYRKASNKAYFEWTIEARD LSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v2	GHGGIRNNLNTENPLWD AIVGLGFLKDGVKNIPSF LSTDNIGTRETFLAGLIDS DGYVTDEHGIKATIKTIHT SVRDGLVSLARSLGLVV SVNAEPAKVDMNVTKHK ISYAIYMSGGDVLLNVLS KCAGSKKFRPAPAAAFA RECRGFYFELQELKEDD YYGITLSDDSDHQFLLGS QVVVQNCGERGNGSG	snapgene common features

Sce-VMA	Sce-VMA-5'-v8	GGFQTVGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v3	GHGGIRNNLNTENPLWDAIVGLGFLKDGVKNIPSFLSTDNIGTRETFLAGLIDSDGYVTDEHGIKATIKTIHTSVRDGLVSLARSLGLVVSVNAEPAKVDMNVTKHKISYAIYMSGGDVLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLGSQVVVQNCTMTEKGSG	snapgene common features
Sce-VMA	Sce-VMA-5'-v8	GGFQTVGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v4	GHGGIRNNLNTENPLWDAIVGLGFLKDGVKNIPSFLSTDNIGTRETFLAGLIDSDGYVTDEHGIKATIKTIHTSVRDGLVSLARSLGLVVSVNAEPAKVDMNVTKHKISYAIYMSGGDVLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLGSQVVVQNCGEKSMGSG	snapgene common features
Sce-VMA	Sce-VMA-5'-v8	GGFQTVGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSREVPELLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v5	VLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDSDHQFLLANQVVVHN	https://www.nature.com/articles/nmeth.3585
Sce-VMA	Sce-VMA-5'-v8	GGFQTVGCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQHRAHKSDSSR	https://www.nature.com/articl	Sce-VMA-3'-v6	VLLNVLSKCAGSKKFRPAPAAAFARECRGFYFELQELKEDDYYGITLSDDS	https://www.nature.com/articl

		EVPELLKFTCNATHELVVRTPRSV RRLSRTIKGVEYFEVITFEMGQKKA PDGRIVELVKEVSKSYPISEGPERA NELVESYRKASNKAYFEWTIEARD LSLLGSHVRKATYQTYAPI	es/nmeth. 3585		DHQFLLANQVVVHNCGE RGNGSG	es/nmeth. 3585
Sce-VMA	Sce-VMA-5'-v8	GGFQTVGCFAKGTNVLMADGSIEC IENIEVGNKVMGKDGRPREVIKLPR GRETMYSVVQKSQHRAHKSDSSR EVPELLKFTCNATHELVVRTPRSV RRLSRTIKGVEYFEVITFEMGQKKA PDGRIVELVKEVSKSYPISEGPERA NELVESYRKASNKAYFEWTIEARD LSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v7	VLLNVLSKCAGSKKFRP APAAAFARECRGFYFEL QELKEDDYYGITLSDDS DHQFLLANQVVVHNCTM TEKGSG	https://www.nature.com/articles/nmeth.3585
Sce-VMA	Sce-VMA-5'-v8	GGFQTVGCFAKGTNVLMADGSIEC IENIEVGNKVMGKDGRPREVIKLPR GRETMYSVVQKSQHRAHKSDSSR EVPELLKFTCNATHELVVRTPRSV RRLSRTIKGVEYFEVITFEMGQKKA PDGRIVELVKEVSKSYPISEGPERA NELVESYRKASNKAYFEWTIEARD LSLLGSHVRKATYQTYAPI	https://www.nature.com/articles/nmeth.3585	Sce-VMA-3'-v8	VLLNVLSKCAGSKKFRP APAAAFARECRGFYFEL QELKEDDYYGITLSDDS DHQFLLANQVVVHNCGE KSMGSG	https://www.nature.com/articles/nmeth.3585

Additional domains

The gene modifying polypeptide can bind a target DNA sequence and template nucleic acid (e.g., template RNA), nick the target site, and write (e.g., reverse transcribe) the template into DNA, resulting in a modification of the target site. In some embodiments, additional domains may be added to the polypeptide to enhance the efficiency of the process. In some embodiments, the gene modifying polypeptide may contain an additional DNA ligation domain to join reverse transcribed DNA to the DNA of the target site. In some embodiments, the polypeptide may comprise a heterologous RNA-binding domain. In some embodiments, the polypeptide may comprise a domain having 5′ to 3′ exonuclease activity (e.g., wherein the 5′ to 3′ exonuclease activity increases repair of the alteration of the target site, e.g., in favor of alteration over the original genomic sequence). In some embodiments, the polypeptide may comprise a domain having 3′ to 5′ exonuclease activity, e.g., proof-reading activity. In some embodiments, the writing domain, e.g., RT domain, has 3′ to 5′ exonuclease activity, e.g., proof-reading activity.

Template nucleic acids

The gene modifying systems described herein can modify a host target DNA site using a template nucleic acid sequence. In some embodiments, the gene modifying systems described herein transcribe an RNA sequence template into host target DNA sites by target-primed reverse transcription (TPRT). By modifying DNA sequence(s) via reverse transcription of the RNA sequence template directly into the host genome, the gene modifying system can insert an object sequence into a target genome without the need for exogenous DNA sequences to be introduced into the host cell (unlike, for example, CRISPR systems), as well as eliminate an exogenous DNA insertion step. The gene modifying system can also delete a sequence from the target genome or introduce a substitution using an object sequence. Therefore, the gene modifying system provides a platform for the use of customized RNA sequence templates containing object sequences, e.g., sequences comprising heterologous gene coding and/or function information.

In some embodiments, the template nucleic acid comprises one or more sequence (e.g., 2 sequences) that binds the gene modifying polypeptide.

In some embodiments, the template RNA comprises a nucleic acid sequence as listed in Table S4, or a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, the template RNA comprises a 5' end block sequence of a template sequence as listed in Table S4, or a nucleic acid sequence having at least 70%, 75%, 80%, 85%,

90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, the template RNA comprises a PBS sequence of a template sequence as listed in Table S4, or a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, the template RNA comprises a linker sequence of a template sequence as listed in Table S4, or a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, the template RNA comprises one or more (e.g., 1, 2, 3, or 4) RRS sequences of a template sequence as listed in Table S4, or nucleic acid sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, the template RNA comprises a 3' end block sequence of a template sequence as listed in Table S4, or a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, the template RNA comprises (e.g., in 5' to 3' order) a 5' end block sequence, PBS sequence, one or more RRS sequences, and a 3' end block sequence of a template sequence as listed in Table S4, or nucleic acid sequences having at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

In some embodiments a system or method described herein comprises a single template nucleic acid (e.g., template RNA). In some embodiments a system or method described herein comprises a plurality of template nucleic acids (e.g., template RNAs). For example, a system described herein comprises a first RNA comprising (e.g., from 5′ to 3′) a sequence that binds the gene modifying polypeptide (e.g., the DNA-binding domain and/or the endonuclease domain, e.g., a gRNA) and a sequence that binds a target site (e.g., a second strand of a site in a target genome), and a second RNA (e.g., a template RNA) comprising (e.g., from 5′ to 3′) optionally a sequence that binds the gene modifying polypeptide (e.g., that specifically binds the RT domain), a heterologous object sequence, and a PBS sequence. In some embodiments, when the system comprises a plurality of nucleic acids, each nucleic acid comprises a conjugating domain. In some embodiments, a conjugating domain enables association of nucleic acid molecules, e.g., by hybridization of complementary sequences. For example, in some embodiments a first RNA comprises a first conjugating domain and a second RNA comprises a second conjugating domain, and the first and second conjugating domains are capable of hybridizing to one another, e.g., under stringent conditions. In some embodiments, the stringent conditions for hybridization include hybridization in 4x sodium chloride/sodium citrate (SSC), at about 65 C, followed by a wash in 1xSSC, at about 65 C.

In some embodiments, the template nucleic acid comprises RNA. In some embodiments, the template nucleic acid comprises DNA (e.g., single stranded or double stranded DNA).

In some embodiments, the template nucleic acid comprises one or more (e.g., 2) homology domains that have homology to the target sequence. In some embodiments, the homology domains are about 10-20, 20-50, or 50-100 nucleotides in length.

In some embodiments, a template RNA can comprise a gRNA sequence, e.g., to direct the gene modifying polypeptide to a target site of interest. In some embodiments, a template RNA comprises (e.g., from 5′ to 3′) (i) optionally a gRNA spacer that binds a target site (e.g., a second strand of a site in a target genome), (ii) optionally a gRNA scaffold that binds a polypeptide described herein (e.g., a gene modifying polypeptide or a Cas polypeptide), (iii) a heterologous object sequence comprising a mutation region (optionally the heterologous object sequence comprises, from 5' to 3', a first homology region, a mutation region, and a second homology region), and (iv) a primer binding site (PBS) sequence comprising a 3′ target homology domain.

The template nucleic acid (e.g., template RNA) component of a genome editing system described herein typically is able to bind the gene modifying polypeptide of the system. In some embodiments the template nucleic acid (e.g., template RNA) has a 3′ region that is capable of binding a gene modifying polypeptide. The binding region, e.g., 3′ region, may be a structured RNA region, e.g., having at least 1, 2 or 3 hairpin loops, capable of binding the gene modifying polypeptide of the system. The binding region may associate the template nucleic acid (e.g., template RNA) with any of the polypeptide modules. In some embodiments, the binding region of the template nucleic acid (e.g., template RNA) may associate with an RNA-binding domain in the polypeptide. In some embodiments, the binding region of the template nucleic acid (e.g., template RNA) may associate with the reverse transcription domain of the gene modifying polypeptide (e.g., specifically bind to the RT domain). In some embodiments, the template nucleic acid (e.g., template RNA) may associate with the DNA binding domain of the polypeptide, e.g., a gRNA associating with a Cas9-derived DNA binding domain. In some embodiments, the binding region may also provide DNA target recognition, e.g., a gRNA hybridizing to the target DNA sequence and binding the polypeptide, e.g., a Cas9 domain. In some embodiments, the template nucleic acid (e.g., template RNA) may associate with multiple components of the polypeptide, e.g., DNA binding domain and reverse transcription domain.

In some embodiments the template RNA has a poly-A tail at the 3′ end. In some embodiments the template RNA does not have a poly-A tail at the 3′ end.

In some embodiments, a template RNA may be customized to correct a given mutation in the genomic DNA of a target cell (e.g., *ex vivo* or *in vivo*, e.g., in a target tissue or organ, e.g., in a subject). For example, the mutation may be a disease-associated mutation relative to the wild-type sequence. Without wishing to be bound by theory, any given target site and edit will have a large number of

possible template RNA molecules for use in a gene modifying system that will result in a range of editing efficiencies and fidelities. To partially reduce this screening burden, sets of empirical parameters help ensure optimal initial *in silico* designs of template RNAs or portions thereof. As a non-limiting illustrative example, for a selected mutation, the following design parameters may be employed. In some embodiments, design is initiated by acquiring approximately 500 bp (e.g., up to 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, or 700 bp, and optionally at least 20, 30, 40, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, or 650 bp) flanking sequence on either side of the mutation to serve as the target region. In some embodiments, a template nucleic acid comprises a gRNA. In some embodiments, a gRNA comprises a sequence (e.g., a CRISPR spacer) that binds a target site. In some embodiments, the sequence (e.g., a CRISPR spacer) that binds a target site for use in targeting a template nucleic acid to a target region is selected by considering the particular gene modifying polypeptide (e.g., endonuclease domain or writing domain, e.g., comprising a CRISPR/Cas domain) being used (e.g., for Cas9, a protospacer-adjacent motif (PAM) of NGG immediately 3′ of a 20 nucleotide gRNA binding region). In some embodiments, the CRISPR spacer is selected by ranking first by whether the PAM will be disrupted by the gene modifying system induced edit. In some embodiments, disruption of the PAM may increase edit efficiency. In some embodiments, the PAM can be disrupted by also introducing (e.g., as part of or in addition to another modification to a target site in genomic DNA) a silent mutation (e.g., a mutation that does not alter an amino acid residue encoded by the target nucleic acid sequence, if any) in the target site during gene modification. In some embodiments, the CRISPR spacer is selected by ranking sequences by the proximity of their corresponding genomic site to the desired edit location. In some embodiments, the gRNA comprises a gRNA scaffold. In some embodiments, the gRNA scaffold used may be a standard scaffold (e.g., for Cas9, 5′-GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC-3′), or may contain one or more nucleotide substitutions. In some embodiments, the heterologous object sequence has at least 90% identity, e.g., at least 90%, 95%, 98%, 99%, or 100% identity, or comprises no more than 1, 2, 3, 4, or 5 positions of non-identity to the target site 3′ of the first strand nick (e.g., immediately 3′ of the first strand nick or up to 1, 2, 3, 4, or 5 nucleotides 3′ of the first strand nick), with the exception of any insertion, substitution, or deletion that may be written into the target site by the gene modifying. In some embodiments, the 3′ target homology domain contains at least 90% identity, e.g., at least 90%, 95%, 98%, 99%, or 100% identity, or comprises no more than 1, 2, 3, 4, or 5 positions of non-identity to the target site 5′ of the first strand nick (e.g., immediately 5′ of the first strand nick or up to 1, 2, 3, 4, or 5 nucleotides 3′ of the first strand nick).

In some embodiments, the template nucleic acid is a template RNA. In some embodiments, the template RNA comprises one or more modified nucleotides. For example, in some embodiments, the

template RNA comprises one or more deoxyribonucleotides. In some embodiments, regions of the template RNA are replaced by DNA nucleotides, e.g., to enhance stability of the molecule. For example, the 3′ end of the template may comprise DNA nucleotides, while the rest of the template comprises RNA nucleotides that can be reverse transcribed. For instance, in some embodiments, the heterologous object sequence is primarily or wholly made up of RNA nucleotides (e.g., at least 90%, 95%, 98%, or 99% RNA nucleotides). In some embodiments, the PBS sequence is primarily or wholly made up of DNA nucleotides (e.g., at least 90%, 95%, 98%, or 99% DNA nucleotides). In other embodiments, the heterologous object sequence for writing into the genome may comprise DNA nucleotides. In some embodiments, the DNA nucleotides in the template are copied into the genome by a domain capable of DNA-dependent DNA polymerase activity. In some embodiments, the DNA-dependent DNA polymerase activity is provided by a DNA polymerase domain in the polypeptide. In some embodiments, the DNA-dependent DNA polymerase activity is provided by a reverse transcriptase domain that is also capable of DNA-dependent DNA polymerization, e.g., second strand synthesis. In some embodiments, the template molecule is composed of only DNA nucleotides.

In some embodiments, a system described herein comprises two nucleic acids which together comprise the sequences of a template RNA described herein. In some embodiments, the two nucleic acids are associated with each other non-covalently, e.g., directly associated with each other (e.g., via base pairing), or indirectly associated as part of a complex comprising one or more additional molecule.

A template RNA described herein may comprise, from 5' to 3': (1) a gRNA spacer; (2) a gRNA scaffold; (3) heterologous object sequence (4) a primer binding site (PBS) sequence. Each of these components is now described in more detail.

gRNA spacer and gRNA scaffold

A template RNA described herein may comprise a gRNA spacer that directs the gene modifying system to a target nucleic acid, and a gRNA scaffold that promotes association of the template RNA with the Cas domain of the gene modifying polypeptide. The systems described herein can also comprise a gRNA that is not part of a template nucleic acid. For example, a gRNA that comprises a gRNA spacer and gRNA scaffold, but not a heterologous object sequence or a PBS sequence, can be used, e.g., to promote unwinding of the target nucleic acid or to reduce MMR reversal of a desired edit by the host cell (e.g., as described in the End Block Sequences and Additional Guide RNA sections herein), or to induce second strand nicking, e.g., as described in the section herein entitled "Second Strand Nicking".

In some embodiments, the gRNA is a short synthetic RNA composed of a scaffold sequence that participates in CRISPR-associated protein binding and a user-defined ~20 nucleotide targeting sequence for a genomic target. The structure of a complete gRNA was described by Nishimasu et al. Cell 156,

P935-949 (2014). The gRNA (also referred to as sgRNA for single-guide RNA) consists of crRNA- and tracrRNA-derived sequences connected by an artificial tetraloop. The crRNA sequence can be divided into guide (20 nt) and repeat (12 nt) regions, whereas the tracrRNA sequence can be divided into anti-repeat (14 nt) and three tracrRNA stem loops (Nishimasu et al. Cell 156, P935-949 (2014)). In practice, guide RNA sequences are generally designed to have a length of between 17 – 24 nucleotides (e.g., 19, 20, or 21 nucleotides) and be complementary to a targeted nucleic acid sequence. Custom gRNA generators and algorithms are available commercially for use in the design of effective guide RNAs. In some embodiments, the gRNA comprises two RNA components from the native CRISPR system, e.g. crRNA and tracrRNA. As is well known in the art, the gRNA may also comprise a chimeric, single guide RNA (sgRNA) containing sequence from both a tracrRNA (for binding the nuclease) and at least one crRNA (to guide the nuclease to the sequence targeted for editing/binding). Chemically modified sgRNAs have also been demonstrated to be effective for use with CRISPR-associated proteins; see, for example, Hendel et al. (2015) Nature Biotechnol., 985 – 991. In some embodiments, a gRNA spacer comprises a nucleic acid sequence that is complementary to a DNA sequence associated with a target gene.

In some embodiments, the region of the template nucleic acid, e.g., template RNA, comprising the gRNA adopts an underwound ribbon-like structure of gRNA bound to target DNA (e.g., as described in Mulepati et al. Science 19 Sep 2014:Vol. 345, Issue 6203, pp. 1479-1484). Without wishing to be bound by theory, this non-canonical structure is thought to be facilitated by rotation of every sixth nucleotide out of the RNA-DNA hybrid. Thus, in some embodiments, the region of the template nucleic acid, e.g., template RNA, comprising the gRNA may tolerate increased mismatching with the target site at some interval, e.g., every sixth base. In some embodiments, the region of the template nucleic acid, e.g., template RNA, comprising the gRNA comprising homology to the target site may possess wobble positions at a regular interval, e.g., every sixth base, that do not need to base pair with the target site.

In some embodiments, the template nucleic acid (e.g., template RNA) has at least 15, 16, 17, 18, 19, 20, 21, 22, 23, or 24 bases of at least 80%, 85%, 90%, 95%, 99%, or 100% homology to the target site, e.g., at the 5' end, e.g., comprising a gRNA spacer sequence of length appropriate to the Cas9 domain of the gene modifying polypeptide (Table 8).

Table 12 provides parameters to define components for designing gRNA and/or Template RNAs to apply Cas variants listed in Table 8 for gene modifying. The cut site indicates the validated or predicted protospacer adjacent motif (PAM) requirements, validated or predicted location of cut site (relative to the most upstream base of the PAM site). The gRNA for a given enzyme can be assembled by concatenating the crRNA, Tetraloop, and tracrRNA sequences, and further adding a 5′ spacer of a length within Spacer

(min) and Spacer (max) that matches a protospacer at a target site. Further, the predicted location of the ssDNA nick at the target is important for designing a PBS sequence of a Template RNA that can anneal to the sequence immediately 5′ of the nick in order to initiate target primed reverse transcription. In some embodiments, a gRNA scaffold described herein comprises a nucleic acid sequence comprising, in the 5' to 3' direction, a crRNA of Table 12, a tetraloop from the same row of Table 12, and a tracrRNA from the same row of Table 12, or a sequence having at least 70%, 80%, 85%, 90%, 95%, or 99% identity thereto. In some embodiments, the gRNA or template RNA comprising the scaffold further comprises a gRNA spacer having a length within the Spacer (min) and Spacer (max) indicated in the same row of Table 12. In some embodiments, the gRNA or template RNA having a sequence according to Table 12 is comprised by a system that further comprises a gene modifying polypeptide, wherein the gene modifying polypeptide comprises a Cas domain described in the same row of Table 12.

Table 12. Parameters to define components for designing gRNA and/or Template RNAs to apply Cas variants listed in Table 8 in gene modifying systems

Variant	PAM(s)	Cut	Tier	Spacer (min)	Spacer (max)	crRNA	Tetraloop	tracrRNA
Nme2Cas9	NNNNCC	-3	1	22	24	GTTGTAGC TCCCTTTCT CATTTCG	GAAA	CGAAATGAGAACCGTTGCTACAATAAGGC CGTCTGAAAAGATGTGCCGCAACGCTCTG CCCCTTAAAGCTTCTGCTTTAAGGGGCATC GTTTA
PpnCas9	NNNNRTT		1	21	24	GTTGTAGC TCCCTTTTT CATTTCGC	GAAA	GCGAAATGAAAAACGTTGTTACAATAAGA GATGAATTTCTCGCAAAGCTCTGCCTCTTG AAATTTCGGTTTCAAGAGGCATC
SauCas9	NNGRR;N NGRRT	-3	1	21	23	GTTTTAGT ACTCTG	GAAA	CAGAATCTACTAAAACAAGGCAAAATGCC GTGTTTATCTCGTCAACTTGTTGGCGAGA
SauCas9-KKH	NNNRR;N NNRRT	-3	1	21	21	GTTTTAGT ACTCTG	GAAA	CAGAATCTACTAAAACAAGGCAAAATGCC GTGTTTATCTCGTCAACTTGTTGGCGAGA
SauriCas9	NNGG	-3	1	21	21	GTTTTAGT ACTCTG	GAAA	CAGAATCTACTAAAACAAGGCAAAATGCC GTGTTTATCTCGTCAACTTGTTGGCGAGA
SauriCas9-KKH	NNRG	-3	1	21	21	GTTTTAGT ACTCTG	GAAA	CAGAATCTACTAAAACAAGGCAAAATGCC GTGTTTATCTCGTCAACTTGTTGGCGAGA
ScaCas9-Sc++	NNG	-3	1	20	20	GTTTTAGA GCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTA TCAACTTGAAAAAGTGGCACCGAGTCGGT GC
SpyCas9	NGG	-3	1	20	20	GTTTTAGA GCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTA TCAACTTGAAAAAGTGGCACCGAGTCGGT GC

SpyCas9-NG	NG (NGG=NGA=NGT>NGC)	-3	1	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-SpRY	NRN>NYN	-3	1	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
St1Cas9	NNAGAAW>NNAGGAW=NNGGAAW	-3	1	20	20	GTCTTTGTACTCTG	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT
BlatCas9	NNNNCNAA>NNNNCNDD>NNNNNC	-3	1	19	23	GCTATAGTTCCTTACT	GAAA	GGTAAGTTGCTATAGTAAGGGCAACAGACCCGAGGCGTTGGGGATCGCCTAGCCCGTGTTTACGGGCTCTCCCCATATTCAAAATAATGACAGACGAGCACCTTGGAGCATTTATCTCCGAGGTGCT
cCas9-v16	NNVACT;NNVATGM;NNVATT;NNVGCT;NNVGTG;NNVGTT	-3	2	21	21	GTCTTAGTACTCTG	GAAA	CAGAATCTACTAAGACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
cCas9-v17	NNVRRN	-3	2	21	21	GTCTTAGTACTCTG	GAAA	CAGAATCTACTAAGACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
cCas9-v21	NNVACT;NNVATGM;NNVATT;NNVGCT;NNVGTG;NNVGTT	-3	2	21	21	GTCTTAGTACTCTG	GAAA	CAGAATCTACTAAGACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
cCas9-v42	NNVRRN	-3	2	21	21	GTCTTAGTACTCTG	GAAA	CAGAATCTACTAAGACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
CdiCas9	NNRHHHY;NNRAAAY		2	22	22	ACTGGGGTTCAG	GAAA	CTGAACCTCAGTAAGCATTGGCTCGTTTCCAATGTTGATTGCTCCGCCGGTGCTCCTTATTTTTAAGGGCGCCGGC
CjeCas9	NNNNRYAC	-3	2	21	23	GTTTTAGTCCCT	GAAA	AGGGACTAAAATAAAGAGTTTGCGGGACTCTGCGGGGTTACAATCCCCTAAAACCGC
GeoCas9	NNNNCRAA		2	21	23	GTCATAGTTCCCCTGA	GAAA	TCAGGGTTACTATGATAAGGGCTTTCTGCCTAAGGCAGACTGACCCGCGGCGTTGGGGATCGCCTGTCGCCCGCTTTTGGCGGGCATTCCCCATCCTT
iSpyMacCas9	NAAN	-3	2	19	21	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
NmeCas9	NNNNGAYT;NNNNGYTT;NNNNGAYA;NNNNGTCT	-3	2	20	24	GTTGTAGCTCCCTTTCTCATTTCG	GAAA	CGAAATGAGAACCGTTGCTACAATAAGGCCGTCTGAAAAGATGTGCCGCAACGCTCTGCCCCTTAAAGCTTCTGCTTTAAGGGGCATCGTTTA

ScaCas9	NNG	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
ScaCas9-HiFi-Sc++	NNG	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-3var-NRRH	NRRH	-3	2	20	20	GTTTAAGAGCTATGCTG	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-3var-NRTH	NRTH	-3	2	20	20	GTTTAAGAGCTATGCTG	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-3var-NRCH	NRCH	-3	2	20	20	GTTTAAGAGCTATGCTG	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-HF1	NGG	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-QQR1	NAAG	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-SpG	NGN	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-VQR	NGAN	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-VRER	NGCG	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-xCas	NG;GAA;GAT	-3	2	20	20	GTTTAAGAGCTATGCTG	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-xCas-NG	NG	-3	2	20	20	GTTTAAGAGCTATGCTG	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
St1Cas9-CNRZ1066	NNACAA	-3	2	20	20	GTCTTTGTACTCTG	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT
St1Cas9-LMG1831	NNGCAA	-3	2	20	20	GTCTTTGTACTCTG	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT
St1Cas9-MTH17CL396	NNAAAA	-3	2	20	20	GTCTTTGTACTCTG	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT
St1Cas9-TH1477	NNGAAA	-3	2	20	20	GTCTTTGTACTCTG	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT

sRGN3.1	NNGG		1	21	23	GTTTTAGT ACTCTG	GAAA	CAGAATCTACTGAAACAAGACAATATGTC GTGTTTATCCCATCAATTTATTGGTGGGAT TTT
sRGN3.3	NNGG		1	21	23	GTTTTAGT ACTCTG	GAAA	CAGAATCTACTGAAACAAGACAATATGTC GTGTTTATCCCATCAATTTATTGGTGGGAT TTT

Herein, when an RNA sequence (e.g., a template RNA sequence) is said to comprise a particular sequence (e.g., a sequence of Table 12 or a portion thereof) that comprises thymine (T), it is of course understood that the RNA sequence may (and frequently does) comprise uracil (U) in place of T. For instance, the RNA sequence may comprise U at every position shown as T in the sequence in Table 12. More specifically, the present disclosure provides an RNA sequence according to every gRNA scaffold sequence of Table 12, wherein the RNA sequence has a U in place of each T in the sequence in Table 12. Additionally, it is understood that terminal Us and Ts may optionally be added or removed from tracrRNA sequences and may be modified or unmodified when provided as RNA. Without wishing to be bound by example, versions of gRNA scaffold sequences alternative to those exemplified in Table 12 may also function with the different Cas9 enzymes or derivatives thereof exemplified in Table 8, e.g., alternate gRNA scaffold sequences with nucleotide additions, substitutions, or deletions, e.g., sequences with stem-loop structures added or removed. It is contemplated herein that the gRNA scaffold sequences represent a component of gene modifying systems that can be similarly optimized for a given system, Cas-RT fusion polypeptide, indication, target mutation, template RNA, or delivery vehicle.

RNA binding domain recruitment sites (RRS)

In some embodiments, a template RNA described herein comprises an RNA binding domain (RBD) recruitment site (RRS), capable of binding to an RBD as described herein. In some embodiments, an RRS binds to the RBD of a gene modifying polypeptide or complex as described herein. In some embodiments, the RRS is located at the 5' end of the template RNA. In some embodiments, the RRS is located within 5, 10, 15, 20, 25, or 30 nucleotides of the 5' end of the template RNA. In some embodiments, the RRS comprises one or more (e.g., 1 or 2) stem-loop sequences.

In some embodiments, a template nucleic acid comprises a plurality of RRS sequences (e.g., a plurality of the same RRS sequence, or a plurality of different RRS sequences). In some embodiments, the RRS sequence is repeated at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 times. In some embodiments, the plurality of RRS sequences is separated by one or more linker sequences. In some embodiments, the

plurality of RRS sequences are positioned adjacent to each other (e.g., without an intervening linker sequence).

In some embodiments, the RRS is not located between a PBS and a heterologous object sequence. In some embodiments, the RRS is located between a PBS and a heterologous object sequence.

In some embodiments, an RRS comprises the nucleic acid sequence of an RRS as listed in Table 40, or a nucleic acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, an RRS comprises the nucleic acid sequence of an RRS as listed in Table 40, or a nucleic acid sequence having no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotide differences therefrom. Herein, when an RNA sequence (e.g., an RRS) is said to comprise a particular sequence (e.g., a sequence of Table 40 or a portion thereof) that comprises thymine (T), it is of course understood that the RNA sequence may (and frequently does) comprise uracil (U) in place of T. For instance, the RNA sequence may comprise U at every position shown as T in the sequence in Table 40. More specifically, the present disclosure provides an RNA sequence according to every RRS sequence of Table 40, wherein the RNA sequence has a U in place of each T in the sequence in Table 40.

Table 40. Exemplary RNA binding domain recruitment sites (RRS)

RBP recognition site (RRS)	RBP binding partner	Sequence (5' to 3')
MS2	MCP	gcACATGAGGATCACCCATGTgc
PP7	PCP	caTAAGGAGTTTATATGGAAACCCTTAtg
com	Com	CTGAATGCCTGCGAGCATC
LS4-1	LS4	GGCAGAGAAAGGCCATACAATCATTGGCCTTGTGAGGCC GTGTGTCTTCCAGTGGC
LS12-1	LS12	GGCAGAGAAAGGCCATACAATCATTGGCTTTTCCATGACG CCAGTTCCAGTGGC
BoxB	lambdaN(1-22)	GGGCCCTGAAGAAGGGCCC
Kt	L7Ae	GGATCCGTGATCGGAAACGTGAGATCC
CS1	LS4	GGTGGCAGAGAAAGGCGAAAGCCTTGTGAGGCCATCAA
CS2	LS12	GGATGCAGAGAACGAAAGTTCCATGACGCATCCAA

End block sequences

In some embodiments, a template RNA as described herein comprises one or more end block sequences. In some instances, an end block sequence or end protection sequence, as described herein,

may protect the template RNA from exonuclease degradation (e.g., reduces exonuclease degradation of the template RNA by at least 25%, 50%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% relative to an otherwise similar template RNA lacking the end block sequence). In some instances, an end block sequence or end protection sequence, as described herein, may act to terminate a reverse transcriptase reaction. In some embodiments, an end block sequence is positioned adjacent to, or within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, or 30 nucleotides of a 5' pro-spacer sequence (e.g., which pairs with the nicked target nucleic acid strand). In embodiments, the 5' pro-spacer sequence has 100% complementarity to the nicked target nucleic acid strand and/or directs nicking activity by a Cas domain (e.g., a Cas9 domain, e.g., an nCas9). In embodiments, the 5' pro-spacer sequence has less than or equal to 17 nucleotides of complementarity (e.g., about 5, 10, 11, 12, 13, 14, 15, 16, or 17 nucleotides of complementarity) to the target nucleic acid strand, e.g., and promotes unwinding of the target nucleic acid without nicking. In some embodiments, an end block sequence (e.g., a 5' end block sequence) comprises a gRNA spacer (e.g., a pro-spacer) as described herein. In some embodiments, an end block sequence (e.g., a 5' end blocksequence) comprises a gRNA scaffold as described herein. In some embodiments, a pro-spacer as described herein does not have a length sufficient for full nicking, or has a length suitable for limited nicking. In some embodiments, a gRNA spacer as described herein has a length suitable for full nicking.

In some embodiments, an end block sequence comprises the nucleic acid sequence of an end block sequence as listed in Table 41, or a nucleic acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto, or the reverse complement thereof. In some embodiments, an end block sequence comprises the nucleic acid sequence of an end block sequence as listed in Table 41, or a nucleic acid sequence having no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotide differences therefrom, or the reverse complement thereof. Herein, when an RNA sequence (e.g., a end block sequence) is said to comprise a particular sequence (e.g., a sequence of Table 41 or a portion thereof) that comprises thymine (T), it is of course understood that the RNA sequence may (and frequently does) comprise uracil (U) in place of T. For instance, the RNA sequence may comprise U at every position shown as T in the sequence in Table 41. More specifically, the present disclosure provides an RNA sequence according to every end block sequence of Table 41, wherein the RNA sequence has a U in place of each T in the sequence in Table 41.

Table 41. Exemplary end block sequences

End-block	Sequence (5' to 3')
G-quadruplex	GGTGGTGGTGG
Tinoco hairpin	GGACTTCGGTCC

GC-Geo hairpin	CTCATAGTTCCCCTGAGAAATCAGGGTTACTATGAG
Nme2Cas9 scaffold	GTTGTAGCTCCCTTTCTCATTTCGGAAACGAAATGAGAACCGTTGCTAC AATAAGGCCGTCTGAAAAGATGTGCCGCAACGCTCTGCCCCTTAAAGC TTCTGCTTTAAGGGGCATCGTTTA
Nme2Cas9 **spacer**+scaffold	**CAGTACATGACCTTACGGGA**GTTGTAGCTCCCTTTCTCATTTCG GAAACGAAATGAGAACCGTTGCTACAATAAGGCCGTCTGAAAAGATGT GCCGCAACGCTCTGCCCCTTAAAGCTTCTGCTTTAAGGGGCATCGTTTA
Nme2Cas9 **16 nt spacer**+scaffold	**ACATGACCTTACGGGA**GTTGTAGCTCCCTTTCTCATTTCGGAAAC GAAATGAGAACCGTTGCTACAATAAGGCCGTCTGAAAAGATGTGCCGC AACGCTCTGCCCCTTAAAGCTTCTGCTTTAAGGGGCATCGTTTA
BlatCas9 scaffold	GCTATAGTTCCTTACTGAAAGGTAAGTTGCTATAGTAAGGGCAACAGA CCCGAGGCGTTGGGGATCGCCTAGCCCGTGTTTACGGGCTCTCCCCAT ATTCAAAATAATGACAGACGAGCACCTTGGAGCATTTATCTCCGAGGT GCT
GeoCas9	GTCATAGTTCCCCTGAGAAATCAGGGTTACTATGATAAGGGCTTTCTGC CTAAGGCAGACTGACCCGCGGCGTTGGGGATCGCCTGTCGCCCGCTTT TGGCGGGCATTCCCCATCCTT
PpnCas9 scaffold	GTTGTAGCTCCCTTTTTCATTTCGCGAAAGCGAAATGAAAAACGTTGTT ACAATAAGAGATGAATTTCTCGCAAAGCTCTGCCTCTTGAAATTTCGGT TTCAAGAGGCATC
CdiCas9scaffold	ACTGGGGTTCAGGAAACTGAACCTCAGTAAGCATTGGCTCGTTTCCAAT GTTGATTGCTCCGCCGGTGCTCCTTATTTTTAAGGGCGCCGGC
SpyCas9+**hairpin** scaffold	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCA ACTTGAAAAAGTGGCACCG**GGACTTCGGTCC**CGGTGC
St1Cas9 scaffold	GTCTTTGTACTCTGGTACCAGAAGCTACAAAGATAAGGCTTCATGCCGA AATCAACACCCTGTCATTTTATGGCAGGGTGTTTT
cCas9-v16 scaffold	GTCTTAGTACTCTGGAAACAGAATCTACTAAGACAAGGCAAAATGCCG TGTTTATCTCGTCAACTTGTTGGCGAGA
SpyCas9-3var-NRRH scaffold	GTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAG TCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SauCas9 scaffold	GTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCG TGTTTATCTCGTCAACTTGTTGGCGAGA
CjeCas9 scaffold	GTTTTAGTCCCTGAAAAGGGACTAAAATAAAGAGTTTGCGGGACTCTG CGGGGTTACAATCCCCTAAAACCGC
SpyCas9 scaffold	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCA ACTTGAAAAAGTGGCACCGAGTCGGTGC

In some embodiments, an end block comprises a pro-spacer sequence (e.g., a 5' protospacer sequence), e.g., as described herein. In certain embodiments, the pro-spacer sequence has greater than or equal to 17 nucleotides of complementarity (e.g., about 17, 18, 19, 20, 21, 22, or 23 nucleotides of

complementarity) to the target nucleic acid strand. In certain embodiments, the pro-spacer sequence promotes unwinding and nicking of the target nucleic acid.

Heterologous object sequence

A template RNA described herein may comprise a heterologous object sequence that the gene modifying polypeptide can use as a template for reverse transcription, to write a desired sequence into the target nucleic acid. In some embodiments, the heterologous object sequence comprises, from 5' to 3', a post-edit homology region, the mutation region, and a pre-edit homology region. Without wishing to be bound by theory, an RT performing reverse transcription on the template RNA first reverse transcribes the pre-edit homology region, then the mutation region, and then the post-edit homology region, thereby creating a DNA strand comprising the desired mutation with a homology region on either side.

In some embodiments, the heterologous object sequence is at least 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 120, 140, 160, 180, 200, 500, or 1,000 nucleotides (nts) in length, or at least 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, or 10 kilobases in length. In some embodiments, the heterologous object sequence is no more than 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 120, 140, 160, 180, 200, 500, 1,000, or 2000 nucleotides (nts) in length, or no more than 20, 15, 10, 9, 8, 7, 6, 5, 4, or 3 kilobases in length. In some embodiments, the heterologous object sequence is 30-1000, 40-1000, 50-1000, 60-1000, 70-1000, 74-1000, 75-1000, 76-1000, 77-1000, 78-1000, 79-1000, 80-1000, 85-1000, 90-1000, 100-1000, 120-1000, 140-1000, 160-1000, 180-1000, 200-1000, 500-1000, 30-500, 40-500, 50-500, 60-500, 70-500, 74-500, 75-500, 76-500, 77-500, 78-500, 79-500, 80-500, 85-500, 90-500, 100-500, 120-500, 140-500, 160-500, 180-500, 200-500, 30-200, 40-200, 50-200, 60-200, 70-200, 74-200, 75-200, 76-200, 77-200, 78-200, 79-200, 80-200, 85-200, 90-200, 100-200, 120-200, 140-200, 160-200, 180-200, 30-100, 40-100, 50-100, 60-100, 70-100, 74-100, 75-100, 76-100, 77-100, 78-100, 79-100, 80-100, 85-100, or 90-100 nucleotides (nts) in length, or 1-20, 1-15, 1-10, 1-9, 1-8, 1-7, 1-6, 1-5, 1-4, 1-3, 1-2, 2-20, 2-15, 2-10, 2-9, 2-8, 2-7, 2-6, 2-5, 2-4, 2-3, 3-20, 3-15, 3-10, 3-9, 3-8, 3-7, 3-6, 3-5, 3-4, 4-20, 4-15, 4-10, 4-9, 4-8, 4-7, 4-6, 4-5, 5-20, 5-15, 5-10, 5-9, 5-8, 5-7, 5-6, 6-20, 6-15, 6-10, 6-9, 6-8, 6-7, 7-20, 7-15, 7-10, 7-9, 7-8, 8-20, 8-15, 8-10, 8-9, 9-20, 9-15, 9-10, 10-15, 10-20, or 15-20 kilobases in length. In some embodiments, the heterologous object sequence is 10-100, 10-90, 10-80, 10-70, 10-60, 10-50, 10-40, 10-30, or 10-20 nt in length, e.g., 10-80, 10-50, or 10-20 nt in length, e.g., about10-20 nt in length. In some

embodiments, the heterologous object sequence is 8-30, 9-25, 10-20, 11-16, or 12-15 nucleotides in length, e.g., is 11-16 nt in length. Without wishing to be bound by theory, in some embodiments, a larger insertion size, larger region of editing (e.g., the distance between a first edit/substitution and a second edit/substitution in the target region), and/or greater number of desired edits (e.g., mismatches of the heterologous object sequence to the target genome), may result in a longer optimal heterologous object sequence.

In certain embodiments, the template nucleic acid comprises a customized RNA sequence template which can be identified, designed, engineered and constructed to contain sequences altering or specifying host genome function, for example by introducing a heterologous coding region into a genome; affecting or causing exon structure/alternative splicing, e.g., leading to exon skipping of one or more exons; causing disruption of an endogenous gene, e.g., creating a genetic knockout; causing transcriptional activation of an endogenous gene; causing epigenetic regulation of an endogenous DNA; causing up-regulation of one or more operably linked genes, e.g., leading to gene activation or overexpression; causing down-regulation of one or more operably linked genes, e.g., creating a genetic knock-down; etc. In certain embodiments, a customized RNA sequence template can be engineered to contain sequences coding for exons and/or transgenes, provide binding sites for transcription factor activators, repressors, enhancers, etc., and combinations thereof. In some embodiments, a customized template can be engineered to encode a nucleic acid or peptide tag to be expressed in an endogenous RNA transcript or endogenous protein operably linked to the target site. In other embodiments, the coding sequence can be further customized with splice donor sites, splice acceptor sites, or poly-A tails.

The template nucleic acid (e.g., template RNA) of the system typically comprises an object sequence (e.g., a heterologous object sequence) for writing a desired sequence into a target DNA. The object sequence may be coding or non-coding. The template nucleic acid (e.g., template RNA) can be designed to result in insertions, mutations, or deletions at the target DNA locus. In some embodiments, the template nucleic acid (e.g., template RNA) may be designed to cause an insertion in the target DNA. For example, the template nucleic acid (e.g., template RNA) may contain a heterologous sequence, wherein the reverse transcription will result in insertion of the heterologous sequence into the target DNA. In other embodiments, the RNA template may be designed to introduce a deletion into the target DNA. For example, the template nucleic acid (e.g., template RNA) may match the target DNA upstream and downstream of the desired deletion, wherein the reverse transcription will result in the copying of the upstream and downstream sequences from the template nucleic acid (e.g., template RNA) without the intervening sequence, e.g., causing deletion of the intervening sequence. In other embodiments, the template nucleic acid (e.g., template RNA) may be designed to introduce an edit into the target DNA. For example, the template RNA may match the target DNA sequence with the exception of one or more

nucleotides, wherein the reverse transcription will result in the copying of these edits into the target DNA, e.g., resulting in mutations, e.g., transition or transversion mutations.

In some embodiments, writing of an object sequence into a target site results in the substitution of nucleotides, e.g., where the full length of the object sequence corresponds to a matching length of the target site with one or more mismatched bases. In some embodiments, a heterologous object sequence may be designed such that a combination of sequence alterations may occur, e.g., a simultaneous addition and deletion, addition and substitution, or deletion and substitution.

In some embodiments, the heterologous object sequence may contain an open reading frame or a fragment of an open reading frame. In some embodiments the heterologous object sequence has a Kozak sequence. In some embodiments the heterologous object sequence has an internal ribosome entry site. In some embodiments the heterologous object sequence has a self-cleaving peptide such as a T2A or P2A site. In some embodiments the heterologous object sequence has a start codon. In some embodiments the template RNA has a splice acceptor site. In some embodiments the template RNA has a splice donor site. Exemplary splice acceptor and splice donor sites are described in WO2016044416, incorporated herein by reference in its entirety. Exemplary splice acceptor site sequences are known to those of skill in the art. In some embodiments the template RNA has a microRNA binding site downstream of the stop codon. In some embodiments the template RNA has a polyA tail downstream of the stop codon of an open reading frame. In some embodiments the template RNA comprises one or more exons. In some embodiments the template RNA comprises one or more introns. In some embodiments the template RNA comprises a eukaryotic transcriptional terminator. In some embodiments the template RNA comprises an enhanced translation element or a translation enhancing element. In some embodiments the RNA comprises the human T-cell leukemia virus (HTLV-1) R region. In some embodiments the RNA comprises a posttranscriptional regulatory element that enhances nuclear export, such as that of Hepatitis B Virus (HPRE) or Woodchuck Hepatitis Virus (WPRE).

In some embodiments, the heterologous object sequence may contain a non-coding sequence. For example, the template nucleic acid (e.g., template RNA) may comprise a regulatory element, e.g., a promoter or enhancer sequence or miRNA binding site. In some embodiments, integration of the object sequence at a target site will result in upregulation of an endogenous gene. In some embodiments, integration of the object sequence at a target site will result in downregulation of an endogenous gene. In some embodiments the template nucleic acid (e.g., template RNA) comprises a tissue specific promoter or enhancer, each of which may be unidirectional or bidirectional. In some embodiments the promoter is an RNA polymerase I promoter, RNA polymerase II promoter, or RNA polymerase III promoter. In some embodiments the promoter comprises a TATA element. In some embodiments the promoter comprises a

B recognition element. In some embodiments the promoter has one or more binding sites for transcription factors.

In some embodiments, the template nucleic acid (e.g., template RNA) comprises a site that coordinates epigenetic modification. In some embodiments, the template nucleic acid (e.g., template RNA) comprises a chromatin insulator. For example, the template nucleic acid (e.g., template RNA) comprises a CTCF site or a site targeted for DNA methylation.

In some embodiments, the template nucleic acid (e.g., template RNA) comprises a gene expression unit composed of at least one regulatory region operably linked to an effector sequence. The effector sequence may be a sequence that is transcribed into RNA (e.g., a coding sequence or a non-coding sequence such as a sequence encoding a micro RNA).

In some embodiments, the heterologous object sequence of the template nucleic acid (e.g., template RNA) is inserted into a target genome in an endogenous intron. In some embodiments, the heterologous object sequence of the template nucleic acid (e.g., template RNA) is inserted into a target genome and thereby acts as a new exon. In some embodiments, the insertion of the heterologous object sequence into the target genome results in replacement of a natural exon or the skipping of a natural exon.

In some embodiments, the heterologous object sequence of the template nucleic acid (e.g., template RNA) is inserted into the target genome in a genomic safe harbor site, such as AAVS1, CCR5, ROSA26, or albumin locus. In some embodiments, a gene modifying is used to integrate a CAR into the T-cell receptor α constant (*TRAC*) locus (Eyquem et al Nature 543, 113-117 (2017)). In some embodiments, a gene modifying system is used to integrate a CAR into a T-cell receptor β constant (*TRBC*) locus. Many other safe harbors have been identified by computational approaches (Pellenz et al Hum Gen Ther 30, 814-828 (2019)) and could be used for gene modifying system-mediated integration. In some embodiments, the heterologous object sequence of the template nucleic acid (e.g., template RNA) is added to the genome in an intergenic or intragenic region. In some embodiments, the heterologous object sequence of the template nucleic acid (e.g., template RNA) is added to the genome 5′ or 3′ within 0.1 kb, 0.25 kb, 0.5 kb, 0.75, kb, 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 7.5 kb, 10 kb, 15 kb, 20 kb, 25 kb, 50, 75 kb, or 100 kb of an endogenous active gene. In some embodiments, the heterologous object sequence of the template nucleic acid (e.g., template RNA) is added to the genome 5′ or 3′ within 0.1 kb, 0.25 kb, 0.5 kb, 0.75, kb, 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 7.5 kb, 10 kb, 15 kb, 20 kb, 25 kb, 50, 75 kb, or 100 kb of an endogenous promoter or enhancer. In some embodiments, the heterologous object sequence of the template nucleic acid (e.g., template RNA) can be, e.g., 50-50,000 base pairs (e.g., between 50-40,000 bp, between 500-30,000 bp between 500-20,000 bp, between 100-15,000 bp, between 500-10,000 bp, between 50-10,000 bp, between 50-5,000 bp.

The template nucleic acid (e.g., template RNA) can be designed to result in insertions, mutations, or deletions at the target DNA locus. In some embodiments, the template nucleic acid (e.g., template RNA) may be designed to cause an insertion in the target DNA. For example, the template nucleic acid (e.g., template RNA) may contain a heterologous object sequence, wherein the reverse transcription will result in insertion of the heterologous object sequence into the target DNA. In other embodiments, the RNA template may be designed to write a deletion into the target DNA. For example, the template nucleic acid (e.g., template RNA) may match the target DNA upstream and downstream of the desired deletion, wherein the reverse transcription will result in the copying of the upstream and downstream sequences from the template nucleic acid (e.g., template RNA) without the intervening sequence, e.g., causing deletion of the intervening sequence. In other embodiments, the template nucleic acid (e.g., template RNA) may be designed to write an edit into the target DNA. For example, the template RNA may match the target DNA sequence with the exception of one or more nucleotides, wherein the reverse transcription will result in the copying of these edits into the target DNA, e.g., resulting in mutations, e.g., transition or transversion mutations.

In some embodiments, the pre-edit homology domain comprises a nucleic acid sequence having 100% sequence identity with a nucleic acid sequence comprised in a target nucleic acid molecule.

In some embodiments, the post-edit homology domain comprises a nucleic acid sequence having 100% sequence identity with a nucleic acid sequence comprised in a target nucleic acid molecule.

In some embodiments, a homology domain (e.g., a pre-edit homology domain) comprises the nucleic acid sequence of a homology 1 sequence as listed in Table 38 below, or a nucleic acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a homology domain (e.g., a pre-edit homology domain) comprises the nucleic acid sequence of a homology 1 sequence as listed in Table 38 below, or a nucleic acid sequence having no more than 1, 2, 3, 4, or 5 nucleotide differences relative thereto. In some embodiments, a homology domain has a length of 0-30 nucleotides (e.g., about 0-10, 10-20, or 20-30 nucleotides). Herein, when an RNA sequence (e.g., a homology domain sequence) is said to comprise a particular sequence (e.g., a sequence of Table 38 or a portion thereof) that comprises thymine (T), it is of course understood that the RNA sequence may (and frequently does) comprise uracil (U) in place of T. For instance, the RNA sequence may comprise U at every position shown as T in the sequence in Table 38. More specifically, the present disclosure provides an RNA sequence according to every homology domain sequence of Table 38, wherein the RNA sequence has a U in place of each T in the sequence in Table 38. In certain embodiments, the homology domain has a length between 0-5, 5-10, 10-15, 15-20, 20-25, 25-30, 30-35, 35-40, 40-45, or 45-50 nucleotides. In certain embodiments, the homology domain has a length between

50-100, 100-150, 150-200, 200-250, 250-300, 300-350, 350-400, 400-450, or 450-550 nucleotides. In certain embodiments, the homology domain has a length of about 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, or 500 nucleotides.

5

Table 38. Exemplary homology 1 sequences

eporter	Edit type	Edit Sequence (5' to 3')	Homology 1 length	Homology 1 Sequence (5' to 3')
BFP to GFP	SNP	GT	3 nt	ACG
	SNP	GT	3 nt	ACG
	SNP	GT	3 nt	ACG
	SNP	GT	3 nt	ACG
	SNP	GT	3 nt	ACG
	SNP	GT	3 nt	ACG
	SNP	GT	3 nt	ACG
	SNP	GT	3 nt	ACG
	SNP	GT	3 nt	ACG
250 bp GFP insertion	250 bp insertion	AGAATTTTGTAATACGACTCACTATAGGGCGGCCGGGAATTCGC CACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGT GCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAG TTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCA AGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTG CCCTGGCCCACCCTCGTGACCACCCTGACGTACG	0 nt	
	250 bp insertion	CAGAATTTTGTAATACGACTCACTATAGGGCGGCCGGGAATTCG CCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGT GCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAG TTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCA AGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTG CCCTGGCCCACCCTCGTGACCACCCTGACGTAC	1 nt	G
	250 bp insertion	TCAGAATTTTGTAATACGACTCACTATAGGGCGGCCGGGAATTC GCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTG GTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACA AGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGG	2 nt	CG

		CAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCG TGCCCTGGCCCACCCTCGTGACCACCCTGACGTA		
	250 bp insert ion	GTCAGAATTTTGTAATACGACTCACTATAGGGCGGCCGGGAATT CGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTG GTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACA AGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGG CAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCG TGCCCTGGCCCACCCTCGTGACCACCCTGACGT	3 nt	ACG
	250 bp insert ion	CGTCAGAATTTTGTAATACGACTCACTATAGGGCGGCCGGGAAT TCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGT GGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCAC AAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACG GCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCC GTGCCCTGGCCCACCCTCGTGACCACCCTGACG	4 nt	TACG
	250 bp insert ion	CCGTCAGAATTTTGTAATACGACTCACTATAGGGCGGCCGGGAA TTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGG TGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCA CAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTAC GGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCC CGTGCCCTGGCCCACCCTCGTGACCACCCTGAC	5 nt	GTACG
	250 bp insert ion	ACCGTCAGAATTTTGTAATACGACTCACTATAGGGCGGCCGGGA ATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGG GTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCC ACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTA CGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGC CCGTGCCCTGGCCCACCCTCGTGACCACCCTGA	6 nt	CGTACG
	250 bp insert ion	AACCGTCAGAATTTTGTAATACGACTCACTATAGGGCGGCCGGG AATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGG GGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGC CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCT ACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTG CCCGTGCCCTGGCCCACCCTCGTGACCACCCTG	7 nt	ACGTACG

	250 bp insertion	GAACCGTCAGAATTTTGTAATACGACTCACTATAGGGCGGCCGG GAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCG GGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGG CCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACC TACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCT GCCCGTGCCCTGGCCCACCCTCGTGACCACCCT	8 nt	GACGTACG
	250 bp insertion	TGAACCGTCAGAATTTTGTAATACGACTCACTATAGGGCGGCCG GGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACC GGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACG GCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCAC CTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGC TGCCCGTGCCCTGGCCCACCCTCGTGACCACCC	9 nt	TGACGTACG
	250 bp insertion	GTGAACCGTCAGAATTTTGTAATACGACTCACTATAGGGCGGCC GGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCAC CGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAAC GGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCA CCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAG CTGCCCGTGCCCTGGCCCACCCTCGTGACCACC	10 nt	CTGACGTACG
	250 bp insertion	AGTGAACCGTCAGAATTTTGTAATACGACTCACTATAGGGCGGC CGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCA CCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAA CGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCC ACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAA GCTGCCCGTGCCCTGGCCCACCCTCGTGACCAC	11 nt	CCTGACGTACG
	250 bp insertion	TAGTGAACCGTCAGAATTTTGTAATACGACTCACTATAGGGCGG CCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTC ACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAA ACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGC CACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCA AGCTGCCCGTGCCCTGGCCCACCCTCGTGACCA	12 nt	CCCTGACGTACG
	250 bp	TTAGTGAACCGTCAGAATTTTGTAATACGACTCACTATAGGGCG GCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTT CACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA	13 nt	ACCCTGACGTACG

	insert ion	AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGAT GCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGG CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACC		
	250 bp insert ion	TTTAGTGAACCGTCAGAATTTTGTAATACGACTCACTATAGGGC GGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGT TCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGT AAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGAT GCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGG CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGAC	14 nt	CACCCTGACGTACG
	250 bp insert ion	GTTTAGTGAACCGTCAGAATTTTGTAATACGACTCACTATAGGG CGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCT GTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGAC GTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGC GATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCAC CGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGA	15 nt	CCACCCTGACGTACG
	250 bp insert ion	CGTTTAGTGAACCGTCAGAATTTTGTAATACGACTCACTATAGG GCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGC TGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGA CGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGC GATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCAC CGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTG	16 nt	ACCACCCTGACGTACG
	250 bp insert ion	TCGTTTAGTGAACCGTCAGAATTTTGTAATACGACTCACTATAGG GCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGC TGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGA CGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGC GATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCAC CGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGT	17 nt	GACCACCCTGACGTACG
	250 bp insert ion	CTCGTTTAGTGAACCGTCAGAATTTTGTAATACGACTCACTATAG GGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAG CTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCG ACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGG CGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCA CCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCG	18 nt	TGACCACCCTGACGTACG

	250 bp insertion	GCTCGTTTAGTGAACCGTCAGAATTTTGTAATACGACTCACTATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTC	19 nt	GTGACCACCCTGACGTACG
	250 bp insertion	AGCTCGTTTAGTGAACCGTCAGAATTTTGTAATACGACTCACTATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCT	20 nt	CGTGACCACCCTGACGTACG
mCherry insertion	>750 bp insertion	[mCherry-expressing cassette]	0 nt	
	>750 bp insertion	[mCherry-expressing cassette]	1 nt	G
	>750 bp insertion	[mCherry-expressing cassette]	2 nt	CG
	>750 bp insertion	[mCherry-expressing cassette]	3 nt	ACG
	>750 bp insertion	[mCherry-expressing cassette]	4 nt	TACG

	>750 bp insertion	[mCherry-expressing cassette]	5 nt	GTACG
	>750 bp insertion	[mCherry-expressing cassette]	6 nt	CGTACG
	>750 bp insertion	[mCherry-expressing cassette]	7 nt	ACGTACG
	>750 bp insertion	[mCherry-expressing cassette]	8 nt	GACGTACG
	>750 bp insertion	[mCherry-expressing cassette]	9 nt	TGACGTACG
	>750 bp insertion	[mCherry-expressing cassette]	10 nt	CTGACGTACG
	>750 bp insertion	[mCherry-expressing cassette]	11 nt	CCTGACGTACG
	>750 bp insertion	[mCherry-expressing cassette]	12 nt	CCCTGACGTACG

	>750 bp insertion	[mCherry-expressing cassette]	13 nt	ACCCTGACGTACG
	>750 bp insertion	[mCherry-expressing cassette]	14 nt	CACCCTGACGTACG
	>750 bp insertion	[mCherry-expressing cassette]	15 nt	CCACCCTGACGTACG
	>750 bp insertion	[mCherry-expressing cassette]	16 nt	ACCACCCTGACGTACG
	>750 bp insertion	[mCherry-expressing cassette]	17 nt	GACCACCCTGACGTACG
	>750 bp insertion	[mCherry-expressing cassette]	18 nt	TGACCACCCTGACGTACG
	>750 bp insertion	[mCherry-expressing cassette]	19 nt	GTGACCACCCTGACGTACG
	>750 bp insertion	[mCherry-expressing cassette]	20 nt	CGTGACCACCCTGACGTACG

In some embodiments, a homology domain (e.g., a pre-edit homology domain) comprises the nucleic acid sequence of a homology 2 sequence as listed in Table 39 below, or a nucleic acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, a homology domain (e.g., a pre-edit homology domain) comprises the nucleic acid sequence of a homology 2 sequence as listed in Table 39 below, or a nucleic acid sequence having no more than 1, 2, 3, 4, or 5 nucleotide differences relative thereto. In some embodiments, a homology domain has a length of 0-1000 nucleotides (e.g., about 0-5, 5-10, 10-50, 50-100, 100-500, or 500-1000 nucleotides). Herein, when an RNA sequence (e.g., a homology domain sequence) is said to comprise a particular sequence (e.g., a sequence of Table 39 or a portion thereof) that comprises thymine (T), it is of course understood that the RNA sequence may (and frequently does) comprise uracil (U) in place of T. For instance, the RNA sequence may comprise U at every position shown as T in the sequence in Table 39. More specifically, the present disclosure provides an RNA sequence according to every homology domain sequence of Table 39, wherein the RNA sequence has a U in place of each T in the sequence in Table 39.

Table 39. Exemplary homology 2 sequences

Reporter	Homology 2 length	Homology 2 Sequence (5' to 3')	Homology 1 pair
BFP to GFP	8 nt	ACCCTGAC	
	11 nt	ACCACCCTGAC	
	12 nt	GACCACCCTGAC	
	13 nt	TGACCACCCTGAC	
	14 nt	GTGACCACCCTGAC	
	16 nt	TCGTGACCACCCTGAC	
	20 nt	ACCCTCGTGACCACCCTGAC	
	24 nt	GCCCACCCTCGTGACCACCCTGAC	
	25 nt	GGCCCACCCTCGTGACCACCCTGAC	
250 bp GFP insertion	500 nt	CCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAAC GCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTAC ATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATT ATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTAC CATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAA GTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTC GTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAG AGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	499 nt	CCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACG CCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACA TCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTA TGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACC ATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAA GTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTC GTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAG AGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	498 nt	CGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGC CAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACAT CAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTAT GCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCA TGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAG TCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCG TAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGA GCTCGTTTAGTGAACCGTC	0 nt Homology 1
	497 nt	GCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCC AATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC AAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTAT GCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCA TGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAG TCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCG TAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGA GCTCGTTTAGTGAACCGTC	0 nt Homology 1
	496 nt	CCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCA ATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCA AGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGC CCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATG GTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCT CCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAA TAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCT CGTTTAGTGAACCGTC	0 nt Homology 1
	495 nt	CTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAA TAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAA GTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCC CAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATG GTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCT CCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAA TAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCT CGTTTAGTGAACCGTC	0 nt Homology 1
	494 nt	TGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAAT AGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAG	0 nt Homology 1

		TGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCC AGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGT GATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCC ACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATA ACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCG TTTAGTGAACCGTC	
	493 nt	GGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAAT AGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAG TGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCC AGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGT GATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCC ACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATA ACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCG TTTAGTGAACCGTC	0 nt Homology 1
	492 nt	GCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATA GGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGT GTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCA GTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGT GATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCC ACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATA ACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCG TTTAGTGAACCGTC	0 nt Homology 1
	491 nt	CTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAG GGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTG TATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAG TACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGA TGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCAC CCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAAC CCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTT TAGTGAACCGTC	0 nt Homology 1
	490 nt	TGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGG GACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGT ATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGT ACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGAT	0 nt Homology 1

		GCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACC CCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACC CCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTT AGTGAACCGTC	
	489 nt	GACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGG ACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTAT CATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTAC ATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATG CGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCC CATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCC CGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTA GTGAACCGTC	0 nt Homology 1
	488 nt	ACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGA CTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATC ATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTAC ATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATG CGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCC CATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCC CGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTA GTGAACCGTC	0 nt Homology 1
	487 nt	CCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGAC TTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCA TATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACAT GACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCG GTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCA TTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCG CCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGT GAACCGTC	0 nt Homology 1
	486 nt	CGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTT TCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATA TGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATG ACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGG TTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATT GACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGC	0 nt Homology 1

		CCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTG AACCGTC	
	485 nt	GCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTT CCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATAT GCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGA CCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGT TTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATT GACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGC CCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTG AACCGTC	0 nt Homology 1
	484 nt	CCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTC CATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATAT GCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGA CCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGT TTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATT GACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGC CCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTG AACCGTC	0 nt Homology 1
	483 nt	CCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC ATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATG CCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGAC CTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTT TGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGA CGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCC GTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAAC CGTC	0 nt Homology 1
	482 nt	CAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCA TTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGC CAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCT TACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTT GGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGAC GTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCC GTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAAC CGTC	0 nt Homology 1

	481 nt	AACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCAT TGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCC AAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTT ACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTG GCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACG TCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGT TGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCG TC	0 nt Homology 1
	480 nt	ACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATT GACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCA AGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTA CGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGG CAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGT CAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTT GACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGT C	0 nt Homology 1
	479 nt	CGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTG ACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAA GTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTAC GGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGC AGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTC AATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTG ACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	478 nt	GACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGA CGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAG TCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACG GGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCA GTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCA ATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGA CGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	477 nt	ACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGAC GTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGT CCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACG GGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCA	0 nt Homology 1

		GTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCA ATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGA CGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	
	476 nt	CCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACG TCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTC CGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGG GACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAG TACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAA TGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGAC GCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	475 nt	CCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGT CAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCC GCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGG ACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGT ACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAAT GGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGAC GCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	474 nt	CCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTC AATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCG CCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGA CTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTAC ACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGG GAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCA AATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	473 nt	CCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCA ATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGC CCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGAC TTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTAC ACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGG GAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCA AATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	472 nt	CGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAA TGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCC CCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTT	0 nt Homology 1

		TCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACAC CAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGA GTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAA TGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	
	471 nt	GCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAAT GGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCC CCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTT CCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACC AATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAG TTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAAT GGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	470 nt	CCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATG GGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCC CTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTC CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCA ATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGT TTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATG GGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	469 nt	CCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGG GTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCT ATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCT ACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAAT GGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTT GTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGG GCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	468 nt	CATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGG TGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTA TTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTA CTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATG GGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGG CGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	467 nt	ATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGT GGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTAT	0 nt Homology 1

		TGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTAC TTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGG GCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGT TTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGC GGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	
	466 nt	TTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTG GAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATT GACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACT TGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGG CGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTT TGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGG TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	465 nt	TGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTG GAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATT GACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACT TGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGG CGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTT TGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGG TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	464 nt	GACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGG AGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTG ACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTT GGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGG CGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTT TGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGG TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	463 nt	ACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGA GTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGA CGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTG GCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGC GTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTT GGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGT AGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	462 nt	CGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAG TATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGAC GTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGG CAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGT GGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTG GCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTA GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	461 nt	GTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGT ATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACG TCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGC AGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTG GATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC ACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGG CGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	460 nt	TCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTA TTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGT CAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCA GTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTG GATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC ACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGG CGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	459 nt	CAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTAT TTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTC AATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCA GTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTG GATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC ACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGG CGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	458 nt	AATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTT ACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAA TGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTA CATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATA GCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCA	0 nt Homology 1

		AAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTG TACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	
	457 nt	ATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTA CGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAAT GACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTA CATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATA GCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCA AAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTG TACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	456 nt	TAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTAC GGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATG ACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACA TCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAG CGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAA AATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGT ACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	455 nt	AATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTAC GGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATG ACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACA TCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAG CGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAA AATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGT ACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	454 nt	ATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACG GTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGA CGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACAT CTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGC GGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAA ATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTA CGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	453 nt	TGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGG TAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGAC GGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATC TACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCG	0 nt Homology 1

		GTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAA TCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTAC GGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	
	452 nt	GACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGT AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACG GTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCT ACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGG TTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAAT CAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACG GTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	451 nt	ACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTA AACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGG TAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTA CGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGT TTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATC AACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACG GTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	450 nt	CGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAA ACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGT AAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTAC GTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTT TGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCA ACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGT GGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	449 nt	GTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAA CTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTA AATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACG TATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTT GACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAA CGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGT GGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	448 nt	TATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACT GCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAA TGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTA	0 nt Homology 1

		TTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGA CTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACG GGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGG GAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	
	447 nt	ATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACT GCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAA TGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTA TTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGA CTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACG GGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGG GAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	446 nt	TGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTG CCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAAT GGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTAT TAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGAC TCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGG GACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGG AGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	445 nt	GTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGC CCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATG GCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATT AGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACT CACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGG GACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGG AGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	444 nt	TTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCC ACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGG CCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTA GTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTC ACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGG ACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGA GGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	443 nt	TCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCC ACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGG	0 nt Homology 1

		CCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTA GTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTC ACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGG ACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGA GGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	
	442 nt	CCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCA CTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCC CGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGT CATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCAC GGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGAC TTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGG TCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	441 nt	CCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCAC TTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCC GCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTC ATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACG GGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTT TCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTC TATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	440 nt	CATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACT TGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCC GCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTC ATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACG GGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTT TCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTC TATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	439 nt	ATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTT GGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCG CCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCAT CGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGG GATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTC CAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTA TATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	438 nt	TAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTG GCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGC CTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATC GCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGG GATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTC CAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTA TATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	437 nt	AGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGG CAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCT GGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCG CTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGA TTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCA AAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATA TAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	436 nt	GTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGC AGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCT GGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCG CTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGA TTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCA AAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATA TAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	435 nt	TAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCA GTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTG GCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCT ATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATT TCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAA ATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATA AGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	434 nt	AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAG TACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGC ATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTAT TACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTC CAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAAT	0 nt Homology 1

		GTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAG CAGAGCTCGTTTAGTGAACCGTC	
	433 nt	ACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGT ACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGC ATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTAT TACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTC CAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAAT GTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAG CAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	432 nt	CGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTA CATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCAT TATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTA CCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCA AGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGT CGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCA GAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	431 nt	GCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTAC ATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATT ATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTAC CATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAA GTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTC GTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAG AGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	430 nt	CCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACA TCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTA TGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACC ATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAA GTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTC GTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAG AGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	429 nt	CAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACAT CAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTAT GCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCA TGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAG	0 nt Homology 1

		TCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCG TAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGA GCTCGTTTAGTGAACCGTC	
	428 nt	AATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATC AAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTAT GCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCA TGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAG TCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCG TAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGA GCTCGTTTAGTGAACCGTC	0 nt Homology 1
	427 nt	ATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCA AGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGC CCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATG GTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCT CCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAA TAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCT CGTTTAGTGAACCGTC	0 nt Homology 1
	426 nt	TAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAA GTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCC CAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATG GTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCT CCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAA TAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCT CGTTTAGTGAACCGTC	0 nt Homology 1
	425 nt	AGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAG TGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCC AGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGT GATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCC ACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATA ACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCG TTTAGTGAACCGTC	0 nt Homology 1
	424 nt	GGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGT GTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCA GTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGT	0 nt Homology 1

		GATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCC ACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATA ACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCG TTTAGTGAACCGTC	
	423 nt	GGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTG TATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAG TACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGA TGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCAC CCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAAC CCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTT TAGTGAACCGTC	0 nt Homology 1
	422 nt	GACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGT ATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGT ACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGAT GCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACC CCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACC CCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTT AGTGAACCGTC	0 nt Homology 1
	421 nt	ACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTAT CATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTAC ATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATG CGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCC CATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCC CGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTA GTGAACCGTC	0 nt Homology 1
	420 nt	CTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATC ATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTAC ATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATG CGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCC CATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCC CGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTA GTGAACCGTC	0 nt Homology 1
	419 nt	TTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCA TATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACAT	0 nt Homology 1

		GACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCG GTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCA TTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCG CCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGT GAACCGTC	
	418 nt	TTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCAT ATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACAT GACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCG GTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCA TTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCG CCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGT GAACCGTC	0 nt Homology 1
	417 nt	TCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATA TGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATG ACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGG TTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATT GACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGC CCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTG AACCGTC	0 nt Homology 1
	416 nt	CCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATAT GCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGA CCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGT TTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATT GACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGC CCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTG AACCGTC	0 nt Homology 1
	415 nt	CATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATAT GCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGA CCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGT TTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATT GACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGC CCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTG AACCGTC	0 nt Homology 1

	414 nt	ATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATG CCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGAC CTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTT TGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGA CGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCC GTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAAC CGTC	0 nt Homology 1
	413 nt	TTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGC CAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCT TACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTT GGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGAC GTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCC GTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAAC CGTC	0 nt Homology 1
	412 nt	TGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCC AAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTT ACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTG GCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACG TCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGT TGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCG TC	0 nt Homology 1
	411 nt	GACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCA AGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTA CGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGG CAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGT CAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTT GACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGT C	0 nt Homology 1
	410 nt	ACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAA GTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTAC GGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGC AGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTC AATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTG ACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	409 nt	CGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAG TCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACG GGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCA GTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCA ATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGA CGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	408 nt	GTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGT CCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACG GGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCA GTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCA ATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGA CGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	407 nt	TCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTC CGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGG GACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAG TACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAA TGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGAC GCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	406 nt	CAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCC GCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGG ACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGT ACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAAT GGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGAC GCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	405 nt	AATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCG CCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGA CTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTAC ACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGG GAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCA AATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	404 nt	ATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGC CCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGAC TTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTAC ACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGG	0 nt Homology 1

		GAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCA AATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	
	403 nt	TGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCC CCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTT TCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACAC CAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGA GTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAA TGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	402 nt	GGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCC CCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTT CCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACC AATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAG TTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAAT GGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	401 nt	GGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCC CTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTC CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCA ATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGT TTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATG GGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	400 nt	GTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCT ATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCT ACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAAT GGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTT GTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGG GCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	399 nt	TGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTA TTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTA CTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATG GGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGG CGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	398 nt	GGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTAT TGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTAC	0 nt Homology 1

		TTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGG GCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGT TTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGC GGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	
	397 nt	GAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATT GACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACT TGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGG CGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTT TGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGG TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	396 nt	AGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTG ACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTT GGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGG CGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTT TGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGG TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	395 nt	GTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGA CGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTG GCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGC GTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTT GGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGT AGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	394 nt	TATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGAC GTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGG CAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGT GGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTG GCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTA GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	393 nt	ATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACG TCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGC AGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTG GATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC ACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGG CGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	392 nt	TTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGT CAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCA GTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTG GATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC ACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGG CGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	391 nt	TTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTC AATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCA GTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTG GATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC ACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGG CGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	390 nt	TACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCA ATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGT ACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGAT AGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACC AAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGT GTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	389 nt	ACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAA TGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTA CATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATA GCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCA AAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTG TACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	388 nt	CGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAAT GACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTA CATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATA GCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCA AAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTG TACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	387 nt	GGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATG ACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACA TCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAG CGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAA	0 nt Homology 1

		AATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGT ACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	
	386 nt	GTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGA CGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACAT CTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGC GGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAA ATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTA CGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	385 nt	TAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGAC GGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATC TACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCG GTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAA TCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTAC GGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	384 nt	AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACG GTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCT ACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGG TTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAAT CAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACG GTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	383 nt	AACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGG TAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTA CGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGT TTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATC AACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACG GTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	382 nt	ACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGT AAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTAC GTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTT TGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCA ACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGT GGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	381 nt	CTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTA AATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACG	0 nt Homology 1

		TATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTT GACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAA CGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGT GGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	
	380 nt	TGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAA ATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGT ATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTG ACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAAC GGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTG GGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	379 nt	GCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAA TGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTA TTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGA CTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACG GGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGG GAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	378 nt	CCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAAT GGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTAT TAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGAC TCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGG GACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGG AGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	377 nt	CCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATG GCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATT AGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACT CACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGG GACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGG AGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	376 nt	CACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATG GCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATT AGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACT CACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGG GACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGG AGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	375 nt	ACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGG CCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTA GTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTC ACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGG ACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGA GGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	374 nt	CTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCC CGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGT CATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCAC GGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGAC TTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGG TCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	373 nt	TTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCC GCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTC ATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACG GGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTT TCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTC TATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	372 nt	TGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCC GCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTC ATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACG GGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTT TCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTC TATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	371 nt	GGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCG CCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCAT CGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGG GATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTC CAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTA TATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	370 nt	GCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGC CTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATC GCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGG GATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTC	0 nt Homology 1

		CAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTA TATAAGCAGAGCTCGTTTAGTGAACCGTC	
	369 nt	CAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCT GGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCG CTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGA TTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCA AAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATA TAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	368 nt	AGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCT GGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCG CTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGA TTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCA AAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATA TAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	367 nt	GTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTG GCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCT ATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATT TCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAA ATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATA AGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	366 nt	TACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGC ATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTAT TACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTC CAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAAT GTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAG CAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	365 nt	ACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGC ATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTAT TACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTC CAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAAT GTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAG CAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	364 nt	CATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCAT TATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTA	0 nt Homology 1

		CCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCA AGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGT CGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCA GAGCTCGTTTAGTGAACCGTC	
	363 nt	ATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATT ATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTAC CATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAA GTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTC GTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAG AGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	362 nt	TCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTA TGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACC ATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAA GTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTC GTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAG AGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	361 nt	CAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTAT GCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCA TGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAG TCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCG TAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGA GCTCGTTTAGTGAACCGTC	0 nt Homology 1
	360 nt	AAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTAT GCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCA TGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAG TCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCG TAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGA GCTCGTTTAGTGAACCGTC	0 nt Homology 1
	359 nt	AGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGC CCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATG GTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCT CCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAA TAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCT CGTTTAGTGAACCGTC	0 nt Homology 1

	358 nt	GTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCC CAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATG GTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCT CCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAA TAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCT CGTTTAGTGAACCGTC	0 nt Homology 1
	357 nt	TGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCC AGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGT GATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCC ACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATA ACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCG TTTAGTGAACCGTC	0 nt Homology 1
	356 nt	GTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCA GTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGT GATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCC ACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATA ACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCG TTTAGTGAACCGTC	0 nt Homology 1
	355 nt	TATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAG TACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGA TGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCAC CCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAAC CCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTT TAGTGAACCGTC	0 nt Homology 1
	354 nt	ATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGT ACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGAT GCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACC CCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACC CCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTT AGTGAACCGTC	0 nt Homology 1
	353 nt	TCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTA CATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGAT GCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACC CCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACC	0 nt Homology 1

		CCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	
	352 nt	CATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	351 nt	ATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	350 nt	TATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	349 nt	ATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	348 nt	TGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	347 nt	GCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGT	0 nt Homology 1

		TTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATT GACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGC CCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTG AACCGTC	
	346 nt	CCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGAC CTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTT TGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGA CGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCC GTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAAC CGTC	0 nt Homology 1
	345 nt	CAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCT TACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTT GGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGAC GTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCC GTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAAC CGTC	0 nt Homology 1
	344 nt	AAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTT ACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTG GCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACG TCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGT TGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCG TC	0 nt Homology 1
	343 nt	AGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTA CGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGG CAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGT CAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTT GACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGT C	0 nt Homology 1
	342 nt	GTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTAC GGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGC AGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTC AATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTG ACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	341 nt	TCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACG GGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCA GTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCA ATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGA CGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	340 nt	CCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACG GGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCA GTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCA ATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGA CGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	339 nt	CGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGG GACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAG TACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAA TGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGAC GCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	338 nt	GCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGG ACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGT ACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAAT GGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGAC GCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	337 nt	CCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGA CTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTAC ACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGG GAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCA AATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	336 nt	CCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGAC TTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTAC ACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGG GAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCA AATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	335 nt	CCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTT TCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACAC CAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGA	0 nt Homology 1

		GTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAA TGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	
	334 nt	CCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTT CCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACC AATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAG TTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAAT GGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	333 nt	CTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTC CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCA ATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGT TTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATG GGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	332 nt	TATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCC TACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAA TGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTT TGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATG GGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	331 nt	ATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCT ACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAAT GGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTT GTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGG GCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	330 nt	TTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTA CTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATG GGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGG CGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	329 nt	TGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTAC TTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGG GCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGT TTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGC GGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	328 nt	GACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACT TGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGG	0 nt Homology 1

		CGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTT TGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGG TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	
	327 nt	ACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTT GGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGG CGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTT TGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGG TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	326 nt	CGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTG GCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGC GTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTT GGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGT AGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	325 nt	GTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGG CAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGT GGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTG GCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTA GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	324 nt	TCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGC AGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTG GATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC ACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGG CGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	323 nt	CAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCA GTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTG GATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC ACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGG CGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	322 nt	AATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCA GTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTG GATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC ACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGG CGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	321 nt	ATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGT ACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGAT AGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACC AAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGT GTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	320 nt	TGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTA CATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATA GCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCA AAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTG TACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	319 nt	GACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTA CATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATA GCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCA AAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTG TACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	318 nt	ACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACA TCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAG CGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAA AATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGT ACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	317 nt	CGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACAT CTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGC GGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAA ATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTA CGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	316 nt	GGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATC TACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCG GTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAA TCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTAC GGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	315 nt	GTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCT ACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGG TTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAAT	0 nt Homology 1

		CAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACG GTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	
	314 nt	TAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTA CGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGT TTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATC AACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACG GTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	313 nt	AAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTAC GTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTT TGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCA ACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGT GGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	312 nt	AATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACG TATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTT GACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAA CGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGT GGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	311 nt	ATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGT ATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTG ACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAAC GGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTG GGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	310 nt	TGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTA TTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGA CTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACG GGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGG GAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	309 nt	GGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTAT TAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGAC TCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGG GACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGG AGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	308 nt	GCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATT AGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACT	0 nt Homology 1

		CACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGG GACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGG AGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	
	307 nt	CCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTA GTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTC ACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGG ACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGA GGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	306 nt	CCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAG TCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCA CGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGA CTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAG GTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	305 nt	CGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGT CATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCAC GGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGAC TTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGG TCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	304 nt	GCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTC ATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACG GGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTT TCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTC TATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	303 nt	CCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCAT CGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGG GATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTC CAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTA TATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	302 nt	CTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATC GCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGG GATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTC CAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTA TATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	301 nt	TGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATC GCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGG GATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTC CAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTA TATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	300 nt	GGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCG CTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGA TTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCA AAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATA TAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	299 nt	GCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCT ATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATT TCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAA ATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATA AGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	298 nt	CATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTA TTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTT CCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAA TGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAA GCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	297 nt	ATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTAT TACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTC CAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAAT GTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAG CAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	296 nt	TTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATT ACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCC AAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATG TCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGC AGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	295 nt	TATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTA CCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCA AGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGT	0 nt Homology 1

		CGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCA GAGCTCGTTTAGTGAACCGTC	
	294 nt	ATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTAC CATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAA GTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTC GTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAG AGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	293 nt	TGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACC ATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAA GTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTC GTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAG AGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	292 nt	GCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCA TGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAG TCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCG TAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGA GCTCGTTTAGTGAACCGTC	0 nt Homology 1
	291 nt	CCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCAT GGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGT CTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGT AATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAG CTCGTTTAGTGAACCGTC	0 nt Homology 1
	290 nt	CCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATG GTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCT CCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAA TAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCT CGTTTAGTGAACCGTC	0 nt Homology 1
	289 nt	CAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATG GTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCT CCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAA TAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCT CGTTTAGTGAACCGTC	0 nt Homology 1
	288 nt	AGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGT GATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCC	0 nt Homology 1

		ACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATA ACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCG TTTAGTGAACCGTC	
	287 nt	GTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGT GATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCC ACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATA ACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCG TTTAGTGAACCGTC	0 nt Homology 1
	286 nt	TACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGA TGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCAC CCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAAC CCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTT TAGTGAACCGTC	0 nt Homology 1
	285 nt	ACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGAT GCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACC CCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACC CCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTT AGTGAACCGTC	0 nt Homology 1
	284 nt	CATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGAT GCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACC CCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACC CCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTT AGTGAACCGTC	0 nt Homology 1
	283 nt	ATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATG CGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCC CATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCC CGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTA GTGAACCGTC	0 nt Homology 1
	282 nt	TGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGC GGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCC ATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCC GCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAG TGAACCGTC	0 nt Homology 1

	281 nt	GACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCG GTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCA TTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCG CCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGT GAACCGTC	0 nt Homology 1
	280 nt	ACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGG TTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATT GACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGC CCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTG AACCGTC	0 nt Homology 1
	279 nt	CCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGT TTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATT GACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGC CCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTG AACCGTC	0 nt Homology 1
	278 nt	CTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTT TGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGA CGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCC GTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAAC CGTC	0 nt Homology 1
	277 nt	TTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTT GGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGAC GTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCC GTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAAC CGTC	0 nt Homology 1
	276 nt	TACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTT GGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGAC GTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCC GTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAAC CGTC	0 nt Homology 1
	275 nt	ACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTG GCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACG TCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGT	0 nt Homology 1

		TGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCG TC	
	274 nt	CGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGG CAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGT CAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTT GACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGT C	0 nt Homology 1
	273 nt	GGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGC AGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTC AATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTG ACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	272 nt	GGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCA GTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCA ATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGA CGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	271 nt	GACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAG TACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAA TGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGAC GCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	270 nt	ACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGT ACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAAT GGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGAC GCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	269 nt	CTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTAC ACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGG GAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCA AATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	268 nt	TTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTAC ACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGG GAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCA AATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	267 nt	TTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACA CCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGG	0 nt Homology 1

		AGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAA ATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	
	266 nt	TCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACAC CAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGA GTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAA TGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	265 nt	CCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACC AATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAG TTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAAT GGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	264 nt	CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCA ATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGT TTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATG GGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	263 nt	TACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAA TGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTT TGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATG GGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	262 nt	ACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAAT GGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTT GTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGG GCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	261 nt	CTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATG GGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGG CGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	260 nt	TTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGG GCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGT TTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGC GGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	259 nt	TGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGG CGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTT TGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGG TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	258 nt	GGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGG CGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTT TGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGG TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	257 nt	GCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGC GTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTT GGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGT AGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	256 nt	CAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGT GGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTG GCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTA GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	255 nt	AGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTG GATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC ACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGG CGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	254 nt	GTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTG GATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC ACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGG CGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	253 nt	TACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGA TAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCAC CAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCG TGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	252 nt	ACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGAT AGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACC AAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGT GTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	251 nt	CATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATA GCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCA AAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTG TACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	250 nt	ATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATA GCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCA	0 nt Homology 1

		AAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTG TACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	
	249 nt	TCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAG CGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAA AATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGT ACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	248 nt	CTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGC GGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAA ATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTA CGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	247 nt	TACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCG GTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAA TCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTAC GGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	246 nt	ACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGG TTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAAT CAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACG GTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	245 nt	CGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGT TTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATC AACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACG GTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	244 nt	GTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTT TGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCA ACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGT GGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	243 nt	TATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTT GACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAA CGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGT GGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	242 nt	ATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTG ACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAAC GGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTG GGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	241 nt	TTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	240 nt	TAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	239 nt	AGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	238 nt	GTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	237 nt	TCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	236 nt	CATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	235 nt	ATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	234 nt	TCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	233 nt	CGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTC	0 nt Homology 1

		CAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTA TATAAGCAGAGCTCGTTTAGTGAACCGTC	
	232 nt	GCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGG GATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTC CAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTA TATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	231 nt	CTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGA TTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCA AAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATA TAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	230 nt	TATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGAT TTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAA AATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATAT AAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	229 nt	ATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATT TCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAA ATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATA AGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	228 nt	TTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTT CCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAA TGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAA GCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	227 nt	TACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTC CAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAAT GTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAG CAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	226 nt	ACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCC AAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATG TCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGC AGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	225 nt	CCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCA AGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGT CGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCA GAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	224 nt	CATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	223 nt	ATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	222 nt	TGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	221 nt	GGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	220 nt	GTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	219 nt	TGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	218 nt	GATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	217 nt	ATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	216 nt	TGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAAC	0 nt Homology 1

		CCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTT TAGTGAACCGTC	
	215 nt	GCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACC CCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACC CCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTT AGTGAACCGTC	0 nt Homology 1
	214 nt	CGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCC CATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCC CGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTA GTGAACCGTC	0 nt Homology 1
	213 nt	GGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCC ATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCC GCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAG TGAACCGTC	0 nt Homology 1
	212 nt	GTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCA TTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCG CCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGT GAACCGTC	0 nt Homology 1
	211 nt	TTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATT GACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGC CCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTG AACCGTC	0 nt Homology 1
	210 nt	TTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATT GACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGC CCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTG AACCGTC	0 nt Homology 1
	209 nt	TTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTG ACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCC CGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAA CCGTC	0 nt Homology 1
	208 nt	TGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGA CGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCC GTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAAC CGTC	0 nt Homology 1

	207 nt	GGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGAC GTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCC GTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAAC CGTC	0 nt Homology 1
	206 nt	GCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACG TCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGT TGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCG TC	0 nt Homology 1
	205 nt	CAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGT CAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTT GACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGT C	0 nt Homology 1
	204 nt	AGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTC AATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTG ACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	203 nt	GTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCA ATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGA CGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	202 nt	TACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAA TGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGAC GCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	201 nt	ACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAAT GGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGAC GCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	200 nt	CACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATG GGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGC AAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	199 nt	ACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGG GAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCA AATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	198 nt	CCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGG AGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAA ATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	197 nt	CAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	196 nt	AATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	195 nt	ATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	194 nt	TGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	193 nt	GGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	192 nt	GGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	191 nt	GCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	190 nt	CGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	189 nt	GTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	188 nt	TGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	187 nt	GGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	186 nt	GATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	185 nt	ATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	184 nt	TAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	183 nt	AGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	182 nt	GCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	181 nt	CGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	180 nt	GGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	179 nt	GTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	178 nt	TTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	177 nt	TTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	176 nt	TGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	175 nt	GACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	174 nt	ACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	173 nt	CTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	172 nt	TCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	171 nt	CACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	170 nt	ACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	169 nt	CGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	168 nt	GGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	167 nt	GGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	166 nt	GGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	165 nt	GATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	164 nt	ATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	163 nt	TTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	162 nt	TTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	161 nt	TCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	160 nt	CCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	159 nt	CAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	158 nt	AAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	157 nt	AGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	156 nt	GTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	155 nt	TCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	154 nt	CTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	153 nt	TCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	152 nt	CCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	151 nt	CACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	150 nt	ACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	149 nt	CCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	148 nt	CCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	147 nt	CCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	146 nt	CATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	145 nt	ATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	144 nt	TTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	143 nt	TGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	142 nt	GACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	141 nt	ACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	140 nt	CGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	139 nt	GTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	138 nt	TCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	137 nt	CAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	136 nt	AATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	135 nt	ATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	134 nt	TGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	133 nt	GGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	132 nt	GGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	131 nt	GAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	130 nt	AGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	129 nt	GTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	128 nt	TTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	127 nt	TTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	126 nt	TGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	125 nt	GTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	124 nt	TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	123 nt	TTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	122 nt	TTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	121 nt	TGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	120 nt	GGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	119 nt	GCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	118 nt	CACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	117 nt	ACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	116 nt	CCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	115 nt	CAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	114 nt	AAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	113 nt	AAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	112 nt	AATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	111 nt	ATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	110 nt	TCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	109 nt	CAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	108 nt	AACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	107 nt	ACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	106 nt	CGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	105 nt	GGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	104 nt	GGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	103 nt	GACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	102 nt	ACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	101 nt	CTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	100 nt	TTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	99 nt	TTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	98 nt	TCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	97 nt	CCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	96 nt	CAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	95 nt	AAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	94 nt	AAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATA TAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	93 nt	AATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATAT AAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	92 nt	ATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATA AGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	91 nt	TGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAA GCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	90 nt	GTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAG CAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	89 nt	TCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGC AGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	88 nt	CGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCA GAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	87 nt	GTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAG AGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	86 nt	TAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGA GCTCGTTTAGTGAACCGTC	0 nt Homology 1
	85 nt	AATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAG CTCGTTTAGTGAACCGTC	0 nt Homology 1
	84 nt	ATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGC TCGTTTAGTGAACCGTC	0 nt Homology 1
	83 nt	TAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCT CGTTTAGTGAACCGTC	0 nt Homology 1
	82 nt	AACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTC GTTTAGTGAACCGTC	0 nt Homology 1
	81 nt	ACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCG TTTAGTGAACCGTC	0 nt Homology 1
	80 nt	CCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGT TTAGTGAACCGTC	0 nt Homology 1
	79 nt	CCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTT TAGTGAACCGTC	0 nt Homology 1
	78 nt	CCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTT AGTGAACCGTC	0 nt Homology 1

	77 nt	CGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	76 nt	GCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	75 nt	CCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	74 nt	CCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	73 nt	CCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	72 nt	CGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	71 nt	GTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	70 nt	TTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	69 nt	TGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	68 nt	GACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	67 nt	ACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	66 nt	CGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	65 nt	GCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	64 nt	CAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	63 nt	AAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	62 nt	AATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	61 nt	ATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	60 nt	TGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	59 nt	GGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	58 nt	GGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	57 nt	GCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	56 nt	CGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	55 nt	GGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	54 nt	GTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	53 nt	TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	52 nt	AGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	51 nt	GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	50 nt	GCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	49 nt	CGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	48 nt	GTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	47 nt	TGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	46 nt	GTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	45 nt	TACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	44 nt	ACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	43 nt	CGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	42 nt	GGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	41 nt	GTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	40 nt	TGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	39 nt	GGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	38 nt	GGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	37 nt	GAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	36 nt	AGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	35 nt	GGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	34 nt	GTCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	33 nt	TCTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	32 nt	CTATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	31 nt	TATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	30 nt	ATATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	29 nt	TATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1

	28 nt	ATAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	27 nt	TAAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	26 nt	AAGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	25 nt	AGCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	24 nt	GCAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	23 nt	CAGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	22 nt	AGAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	21 nt	GAGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	20 nt	AGCTCGTTTAGTGAACCGTC	0 nt Homology 1
	19 nt	GCTCGTTTAGTGAACCGTC	0 nt Homology 1
	18 nt	CTCGTTTAGTGAACCGTC	0 nt Homology 1
	17 nt	TCGTTTAGTGAACCGTC	0 nt Homology 1
	16 nt	CGTTTAGTGAACCGTC	0 nt Homology 1
	15 nt	GTTTAGTGAACCGTC	0 nt Homology 1
	14 nt	TTTAGTGAACCGTC	0 nt Homology 1
	13 nt	TTAGTGAACCGTC	0 nt Homology 1
	12 nt	TAGTGAACCGTC	0 nt Homology 1
	11 nt	AGTGAACCGTC	0 nt Homology 1
	10 nt	GTGAACCGTC	0 nt Homology 1
	9 nt	TGAACCGTC	0 nt Homology 1
	8 nt	GAACCGTC	0 nt Homology 1
	7 nt	AACCGTC	0 nt Homology 1
	6 nt	ACCGTC	0 nt Homology 1
	5 nt	CCGTC	0 nt Homology 1
	4 nt	CGTC	0 nt Homology 1
	3 nt	GTC	0 nt Homology 1
	2 nt	TC	0 nt Homology 1
	1 nt	C	0 nt Homology 1

	0 nt		0 nt Homology 1
	(0-500 nt)	GCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAA CGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTA CATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCAT TATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTA CCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCA AGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGT CGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCA GAGCTCGTTTAGTGAACCGT	1 nt Homology 1
	(0-500 nt)	GGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAG TACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGC ATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTAT TACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTC CAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAAT GTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAG CAGAGCTCGTTTAGTGAACCG	2 nt Homology 1
	(0-500 nt)	TGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAG TACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGC ATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTAT TACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTC CAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAAT GTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAG CAGAGCTCGTTTAGTGAACC	3 nt Homology 1
	(0-500 nt)	ATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATA GTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGC AGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCT GGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCG CTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGA TTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCA AAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATA TAAGCAGAGCTCGTTTAGTGAAC	4 nt Homology 1

	(0-500 nt)	AATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATA GTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGC AGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCT GGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCG CTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGA TTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCA AAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATA TAAGCAGAGCTCGTTTAGTGAA	5 nt Homology 1
	(0-500 nt)	AAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCAT AGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGG CAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCT GGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCG CTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGA TTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCA AAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATA TAAGCAGAGCTCGTTTAGTGA	6 nt Homology 1
	(0-500 nt)	TAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCA TAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTG GCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGC CTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATC GCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGG GATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTC CAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTA TATAAGCAGAGCTCGTTTAGTG	7 nt Homology 1
	(0-500 nt)	GTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCC CATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACT TGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCC GCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTC ATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACG GGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTT TCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTC TATATAAGCAGAGCTCGTTTAGT	8 nt Homology 1
	(0-500 nt)	GGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTC CCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCAC	9 nt Homology 1

		TTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCC GCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTC ATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACG GGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTT TCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTC TATATAAGCAGAGCTCGTTTAG	
	(0-500 nt)	CGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTT CCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCA CTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCC CGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGT CATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCAC GGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGAC TTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGG TCTATATAAGCAGAGCTCGTTTA	10 nt Homology 1
	(0-500 nt)	ACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGT TCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCC ACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGG CCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTA GTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTC ACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGG ACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGA GGTCTATATAAGCAGAGCTCGTTT	11 nt Homology 1
	(0-500 nt)	TACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATG TTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCC ACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGG CCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTA GTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTC ACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGG ACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGA GGTCTATATAAGCAGAGCTCGTT	12 nt Homology 1
	(0-500 nt)	TTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTAT GTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGC CCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATG GCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATT	13 nt Homology 1

		AGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACT CACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGG GACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGG AGGTCTATATAAGCAGAGCTCGT	
	(0-500 nt)	CTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTA TGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTG CCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAAT GGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTAT TAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGAC TCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGG GACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGG AGGTCTATATAAGCAGAGCTCG	14 nt Homology 1
	(0-500 nt)	ACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGT ATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACT GCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAA TGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTA TTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGA CTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACG GGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGG GAGGTCTATATAAGCAGAGCTC	15 nt Homology 1
	(0-500 nt)	AACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACG TATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACT GCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAA TGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTA TTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGA CTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACG GGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGG GAGGTCTATATAAGCAGAGCT	16 nt Homology 1
	(0-500 nt)	TAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGAC GTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAA CTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTA AATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACG TATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTT GACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAA	17 nt Homology 1

		CGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGT GGGAGGTCTATATAAGCAGAGC	
	(0-500 nt)	ATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGA CGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAA ACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGT AAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTAC GTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTT TGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCA ACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGT GGGAGGTCTATATAAGCAGAG	18 nt Homology 1
	(0-500 nt)	CATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATG ACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTA AACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGG TAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTA CGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGT TTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATC AACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACG GTGGGAGGTCTATATAAGCAGA	19 nt Homology 1
	(0-500 nt)	ACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAAT GACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGT AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGACG GTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCT ACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGG TTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAAT CAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACG GTGGGAGGTCTATATAAGCAG	20 nt Homology 1
mCherry insertion	(0-500 nt)	ATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATA GCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCA AAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTG TACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAATTTTGTAATACGACTCACT ATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCC CATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGG	0 nt Homology 1

		CGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGC CCACCCTCGTGACCACCCTGACGTACG	
	(0-500 nt)	CATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATA GCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCA AAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTG TACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAATTTTGTAATACGACTCACT ATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCC CATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGG CGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGC CCACCCTCGTGACCACCCTGACGTAC	1 nt Homology 1
	(5-500 nt)	ACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGAT AGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACC AAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGT GTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAATTTTGTAATACGACTCAC TATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGC CCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGG GCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGG CCCACCCTCGTGACCACCCTGACGTA	2 nt Homology 1
	(5-500 nt)	TACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGA TAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCAC CAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCG TGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAATTTTGTAATACGACTCA CTATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTG CCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAG GGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTG GCCCACCCTCGTGACCACCCTGACGT	3 nt Homology 1
	(5-500 nt)	GTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTG GATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC ACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGG CGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAATTTTGTAATACGACT CACTATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGG TGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGA GGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCT GGCCCACCCTCGTGACCACCCTGACG	4 nt Homology 1

	(5-500 nt)	AGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGTG GATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGC ACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGG CGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAATTTTGTAATACGACT CACTATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGG TGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGA GGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCT GGCCCACCCTCGTGACCACCCTGAC	5 nt Homology 1
	(5-500 nt)	CAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCGT GGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTG GCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTA GGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAATTTTGTAATACGA CTCACTATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGT GGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGC GAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCC CTGGCCCACCCTCGTGACCACCCTGA	6 nt Homology 1
	(5-500 nt)	GCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGC GTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTT GGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGT AGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAATTTTGTAATACG ACTCACTATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGT GGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGC GAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCC CTGGCCCACCCTCGTGACCACCCTG	7 nt Homology 1
	(5-500 nt)	GGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGG CGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTT TGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGG TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAATTTTGTAATAC GACTCACTATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGG GTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGG GCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGT GCCCTGGCCCACCCTCGTGACCACCCT	8 nt Homology 1
	(5-500 nt)	TGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGG CGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTT	9 nt Homology 1

		TGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGG TAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAATTTTGTAATAC GACTCACTATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGG GTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGG GCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGT GCCCTGGCCCACCCTCGTGACCACCC	
	(5-500 nt)	TTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGG GCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGT TTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGC GGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAATTTTGTAAT ACGACTCACTATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGG GGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAG GGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCG TGCCCTGGCCCACCCTCGTGACCACC	10 nt Homology 1
	(5-500 nt)	CTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATG GGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTG TTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGG CGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAATTTTGTAA TACGACTCACTATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCG GGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGA GGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCC GTGCCCTGGCCCACCCTCGTGACCAC	11 nt Homology 1
	(5-500 nt)	ACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAAT GGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTT GTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGG GCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAATTTTGT AATACGACTCACTATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCAC CGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGC GAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGC CCGTGCCCTGGCCCACCCTCGTGACCA	12 nt Homology 1
	(5-500 nt)	TACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAA TGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTT TGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATG GGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAATTTTG	13 nt Homology 1

		TAATACGACTCACTATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCA CCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGG CGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTG CCCGTGCCCTGGCCCACCCTCGTGACC	
	(5-500 nt)	CTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCA ATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGT TTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATG GGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAATTTTG TAATACGACTCACTATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCA CCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGG CGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTG CCCGTGCCCTGGCCCACCCTCGTGAC	14 nt Homology 1
	(5-500 nt)	CCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACC AATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAG TTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAAT GGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAATTTT GTAATACGACTCACTATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTC ACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCG GCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCT GCCCGTGCCCTGGCCCACCCTCGTGA	15 nt Homology 1
	(5-500 nt)	TCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACAC CAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGA GTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAA TGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAATTT TGTAATACGACTCACTATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTT CACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCC GGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGC TGCCCGTGCCCTGGCCCACCCTCGTG	16 nt Homology 1
	(5-500 nt)	TTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACA CCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGG AGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAA ATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAATT TTGTAATACGACTCACTATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTGT TCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTC	17 nt Homology 1

		CGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAG CTGCCCGTGCCCTGGCCCACCCTCGT	
	(5-500 nt)	TTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTAC ACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGG GAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCA AATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAAT TTTGTAATACGACTCACTATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTG TTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGT CCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAA GCTGCCCGTGCCCTGGCCCACCCTCG	18 nt Homology 1
	(5-500 nt)	CTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTAC ACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGG GAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCA AATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAAT TTTGTAATACGACTCACTATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAGCTG TTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGT CCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAA GCTGCCCGTGCCCTGGCCCACCCTC	19 nt Homology 1
	(5-500 nt)	ACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGT ACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAAT GGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGAC GCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAG AATTTTGTAATACGACTCACTATAGGGCGGCCGGGAATTCGCCACCATGGTGAGCAAGGGCGAGGAG CTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCG TGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGG CAAGCTGCCCGTGCCCTGGCCCACCCT	20 nt Homology 1

PBS sequence

In some embodiments, a template nucleic acid (e.g., template RNA) comprises a PBS sequence. In some embodiments, a PBS sequence is disposed 3′ of the heterologous object sequence and is complementary to a sequence adjacent to a site to be modified by a system described herein, or comprises no more than 1, 2, 3, 4, or 5 mismatches to a sequence complementary to the sequence adjacent to a site to be modified by the system/gene modifying polypeptide. In some embodiments, the PBS sequence binds within 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides of a nick site in the target nucleic acid molecule. In some embodiments, binding of the PBS sequence to the target nucleic acid molecule permits initiation of target-primed reverse transcription (TPRT), e.g., with the 3′ homology domain acting as a primer for TPRT. In some embodiments, the PBS sequence is 3-5, 5-10, 10-30, 10-25, 10-20, 10-19, 10-18, 10-17, 10-16, 10-15, 10-14, 10-13, 10-12, 10-11, 11-30, 11-25, 11-20, 11-19, 11-18, 11-17, 11-16, 11-15, 11-14, 11-13, 11-12, 12-30, 12-25, 12-20, 12-19, 12-18, 12-17, 12-16, 12-15, 12-14, 12-13, 13-30, 13-25, 13-20, 13-19, 13-18, 13-17, 13-16, 13-15, 13-14, 14-30, 14-25, 14-20, 14-19, 14-18, 14-17, 14-16, 14-15, 15-30, 15-25, 15-20, 15-19, 15-18, 15-17, 15-16, 16-30, 16-25, 16-20, 16-19, 16-18, 16-17, 17-30, 17-25, 17-20, 17-19, 17-18, 18-30, 18-25, 18-20, 18-19, 19-30, 19-25, 19-20, 20-30, 20-25, or 25-30 nucleotides in length, e.g., 10-17, 12-16, or 12-14 nucleotides in length. In some embodiments, the PBS sequence is 5-20, 8-16, 8-14, 8-13, 9-13, 9-12, or 10-12 nucleotides in length, e.g., 9-12 nucleotides in length.

The template nucleic acid (e.g., template RNA) may have some homology to the target DNA. In some embodiments, the template nucleic acid (e.g., template RNA) PBS sequence domain may serve as an annealing region to the target DNA, such that the target DNA is positioned to prime the reverse transcription of the template nucleic acid (e.g., template RNA). In some embodiments the template nucleic acid (e.g., template RNA) has at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 175, 200 or more bases of exact homology to the target DNA at the 3′ end of the RNA. In some embodiments the template nucleic acid (e.g., template RNA) has at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 175, 200 or more bases of at least 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 98%, 99% or 100% homology to the target DNA, e.g., at the 5′ end of the template nucleic acid (e.g., template RNA).

In some embodiments, a PBS sequence comprises a nucleic acid sequence as listed in Table 37, or a nucleic acid sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, or 99% identity thereto. In some embodiments, a PBS sequence comprises a nucleic acid sequence as listed in Table 37, or a nucleic acid sequence having no more than 1, 2, 3, 4, or 5 nucleotide differences thereto. Herein, when an RNA sequence (e.g., in a PBS sequence) is said to comprise a particular sequence (e.g., a sequence of Table 37 or a portion thereof) that comprises thymine (T), it is of course understood that the RNA sequence may (and frequently does) comprise uracil (U) in place of T. For instance, the RNA sequence may comprise U at every position shown as T in the sequence in Table 37. More specifically, the present disclosure provides an RNA sequence according to every PBS sequence of Table 37, wherein the RNA sequence has a U in place of each T in the sequence in Table 37. In certain embodiments, the PBS has a length between 1-3, 3-5, 5-8, 8-10, 10-12, 12-15, 15-17, 17-20, 20-25, 25-30, or 30-40 nucleotides. In certain embodiments, the PBS has a length of about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides.

Table 37. Exemplary PBS sequences. The below sequences are listed 5' to 3' direction.

PBS length	Sequence (5' to 3')
1 nt	G
2 nt	GC
3 nt	GCG
4 nt	GCGT
5 nt	GCGTG
6 nt	GCGTGC
7 nt	GCGTGCA
8 nt	GCGTGCAG
9 nt	GCGTGCAGT
10 nt	GCGTGCAGTG
11 nt	GCGTGCAGTGC
12 nt	GCGTGCAGTGCT
13 nt	GCGTGCAGTGCTT
14 nt	GCGTGCAGTGCTTC
15 nt	GCGTGCAGTGCTTCG
16 nt	GCGTGCAGTGCTTCGG
17 nt	GCGTGCAGTGCTTCGGC

gRNAs with inducible activity

In some embodiments, a gRNA described herein (e.g., a gRNA that is part of a template RNA or a gRNA used for second strand nicking) has inducible activity. Inducible activity may be achieved by the template nucleic acid, e.g., template RNA, further comprising (in addition to the gRNA) a blocking domain, wherein the sequence of a portion of or all of the blocking domain is at least partially complementary to a portion or all of the gRNA. The blocking domain is thus capable of hybridizing or substantially hybridizing to a portion of or all of the gRNA. In some embodiments, the blocking domain and inducibly active gRNA are disposed on the template nucleic acid, e.g., template RNA, such that the gRNA can adopt a first conformation where the blocking domain is hybridized or substantially hybridized to the gRNA, and a second conformation where the blocking domain is not hybridized or not substantially hybridized to the gRNA. In some embodiments, in the first conformation the gRNA is unable to bind to the gene modifying polypeptide (e.g., the template nucleic acid binding domain, DNA binding domain, or endonuclease domain (e.g., a CRISPR/Cas protein)) or binds with substantially decreased affinity compared to an otherwise similar template RNA lacking the blocking domain. In some embodiments, in the second conformation the gRNA is able to bind to the gene modifying polypeptide (e.g., the template nucleic acid binding domain, DNA binding domain, or endonuclease domain (e.g., a CRISPR/Cas protein)). In some embodiments, whether the gRNA is in the first or second conformation can influence whether the DNA binding or endonuclease activities of the gene modifying polypeptide (e.g., of the CRISPR/Cas protein the gene modifying polypeptide comprises) are active.

In some embodiments, the gRNA that coordinates the second nick has inducible activity. In some embodiments, the gRNA that coordinates the second nick is induced after the template is reverse transcribed. In some embodiments, hybridization of the gRNA to the blocking domain can be disrupted using an opener molecule. In some embodiments, an opener molecule comprises an agent that binds to a portion or all of the gRNA or blocking domain and inhibits hybridization of the gRNA to the blocking domain. In some embodiments, the opener molecule comprises a nucleic acid, e.g., comprising a sequence that is partially or wholly complementary to the gRNA, blocking domain, or both. By choosing or designing an appropriate opener molecule, providing the opener molecule can promote a change in the conformation of the gRNA such that it can associate with a CRISPR/Cas protein and provide the associated functions of the CRISPR/Cas protein (e.g., DNA binding and/or endonuclease activity). Without wishing to be bound by theory, providing the opener molecule at a selected time and/or location may allow for spatial and temporal control of the activity of the gRNA, CRISPR/Cas protein, or gene modifying system comprising the same. In some embodiments, the opener molecule is exogenous to the cell comprising the gene modifying polypeptide and or template nucleic acid. In some embodiments, the opener molecule comprises an endogenous agent (e.g., endogenous to the cell comprising the gene

modifying polypeptide and or template nucleic acid comprising the gRNA and blocking domain). For example, an inducible gRNA, blocking domain, and opener molecule may be chosen such that the opener molecule is an endogenous agent expressed in a target cell or tissue, e.g., thereby ensuring activity of a gene modifying system in the target cell or tissue. As a further example, an inducible gRNA, blocking domain, and opener molecule may be chosen such that the opener molecule is absent or not substantially expressed in one or more non-target cells or tissues, e.g., thereby ensuring that activity of a gene modifying system does not occur or substantially occur in the one or more non-target cells or tissues, or occurs at a reduced level compared to a target cell or tissue. Exemplary blocking domains, opener molecules, and uses thereof are described in PCT App. Publication WO2020044039A1, which is incorporated herein by reference in its entirety. In some embodiments, the template nucleic acid, e.g., template RNA, may comprise one or more sequences or structures for binding by one or more components of a gene modifying polypeptide, e.g., by a reverse transcriptase or RNA binding domain, and a gRNA. In some embodiments, the gRNA facilitates interaction with the template nucleic acid binding domain (e.g., RNA binding domain) of the gene modifying polypeptide. In some embodiments, the gRNA directs the gene modifying polypeptide to the matching target sequence, e.g., in a target cell genome.

Additional Guide RNAs

In some embodiments, a gene modifying system as described herein comprises an additional guide RNA (gRNA), e.g., for unwinding or nicking of the target nucleic acid (e.g., nicking of the opposite strand of that recognized by a PBS sequence of the template RNA). In embodiments, a gene modifying system as described herein comprises 1, 2, 3, 4, or 5 additional distinct gRNAs. In certain embodiments, the additional guide RNA is a separate molecule from the template RNA. In other embodiments, the additional guide RNA is attached to or incorporated into the template RNA (e.g., at the 3' end of the template RNA). In embodiments, the additional guide RNA is attached to the remainder of the template RNA by a linker region. In some embodiments, the additional guide RNA comprises a stem-loop sequence, e.g., as noted in Table 44.

In some embodiments, the additional gRNA comprises a pro-spacer, e.g., attached to an end block sequence of the template RNA (e.g., as described herein). In some embodiments, the pro-spacer is 17 nucleotides or longer (e.g., at least about 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, or 40 nucleotides). In embodiments, the pro-spacer directs nicking of the target nucleic acid by a Cas domain (e.g., a Cas9 domain, e.g., an nCas9 domain). In other embodiments, the pro-spacer is less than or equal to 17 nucleotides in length (e.g., about 5, 10, 11, 12, 13, 14, 15, 16, or 17 nucleotides long). In embodiments, the pro-spacer less than or equal to 17 nucleotides long directs unwinding of the target nucleic acid (e.g.,

by a Cas domain, e.g., a Cas9 domain, e.g., a dCas9 domain) but does not direct nicking of the target nucleic acid by a Cas domain (e.g., a Cas9 domain, e.g., an nCas9 domain).

In some embodiments, the additional gRNA comprises a scaffold sequence as listed in Table 44, or a nucleic acid sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some embodiments, the additional gRNA comprises a scaffold sequence as listed in Table 44, or a nucleic acid sequence having no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, or 50 nucleotide differences therefrom.

Herein, when an RNA sequence (e.g., a gRNA sequence) is said to comprise a particular sequence (e.g., a sequence of Table 44 or a portion thereof) that comprises thymine (T), it is of course understood that the RNA sequence may (and frequently does) comprise uracil (U) in place of T. For instance, the RNA sequence may comprise U at every position shown as T in the sequence in Table 44. More specifically, the present disclosure provides an RNA sequence according to every gRNA sequence of Table 44, wherein the RNA sequence has a U in place of each T in the sequence in Table 44.

Table 44. Exemplary additional guide RNA (gRNA) sequences

Cas9 variant	Pro-spacer sequence (5' to 3')	Scaffold Sequence (5' to 3')
Nme2Cas9 scaffold		GTTGTAGCTCCCTTTCTCATTTCGGAAACGAAATGAGAACC GTTGCTACAATAAGGCCGTCTGAAAAGATGTGCCGCAACG CTCTGCCCCTTAAAGCTTCTGCTTTAAGGGGCATCGTTTA
Nme2Cas9 **spacer**+scaffold	**CAGTACATGACCTTACGGGA**	GTTGTAGCTCCCTTTCTCATTTCGGAAACGAAATGAGAACC GTTGCTACAATAAGGCCGTCTGAAAAGATGTGCCGCAACG CTCTGCCCCTTAAAGCTTCTGCTTTAAGGGGCATCGTTTA
Nme2Cas9 **16 nt spacer**+scaffold	**ACATGACCTTACGGGA**	GTTGTAGCTCCCTTTCTCATTTCGGAAACGAAATGAGAACC GTTGCTACAATAAGGCCGTCTGAAAAGATGTGCCGCAACG CTCTGCCCCTTAAAGCTTCTGCTTTAAGGGGCATCGTTTA
BlatCas9 scaffold		GCTATAGTTCCTTACTGAAAGGTAAGTTGCTATAGTAAGGG CAACAGACCCGAGGCGTTGGGGATCGCCTAGCCCGTGTTT ACGGGCTCTCCCCATATTCAAAATAATGACAGACGAGCACC TTGGAGCATTTATCTCCGAGGTGCT
GeoCas9		GTCATAGTTCCCCTGAGAAATCAGGGTTACTATGATAAGGG CTTTCTGCCTAAGGCAGACTGACCCGCGGCGTTGGGGATC GCCTGTCGCCCGCTTTTGGCGGGCATTCCCCATCCTT
PpnCas9 scaffold		GTTGTAGCTCCCTTTTTCATTTCGCGAAAGCGAAATGAAAA ACGTTGTTACAATAAGAGATGAATTTCTCGCAAAGCTCTGC CTCTTGAAATTTCGGTTTCAAGAGGCATC

CdiCas9scaffold		ACTGGGGTTCAGGAAACTGAACCTCAGTAAGCATTGGCTCGTTTCCAATGTTGATTGCTCCGCCGGTGCTCCTTATTTTTAAGGGCGCCGGC
SpyCas9+**hairpin** scaffold		GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCG**GGACTTCGGTCCCGGTGC**
St1Cas9 scaffold		GTCTTTGTACTCTGGTACCAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT
cCas9-v16 scaffold		GTCTTAGTACTCTGGAAACAGAATCTACTAAGACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
SpyCas9-3var-NRRH scaffold		GTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SauCas9 scaffold		GTTTTAGTACTCTGGAAACAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
CjeCas9 scaffold		GTTTTAGTCCCTGAAAAGGGACTAAAATAAAGAGTTTGCGGGACTCTGCGGGGTTACAATCCCCTAAAACCGC
SpyCas9 scaffold		GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC

Circular RNAs and Ribozymes in Gene Modifying Systems

It is contemplated that it may be useful to employ circular and/or linear RNA states during the formulation, delivery, or gene modifying reaction within the target cell. Thus, in some embodiments of any of the aspects described herein, a gene modifying system comprises one or more circular RNAs (circRNAs). In some embodiments of any of the aspects described herein, a gene modifying system comprises one or more linear RNAs. In some embodiments, a nucleic acid as described herein (e.g., a template nucleic acid, a nucleic acid molecule encoding a gene modifying polypeptide, or both) is a circRNA. In some embodiments, a circular RNA molecule encodes the gene modifying polypeptide. In some embodiments, the circRNA molecule encoding the gene modifying polypeptide is delivered to a host cell. In some embodiments, a circular RNA molecule encodes a recombinase, e.g., as described herein. In some embodiments, the circRNA molecule encoding the recombinase is delivered to a host cell. In some embodiments, the circRNA molecule encoding the gene modifying polypeptide is linearized (e.g., in the host cell, e.g., in the nucleus of the host cell) prior to translation.

Circular RNAs (circRNAs) have been found to occur naturally in cells and have been found to have diverse functions, including both non-coding and protein coding roles in human cells. It has been shown that a circRNA can be engineered by incorporating a self-splicing intron into an RNA molecule (or

DNA encoding the RNA molecule) that results in circularization of the RNA, and that an engineered circRNA can have enhanced protein production and stability (Wesselhoeft et al. *Nature Communications* 2018). In some embodiments, the gene modifying polypeptide is encoded as circRNA. In certain embodiments, the template nucleic acid is a DNA, such as a dsDNA or ssDNA. In certain embodiments, the circDNA comprises a template RNA.

In some embodiments, the circRNA comprises one or more ribozyme sequences. In some embodiments, the ribozyme sequence is activated for autocleavage, e.g., in a host cell, e.g., thereby resulting in linearization of the circRNA. In some embodiments, the ribozyme is activated when the concentration of magnesium reaches a sufficient level for cleavage, e.g., in a host cell. In some embodiments the circRNA is maintained in a low magnesium environment prior to delivery to the host cell. In some embodiments, the ribozyme is a protein-responsive ribozyme. In some embodiments, the ribozyme is a nucleic acid-responsive ribozyme. In some embodiments, the circRNA comprises a cleavage site. In some embodiments, the circRNA comprises a second cleavage site.

In some embodiments, the circRNA is linearized in the nucleus of a target cell. In some embodiments, linearization of a circRNA in the nucleus of a cell involves components present in the nucleus of the cell, e.g., to activate a cleavage event. In some embodiments, a ribozyme, e.g., a ribozyme from a B2 or ALU element, that is responsive to a nuclear element, e.g., a nuclear protein, e.g., a genome-interacting protein, e.g., an epigenetic modifier, e.g., EZH2, is incorporated into a circRNA, e.g., of a gene modifying system. In some embodiments, nuclear localization of the circRNA results in an increase in autocatalytic activity of the ribozyme and linearization of the circRNA.

In some embodiments, the ribozyme is heterologous to one or more of the other components of the gene modifying system. In some embodiments, an inducible ribozyme (e.g., in a circRNA as described herein) is created synthetically, for example, by utilizing a protein ligand-responsive aptamer design. A system for utilizing the satellite RNA of tobacco ringspot virus hammerhead ribozyme with an MS2 coat protein aptamer has been described (Kennedy et al. Nucleic Acids Res 42(19):12306-12321 (2014), incorporated herein by reference in its entirety) that results in activation of the ribozyme activity in the presence of the MS2 coat protein. In embodiments, such a system responds to protein ligand localized to the cytoplasm or the nucleus. In some embodiments the protein ligand is not MS2. Methods for generating RNA aptamers to target ligands have been described, for example, based on the systematic evolution of ligands by exponential enrichment (SELEX) (Tuerk and Gold, Science 249(4968):505-510 (1990); Ellington and Szostak, Nature 346(6287):818-822 (1990); the methods of each of which are incorporated herein by reference) and have, in some instances, been aided by *in silico* design (Bell et al. PNAS 117(15):8486-8493, the methods of which are incorporated herein by reference). Thus, in some

embodiments, an aptamer for a target ligand is generated and incorporated into a synthetic ribozyme system, e.g., to trigger ribozyme-mediated cleavage and circRNA linearization, e.g., in the presence of the protein ligand. In some embodiments, circRNA linearization is triggered in the cytoplasm, e.g., using an aptamer that associates with a ligand in the cytoplasm. In some embodiments, circRNA linearization is triggered in the nucleus, e.g., using an aptamer that associates with a ligand in the nucleus. In embodiments, the ligand in the nucleus comprises an epigenetic modifier or a transcription factor. In some embodiments the ligand that triggers linearization is present at higher levels in on-target cells than off-target cells.

It is further contemplated that a nucleic acid-responsive ribozyme system can be employed for circRNA linearization. For example, biosensors that sense defined target nucleic acid molecules to trigger ribozyme activation are described, e.g., in Penchovsky (Biotechnology Advances 32(5):1015-1027 (2014), incorporated herein by reference). By these methods, a ribozyme naturally folds into an inactive state and is only activated in the presence of a defined target nucleic acid molecule (e.g., an RNA molecule). In some embodiments, a circRNA of a gene modifying system comprises a nucleic acid-responsive ribozyme that is activated in the presence of a defined target nucleic acid, e.g., an RNA, e.g., an mRNA, miRNA, guide RNA, gRNA, sgRNA, ncRNA, lncRNA, tRNA, snRNA, or mtRNA. In some embodiments the nucleic acid that triggers linearization is present at higher levels in on-target cells than off-target cells.

In some embodiments of any of the aspects herein, a gene modifying system incorporates one or more ribozymes with inducible specificity to a target tissue or target cell of interest, e.g., a ribozyme that is activated by a ligand or nucleic acid present at higher levels in a target tissue or target cell of interest. In some embodiments, the gene modifying system incorporates a ribozyme with inducible specificity to a subcellular compartment, e.g., the nucleus, nucleolus, cytoplasm, or mitochondria. In some embodiments, the ribozyme that is activated by a ligand or nucleic acid present at higher levels in the target subcellular compartment. In some embodiments, an RNA component of a gene modifying system is provided as circRNA, e.g., that is activated by linearization. In some embodiments, linearization of a circRNA encoding a gene modifying polypeptide activates the molecule for translation. In some embodiments, a signal that activates a circRNA component of a gene modifying system is present at higher levels in on-target cells or tissues, e.g., such that the system is specifically activated in these cells.

In some embodiments, an RNA component of a gene modifying system is provided as a circRNA that is inactivated by linearization. In some embodiments, a circRNA encoding the gene modifying polypeptide is inactivated by cleavage and degradation. In some embodiments, a circRNA encoding the gene modifying polypeptide is inactivated by cleavage that separates a translation signal from the coding

sequence of the polypeptide. In some embodiments, a signal that inactivates a circRNA component of a gene modifying system is present at higher levels in off-target cells or tissues, such that the system is specifically inactivated in these cells.

Target Nucleic Acid Site

In some embodiments, after gene modification, the target site surrounding the edited sequence contains a limited number of insertions or deletions, for example, in less than about 50% or 10% of editing events, e.g., as determined by long-read amplicon sequencing of the target site, e.g., as described in Karst et al. (2020) *bioRxiv* doi.org/10.1101/645903 (incorporated by reference herein in its entirety). In some embodiments, the target site does not show multiple consecutive editing events, e.g., head-to-tail or head-to-head duplications, e.g., as determined by long-read amplicon sequencing of the target site, e.g., as described in Karst et al. *bioRxiv* doi.org/10.1101/645903 (2020) (incorporated herein by reference in its entirety). In some embodiments, the target site contains an integrated sequence corresponding to the template RNA. In some embodiments, the target site does not contain insertions resulting from endogenous RNA in more than about 1% or 10% of events, e.g., as determined by long-read amplicon sequencing of the target site, e.g., as described in Karst et al. *bioRxiv* doi.org/10.1101/645903 (2020) (incorporated herein by reference in its entirety). In some embodiments, the target site contains the integrated sequence corresponding to the template RNA.

In certain aspects of the present invention, the host DNA-binding site integrated into by the gene modifying system can be in a gene, in an intron, in an exon, an ORF, outside of a coding region of any gene, in a regulatory region of a gene, or outside of a regulatory region of a gene. In other aspects, the polypeptide may bind to one or more than one host DNA sequence.

In some embodiments, a gene modifying system is used to edit a target locus in multiple alleles. In some embodiments, a gene modifying system is designed to edit a specific allele. For example, a gene modifying polypeptide may be directed to a specific sequence that is only present on one allele, e.g., comprises a template RNA with homology to a target allele, e.g., a gRNA or annealing domain, but not to a second cognate allele. In some embodiments, a gene modifying system can alter a haplotype-specific allele. In some embodiments, a gene modifying system that targets a specific allele preferentially targets that allele, e.g., has at least a 2, 4, 6, 8, or 10-fold preference for a target allele.

Second Strand Nicking

In some embodiments, a gene modifying system described herein comprises a nickase activity (e.g., in the gene modifying polypeptide) that nicks the first strand, and a nickase activity (e.g., in a polypeptide separate from the gene modifying polypeptide) that nicks the second strand of target DNA. As discussed herein, without wishing to be bound by theory, nicking of the first strand of the target site

DNA is thought to provide a 3′ OH that can be used by an RT domain to reverse transcribe a sequence of a template RNA, e.g., a heterologous object sequence. Without wishing to be bound by theory, it is thought that introducing an additional nick to the second strand may bias the cellular DNA repair machinery to adopt the heterologous object sequence-based sequence more frequently than the original genomic sequence. In some embodiments, the additional nick to the second strand is made by the same endonuclease domain (e.g., nickase domain) as the nick to the first strand. In some embodiments, the same gene modifying polypeptide performs both the nick to the first strand and the nick to the second strand. In some embodiments, the gene modifying polypeptide comprises a CRISPR/Cas domain and the additional nick to the second strand is directed by an additional nucleic acid, e.g., comprising a second gRNA directing the CRISPR/Cas domain to nick the second strand. In other embodiments, the additional second strand nick is made by a different endonuclease domain (e.g., nickase domain) than the nick to the first strand. In some embodiments, that different endonuclease domain is situated in an additional polypeptide (e.g., a system of the invention further comprises the additional polypeptide), separate from the gene modifying polypeptide. In some embodiments, the additional polypeptide comprises an endonuclease domain (e.g., nickase domain) described herein. In some embodiments, the additional polypeptide comprises a DNA binding domain, e.g., described herein.

It is contemplated herein that the position at which the second strand nick occurs relative to the first strand nick may influence the extent to which one or more of: desired gene modifying DNA modifications are obtained, undesired double-strand breaks (DSBs) occur, undesired insertions occur, or undesired deletions occur. Without wishing to be bound by theory, second strand nicking may occur in two general orientations: inward nicks and outward nicks.

In some embodiments, in the inward nick orientation, the RT domain polymerizes (e.g., using the template RNA (e.g., the heterologous object sequence)) away from the second strand nick. In some embodiments, in the inward nick orientation, the location of the nick to the first strand and the location of the nick to the second strand are positioned between the first PAM site and second PAM site (e.g., in a scenario wherein both nicks are made by a polypeptide (e.g., a gene modifying polypeptide) comprising a CRISPR/Cas domain). When there are two PAMs on the outside and two nicks on the inside, this inward nick orientation can also be referred to as “PAM-out”. In some embodiments, in the inward nick orientation, the location of the nick to the first strand and the location of the nick to the second strand are between the sites where the polypeptide and the additional polypeptide bind to the target DNA. In some embodiments, in the inward nick orientation, the location of the nick to the second strand is positioned between the binding sites of the polypeptide and additional polypeptide, and the nick to the first strand is also located between the binding sites of the polypeptide and additional polypeptide. In some embodiments, in the inward nick orientation, the location of the nick to the first strand and the location of

the nick to the second strand are positioned between the PAM site and the binding site of the second polypeptide which is at a distance from the target site.

An example of a gene modifying system that provides an inward nick orientation comprises a gene modifying polypeptide comprising a CRISPR/Cas domain, a template RNA comprising a gRNA that directs nicking of the target site DNA on the first strand, and an additional nucleic acid comprising an additional gRNA that directs nicking at a site a distance from the location of the first nick, wherein the location of the first nick and the location of the second nick are between the PAM sites of the sites to which the two gRNAs direct the gene modifying polypeptide. As a further example, another gene modifying system that provides an inward nick orientation comprises a gene modifying polypeptide comprising a zinc finger molecule and a first nickase domain wherein the zinc finger molecule binds to the target DNA in a manner that directs the first nickase domain to nick the first strand of the target site; an additional polypeptide comprising a CRISPR/Cas domain, and an additional nucleic acid comprising a gRNA that directs the additional polypeptide to nick a site a distance from the target site DNA on the second strand, wherein the location of the first nick and the location of the second nick are between the PAM site and the site to which the zinc finger molecule binds. As a further example, another gene modifying system that provides an inward nick orientation comprises a gene modifying polypeptide comprising a zinc finger molecule and a first nickase domain wherein the zinc finger molecule binds to the target DNA in a manner that directs the first nickase domain to nick the first strand of the target site; an additional polypeptide comprising a TAL effector molecule and a second nickase domain wherein the TAL effector molecule binds to a site a distance from the target site in a manner that directs the additional polypeptide to nick the second strand, wherein the location of the first nick and the location of the second nick are between the site to which the TAL effector molecule binds and the site to which the zinc finger molecule binds.

In some embodiments, in the outward nick orientation, the RT domain polymerizes (e.g., using the template RNA (e.g., the heterologous object sequence)) toward the second strand nick. In some embodiments, in the outward nick orientation when both the first and second nicks are made by a polypeptide comprising a CRISPR/Cas domain (e.g., a gene modifying polypeptide), the first PAM site and second PAM site are positioned between the location of the nick to the first strand and the location of the nick to the second strand. When there are two PAMs on the inside and two nicks on the outside, this outward nick orientation also can be referred to as “PAM-in”. In some embodiments, in the outward nick orientation, the polypeptide (e.g., the gene modifying polypeptide) and the additional polypeptide bind to sites on the target DNA between the location of the nick to the first strand and the location of the nick to the second. In some embodiments, in the outward nick orientation, the location of the nick to the second strand is positioned on the opposite side of the binding sites of the polypeptide and additional polypeptide

relative to the location of the nick to the first strand. In some embodiments, in the outward orientation, the PAM site and the binding site of the second polypeptide which is at a distance from the target site are positioned between the location of the nick to the first strand and the location of the nick to the second strand.

An example of a gene modifying system that provides an outward nick orientation comprises a gene modifying polypeptide comprising a CRISPR/Cas domain, a template RNA comprising a gRNA that directs nicking of the target site DNA on the first strand, and an additional nucleic acid comprising an additional gRNA that directs nicking at a site a distance from the location of the first nick, wherein the location of the first nick and the location of the second nick are outside of the PAM sites of the sites to which the two gRNAs direct the gene modifying polypeptide (i.e., the PAM sites are between the location of the first nick and the location of the second nick). As a further example, another gene modifying system that provides an outward nick orientation comprises a gene modifying polypeptide comprising a zinc finger molecule and a first nickase domain wherein the zinc finger molecule binds to the target DNA in a manner that directs the first nickase domain to nick the first strand of the target site; an additional polypeptide comprising a CRISPR/Cas domain, and an additional nucleic acid comprising a gRNA that directs the additional polypeptide to nick a site a distance from the target site DNA on the second strand, wherein the location of the first nick and the location of the second nick are outside the PAM site and the site to which the zinc finger molecule binds (i.e., the PAM site and the site to which the zinc finger molecule binds are between the location of the first nick and the location of the second nick). As a further example, another gene modifying system that provides an outward nick orientation comprises a gene modifying polypeptide comprising a zinc finger molecule and a first nickase domain wherein the zinc finger molecule binds to the target DNA in a manner that directs the first nickase domain to nick the first strand of the target site; an additional polypeptide comprising a TAL effector molecule and a second nickase domain wherein the TAL effector molecule binds to a site a distance from the target site in a manner that directs the additional polypeptide to nick the second strand, wherein the location of the first nick and the location of the second nick are outside the site to which the TAL effector molecule binds and the site to which the zinc finger molecule binds (i.e., the site to which the TAL effector molecule binds and the site to which the zinc finger molecule binds are between the location of the first nick and the location of the second nick).

Without wishing to be bound by theory, it is thought that, for gene modifying systems where a second strand nick is provided, an outward nick orientation is preferred in some embodiments. As is described herein, an inward nick may produce a higher number of double-strand breaks (DSBs) than an outward nick orientation. DSBs may be recognized by the DSB repair pathways in the nucleus of a cell, which can result in undesired insertions and deletions. An outward nick orientation may provide a

decreased risk of DSB formation, and a corresponding lower amount of undesired insertions and deletions. In some embodiments, undesired insertions and deletions are insertions and deletions not encoded by the heterologous object sequence, e.g., an insertion or deletion produced by the double-strand break repair pathway unrelated to the modification encoded by the heterologous object sequence. In some embodiments, a desired gene modification comprises a change to the target DNA (e.g., a substitution, insertion, or deletion) encoded by the heterologous object sequence (e.g., and achieved by the gene modifying writing the heterologous object sequence into the target site). In some embodiments, the first strand nick and the second strand nick are in an outward orientation.

In addition, the distance between the first strand nick and second strand nick may influence the extent to which one or more of: desired gene modifying system DNA modifications are obtained, undesired double-strand breaks (DSBs) occur, undesired insertions occur, or undesired deletions occur. Without wishing to be bound by theory, it is thought the second strand nick benefit, the biasing of DNA repair toward incorporation of the heterologous object sequence into the target DNA, increases as the distance between the first strand nick and second strand nick decreases. However, it is thought that the risk of DSB formation also increases as the distance between the first strand nick and second strand nick decreases. Correspondingly, it is thought that the number of undesired insertions and/or deletions may increase as the distance between the first strand nick and second strand nick decreases. In some embodiments, the distance between the first strand nick and second strand nick is chosen to balance the benefit of biasing DNA repair toward incorporation of the heterologous object sequence into the target DNA and the risk of DSB formation and of undesired deletions and/or insertions. In some embodiments, a system where the first strand nick and the second strand nick are at least a threshold distance apart has an increased level of desired gene modifying system modification outcomes, a decreased level of undesired deletions, and/or a decreased level of undesired insertions relative to an otherwise similar inward nick orientation system where the first nick and the second nick are less than the a threshold distance apart. In some embodiments the threshold distance(s) is given below.

In some embodiments, the first nick and the second nick are at least 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides apart. In some embodiments, the first nick and the second nick are no more than 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, or 250 nucleotides apart. In some embodiments, the first nick and the second nick are 20-200, 30-200, 40-200, 50-200, 60-200, 70-200, 80-200, 90-200, 100-200, 110-200, 120-200, 130-200, 140-200, 150-200, 160-200, 170-200, 180-200, 190-200, 20-190, 30-190, 40-190, 50-190, 60-190, 70-190, 80-190, 90-190, 100-190, 110-190, 120-190, 130-190, 140-190, 150-190, 160-190, 170-190, 180-190, 20-180, 30-180, 40-180, 50-180, 60-180, 70-180, 80-180, 90-180, 100-180, 110-180, 120-180, 130-180, 140-180, 150-180, 160-180, 170-180, 20-

170, 30-170, 40-170, 50-170, 60-170, 70-170, 80-170, 90-170, 100-170, 110-170, 120-170, 130-170, 140-170, 150-170, 160-170, 20-160, 30-160, 40-160, 50-160, 60-160, 70-160, 80-160, 90-160, 100-160, 110-160, 120-160, 130-160, 140-160, 150-160, 20-150, 30-150, 40-150, 50-150, 60-150, 70-150, 80-150, 90-150, 100-150, 110-150, 120-150, 130-150, 140-150, 20-140, 30-140, 40-140, 50-140, 60-140, 70-140, 80-140, 90-140, 100-140, 110-140, 120-140, 130-140, 20-130, 30-130, 40-130, 50-130, 60-130, 70-130, 80-130, 90-130, 100-130, 110-130, 120-130, 20-120, 30-120, 40-120, 50-120, 60-120, 70-120, 80-120, 90-120, 100-120, 110-120, 20-110, 30-110, 40-110, 50-110, 60-110, 70-110, 80-110, 90-110, 100-110, 20-100, 30-100, 40-100, 50-100, 60-100, 70-100, 80-100, 90-100, 20-90, 30-90, 40-90, 50-90, 60-90, 70-90, 80-90, 20-80, 30-80, 40-80, 50-80, 60-80, 70-80, 20-70, 30-70, 40-70, 50-70, 60-70, 20-60, 30-60, 40-60, 50-60, 20-50, 30-50, 40-50, 20-40, 30-40, or 20-30 nucleotides apart. In some embodiments, the first nick and the second nick are 40-100 nucleotides apart.

Without wishing to be bound by theory, it is thought that, for gene modifying systems where a second strand nick is provided and an inward nick orientation is selected, increasing the distance between the first strand nick and second strand nick may be preferred. As is described herein, an inward nick orientation may produce a higher number of DSBs than an outward nick orientation, and may result in a higher amount of undesired insertions and deletions than an outward nick orientation, but increasing the distance between the nicks may mitigate that increase in DSBs, undesired deletions, and/or undesired insertions. In some embodiments, an inward nick orientation wherein the first nick and the second nick are at least a threshold distance apart has an increased level of desired gene modifying system modification outcomes, a decreased level of undesired deletions, and/or a decreased level of undesired insertions relative to an otherwise similar inward nick orientation system where the first nick and the second nick are less than the a threshold distance apart. In some embodiments the threshold distance is given below.

In some embodiments, the first strand nick and the second strand nick are in an inward orientation. In some embodiments, the first strand nick and the second strand nick are in an inward orientation and the first strand nick and second strand nick are at least 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 220, 240, 260, 280, 300, 350, 400, 450, or 500 nucleotides apart, e.g., at least 100 nucleotides apart, (and optionally no more than 500, 400, 300, 200, 190, 180, 170, 160, 150, 140, 130, or 120 nucleotides apart). In some embodiments, the first strand nick and the second strand nick are in an inward orientation and the first strand nick and second strand nick are 100-200, 110-200, 120-200, 130-200, 140-200, 150-200, 160-200, 170-200, 180-200, 190-200, 100-190, 110-190, 120-190, 130-190, 140-190, 150-190, 160-190, 170-190, 180-190, 100-180, 110-180, 120-180, 130-180, 140-180, 150-180, 160-180, 170-180, 100-170, 110-170, 120-170, 130-170, 140-170, 150-170, 160-170, 100-160, 110-160, 120-

160, 130-160, 140-160, 150-160, 100-150, 110-150, 120-150, 130-150, 140-150, 100-140, 110-140, 120-140, 130-140, 100-130, 110-130, 120-130, 100-120, 110-120, or 100-110 nucleotides apart.

Chemically modified nucleic acids and nucleic acid end features

A nucleic acid described herein (e.g., a template nucleic acid, e.g., a template RNA; or a nucleic acid (e.g., mRNA) encoding a gene modifying polypeptide; or a gRNA) can comprise unmodified or modified nucleobases. Naturally occurring RNAs are synthesized from four basic ribonucleotides: ATP, CTP, UTP and GTP, but may contain post-transcriptionally modified nucleotides. Further, approximately one hundred different nucleoside modifications have been identified in RNA (Rozenski, J, Crain, P, and McCloskey, J. (1999). The RNA Modification Database: 1999 update. Nucl Acids Res 27: 196-197). An RNA can also comprise wholly synthetic nucleotides that do not occur in nature.

In some embodiments, the chemical modification is one provided in WO/2016/183482, US Pat. Pub. No. 20090286852, of International Application No. WO/2012/019168, WO/2012/045075, WO/2012/135805, WO/2012/158736, WO/2013/039857, WO/2013/039861, WO/2013/052523, WO/2013/090648, WO/2013/096709, WO/2013/101690, WO/2013/106496, WO/2013/130161, WO/2013/151669, WO/2013/151736, WO/2013/151672, WO/2013/151664, WO/2013/151665, WO/2013/151668, WO/2013/151671, WO/2013/151667, WO/2013/151670, WO/2013/151666, WO/2013/151663, WO/2014/028429, WO/2014/081507, WO/2014/093924, WO/2014/093574, WO/2014/113089, WO/2014/144711, WO/2014/144767, WO/2014/144039, WO/2014/152540, WO/2014/152030, WO/2014/152031, WO/2014/152027, WO/2014/152211, WO/2014/158795, WO/2014/159813, WO/2014/164253, WO/2015/006747, WO/2015/034928, WO/2015/034925, WO/2015/038892, WO/2015/048744, WO/2015/051214, WO/2015/051173, WO/2015/051169, WO/2015/058069, WO/2015/085318, WO/2015/089511, WO/2015/105926, WO/2015/164674, WO/2015/196130, WO/2015/196128, WO/2015/196118, WO/2016/011226, WO/2016/011222, WO/2016/011306, WO/2016/014846, WO/2016/022914, WO/2016/036902, WO/2016/077125, or WO/2016/077123, each of which is herein incorporated by reference in its entirety. It is understood that incorporation of a chemically modified nucleotide into a polynucleotide can result in the modification being incorporated into a nucleobase, the backbone, or both, depending on the location of the modification in the nucleotide. In some embodiments, the backbone modification is one provided in EP 2813570, which is herein incorporated by reference in its entirety. In some embodiments, the modified cap is one provided in US Pat. Pub. No. 20050287539, which is herein incorporated by reference in its entirety.

In some embodiments, the chemically modified nucleic acid (e.g., RNA, e.g., mRNA) comprises one or more of ARCA: anti-reverse cap analog (m27.3′-OGP3G), GP3G (Unmethylated Cap Analog), m7GP3G (Monomethylated Cap Analog), m32.2.7GP3G (Trimethylated Cap Analog), m5CTP (5′-

methyl-cytidine triphosphate), m6ATP (N6-methyl-adenosine-5′-triphosphate), s2UTP (2-thio-uridine triphosphate), and Ψ (pseudouridine triphosphate).

In some embodiments, the chemically modified nucleic acid comprises a 5′ cap, e.g.: a 7-methylguanosine cap (e.g., a O-Me-m7G cap); a hypermethylated cap analog; an NAD+-derived cap analog (e.g., as described in Kiledjian, Trends in Cell Biology 28, 454-464 (2018)); or a modified, e.g., biotinylated, cap analog (e.g., as described in Bednarek et al., Phil Trans R Soc B 373, 20180167 (2018)).

In some embodiments, the chemically modified nucleic acid comprises a 3′ feature selected from one or more of: a polyA tail; a 16-nucleotide long stem-loop structure flanked by unpaired 5 nucleotides (e.g., as described by Mannironi et al., Nucleic Acid Research 17, 9113-9126 (1989)); a triple-helical structure (e.g., as described by Brown et al., PNAS 109, 19202-19207 (2012)); a tRNA, Y RNA, or vault RNA structure (e.g., as described by Labno et al., Biochemica et Biophysica Acta 1863, 3125-3147 (2016)); incorporation of one or more deoxyribonucleotide triphosphates (dNTPs), 2'O-Methylated NTPs, or phosphorothioate-NTPs; a single nucleotide chemical modification (e.g., oxidation of the 3′ terminal ribose to a reactive aldehyde followed by conjugation of the aldehyde-reactive modified nucleotide); or chemical ligation to another nucleic acid molecule.

In some embodiments, the nucleic acid (e.g., template nucleic acid) comprises one or more modified nucleotides, e.g., selected from dihydrouridine, inosine, 7-methylguanosine, 5-methylcytidine (5mC), 5′ Phosphate ribothymidine, 2′-O-methyl ribothymidine, 2′-O-ethyl ribothymidine, 2′-fluoro ribothymidine, C-5 propynyl-deoxycytidine (pdC), C-5 propynyl-deoxyuridine (pdU), C-5 propynyl-cytidine (pC), C-5 propynyl-uridine (pU), 5-methyl cytidine, 5-methyl uridine, 5-methyl deoxycytidine, 5-methyl deoxyuridine methoxy, 2,6-diaminopurine, 5′-Dimethoxytrityl-N4-ethyl-2′-deoxycytidine, C-5 propynyl-f-cytidine (pfC), C-5 propynyl-f-uridine (pfU), 5-methyl f-cytidine, 5-methyl f-uridine, C-5 propynyl-m-cytidine (pmC), C-5 propynyl-f-uridine (pmU), 5-methyl m-cytidine, 5-methyl m-uridine, LNA (locked nucleic acid), MGB (minor groove binder) pseudouridine (Ψ), 1-N-methylpseudouridine (1-Me-Ψ), or 5-methoxyuridine (5-MO-U).

In some embodiments, the nucleic acid comprises a backbone modification, e.g., a modification to a sugar or phosphate group in the backbone. In some embodiments, the nucleic acid comprises a nucleobase modification.

In some embodiments, the nucleic acid comprises one or more chemically modified nucleotides of Table 13, one or more chemical backbone modifications of Table 14, one or more chemically modified caps of Table 15. For instance, in some embodiments, the nucleic acid comprises two or more (e.g., 3, 4, 5, 6, 7, 8, 9, or 10 or more) different types of chemical modifications. As an example, the nucleic acid may comprise two or more (e.g., 3, 4, 5, 6, 7, 8, 9, or 10 or more) different types of modified nucleobases, e.g., as described herein, e.g., in Table 13. Alternatively or in combination, the nucleic acid may

comprise two or more (e.g., 3, 4, 5, 6, 7, 8, 9, or 10 or more) different types of backbone modifications, e.g., as described herein, e.g., in Table 14. Alternatively or in combination, the nucleic acid may comprise one or more modified cap, e.g., as described herein, e.g., in Table 15. For instance, in some embodiments, the nucleic acid comprises one or more type of modified nucleobase and one or more type of backbone modification; one or more type of modified nucleobase and one or more modified cap; one or more type of modified cap and one or more type of backbone modification; or one or more type of modified nucleobase, one or more type of backbone modification, and one or more type of modified cap.

In some embodiments, the nucleic acid comprises one or more (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, or more) modified nucleobases. In some embodiments, all nucleobases of the nucleic acid are modified. In some embodiments, the nucleic acid is modified at one or more (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, or more) positions in the backbone. In some embodiments, all backbone positions of the nucleic acid are modified.

Table 13. Modified nucleotides

5-aza-uridine	N2-methyl-6-thio-guanosine
2-thio-5-aza-midine	N2,N2-dimethyl-6-thio-guanosine
2-thiouridine	pyridin-4-one ribonucleoside
4-thio-pseudouridine	2-thio-5-aza-uridine
2-thio-pseudouridine	2-thiomidine
5-hydroxyuridine	4-thio-pseudomidine
3-methyluridine	2-thio-pseudowidine
5-carboxymethyl-uridine	3-methylmidine
1-carboxymethyl-pseudouridine	1-propynyl-pseudomidine
5-propynyl-uridine	1-methyl-1-deaza-pseudomidine
1-propynyl-pseudouridine	2-thio-1-methyl-1-deaza-pseudouridine
5-taurinomethyluridine	4-methoxy-pseudomidine
1-taurinomethyl-pseudouridine	5'-O-(1-Thiophosphate)-Adenosine
5-taurinomethyl-2-thio-uridine	5'-O-(1-Thiophosphate)-Cytidine
1-taurinomethyl-4-thio-uridine	5'-O-(1-thiophosphate)-Guanosine
5-methyl-uridine	5'-O-(1-Thiophophate)-Uridine
1-methyl-pseudouridine	5'-O-(1-Thiophosphate)-Pseudouridine
4-thio-1-methyl-pseudouridine	2'-O-methyl-Adenosine
2-thio-1-methyl-pseudouridine	2'-O-methyl-Cytidine
1-methyl-1-deaza-pseudouridine	2'-O-methyl-Guanosine
2-thio-1-methyl-1-deaza-pseudomidine	2'-O-methyl-Uridine
dihydrouridine	2'-O-methyl-Pseudouridine
dihydropseudouridine	2'-O-methyl-Inosine
2-thio-dihydromidine	2-methyladenosine
2-thio-dihydropseudouridine	2-methylthio-N6-methyladenosine
2-methoxyuridine	2-methylthio-N6 isopentenyladenosine
2-methoxy-4-thio-uridine	2-methylthio-N6-(cis-hydroxyisopentenyl)adenosine
4-methoxy-pseudouridine	

4-methoxy-2-thio-pseudouridine	N6-methyl-N6-threonylcarbamoyladenosine
5-aza-cytidine	N6-hydroxynorvalylcarbamoyladenosine
pseudoisocytidine	2-methylthio-N6-hydroxynorvalyl carbamoyladenosine
3-methyl-cytidine	2'-O-ribosyladenosine (phosphate)
N4-acetylcytidine	1,2'-O-dimethylinosine
5-formylcytidine	5,2'-O-dimethylcytidine
N4-methylcytidine	N4-acetyl-2'-O-methylcytidine
5-hydroxymethylcytidine	Lysidine
1-methyl-pseudoisocytidine	7-methylguanosine
pyrrolo-cytidine	N2,2'-O-dimethylguanosine
pyrrolo-pseudoisocytidine	N2,N2,2'-O-trimethylguanosine
2-thio-cytidine	2'-O-ribosylguanosine (phosphate)
2-thio-5-methyl-cytidine	Wybutosine
4-thio-pseudoisocytidine	Peroxywybutosine
4-thio-1-methyl-pseudoisocytidine	Hydroxywybutosine
4-thio-1-methyl-1-deaza-pseudoisocytidine	undermodified hydroxywybutosine
1-methyl-1-deaza-pseudoisocytidine	methylwyosine
zebularine	queuosine
5-aza-zebularine	epoxyqueuosine
5-methyl-zebularine	galactosyl-queuosine
5-aza-2-thio-zebularine	mannosyl-queuosine
2-thio-zebularine	7-cyano-7-deazaguanosine
2-methoxy-cytidine	7-aminomethyl-7-deazaguanosine
2-methoxy-5-methyl-cytidine	archaeosine
4-methoxy-pseudoisocytidine	5,2'-O-dimethyluridine
4-methoxy-1-methyl-pseudoisocytidine	4-thiouridine
2-aminopurine	5-methyl-2-thiouridine
2,6-diaminopurine	2-thio-2'-O-methyluridine
7-deaza-adenine	3-(3-amino-3-carboxypropyl)uridine
7-deaza-8-aza-adenine	5-methoxyuridine
7-deaza-2-aminopurine	uridine 5-oxyacetic acid
7-deaza-8-aza-2-aminopurine	uridine 5-oxyacetic acid methyl ester
7-deaza-2,6- diaminopurine	5-(carboxyhydroxymethyl)uridine)
7-deaza-8-aza-2,6-diarninopurine	5-(carboxyhydroxymethyl)uridine methyl ester
1-methyladenosine	5-methoxycarbonylmethyluridine
N6-isopentenyladenosine	5-methoxycarbonylmethyl-2'-O-methyluridine
N6-(cis-hydroxyisopentenyl)adenosine	5-methoxycarbonylmethyl-2-thiouridine
2-methylthio-N6-(cis-hydroxyisopentenyl) adenosine	5-aminomethyl-2-thiouridine
N6-glycinylcarbamoyladenosine	5-methylaminomethyluridine
N6-threonylcarbamoyladenosine	5-methylaminomethyl-2-thiouridine
2-methylthio-N6-threonyl carbamoyladenosine	5-methylaminomethyl-2-selenouridine
N6,N6-dimethyladenosine	5-carbamoylmethyluridine
7-methyladenine	5-carbamoylmethyl-2'-O-methyluridine
2-methylthio-adenine	5-carboxymethylaminomethyluridine
2-methoxy-adenine	5-carboxymethylaminomethyl-2'-O-methyluridine
inosine	5-carboxymethylaminomethyl-2-thiouridine
1-methyl-inosine	N4,2'-O-dimethylcytidine
wyosine	5-carboxymethyluridine
wybutosine	N6,2'-O-dimethyladenosine
7-deaza-guanosine	N,N6,O-2'-trimethyladenosine

7-deaza-8-aza-guanosine	N2,7-dimethylguanosine
6-thio-guanosine	N2,N2,7-trimethylguanosine
6-thio-7-deaza-guanosine	3,2'-O-dimethyluridine
6-thio-7-deaza-8-aza-guanosine	5-methyldihydrouridine
7-methyl-guanosine	5-formyl-2'-O-methylcytidine
6-thio-7-methyl-guanosine	1,2'-O-dimethylguanosine
7-methylinosine	4-demethylwyosine
6-methoxy-guanosine	Isowyosine
1-methylguanosine	N6-acetyladenosine
N2-methylguanosine	
N2,N2-dimethylguanosine	
8-oxo-guanosine	
7-methyl-8-oxo-guanosine	
1-methyl-6-thio-guanosine	

Table 14. Backbone modifications

2'-O-Methyl backbone
Peptide Nucleic Acid (PNA) backbone
phosphorothioate backbone
morpholino backbone
carbamate backbone
siloxane backbone
sulfide backbone
sulfoxide backbone
sulfone backbone
formacetyl backbone
thioformacetyl backbone
methyleneformacetyl backbone
riboacetyl backbone
alkene containing backbone
sulfamate backbone
sulfonate backbone
sulfonamide backbone
methyleneimino backbone
methylenehydrazino backbone
amide backbone

Table 15. Modified caps

m7GpppA
m7GpppC
m2,7GpppG
m2,2,7GpppG
m7Gpppm7G
m7,2'OmeGpppG
m72'dGpppG
m7,3'OmeGpppG
m7,3'dGpppG
GppppG

m7GppppG
m7GpppA
m7GpppC
m2,7GppppG
m2,2,7GppppG
m7Gpppm7G
m7,2'OmeGppppG
m72'dGppppG
m7,3'OmeGppppG
m7,3'dGppppG

The nucleotides comprising the template of the gene modifying system can be natural or modified bases, or a combination thereof. For example, the template may contain pseudouridine, dihydrouridine, inosine, 7-methylguanosine, or other modified bases. In some embodiments, the template may contain locked nucleic acid nucleotides. In some embodiments, the modified bases used in the template do not inhibit the reverse transcription of the template. In some embodiments, the modified bases used in the template may improve reverse transcription, e.g., specificity or fidelity.

In some embodiments, an RNA component of the system (e.g., a template RNA or a gRNA) comprises one or more nucleotide modifications. In some embodiments, the modification pattern of a gRNA can significantly affect *in vivo* activity compared to unmodified or end-modified guides (e.g., as shown in Figure 1D from Finn et al. *Cell Rep* 22(9):2227-2235 (2018); incorporated herein by reference in its entirety). Without wishing to be bound by theory, this process may be due, at least in part, to a stabilization of the RNA conferred by the modifications. Non-limiting examples of such modifications may include 2'-O-methyl (2'-O-Me), 2'-0-(2-methoxyethyl) (2'-0-MOE), 2'- fluoro (2'-F), phosphorothioate (PS) bond between nucleotides, G-C substitutions, and inverted abasic linkages between nucleotides and equivalents thereof.

In some embodiments, the template RNA (e.g., at the portion thereof that binds a target site) or the guide RNA comprises a 5′ terminus region. In some embodiments, the template RNA or the guide RNA does not comprise a 5′ terminus region. In some embodiments, the 5′ terminus region comprises a gRNA spacer region, e.g., as described with respect to sgRNA in Briner AE et al, Molecular Cell 56: 333-339 (2014) (incorporated herein by reference in its entirety; applicable herein, e.g., to all guide RNAs). In some embodiments, the 5′ terminus region comprises a 5′ end modification. In some embodiments, a 5′ terminus region with or without a spacer region may be associated with a crRNA, trRNA, sgRNA and/or dgRNA. The gRNA spacer region can, in some instances, comprise a guide region, guide domain, or targeting domain.

In some embodiments, the template RNAs (e.g., at the portion thereof that binds a target site) or guide RNAs described herein comprises any of the sequences shown in Table 4 of WO2018107028A1, incorporated herein by reference in its entirety. In some embodiments, where a sequence shows a guide and/or spacer region, the composition may comprise this region or not. In some embodiments, a guide RNA comprises one or more of the modifications of any of the sequences shown in Table 4 of WO2018107028A1, e.g., as identified therein by a SEQ ID NO. In embodiments, the nucleotides may be the same or different, and/or the modification pattern shown may be the same or similar to a modification pattern of a guide sequence as shown in Table 4 of WO2018107028A1. In some embodiments, a modification pattern includes the relative position and identity of modifications of the gRNA or a region of the gRNA (e.g. 5′ terminus region, lower stem region, bulge region, upper stem region, nexus region, hairpin 1 region, hairpin 2 region, 3′ terminus region). In some embodiments, the modification pattern contains at least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of the modifications of any one of the sequences shown in the sequence column of Table 4 of WO2018107028A1, and/or over one or more regions of the sequence. In some embodiments, the modification pattern is at least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the modification pattern of any one of the sequences shown in the sequence column of Table 4 of WO2018107028A1. In some embodiments, the modification pattern is at least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over one or more regions of the sequence shown in Table 4 of WO2018107028A1, e.g., in a 5 ' terminus region, lower stem region, bulge region, upper stem region, nexus region, hairpin 1 region, hairpin 2 region, and/or 3′ terminus region. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the modification pattern of a sequence over the 5 ' terminus region. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over the lower stem. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over the bulge. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over the upper stem. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over the nexus. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over the hairpin 1. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over the hairpin 2. In some embodiments, the modification pattern is least 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% identical over the 3 ' terminus. In some embodiments, the

modification pattern differs from the modification pattern of a sequence of Table 4 of WO2018107028A1, or a region (e.g. 5′ terminus, lower stem, bulge, upper stem, nexus, hairpin 1, hairpin 2, 3′ terminus) of such a sequence, e.g., at 0, 1, 2, 3, 4, 5, 6, or more nucleotides. In some embodiments, the gRNA comprises modifications that differ from the modifications of a sequence of Table 4 of WO2018107028A1, e.g., at 0, 1, 2, 3, 4, 5, 6, or more nucleotides. In some embodiments, the gRNA comprises modifications that differ from modifications of a region (e.g. 5 ' terminus, lower stem, bulge, upper stem, nexus, hairpin 1, hairpin 2, 3′ terminus) of a sequence of Table 4 of WO2018107028A1, e.g., at 0, 1, 2, 3, 4, 5, 6, or more nucleotides.

In some embodiments, the template RNAs (e.g., at the portion thereof that binds a target site) or the gRNA comprises a 2'-O-methyl (2'-O-Me) modified nucleotide. In some embodiments, the gRNA comprises a 2'-O-(2-methoxy ethyl) (2'-O-moe) modified nucleotide. In some embodiments, the gRNA comprises a 2'-fluoro (2'- F) modified nucleotide. In some embodiments, the gRNA comprises a phosphorothioate (PS) bond between nucleotides. In some embodiments, the gRNA comprises a 5′ end modification, a 3′ end modification, or 5′ and 3′ end modifications. In some embodiments, the 5′ end modification comprises a phosphorothioate (PS) bond between nucleotides. In some embodiments, the 5′ end modification comprises a 2'-O-methyl (2'-O-Me), 2'-O-(2-methoxy ethyl) (2'-O-MOE), and/or 2'-fluoro (2'-F) modified nucleotide. In some embodiments, the 5′ end modification comprises at least one phosphorothioate (PS) bond and one or more of a 2'-O-methyl (2'-O- Me), 2'-O-(2-methoxyethyl) (2'-O-MOE), and/or 2'-fluoro (2'-F) modified nucleotide. The end modification may comprise a phosphorothioate (PS), 2'-O-methyl (2'-O-Me), 2'-O-(2- methoxyethyl) (2'-O-MOE), and/or 2'-fluoro (2'-F) modification. Equivalent end modifications are also encompassed by embodiments described herein. In some embodiments, the template RNA or gRNA comprises an end modification in combination with a modification of one or more regions of the template RNA or gRNA. Additional exemplary modifications and methods for protecting RNA, e.g., gRNA, and formulae thereof, are described in WO2018126176A1, which is incorporated herein by reference in its entirety.

In some embodiments, structure-guided and systematic approaches are used to introduce modifications (e.g., 2′-OMe-RNA, 2′-F-RNA, and PS modifications) to a template RNA or guide RNA, for example, as described in Mir et al. *Nat Commun* 9:2641 (2018) (incorporated by reference herein in its entirety). In some embodiments, the incorporation of 2′-F-RNAs increases thermal and nuclease stability of RNA:RNA or RNA:DNA duplexes, e.g., while minimally interfering with C3′-endo sugar puckering. In some embodiments, 2′-F may be better tolerated than 2′-OMe at positions where the 2′-OH is important for RNA:DNA duplex stability. In some embodiments, a crRNA comprises one or more modifications that do not reduce Cas9 activity, e.g., C10, C20, or C21 (fully modified), e.g., as described in

Supplementary Table 1 of Mir et al. *Nat Commun* 9:2641 (2018), incorporated herein by reference in its entirety. In some embodiments, a tracrRNA comprises one or more modifications that do not reduce Cas9 activity, e.g., T2, T6, T7, or T8 (fully modified) of Supplementary Table 1 of Mir et al. *Nat Commun* 9:2641 (2018). In some embodiments, a crRNA comprises one or more modifications (e.g., as described herein) may be paired with a tracrRNA comprising one or more modifications, e.g., C20 and T2. In some embodiments, a gRNA comprises a chimera, e.g., of a crRNA and a tracrRNA (e.g., Jinek et al. *Science* 337(6096):816-821 (2012)). In embodiments, modifications from the crRNA and tracrRNA are mapped onto the single-guide chimera, e.g., to produce a modified gRNA with enhanced stability.

In some embodiments, gRNA molecules may be modified by the addition or subtraction of the naturally occurring structural components, e.g., hairpins. In some embodiments, a gRNA may comprise a gRNA with one or more 3′ hairpin elements deleted, e.g., as described in WO2018106727, incorporated herein by reference in its entirety. In some embodiments, a gRNA may contain an added hairpin structure, e.g., an added hairpin structure in the spacer region, which was shown to increase specificity of a CRISPR-Cas system in the teachings of Kocak et al. *Nat Biotechnol* 37(6):657-666 (2019). Additional modifications, including examples of shortened gRNA and specific modifications improving *in vivo* activity, can be found in US20190316121, incorporated herein by reference in its entirety.

In some embodiments, structure-guided and systematic approaches (e.g., as described in Mir et al. *Nat Commun* 9:2641 (2018); incorporated herein by reference in its entirety) are employed to find modifications for the template RNA. In embodiments, the modifications are identified with the inclusion or exclusion of a guide region of the template RNA. In some embodiments, a structure of polypeptide bound to template RNA is used to determine non-protein-contacted nucleotides of the RNA that may then be selected for modifications, e.g., with lower risk of disrupting the association of the RNA with the polypeptide. Secondary structures in a template RNA can also be predicted *in silico* by software tools, e.g., the RNAstructure tool available at rna.urmc.rochester.edu/RNAstructureWeb (Bellaousov et al. *Nucleic Acids Res* 41:W471-W474 (2013); incorporated by reference herein in its entirety), e.g., to determine secondary structures for selecting modifications, e.g., hairpins, stems, and/or bulges.

Production of Compositions and Systems

As will be appreciated by one of skill, methods of designing and constructing nucleic acid constructs and proteins or polypeptides (such as the systems, constructs and polypeptides described herein) are routine in the art. Generally, recombinant methods may be used. See, in general, Smales & James (Eds.), *Therapeutic Proteins: Methods and Protocols* (Methods in Molecular Biology), Humana Press (2005); and Crommelin, Sindelar & Meibohm (Eds.), *Pharmaceutical Biotechnology: Fundamentals and Applications*, Springer (2013). Methods of designing, preparing, evaluating, purifying

and manipulating nucleic acid compositions are described in Green and Sambrook (Eds.), *Molecular Cloning: A Laboratory Manual* (Fourth Edition), Cold Spring Harbor Laboratory Press (2012).

The disclosure provides, in part, a nucleic acid, e.g., vector, encoding a gene modifying polypeptide described herein, a template nucleic acid described herein, or both. In some embodiments, a vector comprises a selective marker, e.g., an antibiotic resistance marker. In some embodiments, the antibiotic resistance marker is a kanamycin resistance marker. In some embodiments, the antibiotic resistance marker does not confer resistance to beta-lactam antibiotics. In some embodiments, the vector does not comprise an ampicillin resistance marker. In some embodiments, the vector comprises a kanamycin resistance marker and does not comprise an ampicillin resistance marker. In some embodiments, a vector encoding a gene modifying polypeptide is integrated into a target cell genome (e.g., upon administration to a target cell, tissue, organ, or subject). In some embodiments, a vector encoding a gene modifying polypeptide is not integrated into a target cell genome (e.g., upon administration to a target cell, tissue, organ, or subject). In some embodiments, a vector encoding a template nucleic acid (e.g., template RNA) is not integrated into a target cell genome (e.g., upon administration to a target cell, tissue, organ, or subject). In some embodiments, if a vector is integrated into a target site in a target cell genome, the selective marker is not integrated into the genome. In some embodiments, if a vector is integrated into a target site in a target cell genome, genes or sequences involved in vector maintenance (e.g., plasmid maintenance genes) are not integrated into the genome. In some embodiments, if a vector is integrated into a target site in a target cell genome, transfer regulating sequences (e.g., inverted terminal repeats, e.g., from an AAV) are not integrated into the genome. In some embodiments, administration of a vector (e.g., encoding a gene modifying polypeptide described herein, a template nucleic acid described herein, or both) to a target cell, tissue, organ, or subject results in integration of a portion of the vector into one or more target sites in the genome(s) of said target cell, tissue, organ, or subject. In some embodiments, less than 99, 95, 90, 80, 70, 60, 50, 40, 30, 20, 10, 5, 4, 3, 2, or 1% of target sites (e.g., no target sites) comprising integrated material comprise a selective marker (e.g., an antibiotic resistance gene), a transfer regulating sequence (e.g., an inverted terminal repeat, e.g., from an AAV), or both from the vector.

Exemplary methods for producing a therapeutic pharmaceutical protein or polypeptide described herein involve expression in mammalian cells, although recombinant proteins can also be produced using insect cells, yeast, bacteria, or other cells under control of appropriate promoters. Mammalian expression vectors may comprise non-transcribed elements such as an origin of replication, a suitable promoter, and other 5' or 3' flanking non-transcribed sequences, and 5' or 3' non-translated sequences such as necessary ribosome binding sites, a polyadenylation site, splice donor and acceptor sites, and termination sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, splice,

and polyadenylation sites may be used to provide other genetic elements required for expression of a heterologous DNA sequence. Appropriate cloning and expression vectors for use with bacterial, fungal, yeast, and mammalian cellular hosts are described in Green & Sambrook, *Molecular Cloning: A Laboratory Manual* (Fourth Edition), Cold Spring Harbor Laboratory Press (2012).

Various mammalian cell culture systems can be employed to express and manufacture recombinant protein. Examples of mammalian expression systems include CHO, COS, HEK293, HeLA, and BHK cell lines. Processes of host cell culture for production of protein therapeutics are described in Zhou and Kantardjieff (Eds.), *Mammalian Cell Cultures for Biologics Manufacturing (Advances in Biochemical Engineering/Biotechnology)*, Springer (2014). Compositions described herein may include a vector, such as a viral vector, e.g., a lentiviral vector, encoding a recombinant protein. In some embodiments, a vector, e.g., a viral vector, may comprise a nucleic acid encoding a recombinant protein.

Purification of protein therapeutics is described in Franks, Protein Biotechnology: *Isolation, Characterization, and Stabilization*, Humana Press (2013); and in Cutler, *Protein Purification Protocols (Methods in Molecular Biology)*, Humana Press (2010).

The disclosure also provides compositions and methods for the production of template nucleic acid molecules (e.g., template RNAs) with specificity for a gene modifying polypeptide and/or a genomic target site. In an aspect, the method comprises production of RNA segments including an upstream homology segment, a heterologous object sequence segment, a gene modifying polypeptide binding motif, and a gRNA segment.

Therapeutic Applications

In some embodiments, a gene modifying system as described herein can be used to modify a cell (e.g., an animal cell, plant cell, or fungal cell). In some embodiments, a gene modifying system as described herein can be used to modify a mammalian cell (e.g., a human cell). In some embodiments, a gene modifying system as described herein can be used to modify a cell from a livestock animal (e.g., a cow, horse, sheep, goat, pig, llama, alpaca, camel, yak, chicken, duck, goose, or ostrich). In some embodiments, a gene modifying system as described herein can be used as a laboratory tool or a research tool, or used in a laboratory method or research method, e.g., to modify an animal cell, e.g., a mammalian cell (e.g., a human cell), a plant cell, or a fungal cell.

By integrating coding genes into a RNA sequence template, the gene modifying system can address therapeutic needs, for example, by providing expression of a therapeutic transgene in individuals with loss-of-function mutations, by replacing gain-of-function mutations with normal transgenes, by providing regulatory sequences to eliminate gain-of-function mutation expression, and/or by controlling the expression of operably linked genes, transgenes and systems thereof. In certain embodiments, the

RNA sequence template encodes a promotor region specific to the therapeutic needs of the host cell, for example a tissue specific promotor or enhancer. In still other embodiments, a promotor can be operably linked to a coding sequence.

In some embodiments, an insertion, deletion, substitution, or combination thereof, increases or decreases expression (e.g. transcription or translation) of a target gene. In some embodiments, an insertion, deletion, substitution, or combination thereof, increases or decreases expression (e.g. transcription or translation) of a target gene by altering, adding, or deleting sequences in a promoter or enhancer, e.g. sequences that bind transcription factors. In some embodiments, an insertion, deletion, substitution, or combination thereof alters translation of a target gene (e.g. alters an amino acid sequence), inserts or deletes a start or stop codon, alters or fixes the translation frame of a gene. In some embodiments, an insertion, deletion, substitution, or combination thereof alters splicing of a target gene, e.g. by inserting, deleting, or altering a splice acceptor or donor site. In some embodiments, an insertion, deletion, substitution, or combination thereof alters transcript or protein half-life. In some embodiments, an insertion, deletion, substitution, or combination thereof, alters, increases, decreases the activity of a target gene, e.g. a protein encoded by the target gene.

Compensatory edits

In some embodiments, the systems or methods provided herein can be used to introduce a compensatory edit. In some embodiments, the compensatory edit is at a position of a gene associated with a disease or disorder, which is different from the position of a disease-causing mutation. In some embodiments, the compensatory mutation is not in the gene containing the causative mutation. In some embodiments, the compensatory edit can negate or compensate for a disease-causing mutation. In some embodiments, the compensatory edit can be introduced by the systems or methods provided herein to suppress or reverse the mutant effect of a disease-causing mutation.

Regulatory edits

In some embodiments, the systems or methods provided herein can be used to introduce a regulatory edit. In some embodiments, the regulatory edit is introduced to a regulatory sequence of a gene, for example, a gene promoter, gene enhancer, gene repressor, or a sequence that regulates gene splicing. In some embodiments, the regulatory edit increases or decreases the expression level of a target gene. In some embodiments, the target gene is the same as the gene containing a disease-causing

mutation. In some embodiments, the target gene is different from the gene containing a disease-causing mutation.

Repeat expansion diseases

In some embodiments, the systems or methods provided herein can be used to treat a repeat expansion disease. In some embodiments, the systems or methods provided herein, for example, those comprising gene modifying polypeptides, can be used to treat repeat expansion diseases by resetting the number of repeats at the locus according to a customized RNA template.

Administration and Delivery

The compositions and systems described herein may be used *in vitro* or *in vivo*. In some embodiments the system or components of the system are delivered to cells (e.g., mammalian cells, e.g., human cells), e.g., *in vitro* or *in vivo*. In some embodiments, the cells are eukaryotic cells, e.g., cells of a multicellular organism, e.g., an animal, e.g., a mammal (e.g., human, swine, bovine), a bird (e.g., poultry, such as chicken, turkey, or duck), or a fish. In some embodiments, the cells are non-human animal cells (e.g., a laboratory animal, a livestock animal, or a companion animal). In some embodiments, the cell is a stem cell (e.g., a hematopoietic stem cell), a fibroblast, or a T cell. In some embodiments, the cell is an immune cell, e.g., a T cell (e.g., a Treg, CD4, CD8, γδ, or memory T cell), B cell (e.g., memory B cell or plasma cell), or NK cell. In some embodiments, the cell is a non-dividing cell, e.g., a non-dividing fibroblast or non-dividing T cell. In some embodiments, the cell is an HSC and p53 is not upregulated or is upregulated by less than 10%, 5%, 2%, or 1%, e.g., as determined according to the method described in Example 30 of PCT/US2019/048607. The skilled artisan will understand that the components of the gene modifying system may be delivered in the form of polypeptide, nucleic acid (e.g., DNA, RNA), and combinations thereof.

In one embodiment the system and/or components of the system are delivered as nucleic acid. For example, the gene modifying polypeptide may be delivered in the form of a DNA or RNA encoding the polypeptide, and the template RNA may be delivered in the form of RNA or its complementary DNA to be transcribed into RNA. In some embodiments the system or components of the system are delivered on 1, 2, 3, 4, or more distinct nucleic acid molecules. In some embodiments the system or components of the system are delivered as a combination of DNA and RNA. In some embodiments the system or components of the system are delivered as a combination of DNA and protein. In some embodiments the system or components of the system are delivered as a combination of RNA and protein. In some embodiments the gene modifying polypeptide is delivered as a protein.

In some embodiments the system or components of the system are delivered to cells, e.g. mammalian cells or human cells, using a vector. The vector may be, e.g., a plasmid or a virus. In some embodiments, delivery is *in vivo*, *in vitro*, *ex vivo*, or *in situ*. In some embodiments the virus is an adeno associated virus (AAV), a lentivirus, or an adenovirus. In some embodiments the system or components of the system are delivered to cells with a viral-like particle or a virosome. In some embodiments the delivery uses more than one virus, viral-like particle or virosome.

In one embodiment, the compositions and systems described herein can be formulated in liposomes or other similar vesicles. Liposomes are spherical vesicle structures composed of a uni- or multilamellar lipid bilayer surrounding internal aqueous compartments and a relatively impermeable outer lipophilic phospholipid bilayer. Liposomes may be anionic, neutral or cationic. Liposomes are biocompatible, nontoxic, can deliver both hydrophilic and lipophilic drug molecules, protect their cargo from degradation by plasma enzymes, and transport their load across biological membranes and the blood brain barrier (BBB) (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

Vesicles can be made from several different types of lipids; however, phospholipids are most commonly used to generate liposomes as drug carriers. Methods for preparation of multilamellar vesicle lipids are known in the art (see for example U.S. Pat. No. 6,693,086, the teachings of which relating to multilamellar vesicle lipid preparation are incorporated herein by reference). Although vesicle formation can be spontaneous when a lipid film is mixed with an aqueous solution, it can also be expedited by applying force in the form of shaking by using a homogenizer, sonicator, or an extrusion apparatus (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review). Extruded lipids can be prepared by extruding through filters of decreasing size, as described in Templeton et al., Nature Biotech, 15:647-652, 1997, the teachings of which relating to extruded lipid preparation are incorporated herein by reference.

A variety of nanoparticles can be used for delivery, such as a liposome, a lipid nanoparticle, a cationic lipid nanoparticle, an ionizable lipid nanoparticle, a polymeric nanoparticle, a gold nanoparticle, a dendrimer, a cyclodextrin nanoparticle, a micelle, or a combination of the foregoing.

Lipid nanoparticles are an example of a carrier that provides a biocompatible and biodegradable delivery system for the pharmaceutical compositions described herein. Nanostructured lipid carriers (NLCs) are modified solid lipid nanoparticles (SLNs) that retain the characteristics of the SLN, improve drug stability and loading capacity, and prevent drug leakage. Polymer nanoparticles (PNPs) are an important component of drug delivery. These nanoparticles can effectively direct drug delivery to specific targets and improve drug stability and controlled drug release. Lipid–polymer nanoparticles (PLNs), a type of carrier that combines liposomes and polymers, may also be employed. These

nanoparticles possess the complementary advantages of PNPs and liposomes. A PLN is composed of a core–shell structure; the polymer core provides a stable structure, and the phospholipid shell offers good biocompatibility. As such, the two components increase the drug encapsulation efficiency rate, facilitate surface modification, and prevent leakage of water-soluble drugs. For a review, see, e.g., Li et al. 2017, Nanomaterials 7, 122; doi:10.3390/nano7060122.

Exosomes can also be used as drug delivery vehicles for the compositions and systems described herein. For a review, see Ha et al. July 2016. Acta Pharmaceutica Sinica B. Volume 6, Issue 4, Pages 287-296; https://doi.org/10.1016/j.apsb.2016.02.001.

Fusosomes interact and fuse with target cells, and thus can be used as delivery vehicles for a variety of molecules. They generally consist of a bilayer of amphipathic lipids enclosing a lumen or cavity and a fusogen that interacts with the amphipathic lipid bilayer. The fusogen component has been shown to be engineerable in order to confer target cell specificity for the fusion and payload delivery, allowing the creation of delivery vehicles with programmable cell specificity (see for example Patent Application WO2020014209, the teachings of which relating to fusosome design, preparation, and usage are incorporated herein by reference).

In some embodiments, the protein component(s) of the gene modifying system may be pre-associated with the template nucleic acid (e.g., template RNA). For example, in some embodiments, the gene modifying polypeptide may be first combined with the template nucleic acid (e.g., template RNA) to form a ribonucleoprotein (RNP) complex. In some embodiments, the RNP may be delivered to cells via, e.g., transfection, nucleofection, virus, vesicle, LNP, exosome, fusosome.

A gene modifying system can be introduced into cells, tissues and multicellular organisms. In some embodiments the system or components of the system are delivered to the cells via mechanical means or physical means.

Formulation of protein therapeutics is described in Meyer (Ed.), *Therapeutic Protein Drug Products: Practical Approaches to formulation in the Laboratory, Manufacturing, and the Clinic*, Woodhead Publishing Series (2012).

Tissue Specific Activity/Administration

In some embodiments, a system described herein can make use of one or more feature (e.g., a promoter or microRNA binding site) to limit activity in off-target cells or tissues.

In some embodiments, a nucleic acid described herein (e.g., a template RNA or a DNA encoding a template RNA) comprises a promoter sequence, e.g., a tissue specific promoter sequence. In some embodiments, the tissue-specific promoter is used to increase the target-cell specificity of a gene modifying system. For instance, the promoter can be chosen on the basis that it is active in a target cell

type but not active in (or active at a lower level in) a non-target cell type. Thus, even if the promoter integrated into the genome of a non-target cell, it would not drive expression (or only drive low level expression) of an integrated gene. A system having a tissue-specific promoter sequence in the template RNA may also be used in combination with a microRNA binding site, e.g., in the template RNA or a nucleic acid encoding a gene modifying protein, e.g., as described herein. A system having a tissue-specific promoter sequence in the template RNA may also be used in combination with a DNA encoding a gene modifying polypeptide, driven by a tissue-specific promoter, e.g., to achieve higher levels of gene modifying protein in target cells than in non-target cells. In some embodiments, e.g., for liver indications, a tissue-specific promoter is selected from Table 3 of WO2020014209, incorporated herein by reference.

In some embodiments, a nucleic acid described herein (e.g., a template RNA or a DNA encoding a template RNA) comprises a microRNA binding site. In some embodiments, the microRNA binding site is used to increase the target-cell specificity of a gene modifying system. For instance, the microRNA binding site can be chosen on the basis that is recognized by a miRNA that is present in a non-target cell type, but that is not present (or is present at a reduced level relative to the non-target cell) in a target cell type. Thus, when the template RNA is present in a non-target cell, it would be bound by the miRNA, and when the template RNA is present in a target cell, it would not be bound by the miRNA (or bound but at reduced levels relative to the non-target cell). While not wishing to be bound by theory, binding of the miRNA to the template RNA may interfere with its activity, e.g., may interfere with insertion of the heterologous object sequence into the genome. Accordingly, the system would edit the genome of target cells more efficiently than it edits the genome of non-target cells, e.g., the heterologous object sequence would be inserted into the genome of target cells more efficiently than into the genome of non-target cells, or an insertion or deletion is produced more efficiently in target cells than in non-target cells. A system having a microRNA binding site in the template RNA (or DNA encoding it) may also be used in combination with a nucleic acid encoding a gene modifying polypeptide, wherein expression of the gene modifying polypeptide is regulated by a second microRNA binding site, e.g., as described herein. In some embodiments, e.g., for liver indications, a miRNA is selected from Table 4 of WO2020014209, incorporated herein by reference.

In some embodiments, the template RNA comprises a microRNA sequence, an siRNA sequence, a guide RNA sequence, or a piwi RNA sequence.

Promoters

In some embodiments, one or more promoter or enhancer elements are operably linked to a nucleic acid encoding a gene modifying protein or a template nucleic acid, e.g., that controls expression

of the heterologous object sequence. In certain embodiments, the one or more promoter or enhancer elements comprise cell-type or tissue specific elements. In some embodiments, the promoter or enhancer is the same or derived from the promoter or enhancer that naturally controls expression of the heterologous object sequence. For example, the ornithine transcarbomylase promoter and enhancer may be used to control expression of the ornithine transcarbomylase gene in a system or method provided by the invention for correcting ornithine transcarbomylase deficiencies. In some embodiments, the promoter is a promoter of Table 16 or 17 or a functional fragment or variant thereof.

Exemplary tissue specific promoters that are commercially available can be found, for example, at a uniform resource locator (e.g., www.invivogen.com/tissue-specific-promoters). In some embodiments, a promoter is a native promoter or a minimal promoter, e.g., which consists of a single fragment from the 5′ region of a given gene. In some embodiments, a native promoter comprises a core promoter and its natural 5′ UTR. In some embodiments, the 5′ UTR comprises an intron. In other embodiments, these include composite promoters, which combine promoter elements of different origins or were generated by assembling a distal enhancer with a minimal promoter of the same origin.

Exemplary cell or tissue specific promoters are provided in the tables, below, and exemplary nucleic acid sequences encoding them are known in the art and can be readily accessed using a variety of resources, such as the NCBI database, including RefSeq, as well as the Eukaryotic Promoter Database (//epd.epfl.ch//index.php).

Table 16. Exemplary cell or tissue-specific promoters

Promoter	**Target cells**
B29 Promoter	B cells
CD14 Promoter	Monocytic Cells
CD43 Promoter	Leukocytes and platelets
CD45 Promoter	Hematopoeitic cells
CD68 promoter	macrophages
Desmin promoter	muscle cells
Elastase-1 promoter	pancreatic acinar cells
Endoglin promoter	endothelial cells
fibronectin promoter	differentiating cells, healing tissue
Flt-1 promoter	endothelial cells
GFAP promoter	Astrocytes
GPIIB promoter	megakaryocytes
ICAM-2 Promoter	Endothelial cells
INF-Beta promoter	Hematopoeitic cells

Mb promoter	muscle cells
Nphs1 promoter	podocytes
OG-2 promoter	Osteoblasts, Odonblasts
SP-B promoter	Lung
Syn1 promoter	Neurons
WASP promoter	Hematopoeitic cells
SV40/bAlb promoter	Liver
SV40/bAlb promoter	Liver
SV40/Cd3 promoter	Leukocytes and platelets
SV40/CD45 promoter	hematopoeitic cells
NSE/RU5' promoter	Mature Neurons

Table 17. Additional exemplary cell or tissue-specific promoters

Promoter	**Gene Description**	**Gene Specificity**
APOA2	Apolipoprotein A-II	Hepatocytes (from hepatocyte progenitors)
SERPINA1 (hAAT)	Serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1 (also named alpha 1 anti-tryps in)	Hepatocytes (from definitive endoderm stage)
CYP3A	Cytochrome P450, family 3, subfamily A, polypeptide	Mature Hepatocytes

MIR122	MicroRNA 122	Hepatocytes (from early stage embryonic liver cells) and endoderm
Pancreatic specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
INS	Insulin	Pancreatic beta cells (from definitive endoderm stage)
IRS2	Insulin receptor substrate 2	Pancreatic beta cells
Pdx1	Pancreatic and duodenal homeobox 1	Pancreas (from definitive endoderm stage)
Alx3	Aristaless-like homeobox 3	Pancreatic beta cells (from definitive endoderm stage)
Ppy	Pancreatic polypeptide	PP pancreatic cells (gamma cells)
Cardiac specific promoters		
Promoter	**Gene Description**	**Gene Specificity**

Myh6 (aMHC)	Myosin, heavy chain 6, cardiac muscle, alpha	Late differentiation marker of cardiac muscle cells (atrial specificity)
MYL2 (MLC-2v)	Myosin, light chain 2, regulatory, cardiac, slow	Late differentiation marker of cardiac muscle cells (ventricular specificity)
ITNNl3 (cTnl)	Troponin I type 3 (cardiac)	Cardiomyocytes (from immature state)
ITNNl3 (cTnl)	Troponin I type 3 (cardiac)	Cardiomyocytes (from immature state)
NPPA (ANF)	Natriuretic peptide precursor A (also named Atrial Natriuretic Factor)	Atrial specificity in adult cells
Slc8a1 (Ncx1)	Solute carrier family 8 (sodium/calcium exchanger), member 1	Cardiomyocytes from early developmental stages
CNS specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
SYN1 (hSyn)	Synapsin I	Neurons
GFAP	Glial fibrillary acidic protein	Astrocytes
INA	lnternexin neuronal intermediate filament protein, alpha (a-internexin)	Neuroprogenitors
NES	Nestin	Neuroprogenitors and ectoderm
MOBP	Myelin-associated oligodendrocyte basic protein	Oligodendrocytes
MBP	Myelin basic protein	Oligodendrocytes
TH	Tyrosine hydroxylase	Dopaminergic neurons
FOXA2 (HNF3 beta)	Forkhead box A2	Dopaminergic neurons (also used as a marker of endoderm)
Skin specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
FLG	Filaggrin	Keratinocytes from granular layer
K14	Keratin 14	Keratinocytes from granular and basal layers

TGM3	Transglutaminase 3	Keratinocytes from granular layer
Immune cell specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
ITGAM (CD11B)	lntegrin, alpha M (complement component 3 receptor 3 subunit)	Monocytes, macrophages , granulocytes , natural killer cells
Urogential cell specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
Pbsn	Probasin	Prostatic epithelium
Upk2	Uroplakin 2	Bladder
Sbp	Spermine binding protein	Prostate
Fer1l4	Fer-1-like 4	Bladder
Endothelial cell specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
ENG	Endoglin	Endothelial cells
Pluripotent and embryonic cell specific promoters		
Promoter	**Gene Description**	**Gene Specificity**
Oct4 (POU5F1)	POU class 5 homeobox 1	Pluripotent cells (germ cells, ES cells, iPS cells)
NANOG	Nanog homeobox	Pluripotent cells (ES cells, iPS cells)
Synthetic Oct4	Synthetic promoter based on a Oct-4 core enhancer element	Pluripotent cells (ES cells, iPS cells)
T brachyury	Brachyury	Mesoderm
NES	Nestin	Neuroprogenitors and Ectoderm
SOX17	SRY (sex determining region Y)-box 17	Endoderm

FOXA2 (HNFJ beta)	Forkhead box A2	Endoderm (also used as a marker of dopaminergic neurons)
MIR122	MicroRNA 122	Endoderm and hepatocytes (from early stage embryonic liver cells~

Depending on the host/vector system utilized, any of a number of suitable transcription and translation control elements, including constitutive and inducible promoters, transcription enhancer elements, transcription terminators, etc. may be used in the expression vector (see e.g., Bitter et al. (1987) *Methods in Enzymology,* 153:516-544; incorporated herein by reference in its entirety).

In some embodiments, a nucleic acid encoding a gene modifying protein or template nucleic acid is operably linked to a control element, e.g., a transcriptional control element, such as a promoter. The transcriptional control element may, in some embodiment, be functional in either a eukaryotic cell, e.g., a mammalian cell; or a prokaryotic cell (e.g., bacterial or archaeal cell). In some embodiments, a nucleotide sequence encoding a polypeptide is operably linked to multiple control elements, e.g., that allow expression of the nucleotide sequence encoding the polypeptide in both prokaryotic and eukaryotic cells.

For illustration purposes, examples of spatially restricted promoters include, but are not limited to, neuron-specific promoters, adipocyte-specific promoters, cardiomyocyte-specific promoters, smooth muscle-specific promoters, photoreceptor-specific promoters, etc. Neuron-specific spatially restricted promoters include, but are not limited to, a neuron-specific enolase (NSE) promoter (see, e.g., EMBL HSENO2, X51956); an aromatic amino acid decarboxylase (AADC) promoter, a neurofilament promoter (see, e.g., GenBank HUMNFL, L04147); a synapsin promoter (see, e.g., GenBank HUMSYNIB, M55301); a thy-1 promoter (see, e.g., Chen et al. (1987) Cell 51:7-19; and Llewellyn, et al. (2010) Nat. Med. 16(10):1161-1166); a serotonin receptor promoter (see, e.g., GenBank S62283); a tyrosine hydroxylase promoter (TH) (see, e.g., Oh et al. (2009) Gene Ther 16:437; Sasaoka et al. (1992) Mol. Brain Res. 16:274; Boundy et al. (1998) J. Neurosci. 18:9989; and Kaneda et al. (1991) Neuron 6:583-594); a GnRH promoter (see, e.g., Radovick et al. (1991) Proc. Natl. Acad. Sci. USA 88:3402-3406); an L7 promoter (see, e.g., Oberdick et al. (1990) Science 248:223-226); a DNMT promoter (see, e.g., Bartge

et al. (1988) Proc. Natl. Acad. Sci. USA 85:3648-3652); an enkephalin promoter (see, e.g., Comb et al. (1988) EMBO J. 17:3793-3805); a myelin basic protein (MBP) promoter; a Ca2+-calmodulin-dependent protein kinase II-alpha (CamKIIα) promoter (see, e.g., Mayford et al. (1996) Proc. Natl. Acad. Sci. USA 93:13250; and Casanova et al. (2001) Genesis 31:37); a CMV enhancer/platelet-derived growth factor-β promoter (see, e.g., Liu et al. (2004) Gene Therapy 11:52-60); and the like.

Adipocyte-specific spatially restricted promoters include, but are not limited to, the aP2 gene promoter/enhancer, e.g., a region from −5.4 kb to +21 bp of a human aP2 gene (see, e.g., Tozzo et al. (1997) Endocrinol. 138:1604; Ross et al. (1990) Proc. Natl. Acad. Sci. USA 87:9590; and Pavjani et al. (2005) Nat. Med. 11:797); a glucose transporter-4 (GLUT4) promoter (see, e.g., Knight et al. (2003) Proc. Natl. Acad. Sci. USA 100:14725); a fatty acid translocase (FAT/CD36) promoter (see, e.g., Kuriki et al. (2002) Biol. Pharm. Bull. 25:1476; and Sato et al. (2002) J. Biol. Chem. 277:15703); a stearoyl-CoA desaturase-1 (SCD1) promoter (Tabor et al. (1999) J. Biol. Chem. 274:20603); a leptin promoter (see, e.g., Mason et al. (1998) Endocrinol. 139:1013; and Chen et al. (1999) Biochem. Biophys. Res. Comm. 262:187); an adiponectin promoter (see, e.g., Kita et al. (2005) Biochem. Biophys. Res. Comm. 331:484; and Chakrabarti (2010) Endocrinol. 151:2408); an adipsin promoter (see, e.g., Platt et al. (1989) Proc. Natl. Acad. Sci. USA 86:7490); a resistin promoter (see, e.g., Seo et al. (2003) Molec. Endocrinol. 17:1522); and the like.

Cardiomyocyte-specific spatially restricted promoters include, but are not limited to, control sequences derived from the following genes: myosin light chain-2, α-myosin heavy chain, AE3, cardiac troponin C, cardiac actin, and the like. Franz et al. (1997) Cardiovasc. Res. 35:560-566; Robbins et al. (1995) Ann. N.Y. Acad. Sci. 752:492-505; Linn et al. (1995) Circ. Res. 76:584-591; Parmacek et al. (1994) Mol. Cell. Biol. 14:1870-1885; Hunter et al. (1993) Hypertension 22:608-617; and Sartorelli et al. (1992) Proc. Natl. Acad. Sci. USA 89:4047-4051.

Smooth muscle-specific spatially restricted promoters include, but are not limited to, an SM22α promoter (see, e.g., Akyürek et al. (2000) Mol. Med. 6:983; and U.S. Pat. No. 7,169,874); a smoothelin promoter (see, e.g., WO 2001/018048); an α-smooth muscle actin promoter; and the like. For example, a 0.4 kb region of the SM22α promoter, within which lie two CArG elements, has been shown to mediate vascular smooth muscle cell-specific expression (see, e.g., Kim, et al. (1997) Mol. Cell. Biol. 17, 2266-2278; Li, et al., (1996) J. Cell Biol. 132, 849-859; and Moessler, et al. (1996) Development 122, 2415-2425).

Photoreceptor-specific spatially restricted promoters include, but are not limited to, a rhodopsin promoter; a rhodopsin kinase promoter (Young et al. (2003) Ophthalmol. Vis. Sci. 44:4076); a beta phosphodiesterase gene promoter (Nicoud et al. (2007) J. Gene Med. 9:1015); a retinitis pigmentosa gene promoter (Nicoud et al. (2007) supra); an interphotoreceptor retinoid-binding protein (IRBP) gene

enhancer (Nicoud et al. (2007) supra); an IRBP gene promoter (Yokoyama et al. (1992) Exp Eye Res. 55:225); and the like.

In some embodiments, a gene modifying system, e.g., DNA encoding a gene modifying polypeptide, DNA encoding a template RNA, or DNA or RNA encoding a heterologous object sequence, is designed such that one or more elements is operably linked to a tissue-specific promoter, e.g., a promoter that is active in T-cells. In further embodiments, the T-cell active promoter is inactive in other cell types, e.g., B-cells, NK cells. In some embodiments, the T-cell active promoter is derived from a promoter for a gene encoding a component of the T-cell receptor, e.g., TRAC, TRBC, TRGC, TRDC. In some embodiments, the T-cell active promoter is derived from a promoter for a gene encoding a component of a T-cell-specific cluster of differentiation protein, e.g., CD3, e.g., CD3D, CD3E, CD3G, CD3Z. In some embodiments, T-cell-specific promoters in gene modifying systems are discovered by comparing publicly available gene expression data across cell types and selecting promoters from the genes with enhanced expression in T-cells. In some embodiments, promoters may be selecting depending on the desired expression breadth, e.g., promoters that are active in T-cells only, promoters that are active in NK cells only, promoters that are active in both T-cells and NK cells.

Cell-specific promoters known in the art may be used to direct expression of a gene modifying protein, e.g., as described herein. Nonlimiting exemplary mammalian cell-specific promoters have been characterized and used in mice expressing Cre recombinase in a cell-specific manner. Certain nonlimiting exemplary mammalian cell-specific promoters are listed in Table 1 of US9845481, incorporated herein by reference.

In some embodiments, a vector as described herein comprises an expression cassette. Typically, an expression cassette comprises the nucleic acid molecule of the instant invention operatively linked to a promoter sequence. For example, a promoter is operatively linked with a coding sequence when it is capable of affecting the expression of that coding sequence (e.g., the coding sequence is under the transcriptional control of the promoter). Encoding sequences can be operatively linked to regulatory sequences in sense or antisense orientation. In certain embodiments, the promoter is a heterologous promoter. In certain embodiments, an expression cassette may comprise additional elements, for example, an intron, an enhancer, a polyadenylation site, a woodchuck response element (WRE), and/or other elements known to affect expression levels of the encoding sequence. A promoter typically controls the expression of a coding sequence or functional RNA. In certain embodiments, a promoter sequence comprises proximal and more distal upstream elements and can further comprise an enhancer element. An enhancer can typically stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. In certain

embodiments, the promoter is derived in its entirety from a native gene. In certain embodiments, the promoter is composed of different elements derived from different naturally occurring promoters. In certain embodiments, the promoter comprises a synthetic nucleotide sequence. It will be understood by those skilled in the art that different promoters will direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions or to the presence or the absence of a drug or transcriptional co-factor. Ubiquitous, cell-type-specific, tissue-specific, developmental stage-specific, and conditional promoters, for example, drug-responsive promoters (e.g., tetracycline-responsive promoters) are well known to those of skill in the art. Exemplary promoters include, but are not limited to, the phosphoglycerate kinase (PKG) promoter, CAG (composite of the CMV enhancer the chicken beta actin promoter (CBA) and the rabbit beta globin intron), NSE (neuronal specific enolase), synapsin or NeuN promoters, the SV40 early promoter, mouse mammary tumor virus LTR promoter; adenovirus major late promoter (Ad MLP), a herpes simplex virus (HSV) promoter, a cytomegalovirus (CMV) promoter such as the CMV immediate early promoter region (CMVIE), SFFV promoter, rous sarcoma virus (RSV) promoter, synthetic promoters, hybrid promoters, and the like. Other promoters can be of human origin or from other species, including from mice. Common promoters include, e.g., the human cytomegalovirus (CMV) immediate early gene promoter, the SV40 early promoter, the Rous sarcoma virus long terminal repeat, [beta]- actin, rat insulin promoter, the phosphoglycerate kinase promoter, the human alpha-1 antitrypsin (hAAT) promoter, the transthyretin promoter, the TBG promoter and other liver-specific promoters, the desmin promoter and similar muscle-specific promoters, the EF1 -alpha promoter, hybrid promoters with multi-tissue specificity, promoters specific for neurons like synapsin and glyceraldehyde-3 -phosphate dehydrogenase promoter, all of which are promoters well known and readily available to those of skill in the art, can be used to obtain high-level expression of the coding sequence of interest. In addition, sequences derived from non-viral genes, such as the murine metallothionein gene, will also find use herein. Such promoter sequences are commercially available from, e.g., Stratagene (San Diego, CA). Additional exemplary promoter sequences are described, for example, in WO2018213786A1 (incorporated by reference herein in its entirety).

In some embodiments, the apolipoprotein E enhancer (ApoE) or a functional fragment thereof is used, e.g., to drive expression in the liver. In some embodiments, two copies of the ApoE enhancer or a functional fragment thereof are used. In some embodiments, the ApoE enhancer or functional fragment thereof is used in combination with a promoter, e.g., the human alpha-1 antitrypsin (hAAT) promoter.

In some embodiments, the regulatory sequences impart tissue-specific gene expression capabilities. In some cases, the tissue-specific regulatory sequences bind tissue-specific transcription factors that induce transcription in a tissue specific manner. Various tissue-specific regulatory sequences (e.g., promoters, enhancers, etc.) are known in the art. Exemplary tissue-specific regulatory sequences

include, but are not limited to, the following tissue-specific promoters: a liver-specific thyroxin binding globulin (TBG) promoter, a insulin promoter, a glucagon promoter, a somatostatin promoter, a pancreatic polypeptide (PPY) promoter, a synapsin-1 (Syn) promoter, a creatine kinase (MCK) promoter, a mammalian desmin (DES) promoter, a α-myosin heavy chain (a-MHC) promoter, or a cardiac Troponin T (cTnT) promoter. Other exemplary promoters include Beta-actin promoter, hepatitis B virus core promoter, Sandig et al., Gene Ther., 3:1002-9 (1996); alpha-fetoprotein (AFP) promoter, Arbuthnot et al., Hum. Gene Ther., 7:1503-14 (1996)), bone osteocalcin promoter (Stein et al., Mol. Biol. Rep., 24:185-96 (1997)); bone sialoprotein promoter (Chen et al., J. Bone Miner. Res., 11:654-64 (1996)), CD2 promoter (Hansal et al., J. Immunol., 161:1063-8 (1998); immunoglobulin heavy chain promoter; T cell receptor α-chain promoter, neuronal such as neuron-specific enolase (NSE) promoter (Andersen et al., Cell. Mol. Neurobiol., 13:503-15 (1993)), neurofilament light-chain gene promoter (Piccioli et al., Proc. Natl. Acad. Sci. USA, 88:5611-5 (1991)), and the neuron-specific vgf gene promoter (Piccioli et al., Neuron, 15:373-84 (1995)), and others. Additional exemplary promoter sequences are described, for example, in U.S. Patent No. 10300146 (incorporated herein by reference in its entirety). In some embodiments, a tissue-specific regulatory element, e.g., a tissue-specific promoter, is selected from one known to be operably linked to a gene that is highly expressed in a given tissue, e.g., as measured by RNA-seq or protein expression data, or a combination thereof. Methods for analyzing tissue specificity by expression are taught in Fagerberg et al. *Mol Cell Proteomics* 13(2):397-406 (2014), which is incorporated herein by reference in its entirety.

In some embodiments, a vector described herein is a multicistronic expression construct. Multicistronic expression constructs include, for example, constructs harboring a first expression cassette, e.g. comprising a first promoter and a first encoding nucleic acid sequence, and a second expression cassette, e.g. comprising a second promoter and a second encoding nucleic acid sequence. Such multicistronic expression constructs may, in some instances, be particularly useful in the delivery of non-translated gene products, such as hairpin RNAs, together with a polypeptide, for example, a gene modifying polypeptide and gene modifying template. In some embodiments, multicistronic expression constructs may exhibit reduced expression levels of one or more of the included transgenes, for example, because of promoter interference or the presence of incompatible nucleic acid elements in close proximity. If a multicistronic expression construct is part of a viral vector, the presence of a self-complementary nucleic acid sequence may, in some instances, interfere with the formation of structures necessary for viral reproduction or packaging.

In some embodiments, the sequence encodes an RNA with a hairpin. In some embodiments, the hairpin RNA is a guide RNA, a template RNA, a shRNA, or a microRNA. In some embodiments, the first promoter is an RNA polymerase I promoter. In some embodiments, the first promoter is an RNA

polymerase II promoter. In some embodiments, the second promoter is an RNA polymerase III promoter. In some embodiments, the second promoter is a U6 or H1 promoter.

Without wishing to be bound by theory, multicistronic expression constructs may not achieve optimal expression levels as compared to expression systems containing only one cistron. One of the suggested causes of lower expression levels achieved with multicistronic expression constructs comprising two or more promoter elements is the phenomenon of promoter interference (see, e.g., Curtin J A, Dane A P, Swanson A, Alexander I E, Ginn S L. *Bidirectional promoter interference between two widely used internal heterologous promoters in a late-generation lentiviral construct*. Gene Ther. 2008 March; 15(5):384-90; and Martin-Duque P, Jezzard S, Kaftansis L, Vassaux G. *Direct comparison of the insulating properties of two genetic elements in an adenoviral vector containing two different expression cassettes*. Hum Gene Ther. 2004 October; 15(10):995-1002; both references incorporated herein by reference for disclosure of promoter interference phenomenon). In some embodiments, the problem of promoter interference may be overcome, e.g., by producing multicistronic expression constructs comprising only one promoter driving transcription of multiple encoding nucleic acid sequences separated by internal ribosomal entry sites, or by separating cistrons comprising their own promoter with transcriptional insulator elements. In some embodiments, single-promoter driven expression of multiple cistrons may result in uneven expression levels of the cistrons. In some embodiments, a promoter cannot efficiently be isolated and isolation elements may not be compatible with some gene transfer vectors, for example, some retroviral vectors.

MicroRNAs

MicroRNAs (miRNAs) and other small interfering nucleic acids generally regulate gene expression via target RNA transcript cleavage/degradation or translational repression of the target messenger RNA (mRNA). miRNAs may, in some instances, be natively expressed, typically as final 19-25 non-translated RNA products. miRNAs generally exhibit their activity through sequence-specific interactions with the 3′ untranslated regions (UTR) of target mRNAs. These endogenously expressed miRNAs may form hairpin precursors that are subsequently processed into an miRNA duplex, and further into a mature single stranded miRNA molecule This mature miRNA generally guides a multiprotein complex, miRISC, which identifies target 3′ UTR regions of target mRNAs based upon their complementarity to the mature miRNA. Useful transgene products may include, for example, miRNAs or miRNA binding sites that regulate the expression of a linked polypeptide. A non-limiting list of miRNA genes; the products of these genes and their homologues are useful as transgenes or as targets for small interfering nucleic acids (e.g., miRNA sponges, antisense oligonucleotides), e.g., in methods such as

those listed in US10300146, 22:25-25:48, are herein incorporated by reference. In some embodiments, one or more binding sites for one or more of the foregoing miRNAs are incorporated in a transgene, e.g., a transgene delivered by a rAAV vector, e.g., to inhibit the expression of the transgene in one or more tissues of an animal harboring the transgene. In some embodiments, a binding site may be selected to control the expression of a transgene in a tissue specific manner. For example, binding sites for the liver-specific miR-122 may be incorporated into a transgene to inhibit expression of that transgene in the liver. Additional exemplary miRNA sequences are described, for example, in U.S. Patent No. 10,300,146 (incorporated herein by reference in its entirety).

An miR inhibitor or miRNA inhibitor is generally an agent that blocks miRNA expression and/or processing. Examples of such agents include, but are not limited to, microRNA antagonists, microRNA specific antisense, microRNA sponges, and microRNA oligonucleotides (double-stranded, hairpin, short oligonucleotides) that inhibit miRNA interaction with a Drosha complex. MicroRNA inhibitors, e.g., miRNA sponges, can be expressed in cells from transgenes (e.g., as described in Ebert, M. S. Nature Methods, Epub Aug. 12, 2007; incorporated by reference herein in its entirety). In some embodiments, microRNA sponges, or other miR inhibitors, are used with the AAVs. microRNA sponges generally specifically inhibit miRNAs through a complementary heptameric seed sequence. In some embodiments, an entire family of miRNAs can be silenced using a single sponge sequence. Other methods for silencing miRNA function (derepression of miRNA targets) in cells will be apparent to one of ordinary skill in the art.

In some embodiments, a gene modifying system, template RNA, or polypeptide described herein is administered to or is active in (e.g., is more active in) a target tissue, e.g., a first tissue. In some embodiments, the gene modifying system, template RNA, or polypeptide is not administered to or is less active in (e.g., not active in) a non-target tissue. In some embodiments, a gene modifying system, template RNA, or polypeptide described herein is useful for modifying DNA in a target tissue, e.g., a first tissue, (e.g., and not modifying DNA in a non-target tissue).

In some embodiments, a gene modifying system comprises (a) a polypeptide described herein or a nucleic acid encoding the same, (b) a template nucleic acid (e.g., template RNA) described herein, and (c) one or more first tissue-specific expression-control sequences specific to the target tissue, wherein the one or more first tissue-specific expression-control sequences specific to the target tissue are in operative association with (a), (b), or (a) and (b), wherein, when associated with (a), (a) comprises a nucleic acid encoding the polypeptide.

In some embodiments, the nucleic acid in (b) comprises RNA.

In some embodiments, the nucleic acid in (b) comprises DNA.

In some embodiments, the nucleic acid in (b): (i) is single-stranded or comprises a single-stranded segment, e.g., is single-stranded DNA or comprises a single-stranded segment and one or more double stranded segments; (ii) has inverted terminal repeats; or (iii) both (i) and (ii).

In some embodiments, the nucleic acid in (b) is double-stranded or comprises a double-stranded segment.

In some embodiments, (a) comprises a nucleic acid encoding the polypeptide.

In some embodiments, the nucleic acid in (a) comprises RNA.

In some embodiments, the nucleic acid in (a) comprises DNA.

In some embodiments, the nucleic acid in (a): (i) is single-stranded or comprises a single-stranded segment, e.g., is single-stranded DNA or comprises a single-stranded segment and one or more double stranded segments; (ii) has inverted terminal repeats; or (iii) both (i) and (ii).

In some embodiments, the nucleic acid in (a) is double-stranded or comprises a double-stranded segment.

In some embodiments, the nucleic acid in (a), (b), or (a) and (b) is linear.

In some embodiments, the nucleic acid in (a), (b), or (a) and (b) is circular, e.g., a plasmid or minicircle.

In some embodiments, the heterologous object sequence is in operative association with a first promoter.

In some embodiments, the one or more first tissue-specific expression-control sequences comprises a tissue specific promoter.

In some embodiments, the tissue-specific promoter comprises a first promoter in operative association with: (i) the heterologous object sequence, (ii) a nucleic acid encoding the retroviral RT, or (iii) (i) and (ii).

In some embodiments, the one or more first tissue-specific expression-control sequences comprises a tissue-specific microRNA recognition sequence in operative association with: (i) the heterologous object sequence, (ii) a nucleic acid encoding the retroviral RT domain, or (iii) (i) and (ii).

In some embodiments, a system comprises a tissue-specific promoter, and the system further comprises one or more tissue-specific microRNA recognition sequences, wherein: (i) the tissue specific promoter is in operative association with: (I) the heterologous object sequence, (II) a nucleic acid encoding the retroviral RT domain, or (III) (I) and (II); and/or (ii) the one or more tissue-specific microRNA recognition sequences are in operative association with: (I) the heterologous object sequence, (II) a nucleic acid encoding the retroviral RT, or (III) (I) and (II).

In some embodiments, wherein (a) comprises a nucleic acid encoding the polypeptide, the nucleic acid comprises a promoter in operative association with the nucleic acid encoding the polypeptide.

In some embodiments, the nucleic acid encoding the polypeptide comprises one or more second tissue-specific expression-control sequences specific to the target tissue in operative association with the polypeptide coding sequence.

In some embodiments, the one or more second tissue-specific expression-control sequences comprises a tissue specific promoter.

In some embodiments, the tissue-specific promoter is the promoter in operative association with the nucleic acid encoding the polypeptide.

In some embodiments, the one or more second tissue-specific expression-control sequences comprises a tissue-specific microRNA recognition sequence.

In some embodiments, the promoter in operative association with the nucleic acid encoding the polypeptide is a tissue-specific promoter, the system further comprising one or more tissue-specific microRNA recognition sequences.

In some embodiments, a nucleic acid component of a system provided by the invention is a sequence (e.g., encoding the polypeptide or comprising a heterologous object sequence) flanked by untranslated regions (UTRs) that modify protein expression levels. Various 5′ and 3′ UTRs can affect protein expression. For example, in some embodiments, the coding sequence may be preceded by a 5′ UTR that modifies RNA stability or protein translation. In some embodiments, the sequence may be followed by a 3′ UTR that modifies RNA stability or translation. In some embodiments, the sequence may be preceded by a 5′ UTR and followed by a 3′ UTR that modify RNA stability or translation. In some embodiments, the 5′ and/or 3′ UTR may be selected from the 5′ and 3′ UTRs of complement factor 3 (C3) (CACTCCTCCCCATCCTCTCCCTCTGTCCCTCTGTCCCTCTGACCCTGCACTGTCCCAGCACC) or orosomucoid 1 (ORM1) (CAGGACACAGCCTTGGATCAGGACAGAGACTTGGGGGCCATCCTGCCCCTCCAACCCGACA TGTGTACCTCAGCTTTTTCCCTCACTTGCATCAATAAAGCTTCTGTGTTTGGAACAGCTAA) (Asrani et al. RNA Biology 2018). In certain embodiments, the 5′ UTR is the 5′ UTR from C3 and the 3′ UTR is the 3′ UTR from ORM1. In certain embodiments, a 5′ UTR and 3′ UTR for protein expression, e.g., mRNA (or DNA encoding the RNA) for a gene modifying polypeptide or heterologous object sequence, comprise optimized expression sequences. In some embodiments, the 5′ UTR comprises GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACC and/or the 3′ UTR comprising UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCCAGCCC

CUCCUCCCCUUCCUGCACCCGUACCCCCGUGGUCUUUGAAUAAAGUCUGA, e.g., as described in Richner et al. *Cell* 168(6): P1114-1125 (2017), the sequences of which are incorporated herein by reference.

In some embodiments, a 5′ and/or 3′ UTR may be selected to enhance protein expression. In some embodiments, a 5′ and/or 3′ UTR may be selected to modify protein expression such that overproduction inhibition is minimized. In some embodiments, UTRs are around a coding sequence, e.g., outside the coding sequence and in other embodiments proximal to the coding sequence, In some embodiments, additional regulatory elements (e.g., miRNA binding sites, cis-regulatory sites) are included in the UTRs.

In some embodiments, an open reading frame of a gene modifying system, e.g., an ORF of an mRNA (or DNA encoding an mRNA) encoding a gene modifying polypeptide or one or more ORFs of an mRNA (or DNA encoding an mRNA) of a heterologous object sequence, is flanked by a 5′ and/or 3′ untranslated region (UTR) that enhances the expression thereof. In some embodiments, the 5′ UTR of an mRNA component (or transcript produced from a DNA component) of the system comprises the sequence 5′-GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACC-3′. In some embodiments, the 3′ UTR of an mRNA component (or transcript produced from a DNA component) of the system comprises the sequence 5′-UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCCAGCCCCUCCUCCCCUUCCUGCACCCGUACCCCCGUGGUCUUUGAAUAAAGUCUGA-3′. This combination of 5′ UTR and 3′ UTR has been shown to result in desirable expression of an operably linked ORF by Richner et al. *Cell* 168(6): P1114-1125 (2017), the teachings and sequences of which are incorporated herein by reference. In some embodiments, a system described herein comprises a DNA encoding a transcript, wherein the DNA comprises the corresponding 5′ UTR and 3′ UTR sequences, with T substituting for U in the above-listed sequence). In some embodiments, a DNA vector used to produce an RNA component of the system further comprises a promoter upstream of the 5′ UTR for initiating *in vitro* transcription, e.g, a T7, T3, or SP6 promoter. The 5′ UTR above begins with GGG, which is a suitable start for optimizing transcription using T7 RNA polymerase. For tuning transcription levels and altering the transcription start site nucleotides to fit alternative 5′ UTRs, the teachings of Davidson et al. *Pac Symp Biocomput* 433-443 (2010) describe T7 promoter variants, and the methods of discovery thereof, that fulfill both of these traits.

Viral vectors and components thereof

Viruses are a useful source of delivery vehicles for the systems described herein, in addition to a source of relevant enzymes or domains as described herein, e.g., as sources of polymerases and

polymerase functions used herein, e.g., DNA-dependent DNA polymerase, RNA-dependent RNA polymerase, RNA-dependent DNA polymerase, DNA-dependent RNA polymerase, reverse transcriptase. Some enzymes, e.g., reverse transcriptases, may have multiple activities, e.g., be capable of both RNA-dependent DNA polymerization and DNA-dependent DNA polymerization, e.g., first and second strand synthesis. In some embodiments, the virus used as a gene modifying delivery system or a source of components thereof may be selected from a group as described by Baltimore *Bacteriol Rev* 35(3):235-241 (1971).

In some embodiments, the virus is selected from a Group I virus, e.g., is a DNA virus and packages dsDNA into virions. In some embodiments, the Group I virus is selected from, e.g., Adenoviruses, Herpesviruses, Poxviruses.

In some embodiments, the virus is selected from a Group II virus, e.g., is a DNA virus and packages ssDNA into virions. In some embodiments, the Group II virus is selected from, e.g., Parvoviruses. In some embodiments, the parvovirus is a dependoparvovirus, e.g., an adeno-associated virus (AAV).

In some embodiments, the virus is selected from a Group III virus, e.g., is an RNA virus and packages dsRNA into virions. In some embodiments, the Group III virus is selected from, e.g., Reoviruses. In some embodiments, one or both strands of the dsRNA contained in such virions is a coding molecule able to serve directly as mRNA upon transduction into a host cell, e.g., can be directly translated into protein upon transduction into a host cell without requiring any intervening nucleic acid replication or polymerization steps.

In some embodiments, the virus is selected from a Group IV virus, e.g., is an RNA virus and packages ssRNA(+) into virions. In some embodiments, the Group IV virus is selected from, e.g., Coronaviruses, Picornaviruses, Togaviruses. In some embodiments, the ssRNA(+) contained in such virions is a coding molecule able to serve directly as mRNA upon transduction into a host cell, e.g., can be directly translated into protein upon transduction into a host cell without requiring any intervening nucleic acid replication or polymerization steps.

In some embodiments, the virus is selected from a Group V virus, e.g., is an RNA virus and packages ssRNA(-) into virions. In some embodiments, the Group V virus is selected from, e.g., Orthomyxoviruses, Rhabdoviruses. In some embodiments, an RNA virus with an ssRNA(-) genome also carries an enzyme inside the virion that is transduced to host cells with the viral genome, e.g., an RNA-dependent RNA polymerase, capable of copying the ssRNA(-) into ssRNA(+) that can be translated directly by the host.

In some embodiments, the virus is selected from a Group VI virus, e.g., is a retrovirus and packages ssRNA(+) into virions. In some embodiments, the Group VI virus is selected from, e.g.,

retroviruses. In some embodiments, the retrovirus is a lentivirus, e.g., HIV-1, HIV-2, SIV, BIV. In some embodiments, the retrovirus is a spumavirus, e.g., a foamy virus, e.g., HFV, SFV, BFV. In some embodiments, the ssRNA(+) contained in such virions is a coding molecule able to serve directly as mRNA upon transduction into a host cell, e.g., can be directly translated into protein upon transduction into a host cell without requiring any intervening nucleic acid replication or polymerization steps. In some embodiments, the ssRNA(+) is first reverse transcribed and copied to generate a dsDNA genome intermediate from which mRNA can be transcribed in the host cell. In some embodiments, an RNA virus with an ssRNA(+) genome also carries an enzyme inside the virion that is transduced to host cells with the viral genome, e.g., an RNA-dependent DNA polymerase, capable of copying the ssRNA(+) into dsDNA that can be transcribed into mRNA and translated by the host. In some embodiments, the reverse transcriptase from a Group VI retrovirus is incorporated as the reverse transcriptase domain of a gene modifying polypeptide.

In some embodiments, the virus is selected from a Group VII virus, e.g., is a retrovirus and packages dsRNA into virions. In some embodiments, the Group VII virus is selected from, e.g., Hepadnaviruses. In some embodiments, one or both strands of the dsRNA contained in such virions is a coding molecule able to serve directly as mRNA upon transduction into a host cell, e.g., can be directly translated into protein upon transduction into a host cell without requiring any intervening nucleic acid replication or polymerization steps. In some embodiments, one or both strands of the dsRNA contained in such virions is first reverse transcribed and copied to generate a dsDNA genome intermediate from which mRNA can be transcribed in the host cell. In some embodiments, an RNA virus with a dsRNA genome also carries an enzyme inside the virion that is transduced to host cells with the viral genome, e.g., an RNA-dependent DNA polymerase, capable of copying the dsRNA into dsDNA that can be transcribed into mRNA and translated by the host. In some embodiments, the reverse transcriptase from a Group VII retrovirus is incorporated as the reverse transcriptase domain of a gene modifying polypeptide.

In some embodiments, virions used to deliver nucleic acid in this invention may also carry enzymes involved in the process of gene modification. For example, a retroviral virion may contain a reverse transcriptase domain that is delivered into a host cell along with the nucleic acid. In some embodiments, an RNA template may be associated with a gene modifying polypeptide within a virion, such that both are co-delivered to a target cell upon transduction of the nucleic acid from the viral particle. In some embodiments, the nucleic acid in a virion may comprise DNA, e.g., linear ssDNA, linear dsDNA, circular ssDNA, circular dsDNA, minicircle DNA, dbDNA, ceDNA. In some embodiments, the nucleic acid in a virion may comprise RNA, e.g., linear ssRNA, linear dsRNA, circular ssRNA, circular dsRNA. In some embodiments, a viral genome may circularize upon transduction into a host cell, e.g., a linear ssRNA molecule may undergo a covalent linkage to form a circular ssRNA, a

linear dsRNA molecule may undergo a covalent linkage to form a circular dsRNA or one or more circular ssRNA. In some embodiments, a viral genome may replicate by rolling circle replication in a host cell. In some embodiments, a viral genome may comprise a single nucleic acid molecule, e.g., comprise a non-segmented genome. In some embodiments, a viral genome may comprise two or more nucleic acid molecules, e.g., comprise a segmented genome. In some embodiments, a nucleic acid in a virion may be associated with one or proteins. In some embodiments, one or more proteins in a virion may be delivered to a host cell upon transduction. In some embodiments, a natural virus may be adapted for nucleic acid delivery by the addition of virion packaging signals to the target nucleic acid, wherein a host cell is used to package the target nucleic acid containing the packaging signals.

In some embodiments, a virion used as a delivery vehicle may comprise a commensal human virus. In some embodiments, a virion used as a delivery vehicle may comprise an anellovirus, the use of which is described in WO2018232017A1, which is incorporated herein by reference in its entirety.

AAV Administration

In some embodiments, an adeno-associated virus (AAV) is used in conjunction with the system, template nucleic acid, and/or polypeptide described herein. In some embodiments, an AAV is used to deliver, administer, or package the system, template nucleic acid, and/or polypeptide described herein. In some embodiments, the AAV is a recombinant AAV (rAAV).

In some embodiments, a system comprises (a) a polypeptide described herein or a nucleic acid encoding the same, (b) a template nucleic acid (e.g., template RNA) described herein, and (c) one or more first tissue-specific expression-control sequences specific to the target tissue, wherein the one or more first tissue-specific expression-control sequences specific to the target tissue are in operative association with (a), (b), or (a) and (b), wherein, when associated with (a), (a) comprises a nucleic acid encoding the polypeptide.

In some embodiments, a system described herein further comprises a first recombinant adeno-associated virus (rAAV) capsid protein; wherein the at least one of (a) or (b) is associated with the first rAAV capsid protein, wherein at least one of (a) or (b) is flanked by AAV inverted terminal repeats (ITRs).

In some embodiments, (a) and (b) are associated with the first rAAV capsid protein.

In some embodiments, (a) and (b) are on a single nucleic acid.

In some embodiments, the system further comprises a second rAAV capsid protein, wherein at least one of (a) or (b) is associated with the second rAAV capsid protein, and wherein the at least one of (a) or (b) associated with the second rAAV capsid protein is different from the at least one of (a) or (b) is associated with the first rAAV capsid protein.

In some embodiments, the at least one of (a) or (b) is associated with the first or second rAAV capsid protein is dispersed in the interior of the first or second rAAV capsid protein, which first or second rAAV capsid protein is in the form of an AAV capsid particle.

In some embodiments, the system further comprises a nanoparticle, wherein the nanoparticle is associated with at least one of (a) or (b).

In some embodiments, (a) and (b), respectively are associated with: a) a first rAAV capsid protein and a second rAAV capsid protein; b) a nanoparticle and a first rAAV capsid protein; c) a first rAAV capsid protein; d) a first adenovirus capsid protein; e) a first nanoparticle and a second nanoparticle; or f) a first nanoparticle.

Viral vectors are useful for delivering all or part of a system provided by the invention, e.g., for use in methods provided by the invention. Systems derived from different viruses have been employed for the delivery of polypeptides or nucleic acids; for example: integrase-deficient lentivirus, adenovirus, adeno-associated virus (AAV), herpes simplex virus, and baculovirus (reviewed in Hodge et al. Hum Gene Ther 2017; Narayanavari et al. Crit Rev Biochem Mol Biol 2017; Boehme et al. Curr Gene Ther 2015).

Adenoviruses are common viruses that have been used as gene delivery vehicles given well-defined biology, genetic stability, high transduction efficiency, and ease of large-scale production (see, for example, review by Lee et al. Genes & Diseases 2017). They possess linear dsDNA genomes and come in a variety of serotypes that differ in tissue and cell tropisms. In order to prevent replication of infectious virus in recipient cells, adenovirus genomes used for packaging are deleted of some or all endogenous viral proteins, which are provided in trans in viral production cells. This renders the genomes helper-dependent, meaning they can only be replicated and packaged into viral particles in the presence of the missing components provided by so-called helper functions. A helper-dependent adenovirus system with all viral ORFs removed may be compatible with packaging foreign DNA of up to ~37 kb (Parks et al. J Virol 1997). In some embodiments, an adenoviral vector is used to deliver DNA corresponding to the polypeptide or template component of the gene modifying system, or both are contained on separate or the same adenoviral vector. In some embodiments, the adenovirus is a helper-dependent adenovirus (HD-AdV) that is incapable of self-packaging. In some embodiments, the adenovirus is a high-capacity adenovirus (HC-AdV) that has had all or a substantial portion of endogenous viral ORFs deleted, while retaining the necessary sequence components for packaging into adenoviral particles. For this type of vector, the only adenoviral sequences required for genome packaging are noncoding sequences: the inverted terminal repeats (ITRs) at both ends and the packaging signal at the 5′-end (Jager et al. Nat Protoc 2009). In some embodiments, the adenoviral genome also comprises stuffer DNA to meet a

minimal genome size for optimal production and stability (see, for example, Hausl et al. Mol Ther 2010). In some embodiments, an adenovirus is used to deliver a gene modifying system to the liver.

In some embodiments, an adenovirus is used to deliver a gene modifying system to HSCs, e.g., HDAd5/35++. HDAd5/35++ is an adenovirus with modified serotype 35 fibers that de-target the vector from the liver (Wang et al. Blood Adv 2019). In some embodiments, the adenovirus that delivers a gene modifying system to HSCs utilizes a receptor that is expressed specifically on primitive HSCs, e.g., CD46.

Adeno-associated viruses (AAV) belong to the parvoviridae family and more specifically constitute the dependoparvovirus genus. The AAV genome is composed of a linear single-stranded DNA molecule which contains approximately 4.7 kilobases (kb) and consists of two major open reading frames (ORFs) encoding the non-structural Rep (replication) and structural Cap (capsid) proteins. A second ORF within the cap gene was identified that encodes the assembly-activating protein (AAP). The DNAs flanking the AAV coding regions are two cis-acting inverted terminal repeat (ITR) sequences, approximately 145 nucleotides in length, with interrupted palindromic sequences that can be folded into energetically stable hairpin structures that function as primers of DNA replication. In addition to their role in DNA replication, the ITR sequences have been shown to be involved in viral DNA integration into the cellular genome, rescue from the host genome or plasmid, and encapsidation of viral nucleic acid into mature virions (Muzyczka, (1992) Curr. Top. Micro. Immunol. 158:97-129). In some embodiments, one or more gene modifying nucleic acid components is flanked by ITRs derived from AAV for viral packaging. See, e.g., WO2019113310.

In some embodiments, one or more components of the gene modifying system are carried via at least one AAV vector. In some embodiments, the at least one AAV vector is selected for tropism to a particular cell, tissue, organism. In some embodiments, the AAV vector is pseudotyped, e.g., AAV2/8, wherein AAV2 describes the design of the construct but the capsid protein is replaced by that from AAV8. It is understood that any of the described vectors could be pseudotype derivatives, wherein the capsid protein used to package the AAV genome is derived from that of a different AAV serotype. Without wishing to be limited in vector choice, a list of exemplary AAV serotypes can be found in Table 18. In some embodiments, an AAV to be employed for gene modifying may be evolved for novel cell or tissue tropism as has been demonstrated in the literature (e.g., Davidsson et al. Proc Natl Acad Sci U S A 2019).

In some embodiments, the AAV delivery vector is a vector which has two AAV inverted terminal repeats (ITRs) and a nucleotide sequence of interest (for example, a sequence coding for a gene modifying polypeptideor a DNA template, or both), each of said ITRs having an interrupted (or noncontiguous) palindromic sequence, i.e., a sequence composed of three segments: a first segment and a

last segment that are identical when read 5'→ 3' but hybridize when placed against each other, and a segment that is different that separates the identical segments. See, for example, WO2012123430.

Conventionally, AAV virions with capsids are produced by introducing a plasmid or plasmids encoding the rAAV or scAAV genome, Rep proteins, and Cap proteins (Grimm et al, 1998). Upon introduction of these helper plasmids in trans, the AAV genome is "rescued" (i.e., released and subsequently recovered) from the host genome, and is further encapsidated to produce infectious AAV. In some embodiments, one or more gene modifying nucleic acids are packaged into AAV particles by introducing the ITR-flanked nucleic acids into a packaging cell in conjunction with the helper functions.

In some embodiments, the AAV genome is a so called self-complementary genome (referred to as scAAV), such that the sequence located between the ITRs contains both the desired nucleic acid sequence (e.g., DNA encoding the gene modifying polypeptide or template, or both) in addition to the reverse complement of the desired nucleic acid sequence, such that these two components can fold over and self-hybridize. In some embodiments, the self-complementary modules are separated by an intervening sequence that permits the DNA to fold back on itself, e.g., forms a stem-loop. An scAAV has the advantage of being poised for transcription upon entering the nucleus, rather than being first dependent on ITR priming and second-strand synthesis to form dsDNA. In some embodiments, one or more gene modifying components is designed as an scAAV, wherein the sequence between the AAV ITRs contains two reverse complementing modules that can self-hybridize to create dsDNA.

In some embodiments, nucleic acid (e.g., encoding a polypeptide, or a template, or both) delivered to cells is closed-ended, linear duplex DNA (CELiD DNA or ceDNA). In some embodiments, ceDNA is derived from the replicative form of the AAV genome (Li et al. PLoS One 2013). In some embodiments, the nucleic acid (e.g., encoding a polypeptide, or a template DNA, or both) is flanked by ITRs, e.g., AAV ITRs, wherein at least one of the ITRs comprises a terminal resolution site and a replication protein binding site (sometimes referred to as a replicative protein binding site). In some embodiments, the ITRs are derived from an adeno-associated virus, e.g., AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, or a combination thereof. In some embodiments, the ITRs are symmetric. In some embodiments, the ITRs are asymmetric. In some embodiments, at least one Rep protein is provided to enable replication of the construct. In some embodiments, the at least one Rep protein is derived from an adeno-associated virus, e.g., AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, or a combination thereof. In some embodiments, ceDNA is generated by providing a production cell with (i) DNA flanked by ITRs, e.g., AAV ITRs, and (ii) components required for ITR-dependent replication, e.g., AAV proteins Rep78 and Rep52 (or nucleic acid encoding the proteins). In some embodiments, ceDNA is free of any

capsid protein, e.g., is not packaged into an infectious AAV particle. In some embodiments, ceDNA is formulated into LNPs (see, for example, WO2019051289A1).

In some embodiments, the ceDNA vector consists of two self-complementary sequences, e.g., asymmetrical or symmetrical or substantially symmetrical ITRs as defined herein, flanking said expression cassette, wherein the ceDNA vector is not associated with a capsid protein. In some embodiments, the ceDNA vector comprises two self-complementary sequences found in an AAV genome, where at least one ITR comprises an operative Rep-binding element (RBE) (also sometimes referred to herein as "RBS") and a terminal resolution site (trs) of AAV or a functional variant of the RBE. See, for example, WO2019113310.

In some embodiments, the AAV genome comprises two genes that encode four replication proteins and three capsid proteins, respectively. In some embodiments, the genes are flanked on either side by 145-bp inverted terminal repeats (ITRs). In some embodiments, the virion comprises up to three capsid proteins (Vp1, Vp2, and/or Vp3), e.g., produced in a 1:1:10 ratio. In some embodiments, the capsid proteins are produced from the same open reading frame and/or from differential splicing (Vp1) and alternative translational start sites (Vp2 and Vp3, respectively). Generally, Vp3 is the most abundant subunit in the virion and participates in receptor recognition at the cell surface defining the tropism of the virus. In some embodiments, Vp1 comprises a phospholipase domain, e.g., which functions in viral infectivity, in the N-terminus of Vp1.

In some embodiments, packaging capacity of the viral vectors limits the size of the gene modifying system that can be packaged into the vector. For example, the packaging capacity of the AAVs can be about 4.5 kb (e.g., about 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, or 6.0 kb), e.g., including one or two inverted terminal repeats (ITRs), e.g., 145 base ITRs.

In some embodiments, recombinant AAV (rAAV) comprises cis-acting 145-bp ITRs flanking vector transgene cassettes, e.g., providing up to 4.5 kb for packaging of foreign DNA. Subsequent to infection, rAAV can, in some instances, express a fusion protein of the invention and persist without integration into the host genome by existing episomally in circular head-to-tail concatemers. rAAV can be used, for example, *in vitro* and *in vivo*. In some embodiments, AAV-mediated gene delivery requires that the length of the coding sequence of the gene is equal or greater in size than the wild-type AAV genome.

AAV delivery of genes that exceed this size and/or the use of large physiological regulatory elements can be accomplished, for example, by dividing the protein(s) to be delivered into two or more fragments. In some embodiments, the N-terminal fragment is fused to an intein-N sequence. In some embodiments, the C- terminal fragment is fused to an intein-C sequence. In embodiments, the fragments are packaged into two or more AAV vectors.

In some embodiments, dual AAV vectors are generated by splitting a large transgene expression cassette in two separate halves (5′ and 3′ ends, or head and tail), e.g., wherein each half of the cassette is packaged in a single AAV vector (of <5 kb). The re-assembly of the full-length transgene expression cassette can, in some embodiments, then be achieved upon co-infection of the same cell by both dual AAV vectors. In some embodiments, co-infection is followed by one or more of: (1) homologous recombination (HR) between 5′ and 3′ genomes (dual AAV overlapping vectors); (2) ITR-mediated tail-to-head concatemerization of 5′ and 3′ genomes (dual AAV trans-splicing vectors); and/or (3) a combination of these two mechanisms (dual AAV hybrid vectors). In some embodiments, the use of dual AAV vectors *in vivo* results in the expression of full-length proteins. In some embodiments, the use of the dual AAV vector platform represents an efficient and viable gene transfer strategy for transgenes of greater than about 4.0, 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, or 5.0 kb in size. In some embodiments, AAV vectors can also be used to transduce cells with target nucleic acids, e.g., in the *in vitro* production of nucleic acids and peptides. In some embodiments, AAV vectors can be used for *in vivo* and *ex vivo* gene therapy procedures (see, e.g., West et al., Virology 160:38-47 (1987); U.S. Patent No. 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest.94:1351 (1994); each of which is incorporated herein by reference in their entirety). The construction of recombinant AAV vectors is described in a number of publications, including U.S. Patent No.5,173,414; Tratschin et al., Mol. Cell. Biol.5:3251- 3260 (1985); Tratschin, et al., Mol. Cell. Biol.4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); and Samulski et al., J. Virol.63:03822-3828 (1989) (incorporated by reference herein in their entirety).

In some embodiments, a gene modifying polypeptide described herein (e.g., with or without one or more guide nucleic acids) can be delivered using AAV, lentivirus, adenovirus or other plasmid or viral vector types, in particular, using formulations and doses from, for example, U.S. Patent No. 8,454,972 (formulations, doses for adenovirus), U.S. Patent No.8,404,658 (formulations, doses for AAV) and U.S. Patent No.5,846,946 (formulations, doses for DNA plasmids) and from clinical trials and publications regarding the clinical trials involving lentivirus, AAV and adenovirus. For example, for AAV, the route of administration, formulation and dose can be as described in U.S. Patent No.8,454,972 and as in clinical trials involving AAV. For adenovirus, the route of administration, formulation and dose can be as described in U.S. Patent No.8,404,658 and as in clinical trials involving adenovirus. For plasmid delivery, the route of administration, formulation and dose can be as described in U.S. Patent No.5,846,946 and as in clinical studies involving plasmids. Doses can be based on or extrapolated to an average 70 kg individual (e.g. a male adult human), and can be adjusted for patients, subjects, mammals of different weight and species. Frequency of administration is within the ambit of the medical or veterinary practitioner (e.g., physician, veterinarian), depending on usual factors including the age, sex, general

health, other conditions of the patient or subject and the particular condition or symptoms being addressed. In some embodiments, the viral vectors can be injected into the tissue of interest. For cell-type specific gene modifying, the expression of the gene modifying polypeptide and optional guide nucleic acid can, in some embodiments, be driven by a cell-type specific promoter.

In some embodiments, AAV allows for low toxicity, for example, due to the purification method not requiring ultracentrifugation of cell particles that can activate the immune response. In some embodiments, AAV allows low probability of causing insertional mutagenesis, for example, because it does not substantially integrate into the host genome.

In some embodiments, AAV has a packaging limit of about 4.4, 4.5, 4.6, 4.7, or 4.75 kb. In some embodiments, a gene modifying polypeptide-encoding sequence, promoter, and transcription terminator can fit into a single viral vector. SpCas9 (4.1 kb) may, in some instances, be difficult to package into AAV. Therefore, in some embodiments, a gene modifying polypeptide coding sequence is used that is shorter in length than other gene modifying polypeptide coding sequences or base editors. In some embodiments, the gene modifying polypeptide encoding sequences are less than about 4.5 kb, 4.4 kb, 4.3 kb, 4.2 kb, 4.1 kb, 4 kb, 3.9 kb, 3.8 kb, 3.7 kb, 3.6 kb, 3.5 kb, 3.4 kb, 3.3 kb, 3.2 kb, 3.1 kb, 3 kb, 2.9 kb, 2.8 kb, 2.7 kb, 2.6 kb, 2.5 kb, 2 kb, or 1.5 kb.

An AAV can be AAV1, AAV2, AAV5 or any combination thereof. In some embodiments, the type of AAV is selected with respect to the cells to be targeted; e.g., AAV serotypes 1, 2, 5 or a hybrid capsid AAV1, AAV2, AAV5 or any combination thereof can be selected for targeting brain or neuronal cells; or AAV4 can be selected for targeting cardiac tissue. In some embodiments, AAV8 is selected for delivery to the liver. Exemplary AAV serotypes as to these cells are described, for example, in Grimm, D. et al, J. Virol.82: 5887-5911 (2008) (incorporated herein by reference in its entirety). In some embodiments, AAV refers all serotypes, subtypes, and naturally-occurring AAV as well as recombinant AAV. AAV may be used to refer to the virus itself or a derivative thereof. In some embodiments, AAV includes AAV1, AAV2, AAV3, AAV3B, AAV4, AAV5, AAV6, AAV6.2, AAV7, AAVrh.64Rl, AAVhu.37, AAVrh.8, AAVrh.32.33, AAV8, AAV9, AAV-DJ, AAV2/8, AAVrhlO, AAVLK03, AV10, AAV11, AAV 12, rhlO, and hybrids thereof, avian AAV, bovine AAV, canine AAV, equine AAV, primate AAV, non-primate AAV, and ovine AAV. The genomic sequences of various serotypes of AAV, as well as the sequences of the native terminal repeats (TRs), Rep proteins, and capsid subunits are known in the art. Such sequences may be found in the literature or in public databases such as GenBank. Additional exemplary AAV serotypes are listed in Table 18.

Table 18. Exemplary AAV serotypes.

Target Tissue	Vehicle	Reference
Liver	AAV (AAV8[1], AAVrh.8[1], AAVhu.37[1], AAV2/8, AAV2/rh10[2], AAV9, AAV2, NP40[3], NP59[2,3], AAV3B[5], AAV-DJ[4], AAV-LK01[4], AAV-LK02[4], AAV-LK03[4], AAV-LK19[4], AAV5[7] Adenovirus (Ad5, HC-AdV[6])	1. Wang *et al.*, *Mol. Ther.* **18**, 118–25 (2010) 2. Ginn *et al.*, *JHEP Reports*, 100065 (2019) 3. Paulk *et al.*, *Mol. Ther.* **26**, 289–303 (2018). 4. L. Lisowski *et al.*, *Nature*. **506**, 382–6 (2014). 5. L. Wang *et al.*, *Mol. Ther.* **23**, 1877–87 (2015). 6. Hausl *Mol Ther* (2010) 7. Davidoff *et al.*, *Mol. Ther.* 11, 875-88 (2005)
Lung	AAV (AAV4, AAV5, AAV6[1], AAV9, H22[2]) Adenovirus (Ad5, Ad3, Ad21, Ad14)[3]	1. Duncan *et al.*, *Mol Ther Methods Clin Dev* (2018) 2. Cooney *et al.*, Am J Respir Cell Mol Biol (2019) 3. Li *et al.*, *Mol Ther Methods Clin Dev* (2019)
Skin	AAV (AAV6[1], AAV-LK19[2])	1. Petek et al., *Mol. Ther.* (2010) 2. L. Lisowski *et al.*, *Nature*. **506**, 382–6 (2014).
HSCs	Adenovirus (HDAd5/35^{++})	Wang et al. *Blood Adv* (2019)

In some embodiments, a pharmaceutical composition (e.g., comprising an AAV as described herein) has less than 10% empty capsids, less than 8% empty capsids, less than 7% empty capsids, less than 5% empty capsids, less than 3% empty capsids, or less than 1 % empty capsids. In some embodiments, the pharmaceutical composition has less than about 5% empty capsids. In some embodiments, the number of empty capsids is below the limit of detection. In some embodiments, it is advantageous for the pharmaceutical composition to have low amounts of empty capsids, e.g., because empty capsids may generate an adverse response (e.g., immune response, inflammatory response, liver response, and/or cardiac response), e.g., with little or no substantial therapeutic benefit.

In some embodiments, the residual host cell protein (rHCP) in the pharmaceutical composition is less than or equal to 100 ng/ml rHCP per 1 x 10^{13} vg/ml, e.g., less than or equal to 40 ng/ml rHCP per 1 x 10^{13} vg/ml or 1-50 ng/ml rHCP per 1 x 10^{13} vg/ml. In some embodiments, the pharmaceutical composition comprises less than 10 ng rHCP per 1.0 x 10^{13} vg, or less than 5 ng rHCP per 1.0 x 10^{13} vg, less than 4 ng rHCP per 1.0 x 10^{13} vg, or less than 3 ng rHCP per 1.0 x 10^{13} vg, or any concentration in between. In some embodiments, the residual host cell DNA (hcDNA) in the pharmaceutical composition is less than or equal to 5 x 10^{6} pg/ml hcDNA per 1 x 10^{13} vg/ml, less than or equal to 1.2 x 10^{6} pg/ml hcDNA per 1 x 10^{13} vg/ml, or 1 x 10^{5} pg/ml hcDNA per 1 x 10^{13} vg/ml. In some embodiments, the residual host cell DNA in said pharmaceutical composition is less than 5.0 x 10^{5} pg per 1 x 10^{13} vg, less than 2.0 x 10^{5} pg per 1.0 x 10^{13} vg, less than 1.1 x 10^{5} pg per 1.0 x 10^{13} vg, less than 1.0 x 10^{5} pg hcDNA per 1.0 x 10^{13} vg, less than 0.9 x 10^{5} pg hcDNA per 1.0 x 10^{13} vg, less than 0.8 x 10^{5} pg hcDNA per 1.0 x 10^{13} vg, or any concentration in between.

In some embodiments, the residual plasmid DNA in the pharmaceutical composition is less than or equal to 1.7 x 10^{5} pg/ml per 1.0 x 10^{13} vg/ml, or 1 x 10^{5} pg/ml per 1 x 1.0 x 10^{13} vg/ml, or 1.7 x 10^{6} pg/ml per 1.0 x 10^{13} vg/ml. In some embodiments, the residual DNA plasmid in the pharmaceutical composition is less than 10.0 x 10^{5} pg by 1.0 x 10^{13} vg, less than 8.0 x 10^{5} pg by 1.0 x 10^{13} vg or less than 6.8 x 10^{5} pg by 1.0 x 10^{13} vg. In embodiments, the pharmaceutical composition comprises less than 0.5 ng per 1.0 x 10^{13} vg, less than 0.3 ng per 1.0 x 10^{13} vg, less than 0.22 ng per 1.0 x 10^{13} vg or less than 0.2 ng per 1.0 x 10^{13} vg or any intermediate concentration of bovine serum albumin (BSA). In embodiments, the benzonase in the pharmaceutical composition is less than 0.2 ng by 1.0 x 10^{13} vg, less than 0.1 ng by 1.0 x 10^{13} vg, less than 0.09 ng by 1.0 x 10^{13} vg, less than 0.08 ng by 1.0 x 10^{13} vg or any intermediate concentration. In embodiments, Poloxamer 188 in the pharmaceutical composition is about 10 to 150 ppm, about 15 to 100 ppm or about 20 to 80 ppm. In embodiments, the cesium in the pharmaceutical composition is less than 50 pg / g (ppm), less than 30 pg / g (ppm) or less than 20 pg / g (ppm) or any intermediate concentration.

In embodiments, the pharmaceutical composition comprises total impurities, e.g., as determined by SDS-PAGE, of less than 10%, less than 8%, less than 7%, less than 6%, less than 5%, less than 4%, less than 3%, less than 2%, or any percentage in between. In embodiments, the total purity, e.g., as determined by SDS-PAGE, is greater than 90%, greater than 92%, greater than 93%, greater than 94%, greater than 95%, greater than 96%, greater than 97%, greater than 98%, or any percentage in between. In embodiments, no single unnamed related impurity, e.g., as measured by SDS-PAGE, is greater than 5%, greater than 4%, greater than 3% or greater than 2%, or any percentage in between. In embodiments, the pharmaceutical composition comprises a percentage of filled capsids relative to total capsids (e.g., peak 1 + peak 2 as measured by analytical ultracentrifugation) of greater than 85%, greater than 86%, greater than 87%, greater than 88%, greater than 89%, greater than 90%, greater than 91%, greater than 91.9%, greater than 92%, greater than 93%, or any percentage in between. In embodiments of the pharmaceutical composition, the percentage of filled capsids measured in peak 1 by analytical ultracentrifugation is 20-80%, 25-75%, 30-75%, 35-75%, or 37.4-70.3%. In embodiments of the pharmaceutical composition, the percentage of filled capsids measured in peak 2 by analytical ultracentrifugation is 20-80%, 20-70%, 22-65%, 24-62%, or 24.9-60.1%.

In one embodiment, the pharmaceutical composition comprises a genomic titer of 1.0 to 5.0 x 10^{13} vg / mL, 1.2 to 3.0 x 10^{13} vg / mL or 1.7 to 2.3 x 10^{13} vg / ml. In one embodiment, the pharmaceutical composition exhibits a biological load of less than 5 CFU / mL, less than 4 CFU / mL, less than 3 CFU / mL, less than 2 CFU / mL or less than 1 CFU / mL or any intermediate contraction. In embodiments, the amount of endotoxin according to USP, for example, USP <85> (incorporated by reference in its entirety) is less than 1.0 EU / mL, less than 0.8 EU / mL or less than 0.75 EU / mL. In embodiments, the osmolarity of a pharmaceutical composition according to USP, for example, USP <785> (incorporated by reference in its entirety) is 350 to 450 mOsm / kg, 370 to 440 mOsm / kg or 390 to 430 mOsm / kg. In embodiments, the pharmaceutical composition contains less than 1200 particles that are greater than 25 μm per container, less than 1000 particles that are greater than 25 μm per container, less than 500 particles that are greater than 25 μm per container or any intermediate value. In embodiments, the pharmaceutical composition contains less than 10,000 particles that are greater than 10 μm per container, less than 8000 particles that are greater than 10 μm per container or less than 600 particles that are greater than 10 pm per container.

In one embodiment, the pharmaceutical composition has a genomic titer of 0.5 to 5.0 x 10^{13} vg / mL, 1.0 to 4.0 x 10^{13} vg / mL, 1.5 to 3.0 x 10^{13} vg / ml or 1.7 to 2.3 x 10^{13} vg / ml. In one embodiment, the pharmaceutical composition described herein comprises one or more of the following: less than about 0.09 ng benzonase per 1.0 x 10^{13} vg, less than about 30 pg / g (ppm) of cesium, about 20 to 80 ppm Poloxamer 188, less than about 0.22 ng BSA per 1.0 x 10^{13} vg, less than about 6.8 x 10^{5} pg of residual

DNA plasmid per 1.0 x 10^{13} vg, less than about 1.1 x 10^{5} pg of residual hcDNA per 1.0 x 10^{13} vg, less than about 4 ng of rHCP per 1.0 x 10^{13} vg, pH 7.7 to 8.3, about 390 to 430 mOsm / kg, less than about 600 particles that are > 25 μm in size per container, less than about 6000 particles that are > 10 μm in size per container, about 1.7 x 10^{13} - 2.3 x 10^{13} vg / mL genomic titer, infectious titer of about 3.9 x 10^{8} to 8.4 x 10^{10} IU per 1.0 x 10^{13} vg, total protein of about 100-300 pg per 1.0 x 10^{13} vg, mean survival of >24 days in A7SMA mice with about 7.5 x 10^{13} vg / kg dose of viral vector, about 70 to 130% relative potency based on an *in vitro* cell based assay and / or less than about 5% empty capsid. In various embodiments, the pharmaceutical compositions described herein comprise any of the viral particles discussed here, retain a potency of between ± 20%, between ± 15%, between ± 10% or within ± 5% of a reference standard. In some embodiments, potency is measured using a suitable *in vitro* cell assay or *in vivo* animal model.

Additional methods of preparation, characterization, and dosing AAV particles are taught in WO2019094253, which is incorporated herein by reference in its entirety.

Additional rAAV constructs that can be employed consonant with the invention include those described in Wang et al 2019, available at: //doi.org/10.1038/s41573-019-0012-9, including Table 1 thereof, which is incorporated by reference in its entirety.

Lipid Nanoparticles

The methods and systems provided herein may employ any suitable carrier or delivery modality, including, in certain embodiments, lipid nanoparticles (LNPs). Lipid nanoparticles, in some embodiments, comprise one or more ionic lipids, such as non-cationic lipids (e.g., neutral or anionic, or zwitterionic lipids); one or more conjugated lipids (such as PEG-conjugated lipids or lipids conjugated to polymers described in Table 5 of WO2019217941; incorporated herein by reference in its entirety); one or more sterols (e.g., cholesterol); and, optionally, one or more targeting molecules (e.g., conjugated receptors, receptor ligands, antibodies); or combinations of the foregoing.

Lipids that can be used in nanoparticle formations (e.g., lipid nanoparticles) include, for example those described in Table 4 of WO2019217941, which is incorporated by reference—e.g., a lipid-containing nanoparticle can comprise one or more of the lipids in Table 4 of WO2019217941. Lipid nanoparticles can include additional elements, such as polymers, such as the polymers described in Table 5 of WO2019217941, incorporated by reference.

In some embodiments, conjugated lipids, when present, can include one or more of PEG-diacylglycerol (DAG) (such as l-(monomethoxy-polyethyleneglycol)-2,3- dimyristoylglycerol (PEG-DMG)), PEG-dialkyloxypropyl (DAA), PEG-phospholipid, PEG- ceramide (Cer), a pegylated phosphatidylethanoloamine (PEG-PE), PEG succinate diacylglycerol (PEGS-DAG) (such as 4-0-(2',3'-

di(tetradecanoyloxy)propyl-l-0-(w- methoxy(polyethoxy)ethyl) butanedioate (PEG-S-DMG)), PEG dialkoxypropylcarbam, N- (carbonyl-methoxypoly ethylene glycol 2000)- 1 ,2-distearoyl-sn-glycero-3-phosphoethanolamine sodium salt, and those described in Table 2 of WO2019051289 (incorporated by reference), and combinations of the foregoing.

In some embodiments, sterols that can be incorporated into lipid nanoparticles include one or more of cholesterol or cholesterol derivatives, such as those in W02009/127060 or US2010/0130588, which are incorporated by reference. Additional exemplary sterols include phytosterols, including those described in Eygeris et al (2020), dx.doi.org/10.1021/acs.nanolett.0c01386, incorporated herein by reference.

In some embodiments, the lipid particle comprises an ionizable lipid, a non-cationic lipid, a conjugated lipid that inhibits aggregation of particles, and a sterol. The amounts of these components can be varied independently and to achieve desired properties. For example, in some embodiments, the lipid nanoparticle comprises an ionizable lipid is in an amount from about 20 mol % to about 90 mol % of the total lipids (in other embodiments it may be 20-70% (mol), 30-60% (mol) or 40-50% (mol); about 50 mol % to about 90 mol % of the total lipid present in the lipid nanoparticle), a non-cationic lipid in an amount from about 5 mol % to about 30 mol % of the total lipids, a conjugated lipid in an amount from about 0.5 mol % to about 20 mol % of the total lipids, and a sterol in an amount from about 20 mol % to about 50 mol % of the total lipids. The ratio of total lipid to nucleic acid (e.g., encoding the gene modifying polypeptide or template nucleic acid) can be varied as desired. For example, the total lipid to nucleic acid (mass or weight) ratio can be from about 10: 1 to about 30: 1.

In some embodiments, an ionizable lipid may be a cationic lipid, an ionizable cationic lipid, e.g., a cationic lipid that can exist in a positively charged or neutral form depending on pH, or an amine-containing lipid that can be readily protonated. In some embodiments, the cationic lipid is a lipid capable of being positively charged, e.g., under physiological conditions. Exemplary cationic lipids include one or more amine group(s) which bear the positive charge. In some embodiments, the lipid particle comprises a cationic lipid in formulation with one or more of neutral lipids, ionizable amine-containing lipids, biodegradable alkyn lipids, steroids, phospholipids including polyunsaturated lipids, structural lipids (e.g., sterols), PEG, cholesterol and polymer conjugated lipids. In some embodiments, the cationic lipid may be an ionizable cationic lipid. An exemplary cationic lipid as disclosed herein may have an effective pKa over 6.0. In embodiments, a lipid nanoparticle may comprise a second cationic lipid having a different effective pKa (e.g., greater than the first effective pKa), than the first cationic lipid. A lipid nanoparticle may comprise between 40 and 60 mol percent of a cationic lipid, a neutral lipid, a steroid, a polymer conjugated lipid, and a therapeutic agent, e.g., a nucleic acid (e.g., RNA) described herein (e.g., a template nucleic acid or a nucleic acid encoding a gene modifying polypeptide), encapsulated within or

associated with the lipid nanoparticle. In some embodiments, the nucleic acid is co-formulated with the cationic lipid. The nucleic acid may be adsorbed to the surface of an LNP, e.g., an LNP comprising a cationic lipid. In some embodiments, the nucleic acid may be encapsulated in an LNP, e.g., an LNP comprising a cationic lipid. In some embodiments, the lipid nanoparticle may comprise a targeting moiety, e.g., coated with a targeting agent. In embodiments, the LNP formulation is biodegradable. In some embodiments, a lipid nanoparticle comprising one or more lipid described herein, e.g., Formula (i), (ii), (ii), (vii) and/or (ix) encapsulates at least 1%, at least 5%, at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 92%, at least 95%, at least 97%, at least 98% or 100% of an RNA molecule, e.g., template RNA and/or a mRNA encoding the gene modifying polypeptide.

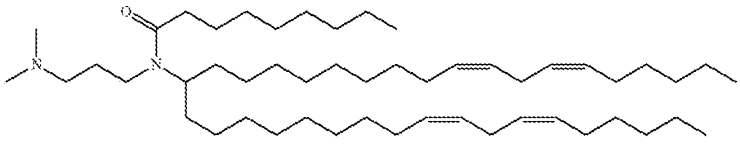
In some embodiments, the lipid to nucleic acid ratio (mass/mass ratio; w/w ratio) can be in the range of from about 1 : 1 to about 25: 1, from about 10: 1 to about 14: 1, from about 3 : 1 to about 15: 1, from about 4: 1 to about 10: 1, from about 5: 1 to about 9: 1, or about 6: 1 to about 9: 1. The amounts of lipids and nucleic acid can be adjusted to provide a desired N/P ratio, for example, N/P ratio of 3, 4, 5, 6, 7, 8, 9, 10 or higher. Generally, the lipid nanoparticle formulation's overall lipid content can range from about 5 mg/ml to about 30 mg/mL.

Exemplary ionizable lipids that can be used in lipid nanoparticle formulations include, without limitation, those listed in Table 1 of WO2019051289, incorporated herein by reference. Additional exemplary lipids include, without limitation, one or more of the following formulae: X of US2016/0311759; I of US20150376115 or in US2016/0376224; I, II or III of US20160151284; I, IA, II, or IIA of US20170210967; I-c of US20150140070; A of US2013/0178541; I of US2013/0303587 or US2013/0123338; I of US2015/0141678; II, III, IV, or V of US2015/0239926; I of US2017/0119904; I or II of WO2017/117528; A of US2012/0149894; A of US2015/0057373; A of WO2013/116126; A of US2013/0090372; A of US2013/0274523; A of US2013/0274504; A of US2013/0053572; A of W02013/016058; A of W02012/162210; I of US2008/042973; I, II, III, or IV of US2012/01287670; I or II of US2014/0200257; I, II, or III of US2015/0203446; I or III of US2015/0005363; I, IA, IB, IC, ID, II, IIA, IIB, IIC, IID, or III-XXIV of US2014/0308304; of US2013/0338210; I, II, III, or IV of W02009/132131; A of US2012/01011478; I or XXXV of US2012/0027796; XIV or XVII of US2012/0058144; of US2013/0323269; I of US2011/0117125; I, II, or III of US2011/0256175; I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII of US2012/0202871; I, II, III, IV, V, VI, VII, VIII, X, XII, XIII, XIV, XV, or XVI of US2011/0076335; I or II of US2006/008378; I of US2013/0123338; I or X-A-Y-Z of US2015/0064242; XVI, XVII, or XVIII of US2013/0022649; I, II, or III of US2013/0116307; I, II, or III of US2013/0116307; I or II of US2010/0062967; I-X of US2013/0189351; I of US2014/0039032; V of US2018/0028664; I of US2016/0317458; I of US2013/0195920; 5, 6, or 10 of US10,221,127; III-3 of

WO2018/081480; I-5 or I-8 of WO2020/081938; 18 or 25 of US9,867,888; A of US2019/0136231; II of WO2020/219876; 1 of US2012/0027803; OF-02 of US2019/0240349; 23 of US10,086,013; cKK-E12/A6 of Miao et al (2020); C12-200 of WO2010/053572; 7C1 of Dahlman et al (2017); 304-O13 or 503-O13 of Whitehead et al; TS-P4C2 of US9,708,628; I of WO2020/106946; I of WO2020/106946.

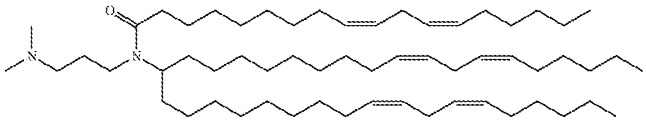
In some embodiments, the ionizable lipid is MC3 (6Z,9Z,28Z,3 lZ)-heptatriaconta- 6,9,28,3 l-tetraen-l9-yl-4-(dimethylamino) butanoate (DLin-MC3-DMA or MC3), e.g., as described in Example 9 of WO2019051289A9 (incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is the lipid ATX-002, e.g., as described in Example 10 of WO2019051289A9 (incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is (l3Z,l6Z)-A,A-dimethyl-3- nonyldocosa-l3, l6-dien-l-amine (Compound 32), e.g., as described in Example 11 of WO2019051289A9 (incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is Compound 6 or Compound 22, e.g., as described in Example 12 of WO2019051289A9 (incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is heptadecan-9-yl 8-((2-hydroxyethyl)(6-oxo-6-(undecyloxy)hexyl)amino)octanoate (SM-102); e.g., as described in Example 1 of US9,867,888(incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is 9Z,12Z)-3-((4,4-bis(octyloxy)butanoyl)oxy)-2-((((3-(diethylamino)propoxy)carbonyl)oxy)methyl)propyl octadeca-9,12-dienoate (LP01) e.g., as synthesized in Example 13 of WO2015/095340(incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is Di((Z)-non-2-en-1-yl) 9-((4-dimethylamino)butanoyl)oxy)heptadecanedioate (L319), e.g. as synthesized in Example 7, 8, or 9 of US2012/0027803(incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is 1,1'-((2-(4-(2-((2-(Bis(2-hydroxydodecyl)amino)ethyl)(2-hydroxydodecyl) amino)ethyl)piperazin-1-yl)ethyl)azanediyl)bis(dodecan-2-ol) (C12-200), e.g., as synthesized in Examples 14 and 16 of WO2010/053572(incorporated by reference herein in its entirety). In some embodiments, the ionizable lipid is; Imidazole cholesterol ester (ICE) lipid (3S, 10R, 13R, 17R)-10, 13-dimethyl-17- ((R)-6-methylheptan-2-yl)-2, 3, 4, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17-tetradecahydro-lH-cyclopenta[a]phenanthren-3-yl 3-(1H-imidazol-4-yl)propanoate, e.g., Structure (I) from WO2020/106946 (incorporated by reference herein in its entirety).

Some non-limiting examples of lipid compounds that may be used (e.g., in combination with other lipid components) to form lipid nanoparticles for the delivery of compositions described herein, e.g., nucleic acid (e.g., RNA) described herein (e.g., a template nucleic acid or a nucleic acid encoding a gene modifying polypeptide) includes,



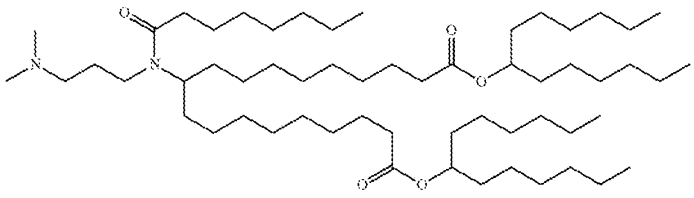
(i)

In some embodiments an LNP comprising Formula (i) is used to deliver a gene modifying composition described herein to the liver and/or hepatocyte cells.



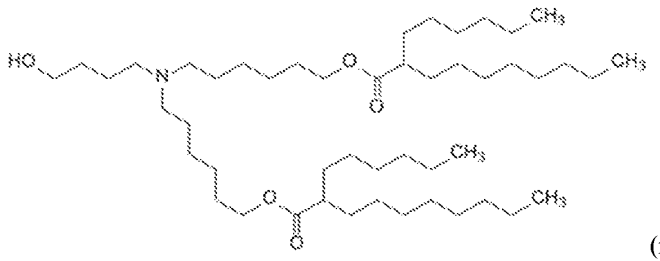
(ii)

In some embodiments an LNP comprising Formula (ii) is used to deliver a gene modifying composition described herein to the liver and/or hepatocyte cells.

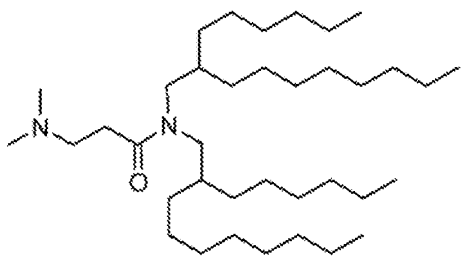


(iii)

In some embodiments an LNP comprising Formula (iii) is used to deliver a gene modifying composition described herein to the liver and/or hepatocyte cells.

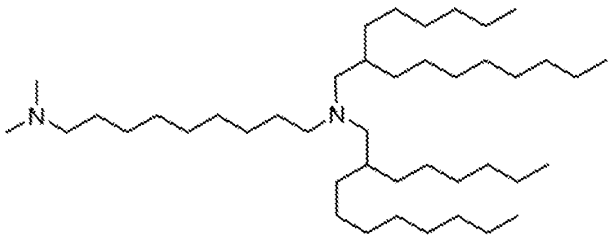


(iv)



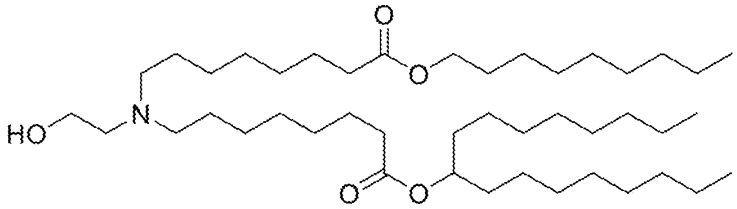
(v)

In some embodiments an LNP comprising Formula (v) is used to deliver a gene modifying composition described herein to the liver and/or hepatocyte cells.

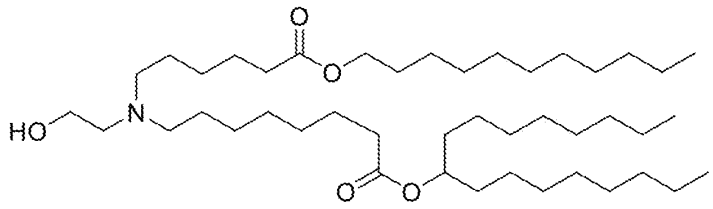


(vi)

In some embodiments an LNP comprising Formula (vi) is used to deliver a gene modifying composition described herein to the liver and/or hepatocyte cells.

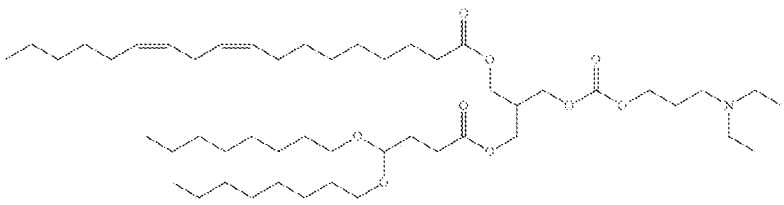


(vii)



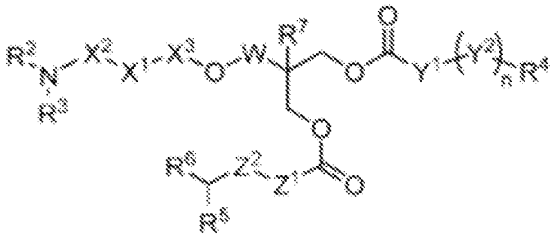
(viii)

In some embodiments an LNP comprising Formula (viii) is used to deliver a gene modifying composition described herein to the liver and/or hepatocyte cells.



(ix)

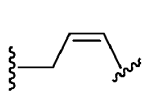
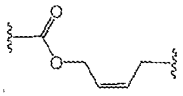
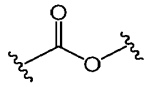
In some embodiments an LNP comprising Formula (ix) is used to deliver a gene modifying composition described herein to the liver and/or hepatocyte cells.



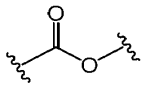
(x)

wherein

X^1 is O, NR^1, or a direct bond, X^2 is C2-5 alkylene, X^3 is C(=0) or a direct bond, R^1 is H or Me, R^3 is Ci-3 alkyl, R^2 is Ci-3 alkyl, or R^2 taken together with the nitrogen atom to which it is attached and 1-3 carbon atoms of X^2 form a 4-, 5-, or 6-membered ring, or X^1 is NR^1, R^1 and R^2 taken together with the nitrogen atoms to which they are attached form a 5- or 6-membered ring, or R^2 taken together with R^3 and the nitrogen atom to which they are attached form a 5-, 6-, or 7-membered ring, Y^1 is C2-12 alkylene, Y^2 is selected from

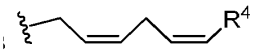
 (in either orientation),  (in either orientation),  (in either orientation),

n is 0 to 3, R^4 is Ci-15 alkyl, Z^1 is Ci-6 alkylene or a direct bond,

Z^2 is 

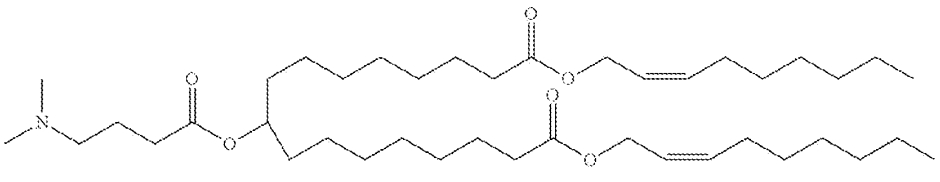
(in either orientation) or absent, provided that if Z^1 is a direct bond, Z^2 is absent;

R^5 is C5-9 alkyl or C6-10 alkoxy, R^6 is C5-9 alkyl or C6-10 alkoxy, W is methylene or a direct bond, and R^7 is H or Me, or a salt thereof, provided that if R^3 and R^2 are C2 alkyls, X^1 is O, X^2 is linear C3 alkylene, X^3 is C(=0), Y^1 is linear Ce alkylene, (Y^2)n-R^4 is



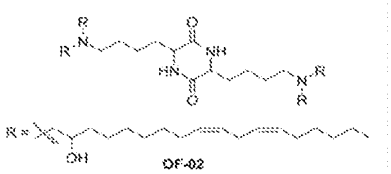
, R^4 is linear C5 alkyl, Z^1 is C2 alkylene, Z^2 is absent, W is methylene, and R^7 is H, then R^5 and R^6 are not Cx alkoxy.

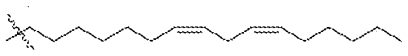
In some embodiments an LNP comprising Formula (xii) is used to deliver a gene modifying composition described herein to the liver and/or hepatocyte cells.



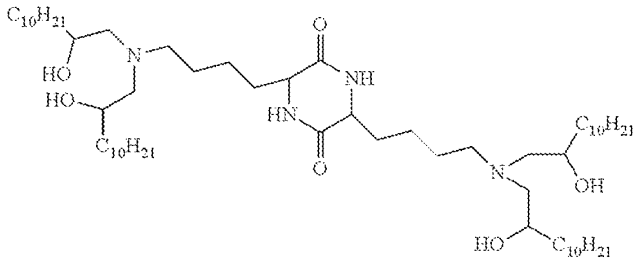
(xi)

In some embodiments an LNP comprising Formula (xi) is used to deliver a gene modifying composition described herein to the liver and/or hepatocyte cells.

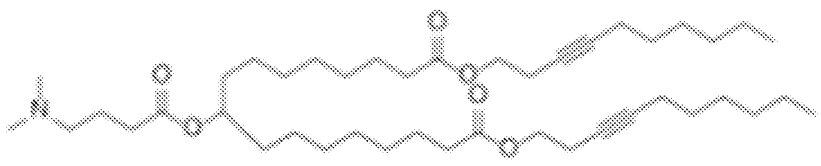


where R= 

(xii)

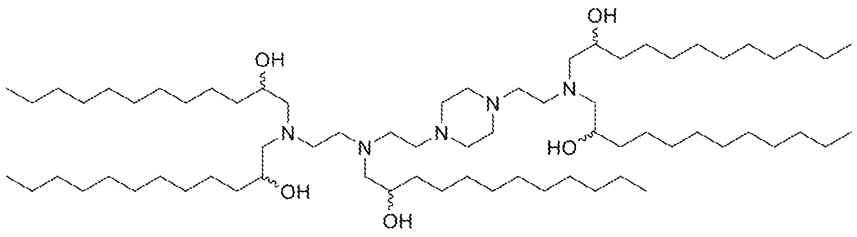


(xiii)



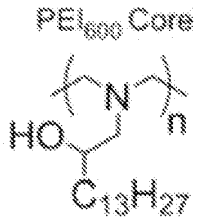
(xiv)

In some embodiments an LNP comprises a compound of Formula (xiii) and a compound of Formula (xiv).



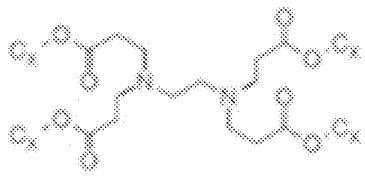
(xv)

In some embodiments an LNP comprising Formula (xv) is used to deliver a gene modifying composition described herein to the liver and/or hepatocyte cells.

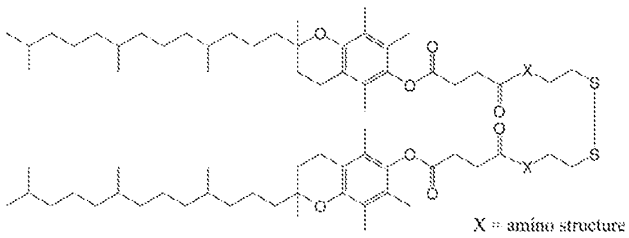


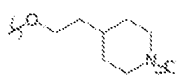
(xvi)

In some embodiments an LNP comprising a formulation of Formula (xvi) is used to deliver a gene modifying composition described herein to the lung endothelial cells.

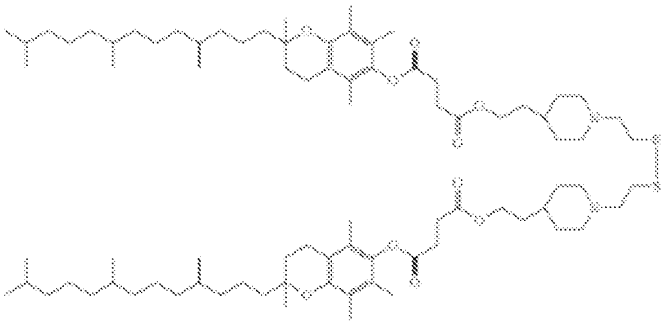


(xvii)

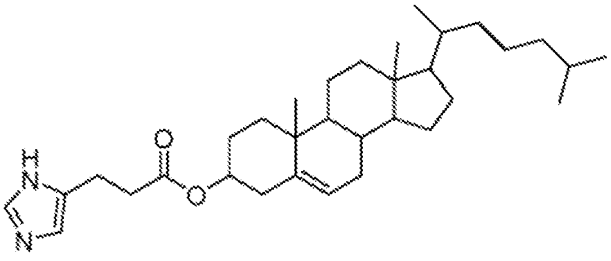


X = amino structure where X=  (xviii)

(a)

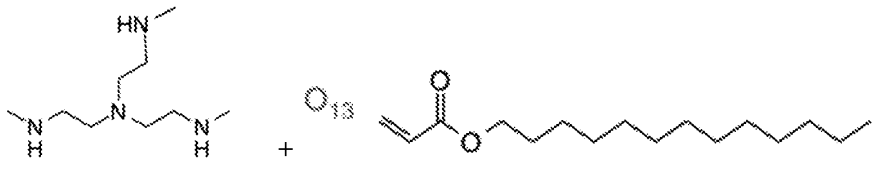


(xviii)(b)

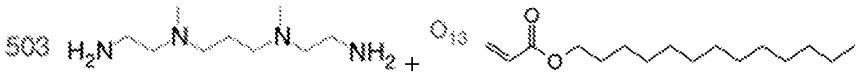


(xix)

In some embodiments, a lipid compound used to form lipid nanoparticles for the delivery of compositions described herein, e.g., nucleic acid (e.g., RNA) described herein (e.g., a template nucleic acid or a nucleic acid encoding a gene modifying polypeptide) is made by one of the following reactions:



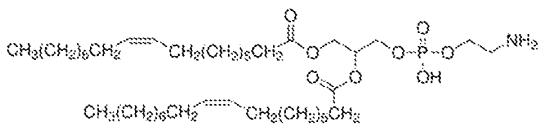
(xx) (a)



(xx)(b)

Exemplary non-cationic lipids include, but are not limited to, distearoyl-sn-glycero-phosphoethanolamine, distearoylphosphatidylcholine (DSPC), dioleoylphosphatidylcholine (DOPC), dipalmitoylphosphatidylcholine (DPPC), dioleoylphosphatidylglycerol (DOPG), dipalmitoylphosphatidylglycerol (DPPG), dioleoyl-phosphatidylethanolamine (DOPE), 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE), palmitoyloleoylphosphatidylcholine (POPC),

palmitoyloleoylphosphatidylethanolamine (POPE), dioleoyl-phosphatidylethanolamine 4-(N-maleimidomethyl)-cyclohexane- 1 - carboxylate (DOPE-mal), dipalmitoyl phosphatidyl ethanolamine (DPPE), dimyristoylphosphoethanolamine (DMPE), distearoyl-phosphatidyl-ethanolamine (DSPE), monomethyl-phosphatidylethanolamine (such as 16-O-monomethyl PE), dimethyl-phosphatidylethanolamine (such as 16-O-dimethyl PE), l8-l-trans PE, l-stearoyl-2-oleoyl-phosphatidyethanolamine (SOPE), hydrogenated soy phosphatidylcholine (HSPC), egg phosphatidylcholine (EPC), dioleoylphosphatidylserine (DOPS), sphingomyelin (SM), dimyristoyl phosphatidylcholine (DMPC), dimyristoyl phosphatidylglycerol (DMPG), distearoylphosphatidylglycerol (DSPG), dierucoylphosphatidylcholine (DEPC), palmitoyloleyolphosphatidylglycerol (POPG), dielaidoyl-phosphatidylethanolamine (DEPE), lecithin, phosphatidylethanolamine, lysolecithin, lysophosphatidylethanolamine, phosphatidylserine, phosphatidylinositol, sphingomyelin, egg sphingomyelin (ESM), cephalin, cardiolipin, phosphatidicacid,cerebrosides, dicetylphosphate, lysophosphatidylcholine, dilinoleoylphosphatidylcholine, or mixtures thereof. It is understood that other diacylphosphatidylcholine and diacylphosphatidylethanolamine phospholipids can also be used. The acyl groups in these lipids are preferably acyl groups derived from fatty acids having C10-C24 carbon chains, e.g., lauroyl, myristoyl, paimitoyl, stearoyl, or oleoyl. Additional exemplary lipids, in certain embodiments, include, without limitation, those described in Kim et al. (2020) dx.doi.org/10.1021/acs.nanolett.0c01386, incorporated herein by reference. Such lipids include, in some embodiments, plant lipids found to improve liver transfection with mRNA (e.g., DGTS). In some embodiments, the non-cationic lipid may have the following structure,



(xxi)

Other examples of non-cationic lipids suitable for use in the lipid nanopartieles include, without limitation, nonphosphorous lipids such as, e.g., stearylamine, dodeeylamine, hexadecylamine, acetyl palmitate, glycerol ricinoleate, hexadecyl stereate, isopropyl myristate, amphoteric acrylic polymers, triethanolamine-lauryl sulfate, alkyl-aryl sulfate polyethyloxylated fatty acid amides, dioctadecyl dimethyl ammonium bromide, ceramide, sphingomyelin, and the like. Other non-cationic lipids are described in WO2017/099823 or US patent publication US2018/0028664, the contents of which is incorporated herein by reference in their entirety.

In some embodiments, the non-cationic lipid is oleic acid or a compound of Formula I, II, or IV of US2018/0028664, incorporated herein by reference in its entirety. The non-cationic lipid can

comprise, for example, 0-30% (mol) of the total lipid present in the lipid nanoparticle. In some embodiments, the non-cationic lipid content is 5-20% (mol) or 10-15% (mol) of the total lipid present in the lipid nanoparticle. In embodiments, the molar ratio of ionizable lipid to the neutral lipid ranges from about 2:1 to about 8:1 (e.g., about 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, or 8:1).

In some embodiments, the lipid nanoparticles do not comprise any phospholipids.

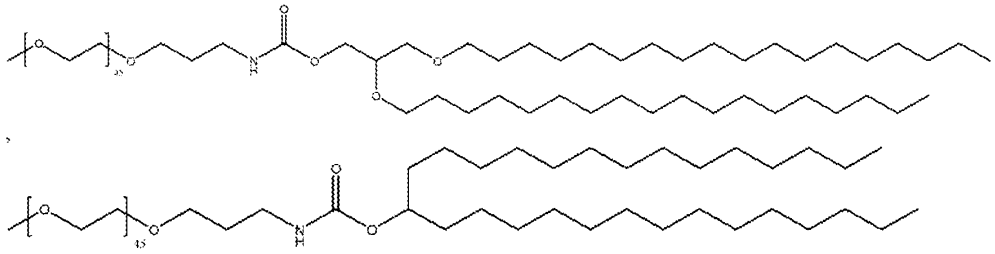
In some aspects, the lipid nanoparticle can further comprise a component, such as a sterol, to provide membrane integrity. One exemplary sterol that can be used in the lipid nanoparticle is cholesterol and derivatives thereof. Non-limiting examples of cholesterol derivatives include polar analogues such as 5a-choiestanol, 53-coprostanol, choiesteryl-(2'-hydroxy)-ethyl ether, choiesteryl-(4'- hydroxy)-butyl ether, and 6-ketocholestanol; non-polar analogues such as 5a-cholestane, cholestenone, 5a-cholestanone, 5p-cholestanone, and cholesteryl decanoate; and mixtures thereof. In some embodiments, the cholesterol derivative is a polar analogue, e.g., choiesteryl-(4'-hydroxy)-buty1 ether. Exemplary cholesterol derivatives are described in PCT publication W02009/127060 and US patent publication US2010/0130588, each of which is incorporated herein by reference in its entirety.

In some embodiments, the component providing membrane integrity, such as a sterol, can comprise 0-50% (mol) (e.g., 0-10%, 10-20%, 20-30%, 30-40%, or 40-50%) of the total lipid present in the lipid nanoparticle. In some embodiments, such a component is 20-50% (mol) 30-40% (mol) of the total lipid content of the lipid nanoparticle.

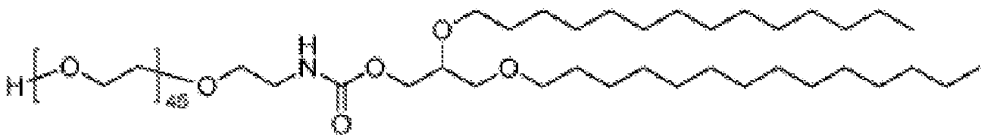
In some embodiments, the lipid nanoparticle can comprise a polyethylene glycol (PEG) or a conjugated lipid molecule. Generally, these are used to inhibit aggregation of lipid nanoparticles and/or provide steric stabilization. Exemplary conjugated lipids include, but are not limited to, PEG-lipid conjugates, polyoxazoline (POZ)-lipid conjugates, polyamide-lipid conjugates (such as ATTA-lipid conjugates), cationic-polymer lipid (CPL) conjugates, and mixtures thereof. In some embodiments, the conjugated lipid molecule is a PEG-lipid conjugate, for example, a (methoxy polyethylene glycol)-conjugated lipid.

Exemplary PEG-lipid conjugates include, but are not limited to, PEG-diacylglycerol (DAG) (such as l-(monomethoxy-polyethyleneglycol)-2,3-dimyristoylglycerol (PEG-DMG)), PEG-dialkyloxypropyl (DAA), PEG-phospholipid, PEG-ceramide (Cer), a pegylated phosphatidylethanoloamine (PEG-PE), 1,2-dimyristoyl-sn-glycerol, methoxypoly ethylene glycol (DMG-PEG-2K), PEG succinate diacylglycerol (PEGS-DAG) (such as 4-0-(2',3'-di(tetradecanoyloxy)propyl-l-0-(w-methoxy(polyethoxy)ethyl) butanedioate (PEG-S-DMG)), PEG dialkoxypropylcarbam, N-(carbonyl-methoxypolyethylene glycol 2000)-l,2-distearoyl-sn-glycero-3-phosphoethanolamine sodium salt, or a mixture thereof. Additional exemplary PEG-lipid conjugates are described, for example, in US5,885,6l3, US6,287,59l,

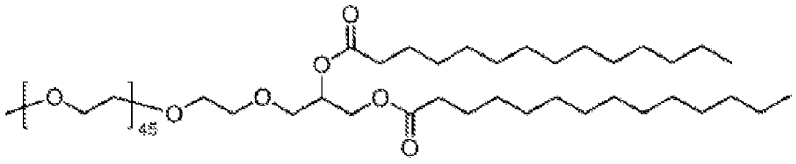
US2003/0077829, US2003/0077829, US2005/0175682, US2008/0020058, US2011/0117125, US2010/0130588, US2016/0376224, US2017/0119904, and US/099823, the contents of all of which are incorporated herein by reference in their entirety. In some embodiments, a PEG-lipid is a compound of Formula III, III-a-I, III-a-2, III-b-1, III-b-2, or V of US2018/0028664, the content of which is incorporated herein by reference in its entirety. In some embodiments, a PEG-lipid is of Formula II of US20150376115 or US2016/0376224, the content of both of which is incorporated herein by reference in its entirety. In some embodiments, the PEG-DAA conjugate can be, for example, PEG-dilauryloxypropyl, PEG- dimyristyloxypropyl, PEG-dipalmityloxypropyl, or PEG-distearyloxypropyl. The PEG-lipid can be one or more of PEG-DMG, PEG-dilaurylglycerol, PEG-dipalmitoylglycerol, PEG-disterylglycerol, PEG-dilaurylglycamide, PEG-dimyristylglycamide, PEG- dipalmitoylglycamide, PEG-disterylglycamide, PEG-cholesterol (l-[8'-(Cholest-5-en-3[beta]- oxy)carboxamido-3',6'-dioxaoctanyl] carbamoyl-[omega]-methyl-poly(ethylene glycol), PEG- DMB (3,4-Ditetradecoxylbenzyl- [omega]-methyl-poly(ethylene glycol) ether), and 1,2- dimyristoyl-sn-glycero-3-phosphoethanolamine-N-[methoxy(polyethylene glycol)-2000]. In some embodiments, the PEG-lipid comprises PEG-DMG, 1,2-dimyristoyl-sn-glycero-3-phosphoethanolamine-N-[methoxy(polyethylene glycol)-2000]. In some embodiments, the PEG-lipid comprises a structure selected from:



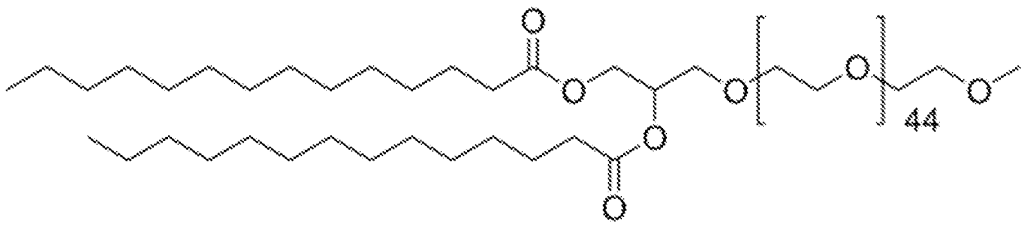
(xxii),



(xxiii),



(xxiv), and



(xxv).

In some embodiments, lipids conjugated with a molecule other than a PEG can also be used in place of PEG-lipid. For example, polyoxazoline (POZ)-lipid conjugates, polyamide-lipid conjugates (such as ATTA-lipid conjugates), and cationic-polymer lipid (GPL) conjugates can be used in place of or in addition to the PEG-lipid.

Exemplary conjugated lipids, i.e., PEG-lipids, (POZ)-lipid conjugates, ATTA-lipid conjugates and cationic polymer-lipids are described in the PCT and LIS patent applications listed in Table 2 of WO2019051289A9 and in WO2020106946A1, the contents of all of which are incorporated herein by reference in their entirety.

In some embodiments an LNP comprises a compound of Formula (xix), a compound of Formula (xxi) and a compound of Formula (xxv). In some embodiments an LNP comprising a formulation of Formula (xix), Formula (xxi) and Formula (xxv)is used to deliver a gene modifying composition described herein to the lung or pulmonary cells.

In some embodiments, a lipid nanoparticle may comprise one or more cationic lipids selected from Formula (i), Formula (ii), Formula (iii), Formula (vii), and Formula (ix). In some embodiments, the LNP may further comprise one or more neutral lipid, e.g., DSPC, DPPC, DMPC, DOPC, POPC, DOPE, SM, a steroid, e.g., cholesterol, and/or one or more polymer conjugated lipid, e.g., a pegylated lipid, e.g., PEG-DAG, PEG-PE, PEG-S-DAG, PEG-cer or a PEG dialkyoxypropylcarbamate.

In some embodiments, the PEG or the conjugated lipid can comprise 0-20% (mol) of the total lipid present in the lipid nanoparticle. In some embodiments, PEG or the conjugated lipid content is 0.5-10% or 2-5% (mol) of the total lipid present in the lipid nanoparticle. Molar ratios of the ionizable lipid, non-cationic-lipid, sterol, and PEG/conjugated lipid can be varied as needed. For example, the lipid particle can comprise 30-70% ionizable lipid by mole or by total weight of the composition, 0-60% cholesterol by mole or by total weight of the composition, 0-30% non-cationic-lipid by mole or by total weight of the composition and 1-10% conjugated lipid by mole or by total weight of the composition. Preferably, the composition comprises 30-40% ionizable lipid by mole or by total weight of the composition, 40-50% cholesterol by mole or by total weight of the composition, and 10- 20% non-

cationic-lipid by mole or by total weight of the composition. In some other embodiments, the composition is 50-75% ionizable lipid by mole or by total weight of the composition, 20-40% cholesterol by mole or by total weight of the composition, and 5 to 10% non-cationic-lipid, by mole or by total weight of the composition and 1-10% conjugated lipid by mole or by total weight of the composition. The composition may contain 60-70% ionizable lipid by mole or by total weight of the composition, 25-35% cholesterol by mole or by total weight of the composition, and 5-10% non-cationic-lipid by mole or by total weight of the composition. The composition may also contain up to 90% ionizable lipid by mole or by total weight of the composition and 2 to 15% non-cationic lipid by mole or by total weight of the composition. The formulation may also be a lipid nanoparticle formulation, for example comprising 8-30% ionizable lipid by mole or by total weight of the composition, 5-30% non- cationic lipid by mole or by total weight of the composition, and 0-20% cholesterol by mole or by total weight of the composition; 4-25% ionizable lipid by mole or by total weight of the composition, 4-25% non-cationic lipid by mole or by total weight of the composition, 2 to 25% cholesterol by mole or by total weight of the composition, 10 to 35% conjugate lipid by mole or by total weight of the composition, and 5% cholesterol by mole or by total weight of the composition; or 2-30% ionizable lipid by mole or by total weight of the composition, 2-30% non-cationic lipid by mole or by total weight of the composition, 1 to 15% cholesterol by mole or by total weight of the composition, 2 to 35% conjugate lipid by mole or by total weight of the composition, and 1-20% cholesterol by mole or by total weight of the composition; or even up to 90% ionizable lipid by mole or by total weight of the composition and 2-10% non-cationic lipids by mole or by total weight of the composition, or even 100% cationic lipid by mole or by total weight of the composition. In some embodiments, the lipid particle formulation comprises ionizable lipid, phospholipid, cholesterol and a PEG-ylated lipid in a molar ratio of 50: 10:38.5: 1.5. In some other embodiments, the lipid particle formulation comprises ionizable lipid, cholesterol and a PEG-ylated lipid in a molar ratio of 60:38.5: 1.5.

In some embodiments, the lipid particle comprises ionizable lipid, non-cationic lipid (e.g. phospholipid), a sterol (e.g., cholesterol) and a PEG-ylated lipid, where the molar ratio of lipids ranges from 20 to 70 mole percent for the ionizable lipid, with a target of 40-60, the mole percent of non-cationic lipid ranges from 0 to 30, with a target of 0 to 15, the mole percent of sterol ranges from 20 to 70, with a target of 30 to 50, and the mole percent of PEG-ylated lipid ranges from 1 to 6, with a target of 2 to 5.

In some embodiments, the lipid particle comprises ionizable lipid / non-cationic- lipid / sterol / conjugated lipid at a molar ratio of 50: 10:38.5: 1.5.

In an aspect, the disclosure provides a lipid nanoparticle formulation comprising phospholipids, lecithin, phosphatidylcholine and phosphatidylethanolamine.

In some embodiments, one or more additional compounds can also be included. Those compounds can be administered separately or the additional compounds can be included in the lipid

nanoparticles of the invention. In other words, the lipid nanoparticles can contain other compounds in addition to the nucleic acid or at least a second nucleic acid, different than the first. Without limitations, other additional compounds can be selected from the group consisting of small or large organic or inorganic molecules, monosaccharides, disaccharides, trisaccharides, oligosaccharides, polysaccharides, peptides, proteins, peptide analogs and derivatives thereof, peptidomimetics, nucleic acids, nucleic acid analogs and derivatives, an extract made from biological materials, or any combinations thereof.

In some embodiments, a lipid nanoparticle (or a formulation comprising lipid nanoparticles) lacks reactive impurities (e.g., aldehydes or ketones), or comprises less than a preselected level of reactive impurities (e.g., aldehydes or ketones). While not wishing to be bound by theory, in some embodiments, a lipid reagent is used to make a lipid nanoparticle formulation, and the lipid reagent may comprise a contaminating reactive impurity (e.g., an aldehyde or ketone). A lipid regent may be selected for manufacturing based on having less than a preselected level of reactive impurities (e.g., aldehydes or ketones). Without wishing to be bound by theory, in some embodiments, aldehydes can cause modification and damage of RNA, e.g., cross-linking between bases and/or covalently conjugating lipid to RNA (e.g., forming lipid-RNA adducts). This may, in some instances, lead to failure of a reverse transcriptase reaction and/or incorporation of inappropriate bases, e.g., at the site(s) of lesion(s), e.g., a mutation in a newly synthesized target DNA.

In some embodiments, a lipid nanoparticle formulation is produced using a lipid reagent comprising less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content. In some embodiments, a lipid nanoparticle formulation is produced using a lipid reagent comprising less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species. In some embodiments, a lipid nanoparticle formulation is produced using a lipid reagent comprising: (i) less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content; and (ii) less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species. In some embodiments, the lipid nanoparticle formulation is produced using a plurality of lipid reagents, and each lipid reagent of the plurality independently meets one or more criterion described in this paragraph. In some embodiments, each lipid reagent of the plurality meets the same criterion, e.g., a criterion of this paragraph.

In some embodiments, the lipid nanoparticle formulation comprises less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content. In some embodiments, the lipid nanoparticle formulation comprises less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species. In some embodiments, the lipid nanoparticle formulation comprises: (i) less than 5%,

4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content; and (ii) less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species.

In some embodiments, one or more, or optionally all, of the lipid reagents used for a lipid nanoparticle as described herein or a formulation thereof comprise less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content. In some embodiments, one or more, or optionally all, of the lipid reagents used for a lipid nanoparticle as described herein or a formulation thereof comprise less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species. In some embodiments, one or more, or optionally all, of the lipid reagents used for a lipid nanoparticle as described herein or a formulation thereof comprise: (i) less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% total reactive impurity (e.g., aldehyde) content; and (ii) less than 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, or 0.1% of any single reactive impurity (e.g., aldehyde) species.

In some embodiments, total aldehyde content and/or quantity of any single reactive impurity (e.g., aldehyde) species is determined by liquid chromatography (LC), e.g., coupled with tandem mass spectrometry (MS/MS), e.g., according to the method described in Example 40 of PCT/US21/20948. In some embodiments, reactive impurity (e.g., aldehyde) content and/or quantity of reactive impurity (e.g., aldehyde) species is determined by detecting one or more chemical modifications of a nucleic acid molecule (e.g., an RNA molecule, e.g., as described herein) associated with the presence of reactive impurities (e.g., aldehydes), e.g., in the lipid reagents. In some embodiments, reactive impurity (e.g., aldehyde) content and/or quantity of reactive impurity (e.g., aldehyde) species is determined by detecting one or more chemical modifications of a nucleotide or nucleoside (e.g., a ribonucleotide or ribonucleoside, e.g., comprised in or isolated from a template nucleic acid, e.g., as described herein) associated with the presence of reactive impurities (e.g., aldehydes), e.g., in the lipid reagents, e.g., according to the method described in Example 41 of PCT/US21/20948. In embodiments, chemical modifications of a nucleic acid molecule, nucleotide, or nucleoside are detected by determining the presence of one or more modified nucleotides or nucleosides, e.g., using LC-MS/MS analysis, e.g., according to the method described in Example 41 of PCT/US21/20948.

In some embodiments, a nucleic acid (e.g., RNA) described herein (e.g., a template nucleic acid or a nucleic acid encoding a gene modifying polypeptide) does not comprise an aldehyde modification, or comprises less than a preselected amount of aldehyde modifications. In some embodiments, on average, a nucleic acid has less than 50, 20, 10, 5, 2, or 1 aldehyde modifications per 1000 nucleotides, e.g., wherein

a single cross-linking of two nucleotides is a single aldehyde modification. In some embodiments, the aldehyde modification is an RNA adduct (e.g., a lipid-RNA adduct). In some embodiments, the aldehyde-modified nucleotide is cross-linking between bases . In some embodiments, a nucleic acid (e.g., RNA) described herein comprises less than 50, 20, 10, 5, 2, or 1 cross-links between nucleotide.

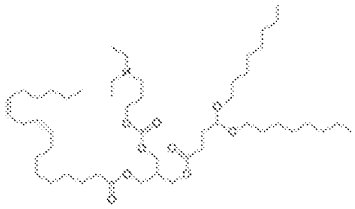
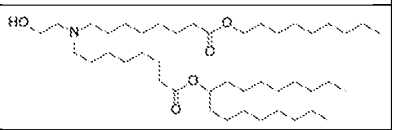
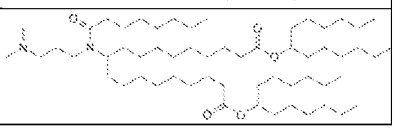
In some embodiments, LNPs are directed to specific tissues by the addition of targeting domains. For example, biological ligands may be displayed on the surface of LNPs to enhance interaction with cells displaying cognate receptors, thus driving association with and cargo delivery to tissues wherein cells express the receptor. In some embodiments, the biological ligand may be a ligand that drives delivery to the liver, e.g., LNPs that display GalNAc result in delivery of nucleic acid cargo to hepatocytes that display asialoglycoprotein receptor (ASGPR). The work of Akinc et al. *Mol Ther* 18(7):1357-1364 (2010) teaches the conjugation of a trivalent GalNAc ligand to a PEG-lipid (GalNAc-PEG-DSG) to yield LNPs dependent on ASGPR for observable LNP cargo effect (see, e.g., Figure 6 therein). Other ligand-displaying LNP formulations, e.g., incorporating folate, transferrin, or antibodies, are discussed in WO2017223135, which is incorporated herein by reference in its entirety, in addition to the references used therein, namely Kolhatkar et al., Curr Drug Discov Technol. 2011 8:197-206; Musacchio and Torchilin, Front Biosci. 2011 16:1388-1412; Yu et al., Mol Membr Biol. 2010 27:286-298; Patil et al., Crit Rev Ther Drug Carrier Syst. 2008 25:1-61 ; Benoit et al., Biomacromolecules. 2011 12:2708-2714; Zhao et al., Expert Opin Drug Deliv. 2008 5:309-319; Akinc et al., Mol Ther. 2010 18:1357-1364; Srinivasan et al., Methods Mol Biol. 2012 820:105-116; Ben-Arie et al., Methods Mol Biol. 2012 757:497-507; Peer 2010 J Control Release. 20:63-68; Peer et al., Proc Natl Acad Sci U S A. 2007 104:4095-4100; Kim et al., Methods Mol Biol. 2011 721:339-353; Subramanya et al., Mol Ther. 2010 18:2028-2037; Song et al., Nat Biotechnol. 2005 23:709-717; Peer et al., Science. 2008 319:627-630; and Peer and Lieberman, Gene Ther. 2011 18:1127-1133.

In some embodiments, LNPs are selected for tissue-specific activity by the addition of a Selective ORgan Targeting (SORT) molecule to a formulation comprising traditional components, such as ionizable cationic lipids, amphipathic phospholipids, cholesterol and poly(ethylene glycol) (PEG) lipids. The teachings of Cheng et al. Nat Nanotechnol 15(4):313-320 (2020) demonstrate that the addition of a supplemental "SORT" component precisely alters the *in vivo* RNA delivery profile and mediates tissue-specific (e.g., lungs, liver, spleen) gene delivery and editing as a function of the percentage and biophysical property of the SORT molecule.

In some embodiments, the LNPs comprise biodegradable, ionizable lipids. In some embodiments, the LNPs comprise (9Z,l2Z)-3-((4,4-bis(octyloxy)butanoyl)oxy)-2-((((3-(diethylamino)propoxy)carbonyl)oxy)methyl)propyl octadeca-9,l2-dienoate, also called 3- ((4,4-bis(octyloxy)butanoyl)oxy)-2-((((3-(diethylamino)propoxy)carbonyl)oxy)methyl)propyl (9Z,l2Z)-

octadeca-9,12-dienoate) or another ionizable lipid. See, e.g, lipids of WO2019/067992, WO/2017/173054, WO2015/095340, and WO2014/136086, as well as references provided therein. In some embodiments, the term cationic and ionizable in the context of LNP lipids is interchangeable, e.g., wherein ionizable lipids are cationic depending on the pH.

In some embodiments, an LNP described herein comprises a lipid described in Table 19

LIPID ID	Chemical Name	Molecular Weight	Structure
LIPIDV003	(9Z,12Z)-3-((4,4-bis(octyloxy)butanoyl)oxy)-2-((((3-(diethylamino)propoxy)carbonyl)oxy)methyl)propyl octadeca-9, 12-dienoate	852.29	
LIPIDV004	Heptadecan-9-yl 8-((2-hydroxyethyl)(8-(nonyloxy)-8-oxooctyl)amino)octanoate	710.18	
LIPIDV005		919.56	

In some embodiments, multiple components of a gene modifying system may be prepared as a single LNP formulation, e.g., an LNP formulation comprises mRNA encoding for the gene modifying polypeptide and an RNA template. Ratios of nucleic acid components may be varied in order to maximize the properties of a therapeutic. In some embodiments, the ratio of RNA template to mRNA encoding a gene modifying polypeptide is about 1:1 to 100:1, e.g., about 1:1 to 20:1, about 20:1 to 40:1, about 40:1 to 60:1, about 60:1 to 80:1, or about 80:1 to 100:1, by molar ratio. In other embodiments, a system of multiple nucleic acids may be prepared by separate formulations, e.g., one LNP formulation comprising a template RNA and a second LNP formulation comprising an mRNA encoding a gene modifying polypeptide. In some embodiments, the system may comprise more than two nucleic acid components formulated into LNPs. In some embodiments, the system may comprise a protein, e.g., a gene modifying polypeptide, and a template RNA formulated into at least one LNP formulation.

In some embodiments, the average LNP diameter of the LNP formulation may be between 10s of nm and 100s of nm, e.g., measured by dynamic light scattering (DLS). In some embodiments, the average LNP diameter of the LNP formulation may be from about 40 nm to about 150 nm, such as about 40 nm, 45 nm, 50 nm, 55 nm, 60 nm, 65 nm, 70 nm, 75 nm, 80 nm, 85 nm, 90 nm, 95 nm, 100 nm, 105 nm, 110 nm, 115 nm, 120 nm, 125 nm, 130 nm, 135 nm, 140 nm, 145 nm, or 150 nm. In some embodiments, the

average LNP diameter of the LNP formulation may be from about 50 nm to about 100 nm, from about 50 nm to about 90 nm, from about 50 nm to about 80 nm, from about 50 nm to about 70 nm, from about 50 nm to about 60 nm, from about 60 nm to about 100 nm, from about 60 nm to about 90 nm, from about 60 nm to about 80 nm, from about 60 nm to about 70 nm, from about 70 nm to about 100 nm, from about 70 nm to about 90 nm, from about 70 nm to about 80 nm, from about 80 nm to about 100 nm, from about 80 nm to about 90 nm, or from about 90 nm to about 100 nm. In some embodiments, the average LNP diameter of the LNP formulation may be from about 70 nm to about 100 nm. In a particular embodiment, the average LNP diameter of the LNP formulation may be about 80 nm. In some embodiments, the average LNP diameter of the LNP formulation may be about 100 nm. In some embodiments, the average LNP diameter of the LNP formulation ranges from about l mm to about 500 mm, from about 5 mm to about 200 mm, from about 10 mm to about 100 mm, from about 20 mm to about 80 mm, from about 25 mm to about 60 mm, from about 30 mm to about 55 mm, from about 35 mm to about 50 mm, or from about 38 mm to about 42 mm.

An LNP may, in some instances, be relatively homogenous. A polydispersity index may be used to indicate the homogeneity of an LNP, e.g., the particle size distribution of the lipid nanoparticles. A small (e.g., less than 0.3) polydispersity index generally indicates a narrow particle size distribution. An LNP may have a polydispersity index from about 0 to about 0.25, such as 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10, 0.11, 0.12, 0.13, 0.14, 0.15, 0.16, 0.17, 0.18, 0.19, 0.20, 0.21, 0.22, 0.23, 0.24, or 0.25. In some embodiments, the polydispersity index of an LNP may be from about 0.10 to about 0.20.

The zeta potential of an LNP may be used to indicate the electrokinetic potential of the composition. In some embodiments, the zeta potential may describe the surface charge of an LNP. Lipid nanoparticles with relatively low charges, positive or negative, are generally desirable, as more highly charged species may interact undesirably with cells, tissues, and other elements in the body. In some embodiments, the zeta potential of an LNP may be from about -10 mV to about +20 mV, from about -10 mV to about +15 mV, from about -10 mV to about +10 mV, from about -10 mV to about +5 mV, from about -10 mV to about 0 mV, from about -10 mV to about -5 mV, from about -5 mV to about +20 mV, from about -5 mV to about +15 mV, from about -5 mV to about +10 mV, from about -5 mV to about +5 mV, from about -5 mV to about 0 mV, from about 0 mV to about +20 mV, from about 0 mV to about +15 mV, from about 0 mV to about +10 mV, from about 0 mV to about +5 mV, from about +5 mV to about +20 mV, from about +5 mV to about +15 mV, or from about +5 mV to about +10 mV.

The efficiency of encapsulation of a protein and/or nucleic acid, e.g., gene modifying polypeptide or mRNA encoding the polypeptide, describes the amount of protein and/or nucleic acid that is encapsulated or otherwise associated with an LNP after preparation, relative to the initial amount provided. The encapsulation efficiency is desirably high (e.g., close to 100%). The encapsulation

efficiency may be measured, for example, by comparing the amount of protein or nucleic acid in a solution containing the lipid nanoparticle before and after breaking up the lipid nanoparticle with one or more organic solvents or detergents. An anion exchange resin may be used to measure the amount of free protein or nucleic acid (e.g., RNA) in a solution. Fluorescence may be used to measure the amount of free protein and/or nucleic acid (e.g., RNA) in a solution. For the lipid nanoparticles described herein, the encapsulation efficiency of a protein and/or nucleic acid may be at least 50%, for example 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%. In some embodiments, the encapsulation efficiency may be at least 80%. In some embodiments, the encapsulation efficiency may be at least 90%. In some embodiments, the encapsulation efficiency may be at least 95%.

An LNP may optionally comprise one or more coatings. In some embodiments, an LNP may be formulated in a capsule, film, or table having a coating. A capsule, film, or tablet including a composition described herein may have any useful size, tensile strength, hardness or density.

Additional exemplary lipids, formulations, methods, and characterization of LNPs are taught by WO2020061457, which is incorporated herein by reference in its entirety.

In some embodiments, *in vitro* or *ex vivo* cell lipofections are performed using Lipofectamine MessengerMax (Thermo Fisher) or TransIT-mRNA Transfection Reagent (Mirus Bio). In certain embodiments, LNPs are formulated using the GenVoy_ILM ionizable lipid mix (Precision NanoSystems). In certain embodiments, LNPs are formulated using 2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]-dioxolane (DLin-KC2-DMA) or dilinoleylmethyl-4-dimethylaminobutyrate (DLin-MC3-DMA or MC3), the formulation and *in vivo* use of which are taught in Jayaraman et al. Angew Chem Int Ed Engl 51(34):8529-8533 (2012), incorporated herein by reference in its entirety.

LNP formulations optimized for the delivery of CRISPR-Cas systems, e.g., Cas9-gRNA RNP, gRNA, Cas9 mRNA, are described in WO2019067992 and WO2019067910, both incorporated by reference.

Additional specific LNP formulations useful for delivery of nucleic acids are described in US8158601 and US8168775, both incorporated by reference, which include formulations used in patisiran, sold under the name ONPATTRO.

Exemplary dosing of gene modifying LNP may include about 0.1, 0.25, 0.3, 0.5, 1, 2, 3, 4, 5, 6, 8, 10, or 100 mg/kg (RNA). Exemplary dosing of AAV comprising a nucleic acid encoding one or more components of the system may include an MOI of about 10^{11}, 10^{12}, 10^{13}, and 10^{14} vg/kg.

Kits, Articles of Manufacture, and Pharmaceutical Compositions

In an aspect the disclosure provides a kit comprising a gene modifying polypeptide or a gene modifying system, e.g., as described herein. In some embodiments, the kit comprises a gene modifying polypeptide (or a nucleic acid encoding the polypeptide) and a template RNA (or DNA encoding the template RNA). In some embodiments, the kit further comprises a reagent for introducing the system into a cell, e.g., transfection reagent, LNP, and the like. In some embodiments, the kit is suitable for any of the methods described herein. In some embodiments, the kit comprises one or more elements, compositions (e.g., pharmaceutical compositions), gene modifying polypeptides, and/or gene modifying systems, or a functional fragment or component thereof, e.g., disposed in an article of manufacture. In some embodiments, the kit comprises instructions for use thereof.

In an aspect, the disclosure provides an article of manufacture, e.g., in which a kit as described herein, or a component thereof, is disposed.

In an aspect, the disclosure provides a pharmaceutical composition comprising a gene modifying polypeptide or a gene modifying system, e.g., as described herein. In some embodiments, the pharmaceutical composition further comprises a pharmaceutically acceptable carrier or excipient. In some embodiments, the pharmaceutical composition comprises a template RNA and/or an RNA encoding the polypeptide. In embodiments, the pharmaceutical composition has one or more (e.g., 1, 2, 3, or 4) of the following characteristics:

(a) less than 1% (e.g., less than 0.5%, 0.4%, 0.3%, 0.2%, or 0.1%) DNA template relative to the template RNA and/or the RNA encoding the polypeptide, e.g., on a molar basis;

(b) less than 1% (e.g., less than 0.5%, 0.4%, 0.3%, 0.2%, or 0.1%) uncapped RNA relative to the template RNA and/or the RNA encoding the polypeptide, e.g., on a molar basis;

(c) less than 1% (e.g., less than 0.5%, 0.4%, 0.3%, 0.2%, or 0.1%) partial length RNAs relative to the template RNA and/or the RNA encoding the polypeptide, e.g., on a molar basis;

(d) substantially lacks unreacted cap dinucleotides.

i)

Chemistry, Manufacturing, and Controls (CMC)

Purification of protein therapeutics is described, for example, in Franks, Protein Biotechnology: *Isolation, Characterization, and Stabilization*, Humana Press (2013); and in Cutler, *Protein Purification Protocols (Methods in Molecular Biology)*, Humana Press (2010).

In some embodiments, a gene modifying system, polypeptide, and/or template nucleic acid (e.g., template RNA) conforms to certain quality standards. In some embodiments, a gene modifying system, polypeptide, and/or template nucleic acid (e.g., template RNA) produced by a method described herein conforms to certain quality standards. Accordingly, the disclosure is directed, in some aspects, to

methods of manufacturing a gene modifying system, polypeptide, and/or template nucleic acid (e.g., template RNA) that conforms to certain quality standards, e.g., in which said quality standards are assayed. The disclosure is also directed, in some aspects, to methods of assaying said quality standards in a gene modifying system, polypeptide, and/or template nucleic acid (e.g., template RNA). In some embodiments, quality standards include, but are not limited to, one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12) of the following:

(i) the length of the template RNA, e.g., whether the template RNA has a length that is above a reference length or within a reference length range, e.g., whether at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% of the template RNA present is greater than 100, 125, 150, 175, or 200 nucleotides long;

(ii) the presence, absence, and/or length of a polyA tail on the template RNA, e.g., whether at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% of the template RNA present contains a polyA tail (e.g., a polyA tail that is at least 5, 10, 20, 30, 50, 70, 100 nucleotides in length);

(iii) the presence, absence, and/or type of a 5′ cap on the template RNA, e.g., whether at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% of the template RNA present contains a 5′ cap, e.g., whether that cap is a 7-methylguanosine cap, e.g., a O-Me-m7G cap;

(iv) the presence, absence, and/or type of one or more modified nucleotides (e.g., selected from pseudouridine, dihydrouridine, inosine, 7-methylguanosine, 1-N-methylpseudouridine (1-Me-Ψ), 5-methoxyuridine (5-MO-U), 5-methylcytidine (5mC), or a locked nucleotide) in the template RNA, e.g., whether at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% of the template RNA present contains one or more modified nucleotides;

(v) the stability of the template RNA (e.g., over time and/or under a pre-selected condition), e.g., whether at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% of the template RNA remains intact (e.g., greater than 100, 125, 150, 175, or 200 nucleotides long) after a stability test;

(vi) the potency of the template RNA in a system for modifying DNA, e.g., whether at least 1% of target sites are modified after a system comprising the template RNA is assayed for potency;

(vii) the length of the polypeptide, first polypeptide, or second polypeptide, e.g., whether the polypeptide, first polypeptide, or second polypeptide has a length that is above a reference length or within a reference length range, e.g., whether at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% of the polypeptide, first polypeptide, or second polypeptide present is greater than 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1600, 1700, 1800, 1900, or 2000 amino acids long (and optionally, no larger than 2500, 2000, 1500, 1400, 1300, 1200, 1100, 1000, 900, 800, 700, or 600 amino acids long);

(viii) the presence, absence, and/or type of post-translational modification on the polypeptide, first polypeptide, or second polypeptide, e.g., whether at least 80, 85, 90, 95, 96, 97, 98, or 99% of the polypeptide, first polypeptide, or second polypeptide contains phosphorylation, methylation, acetylation, myristoylation, palmitoylation, isoprenylation, glipyatyon, or lipoylation, or any combination thereof;

(ix) the presence, absence, and/or type of one or more artificial, synthetic, or non-canonical amino acids (e.g., selected from ornithine, β-alanine, GABA, δ-Aminolevulinic acid, PABA, a D-amino acid (e.g., D-alanine or D-glutamate), aminoisobutyric acid, dehydroalanine, cystathionine, lanthionine, Djenkolic acid, Diaminopimelic acid, Homoalanine, Norvaline, Norleucine, Homonorleucine, homoserine, O-methyl-homoserine and O-ethyl-homoserine, ethionine, selenocysteine, selenohomocysteine, selenomethionine, selenoethionine, tellurocysteine, or telluromethionine) in the polypeptide, first polypeptide, or second polypeptide, e.g., whether at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% of the polypeptide, first polypeptide, or second polypeptide present contains one or more artificial, synthetic, or non-canonical amino acids;

(x) the stability of the polypeptide, first polypeptide, or second polypeptide (e.g., over time and/or under a pre-selected condition), e.g., whether at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% of the polypeptide, first polypeptide, or second polypeptide remains intact (e.g., greater than 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1600, 1700, 1800, 1900, or 2000 amino acids long (and optionally, no larger than 2500, 2000, 1500, 1400, 1300, 1200, 1100, 1000, 900, 800, 700, or 600 amino acids long)) after a stability test;

(xi) the potency of the polypeptide, first polypeptide, or second polypeptide in a system for modifying DNA, e.g., whether at least 1% of target sites are modified after a system comprising the polypeptide, first polypeptide, or second polypeptide is assayed for potency; or

(xii) the presence, absence, and/or level of one or more of a pyrogen, virus, fungus, bacterial pathogen, or host cell protein, e.g., whether the system is free or substantially free of pyrogen, virus, fungus, bacterial pathogen, or host cell protein contamination.

In some embodiments, a system or pharmaceutical composition described herein is endotoxin free.

In some embodiments, the presence, absence, and/or level of one or more of a pyrogen, virus, fungus, bacterial pathogen, and/or host cell protein is determined. In embodiments, whether the system is free or substantially free of pyrogen, virus, fungus, bacterial pathogen, and/or host cell protein contamination is determined.

In some embodiments, a pharmaceutical composition or system as described herein has one or more (e.g., 1, 2, 3, or 4) of the following characteristics:

(a) less than 1% (e.g., less than 0.5%, 0.4%, 0.3%, 0.2%, or 0.1%) DNA template relative to the template RNA and/or the RNA encoding the polypeptide, e.g., on a molar basis;

(b) less than 1% (e.g., less than 0.5%, 0.4%, 0.3%, 0.2%, or 0.1%) uncapped RNA relative to the template RNA and/or the RNA encoding the polypeptide, e.g., on a molar basis;

(c) less than 1% (e.g., less than 0.5%, 0.4%, 0.3%, 0.2%, or 0.1%) partial length RNAs relative to the template RNA and/or the RNA encoding the polypeptide, e.g., on a molar basis;

(d) substantially lacks unreacted cap dinucleotides.

EXAMPLES

Example 1: Trans-recruitment of an RNA template for RNA-based gene modification

This example describes the use of an exemplary three (3)-component RNA gene modifying system for the targeted editing of a sequence in the human genome (FIG. 1). More specifically, in this exemplary system, a (1) gene modifying polypeptide binds (2) a guide RNA molecule and (3) a trans-template RNA molecule, together forming a gene modifying complex, which binds and nicks the target locus and performs templated incorporation of the desired edit in one strand of the genomic DNA. Host repair pathways cause the incorporation of the edit in the second strand, and ligation of the DNA nick. This example further describes the introduction of the 3-component system to mammalian cells for in vitro gene modifying as a means of evaluating the efficacy of various gene modifying system configurations and trans-template RNA recruitment interactions on editing activity in human cells.

1) Gene modifying polypeptide:

The gene modifying polypeptide of this exemplary system includes:

1. a Cas-nuclease with one endonuclease domain inactivated (e.g. Spy N863A Cas9),
2. a reverse transcriptase (RT), and
3. an RNA binding protein (RBP; as in Table 31). The RBP may be repeated 1-5 times in tandem with or without intervening peptide linker sequences.

These 3 domains are brought together either by covalent linkage using (1) peptide linkers, as in Table 32 (FIG. 2A), or (2) intein pairs, as in Table 33, or (3) by the dimerization of two fusion domains (FDs) that (a) is induced by chemical binding or specific wavelengths of light Table 34, or (b) occurs intrinsically in a ligand-free interaction Tables 35 and 36 (FIG. 2B). The FDs may be repeated, in some instances, 1-30 times in tandem with or without intervening peptide linker sequences.

The 3 domains can be assembled in any of the following configurations, where the linkers indicated below can optionally be replaced with intein- or FD-pairs:

- Cas9 – linker – RT – linker – $(RBP)_n$
- RT – linker – Cas9 – linker – $(RBP)_n$
- $(RBP)_n$ – linker – Cas9 – linker – RT
- $(RBP)_n$ – linker – RT – linker – Cas9
- Cas9 – linker $(RBP)_n$ – linker – RT
- RT – linker – $(RBP)_n$ – linker – Cas9
- Cas9 – linker – RT

2) Guide RNA:

The guide RNA (gRNA) molecule (5'-GCCGAAGCACTGCACGCCGTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC -3'; SEQ ID NO: 16,701) binds Cas9 and contains a pro-spacer sequence complementary to the target locus, resulting in localization of the gene modifying complex to the target genomic locus.

3) Trans-template RNA:

An exemplary trans-template RNA (ttRNA) molecule contains a **PBS-template** region, 3' to 5', as below (and as diagrammed in FIG. 3A-3B):

1. a 1-17 nt primer binding site (PBS; as in Table 37) that basepairs with the nicked DNA strand allowing primer extension of the nicked DNA, followed immediately by
2. a 0-20 nt region of homology with the target locus (homology 1; as in Table 38, "homology 1")
3. the desired modification to the genome – insertion, deletion, or substitution (edit; as in Table 38, "edit")

4. a 0-500 nt region of homology with the target locus (homology 2; as in Table 39, "homology 2").

The exemplary ttRNA additionally includes:

5. an RBP recruitment site (**RRS**; as in Table 40). The RRS may be repeated 1-5 times, e.g., 1-4 times, in tandem with or without intervening RNA linker sequences.

The exemplary ttRNA may optionally also include:

6. an end-protecting RNA secondary structure (**end block**; as in Table 41) that protects the ttRNA molecule from exonuclease-mediated degradation.

These components of the ttRNA may be assembled in any of the below configurations, where the RRSs may be joined to the PBS-template by an RNA linker sequence of up to 20 nts (the longest sequence being 5'-ACTAACATACAACTAACATA-3'):

- $(RRS)_n$ – linker – PBS-template
- $(RRS)_n$ – linker – PBS-template – end block
- PBS-template – linker – $(RRS)_n$
- end-block – PBS-template – linker – $(RRS)_n$
- end block - $(RRS)_n$ – linker – PBS-template
- end block - $(RRS)_n$ – linker – PBS-template – end block
- PBS-template – linker – $(RRS)_n$ – end block
- end-block – PBS-template – linker – $(RRS)_n$ – end block

Evaluation of 3-component trans-gene modifying system configurations:

Method 1:

To determine the genome-editing capacity of the various configurations of the trans-gene modifying complex, mammalian HEK293T or U2OS cell lines are generated carrying a genomic landing pad that expresses one of the following (as in Table 42):

i. BFP-GFP SNP reporter line: a BFP-expressing cell line, which is converted to a GFP-expressing cell line via a single base pair substitution;

ii. 250 bp GFP insertion reporter line: a non-fluorescent partially-deleted-GFP cell line, in which GFP-expression is restored via a 250 bp insertion of the deleted coding sequence; or

iii. mCherry insertion reporter line: a GFP-expressing cell line, which loses GFP- expression and gains mCherry expression via the interruption of the GFP coding sequence with a >750 bp insertion that introduces the mCherry coding sequence.

The gene modifying polypeptide is introduced into these cells by transfection of DNA plasmid or mRNA, or by packaging of DNA plasmid into Lentivirus. Cells containing the gene modifying polypeptide are then transfected/electroporated with a ttRNA-expressing DNA plasmid or ncRNA, and a gRNA-expressing DNA plasmid or ncRNA that are together designed to:

i. BFP-GFP SNP reporter line: convert BFP-to-GFP;

ii. 250 bp GFP insertion reporter line: restore the GFP coding sequence; or

iii. mCherry insertion reporter line: interrupt the GFP coding sequence with an mCherry coding sequence.

To assess the genome-editing capacity of the trans-gene modifying system configurations, cells are analyzed by flow cytometry 4-10 days post-transfection/electroporation of ttRNA and gRNA. The fidelity of the edits are also assessed by collecting genomic DNA 3-10 days post-transfection/ electroporation of the ttRNA and gRNA. The frequency of the intended versus unintended mutations at target loci are analyzed by amplicon sequencing.

Method 2:

Evaluation of 3-component trans-gene modifying system configurations:

To determine the genome-editing capacity of the various configurations of the trans-gene modifying complex, mammalian HEK293T or U2OS cell lines are generated carrying a genomic landing pad that expresses one of the following (as in Table 42):

i. BFP-GFP SNP reporter line: a BFP-expressing cell line, which is converted to a GFP-expressing cell line via a single base pair substitution;

ii. 150 or 250 bp GFP insertion reporter line: a non-fluorescent partially-deleted-GFP cell line, in which GFP-expression is restored via a 150 or 250 bp insertion of the deleted coding sequence; or

iii. mKate2 or mCherry insertion reporter line: a GFP-expressing cell line, which loses GFP-expression and gains mKate2 or mCherry expression via the interruption of the GFP coding sequence with a >700 (e.g., >700 or >750) bp insertion that introduces the mKate2 or mCherry coding sequence.

The gene modifying polypeptide is introduced into these cells by transfection of DNA plasmid or mRNA, or by packaging of DNA plasmid into Lentivirus. The cells are additionally co-transfected/electroporated with a ttRNA-expressing DNA plasmid or ncRNA, and a gRNA-expressing DNA plasmid or ncRNA that are together designed to:

i. BFP-GFP SNP reporter line: convert BFP-to-GFP;

ii. 150 or 250 bp GFP insertion reporter line: restore the GFP coding sequence; or

iii. mKate2 or mCherry insertion reporter line: interrupt the GFP coding sequence with an mKate2 or mCherry coding sequence.

To assess the genome-editing capacity of the trans-gene modifying system configurations, cells are analyzed by flow cytometry 4-10 days post-transfection/electroporation of ttRNA and gRNA. The fidelity of the edits are also assessed by collecting genomic DNA 3-10 days post-transfection/ electroporation of the ttRNA and gRNA. The frequency of the intended versus unintended mutations at target loci are analyzed by amplicon sequencing.

Example 2: Trans-recruitment of an RNA template for RNA-based gene modifying via engagement of the second DNA strand by a Cas9-complex

This example describes the use of a four to five (4-5) component RNA-based gene modifying system for the targeted editing of a sequence in the human genome (FIGS 4A-4B). More specifically, a (1) gene modifying polypeptide binds (2) a guide RNA molecule and (3) a trans-template RNA molecule, together forming a gene modifying complex, which binds and nicks the target locus and performs templated incorporation of the desired edit in one strand of the genomic DNA. An additional (4) Cas9 (e.g., a dCas9 or nCas9 domain (FIG. 4A) or such a domain as part of a second gene modifying polypeptide (FIG. 4B) molecule is recruited to the nicked DNA strand by the trans-template RNA (e.g., by an end block containing a 5' pro-spacer), or by (5) its species-matched guide RNA. The second Cas9-complex may extend the bubble and facilitate large insertions, and may also be used to introduce a nick

on the second strand to initiate second strand synthesis, and/or signal to the cell's endogenous repair system (e.g., mismatch repair system) that the edited strand should be copied and maintained. This example further describes the introduction of the 4-5 component system to mammalian cells for in vitro gene modifying as a means of evaluating the efficacy of various gene modifying system configurations and trans-template RNA recruitment interactions on editing activity in human cells.

1) Gene modifying polypeptide:

The gene modifying polypeptide includes:

1. a Cas-nuclease with one endonuclease domain inactivated (e.g. Spy N863A Cas9),
2. a reverse transcriptase (RT), and
3. an RNA binding protein (RBP; as in Table 31). The RBP may be repeated 1-5 times in tandem with or without intervening peptide linker sequences.

These 3 domains are brought together either by covalent linkage using (1) peptide linkers, as in Table 32, or (2) intein pairs, as in Table 33, or by the dimerization of two fusion domains (FDs) that (a) is induced by chemical binding or or specific wavelengths of light Table 34, or (b) occurs intrinsically in a ligand-free interaction Tables 35 and 36. The FDs may, in some instances, be repeated 1-30 times in tandem with or without intervening peptide linker sequences.

The 3 domains can be assembled in any of the following configurations, where the linkers indicated below can, in some instances, be replaced with intein- or FD-pairs.:

- Cas9 – linker – RT – linker – $(RBP)_n$
- RT – linker – Cas9 – linker – $(RBP)_n$
- $(RBP)_n$ – linker – Cas9 – linker – RT
- $(RBP)_n$ – linker – RT – linker – Cas9
- Cas9 – linker $(RBP)_n$ – linker – RT
- RT – linker – $(RBP)_n$ – linker – Cas9
- Cas9 – linker – RT

2) Guide RNA:

The guide RNA (gRNA) molecule (5'-GCCGAAGCACTGCACGCCGTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC -3'; SEQ ID NO: 16,701) binds Cas9 and contains a pro-spacer sequence complementary to the target locus, resulting in localization of the gene modifying complex to the target genomic locus.

3) Trans-template RNA:

A trans-template RNA (ttRNA) molecule contains a **PBS-template** region, 3' to 5', as below:

1. an 1-17 nt primer binding site (PBS; as in Table 37) that basepairs with the nicked DNA strand allowing primer extension of the nicked DNA, followed immediately by
2. a 0-20 nt region of homology with the target locus (homology 1; as in Table 38, "homology 1")
3. the desired modification to the genome – insertion, deletion, or substitution (edit; as in Table 38, "edit")
4. a 0-500 nt region of homology with the target locus (homology 2; as in Table 39, "homology 2").

The exemplary ttRNA additionally includes:

5. an RBP recruitment site (**RRS**; as in Table 40). The RRS may be repeated 1-5 times (e.g., 1-4 times) in tandem with or without intervening RNA linker sequences.

The exemplary ttRNA may optionally also include:

6. an end-protecting RNA secondary structure (**end block**; as in Table 41) that protects the ttRNA molecule from exonuclease-mediated degradation.

These components of the ttRNA may be assembled in any of the below configurations, where the RRSs may be joined to the PBS-template by an RNA linker sequence of up to 20 nts (the longest sequence being 5'-ACTAACATACAACTAACATA-3'; SEQ ID NO: 16,702):

- $(RRS)_n$ – linker – PBS-template
- $(RRS)_n$ – linker – PBS-template – end block

- PBS-template – linker – $(RRS)_n$
- end-block – PBS-template – linker – $(RRS)_n$
- end block - $(RRS)_n$ – linker – PBS-template
- end block - $(RRS)_n$ – linker – PBS-template – end block
- PBS-template – linker – $(RRS)_n$ – end block
- end-block – PBS-template – linker – $(RRS)_n$ – end block

4) Additional (second) Cas9:

An additional nickase (e.g., Spy N863A Cas9, or other orthogonal Cas9) or catalytically inactive ("dead") Cas9 as in Table 43, which binds the nicked DNA strand by pairing with a gRNA scaffold that is encoded in the ttRNA (FIG. 4B), or is introduced as a separate guide RNA (FIG. 4A; see also 5, below). A nickase Cas9 paired with a full spacer is used to introduce a nick on the second, unnicked strand. Alternatively, a dead Cas9 or a nickase paired with a gRNA-2 having a pro-spacer with less than 17 nt complementarity to the target opens up the DNA bubble for long RT reactions, without nicking the second strand. An additional nickase or catalytically inactive Cas9 may be a separate domain (e.g., FIG. 4A) or a part of an additional (second) gene modifying polypeptide (FIG. 4B).

5) Optional 2nd Cas9-paired guideRNA:

The second guide RNA (gRNA-2) molecule with a pro-spacer sequence complementary to the nicked strand, as in Table 44, may be added to recruit the 2nd Cas9 to the nicked DNA strand for DNA bubble extension or second strand nicking. A gRNA-2 pro-spacer may be fully complementary to the target strand to stimulate the nickase activity of the 2nd Cas9, or it may contain 17 or fewer nucleotides of complementarity to inhibit nicking by the 2nd Cas9.

Evaluation of the 4-5 component trans-gene modifying system configurations:

Method 1:

To determine the genome-editing capacity of the various configurations of the trans-gene modifying system complex, mammalian HEK293T or U2OS cell lines are generated carrying a genomic landing pad that expresses one of the following (as in Table 42):

i. BFP-GFP SNP reporter line: a BFP-expressing cell line, which is converted to a GFP-expressing cell line via a single base pair substitution;

ii. 250 bp GFP insertion reporter line: a non-fluorescent partially-deleted-GFP cell line, in which GFP-expression is restored via a 250 bp insertion of the deleted coding sequence; or

iii. mCherry insertion reporter line: a GFP-expressing cell line, which loses GFP- expression and gains mCherry expression via the interruption of the GFP coding sequence with a >750 bp insertion that introduces the mCherry coding sequence.

The gene modifying polypeptide is introduced in these cells by transfection of DNA plasmid or mRNA, or by packaging of DNA plasmid into Lentivirus. The cells are additionally co-transfected/electroporated with 1) a ttRNA-expressing DNA plasmid or ncRNA, 2) a gRNA-expressing DNA plasmid or ncRNA that are together designed to:

i. BFP-GFP SNP reporter line: convert BFP-to-GFP;

ii. 250 bp GFP insertion reporter line: restore a GFP coding sequence; or

iii. mCherry insertion reporter line: interrupt the GFP coding sequence with an mCherry coding sequence,

along with 3) a DNA plasmid or mRNA expressing a 2nd Cas9, and optionally 4) a gRNA-2-expressing DNA plasmid or ncRNA.

To assess the genome-editing capacity of the trans-gene modifying system configurations, cells are analyzed by flow cytometry 4-10 days post-transfection/electroporation of ttRNA and gRNA. The fidelity of the edits are also assessed by collecting genomic DNA 3-10 days post-transfection/ electroporation of the ttRNA and gRNA. The frequency of the intended versus unintended mutations at target loci are analyzed by amplicon sequencing.

Method 2:

To determine the genome-editing capacity of the various configurations of the trans-gene modifying system complex, mammalian HEK293T or U2OS cell lines are generated carrying a genomic landing pad that expresses one of the following (as in Table 42):

i. BFP-GFP SNP reporter line: a BFP-expressing cell line, which is converted to a GFP-expressing cell line via a single base pair substitution;

ii. 150 or 250 bp GFP insertion reporter line: a non-fluorescent partially-deleted-GFP cell line, in which GFP-expression is restored via a 150 or 250 bp insertion of the deleted coding sequence; or

iii. mKate2 or mCherry insertion reporter line: a GFP-expressing cell line, which loses GFP-expression and gains mKate2 or mCherry expression via the interruption of the GFP coding sequence with a >700 (e.g., >750) bp insertion that introduces the mKate2 or mCherry coding sequence.

The gene modifying polypeptide is introduced in these cells by transfection of DNA plasmid or mRNA, or by packaging of DNA plasmid into Lentivirus. The cells are additionally co-transfected/electroporated with 1) a ttRNA-expressing DNA plasmid or ncRNA, 2) a gRNA-expressing DNA plasmid or ncRNA that are together designed to:

iv. BFP-GFP SNP reporter line: convert BFP-to-GFP;

v. 150 or 250 bp GFP insertion reporter line: restore a GFP coding sequence; or

vi. mKate2 or mCherry insertion reporter line: interrupt the GFP coding sequence with an mKate2 or mCherry coding sequence,

along with 3) a DNA plasmid or mRNA expressing a 2nd Cas9, and optionally 4) a gRNA-2-expressing DNA plasmid or ncRNA.

To assess the genome-editing capacity of the trans-gene modifying system configurations, cells are analyzed by flow cytometry 4-10 days post-transfection/electroporation of ttRNA and gRNA. The fidelity of the edits are also assessed by collecting genomic DNA 3-10 days post-transfection/ electroporation of the ttRNA and gRNA. The frequency of the intended versus unintended mutations at target loci are analyzed by amplicon sequencing.

Example 3: Second Cas9-mediated recruitment of reverse transcriptase

This example describes the use of a four or five (4-5) component RNA gene modifying system for the targeted editing of a sequence in the human genome (FIG. 5A-5B). More specifically, a (1)-(2) split gene modifying polypeptide binds (3)-(4) one or two guide RNA molecules and (5) a trans-template RNA molecule, together forming a gene modifying complex, which binds and nicks the target locus and

performs templated incorporation of the desired edit in one strand of the genomic DNA. The first component of the split gene modifying polypeptide is constituted of a nickase Cas9 fused to an RNA binding protein, which binds to and recruits the trans-template RNA molecule. The second component is constituted of a second Cas9 fused to a reverse transcriptase domain, which carries out the templated incorporation of the desired edit in the nicked DNA strand. The Cas9-RT is recruited to the target locus either by the trans-template RNA, or by a separate guide RNA molecule. The presence of the second Cas9-complex may act to extend the bubble and facilitate large insertions, and may also be used to introduce a nick on the second strand to initiate second strand synthesis, and/or signal the cell's endogenous mismatch repair system that the edited strand should be copied and maintained. This example further describes the introduction of the 4-5 component system to mammalian cells for in vitro gene modifying as a means of evaluating the efficacy of various gene modifying system configurations and trans-template RNA recruitment interactions on editing activity in human cells.

1-2) Gene modifying polypeptides:

The gene modifying enzyme includes two polypeptide components – a Cas9-RNA binding protein (RBP) component, and a Cas9-reverse transcriptase (RT) component. The RBP may, in some instances, be repeated 1-5 times in tandem with or without intervening peptide linker sequences.

These domains within these two components are brought together either by covalent linkage using (1) peptide linkers, as in Table 32, or (2) intein pairs, as in Table 33, or by the dimerization of two fusion domains (FDs) that (a) is induced by chemical binding or or specific wavelengths of light Table 34, or (b) occurs intrinsically in a ligand-free interaction Tables 35 and 36. The FDs may, in some instances, be repeated 1-30 times in tandem with or without intervening peptide linker sequences.

The Cas9-RBP component may be assembled in either of the below configurations, where the linkers as indicated below can be replaced, in some instances, with intein- or FD-pairs.:

- Cas9 – linker – $(RBP)_n$
- $(RBP)_n$ – linker – Cas9

The second Cas9-RT component may be assembled in either of the below configurations, where the linkers as indicated below can be replaced, in some instances, with intein- or FD-pairs.:

- Cas9 – linker – RT
- RT – linker – Cas9

3-4) Guide RNA:

The first guide RNA (gRNA) molecule (5'-GCCGAAGCACTGCACGCCGTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTT ATCAACTTGAAAAAGTGGCACCGAGTCGGTGC -3'; SEQ ID NO: 16,703) binds Cas9 and contains a pro-spacer sequence complementary to the target locus, resulting in localization of the Cas9-RBP component to the target genomic locus.

An optional second guide RNA (gRNA-2) molecule may be added when the Cas9-RT is not recruited to the second strand by a guide RNA structure at the 5'-end of the trans-template RNA.

5) Trans-template RNA:

An exemplary trans-template RNA (ttRNA) molecule contains a **PBS-template** region, 3' to 5', as below:

1. an 1-17 nt primer binding site (PBS; as in Table 37) that basepairs with the nicked DNA strand allowing primer extension of the nicked DNA, followed immediately by
2. a 0-20 nt region of homology with the target locus (homology 1; as in Table 38, "homology 1")
3. the desired modification to the genome – insertion, deletion, or substitution (edit; as in Table 38, "edit")
4. a 0-500 nt region of homology with the target locus (homology 2; as in Table 39, "homology 2").

The exemplary ttRNA additionally includes:

5. an RBP recruitment site (**RRS**; as in Table 40). The RRS may be repeated 1-5 times in tandem with or without intervening RNA linker sequences.

The exemplary ttRNA may optionally also include:

6. an end-protecting RNA secondary structure (**end block**; as in Table 41) that protects the ttRNA molecule from exonuclease-mediated degradation.

These components of the ttRNA may be assembled in any of the below configurations, where the RRSs may be joined to the PBS-template by an RNA linker sequence of up to 20 nts (the longest sequence being 5'-ACTAACATACAACTAACATA-3'):

- $(RRS)_n$ – linker – PBS-template
- $(RRS)_n$ – linker – PBS-template – end block
- PBS-template – linker – $(RRS)_n$
- end-block – PBS-template – linker – $(RRS)_n$
- end block - $(RRS)_n$ – linker – PBS-template
- end block - $(RRS)_n$ – linker – PBS-template – end block
- PBS-template – linker – $(RRS)_n$ – end block
- end-block – PBS-template – linker – $(RRS)_n$ – end block

Evaluation of the 4-5 component trans-gene modifying system configurations:

To determine the genome-editing capacity of the various configurations of the trans-gene modifying complex, mammalian HEK293T or U2OS cell lines are generated carrying a genomic landing pad that expresses one of the following (as in Table 42):

i. BFP-GFP SNP reporter line: a BFP-expressing cell line, which is converted to a GFP-expressing cell line via a single base pair substitution;

ii. 250 bp GFP insertion reporter line: a non-fluorescent partially-deleted-GFP cell line, in which GFP-expression is restored via a 250 bp insertion of the deleted coding sequence; or

iii. mCherry insertion reporter line: a GFP-expressing cell line, which loses GFP- expression and gains mCherry expression via the interruption of the GFP coding sequence with a >750 bp insertion that introduces the mCherry coding sequence.

The gene modifying polypeptides are introduced in these cells by transfection of DNA plasmid or mRNA, or by packaging of DNA plasmid into Lentivirus. Cells containing the gene modifying polypeptide are then transfected/electroporated with 1) a ttRNA-expressing DNA plasmid or ncRNA, 2) a gRNA-expressing DNA plasmid or ncRNA, and optionally 3) a gRNA-2-expressing DNA plasmid or ncRNA that are together designed to:

vii. BFP-GFP SNP reporter line: convert BFP-to-GFP;

viii. 250 bp GFP insertion reporter line: restore a GFP coding sequence; or

ix. mCherry insertion reporter line: interrupt the GFP coding sequence with an mCherry coding sequence.

To assess the genome-editing capacity of the trans-gene modifying system configurations, cells are analyzed by flow cytometry 4-10 days post-transfection/electroporation of ttRNA and gRNA. The fidelity of the edits are also assessed by collecting genomic DNA 3-10 days post-transfection/ electroporation of the ttRNA and gRNA. The frequency of the intended versus unintended mutations at target loci are analyzed by amplicon sequencing.

Example 4: Quantifying activity of a gene editing polypeptide using a GFP/BFP assay in human cells

Method 1

This example describes the use of gene modifying system containing a gene modifying polypeptide and a template RNA. In this example, a template RNA contains:

(1) a gRNA spacer;
(2) a gRNA scaffold;
(3) a heterologous object sequence; and
(4) a primer binding site (PBS) sequence.

More specifically, the template RNA comprises the following sequence: GCCGAAGCACTGCACGCCGTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGT CCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCACCCTGACGTACGGCGTGCAG TGCTT (SEQ ID NO: 16,704).

A gene modifying system comprising a given gene modifying polypeptide (e.g., one described herein) and the template RNA is transfected into the HEK293T BFP-expressing cell line. The gene modifying polypeptide and the template RNA are delivered by nucleofection in DNA format. Specifically, 800 ng of gene modifying polypeptide plasmid DNA is combined with 200 ng template RNA in plasmid format. The modifying polypeptide and template RNA in plasmid DNA format are added to 25μL SF buffer containing 250,000 HEK293T BFP-expressing cells, and cells are nucleofected using program DS-150. After nucleofection, cells are grown at 37°C, 5% CO_2 for 3 days prior to cell lysis and genomic DNA extraction. To analyze gene editing activity, primers flanking the BFP locus can be

used to amplify across the locus. Amplicons are analyzed via short read sequencing using an Illumina MiSeq. Conversion of the BFP gene sequence to the GFP gene sequence indicate successful editing. In some embodiments, the assay will indicate that at least 10%, 20%, 30%, 40%, 50%, 60%, or 70% of copies of the BFP gene in the sample are converted to the GFP gene.

Method 2

This example describes the use of gene modifying system containing a gene modifying polypeptide and a template RNA. In this example, a template RNA contains:

(1) a gRNA spacer;
(2) a gRNA scaffold;
(3) a heterologous object sequence; and
(4) a primer binding site (PBS) sequence.

More specifically, the template RNA comprises the following sequence: GCCGAAGCACTGCACGCCGTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGT CCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCACCCTGACGTACGGCGTGCAG TGCTT (SEQ ID NO: 16,705).

A gene modifying system comprising a given gene modifying polypeptide (e.g., one described herein) the gRNA, and a template RNA are transfected/electroporated into the HEK293T/U2OS BFP-expressing cell line. The gene modifying polypeptide and the template RNA are delivered by nucleofection in DNA format. Specifically, 300-800 ng of gene modifying polypeptide plasmid DNA is combined with 50-200 ng template RNA and 50-200 ng gRNA in plasmid format. The modifying polypeptide, template RNA and gRNA in plasmid DNA format are added to 25µL SF/SE buffer containing 250,000 HEK293T/U2OS BFP-expressing cells, and cells are nucleofected using program DS-150/DN-100. After nucleofection, cells are grown at 37°C, 5% CO_2 for at least 3 days prior to cell lysis and genomic DNA extraction. To analyze gene editing activity, primers flanking the BFP locus can be used to amplify across the locus. Amplicons are analyzed via short read sequencing using an Illumina MiSeq. Conversion of the BFP gene sequence to the GFP gene sequence indicate successful editing. In some embodiments, the assay will indicate that at least 10%, 20%, 30%, 40%, 50%, 60%, or 70% of copies of the BFP gene in the sample are converted to the GFP gene.

Example 5: Rewriter polypeptide selection by pooled screening in HEK293T & U2OS cells

This example describes the use of an RNA gene modifying system for the targeted editing of a coding sequence in the human genome. More specifically, this example describes the infection of HEK293T and U2OS cells with a library of gene modifying candidates, followed by transfection of a template guide

RNA (tgRNA) for *in vitro* gene modifying in the cells, e.g., as a means of evaluating a new gene modifying polypeptide for editing activity in human cells by a pooled screening approach.

The gene modifying polypeptide library candidates assayed herein each comprise: 1) a *S. pyogenes* (Spy) Cas9 nickase containing an N863A mutation that inactivates one endonuclease active site; 2) one of the 122 peptide linkers depicted at **Table 10**; and 3) a reverse transcriptase (RT) domain from **Table 6** of retroviral origin. The particular retroviral RT domains utilized were selected if they were expected to function as a monomer. For each selected RT domain, the wild-type sequences were tested, as well as versions with point mutations installed in the primary wild-type sequence. In particular, 143 RT domains were tested, either wild type or containing various mutations. In total, 17,446 Cas-linker-RT gene modifying polypeptides were tested.

The system described here is a two-component system comprising: 1) an expression plasmid encoding a human codon-optimized gene modifying polypeptide library candidate within a lentiviral cassette, and 2) a tgRNA expression plasmid expressing a non-coding tgRNA sequence that is recognized by Cas and localizes it to the genomic locus of interest, and that also templates reverse transcription of the desired edit into the genome by the RT domain, driven by a U6 promoter. The lentiviral cassette comprises: (i) a CMV promoter for expression in mammalian cells; (ii) a gene modifying polypeptide library candidate as shown; (iii) a self-cleaving T2A polypeptide; (iv) a puromycin resistance gene enabling selection in mammalian cells; and (v) a polyA tail termination signal.

To prepare a pool of cells expressing gene modifying polypeptide library candidates, HEK293T or U2OS cells were transduced with pooled lentiviral preparations of the gene modifying candidate plasmid library. HEK293 Lenti-X cells were seeded in 15 cm plates ($12x10^6$ cells) prior to lentiviral plasmid transfection. Lentiviral plasmid transfection using the Lentiviral Packaging Mix (Biosettia, 27 ug) and the plasmid DNA for the gene modifying candidate library (27 ug) was performed the following day using Lipofectamine 2000 and Opti-MEM media according to the manufacturer's protocol. Extracellular DNA was removed by a full media change the next day and virus-containing media was harvested 48 hours after. Lentiviral media was concentrated using Lenti-X Concentrator (TaKaRa Biosciences) and 5 mL lentiviral aliquots were made and stored at -80°C. Lentiviral titering was performed by enumerating colony forming units post Puromycin selection. HEK293T or U2OS cells carrying a BFP-expressing genomic landing pad were seeded at $6x10^7$ cells in culture plates and transduced at a 0.3 multiplicity of infection (MOI) to minimize multiple infections per cell. Puromycin (2.5 ug/mL) was added 48 hours post infection to allow for selection of infected cells. Cells were kept under puromycin selection for at least 7 days and then scaled up for tgRNA electroporation.

To determine the genome-editing capacity of the gene modifying library candidates in the assay, infected BFP-expressing HEK293T or U2OS cells were then transfected by electroporation of 250,000 cells/well with 200 ng of a tgRNA (either g4 or g10) plasmid, designed to convert BFP to GFP, at sufficient cell count for >1000x coverage per library candidate.

The g4 tgRNA (5′ to 3′) is as follows: 20 nucleotide spacer region (GCCGAAGCACTGCACGCCGT), a scaffold region (GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG GCACCGAGTCGGTGC), the template region encoding the single base pair substitution to change BFP to GFP (**bold**) and a PAM inactivation that introduces a synonymous point mutation in the SpyCas9 PAM (NGG to NCG) that prevents re-engagement of the gene modifying polypeptide upon completion of a functional gene modifying reaction (underline) (ACCCTGAC**G**TACG), and the 13 nucleotide PBS (GCGTGCAGTGCTT).

Similarly, the g10 tgRNA (5′ to 3′) is as follows: 20 nucleotide spacer region (AGAAGTCGTGCTGCTTCATG), a scaffold region (GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTG GCACCGAGTCGGTGC), the template region encoding the single base pair substitution to change BFP to GFP (**bold**) and a PAM inactivation that introduces a synonymous point mutation in the SpyCas9 PAM (NGG to NGA) that prevents re-engagement of the gene modifying polypeptide upon completion of a functional gene modifying reaction (underline) (ACCCTGACC**T**ACGGCGTGCAGTGCTTCGGCCGCTACCCCGAT**C**ACAT), and 13 nucleotide PBS (GAAGCAGCACGAC).

.

To assess the genome-editing capacity of the various constructs in the assay, cells were sorted by Fluorescence-Activated Cell Sorting (FACS) for GFP expression 6-7 days post-electroporation. Cells were sorted and harvested as distinct populations of unedited (BFP+) cells, edited (GFP+) cells and imperfect edit (BFP-, GFP-) cells. A sample of unsorted cells was also harvested as the input population to determine enrichment during analysis.

To determine which gene modifying library candidates have genome-editing capacity in this assay, genomic DNA (gDNA) was harvested from sorted and unsorted cell populations, and analyzed by sequencing the gene modifying library candidates in each population. Briefly, gene modifying sequences were amplified from the genome using primers specific to the lentiviral cassette, amplified in a second

round of PCR to dilute genomic DNA, and then sequenced using Oxford Nanopore Sequencing Technology according to the manufacturer's protocol.

After quality control of sequencing reads, reads of at least 1500 and no more than 3200 nucleotides were mapped to the gene modifying polypeptide library sequences and those containing a minimum of an 80% match to a library sequence were considered to be successfully aligned to a given candidate. To identify gene modifying candidates capable of performing gene editing in the assay, the read count of each library candidate in the edited population was compared to its read count in the initial, unsorted population. For purposes of this pooled screen, gene modifying candidates with genome-editing capacity were selected as those candidates that were enriched in the converted (GFP+) population relative to unsorted (input) cells and wherein the enrichment was determined to be at or above the enrichment level of a reference (Element ID No: 17380)..

A large number of gene modifying polypeptide candidates were determined to be enriched in the GFP+ cell populations. For example, of the 17,446 candidates tested, over 3,300 exhibited enrichment in GFP+ sorted populations (relative to unsorted) that was at least equivalent to that of the reference under similar experimental conditions (HEK293T using g4 tgRNA; HEK293T cells using g10 tgRNA; or U2OS cells using g4 tgRNA), shown in Table 1. Although the 17,446 candidates were also tested in U2OS cells using g10 tgRNA, the pooled screen did not yield candidates that were enriched in the converted (GFP+) population relative to unsorted (input) cells under that experimental condition; further investigation is required to explain these results.

Table 1. Combinations of linker and RT sequences ordered by ranked performance in screen. The amino acid sequence of each RT in this table is provided in Table 6.

Linker amino acid sequence	Linker SEQ ID NO:	RT domain name
EAAAKGSS	16,901	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAKEAAAK	16,902	MLVMS_P03355_PLV919
PAPEAAAK	16,903	MLVFF_P26809_3mutA
EAAAKPAPGGG	16,904	MLVFF_P26809_3mutA
GSSGSSGSSGSSGSSGSS	16,905	PERV_Q4VFZ2_3mut
PAPGGGEAAAK	16,906	MLVAV_P03356_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	16,907	MLVMS_P03355_PLV919
GSSEAAAK	16,908	MLVFF_P26809_3mutA
EAAAKPAPGGS	16,909	MLVFF_P26809_3mutA
GGSGGSGGSGGSGGSGGS	16,910	MLVFF_P26809_3mutA

AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	16,911	XMRV6_A1Z651_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	16,912	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAK	16,913	MLVFF_P26809_3mutA
PAPEAAAKGSS	16,914	MLVFF_P26809_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	16,915	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAK	16,916	PERV_Q4VFZ2_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	16,917	AVIRE_P03360_3mutA
PAPAPAPAPAP	16,918	MLVCB_P08361_3mutA
PAPAPAPAPAP	16,919	MLVFF_P26809_3mutA
EAAAKGGSPAP	16,920	PERV_Q4VFZ2_3mutA_WS
PAP	16,921	MLVMS_P03355_PLV919
PAPGGGGSS	16,922	WMSV_P03359_3mutA
SGSETPGTSESATPES	16,923	MLVFF_P26809_3mutA
PAPEAAAKGSS	16,924	XMRV6_A1Z651_3mutA
EAAAKGGSGGG	16,925	MLVMS_P03355_PLV919
GGGGSGGGGS	16,926	MLVFF_P26809_3mutA
GGGPAPGSS	16,927	MLVAV_P03356_3mutA
GGSGGSGGSGGSGGSGGS	16,928	XMRV6_A1Z651_3mut
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	16,929	MLVCB_P08361_3mutA
GSSPAP	16,930	AVIRE_P03360_3mutA
EAAAKGSSPAP	16,931	MLVFF_P26809_3mutA
GSSGGGEAAAK	16,932	MLVFF_P26809_3mutA
GGSGGSGGSGGSGGSGGS	16,933	MLVMS_P03355_3mutA_WS
PAPAPAPAP	16,934	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAK	16,935	XMRV6_A1Z651_3mutA
EAAAKGGSPAP	16,936	MLVMS_P03355_3mutA_WS
PAPGGSEAAAK	16,937	AVIRE_P03360_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	16,938	AVIRE_P03360_3mutA
EAAAKGGGGSEAAAK	16,939	MLVCB_P08361_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	16,940	WMSV_P03359_3mutA
GSS	16,941	MLVMS_P03355_PLV919
GSSGSSGSSGSS	16,942	MLVMS_P03355_PLV919
GSSPAPEAAAK	16,943	XMRV6_A1Z651_3mutA
GGSPAPEAAAK	16,944	MLVFF_P26809_3mutA
GGGEAAAKGGS	16,945	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	16,946	PERV_Q4VFZ2_3mutA_WS
GGGGGGGG	16,947	PERV_Q4VFZ2_3mut

GGGPAP	16,948	MLVCB_P08361_3mutA
PAPAPAPAPAPAP	16,949	MLVCB_P08361_3mutA
GGSGGSGGSGGSGGSGGS	16,950	MLVCB_P08361_3mutA
PAP	16,951	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGSGGSGGS	16,952	PERV_Q4VFZ2_3mutA_WS
PAPAPAPAPAPAP	16,953	MLVMS_P03355_PLV919
EAAAKPAPGSS	16,954	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAK	16,955	MLVMS_P03355_3mutA_WS
EAAAKGGS	16,956	MLVMS_P03355_3mutA_WS
GGGGSEAAAKGGGGS	16,957	MLVFF_P26809_3mutA
EAAAKPAPGSS	16,958	MLVFF_P26809_3mutA
GGGGSGGGGSGGGGSGGGGS	16,959	MLVMS_P03355_PLV919
EAAAKGGGGGS	16,960	MLVMS_P03355_PLV919
GGSPAP	16,961	XMRV6_A1Z651_3mutA
EAAAKGGGPAP	16,962	MLVMS_P03355_PLV919
EAAAKEAAAKEAAAKEAAAKEAAAK	16,963	MLVFF_P26809_3mutA
PAP	16,964	MLVCB_P08361_3mutA
EAAAK	16,965	XMRV6_A1Z651_3mutA
GGSGSSPAP	16,966	PERV_Q4VFZ2_3mutA_WS
GSSGSSGSSGSSGSSGSS	16,967	MLVMS_P03355_PLV919
GSSEAAAKGGG	16,968	MLVAV_P03356_3mutA
GGGEAAAKGGS	16,969	XMRV6_A1Z651_3mutA
EAAAKGGGGSEAAAK	16,970	MLVAV_P03356_3mutA
GGGGSGGGGSGGGGS	16,971	MLVFF_P26809_3mutA
GGGGSGGGGSGGGGSGGGGS	16,972	AVIRE_P03360_3mutA
SGSETPGTSESATPES	16,973	AVIRE_P03360_3mutA
GGGEAAAKPAP	16,974	MLVFF_P26809_3mutA
EAAAKGSSGGG	16,975	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAK	16,976	WMSV_P03359_3mut
GGSGGSGGSGGS	16,977	XMRV6_A1Z651_3mutA
GGSEAAAKPAP	16,978	MLVFF_P26809_3mutA
EAAAKGSSGGG	16,979	XMRV6_A1Z651_3mutA
GGGGS	16,980	MLVFF_P26809_3mutA
GGGEAAAKGSS	16,981	MLVMS_P03355_PLV919
PAPAPAPAPAPAP	16,982	MLVAV_P03356_3mutA
GGGGSGGGGSGGGGSGGGGS	16,983	MLVCB_P08361_3mutA
GGGEAAAKGSS	16,984	MLVCB_P08361_3mutA
PAPGGSGSS	16,985	MLVFF_P26809_3mutA
GSAGSAAGSGEF	16,986	MLVCB_P08361_3mutA
PAPGGSEAAAK	16,987	MLVMS_P03355_3mutA_WS
GGSGSS	16,988	XMRV6_A1Z651_3mutA

PAPGGGGSS	16,989	MLVMS_P03355_PLV919
GSSGSSGSS	16,990	XMRV6_A1Z651_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	16,991	MLVMS_P03355_3mutA_WS
EAAAK	16,992	MLVMS_P03355_PLV919
GSSGSSGSSGSS	16,993	MLVFF_P26809_3mutA
PAPGGGGSS	16,994	MLVCB_P08361_3mutA
GGGEAAAKGGS	16,995	MLVCB_P08361_3mutA
PAPGGGEAAAK	16,996	MLVMS_P03355_PLV919
GGGGGSPAP	16,997	XMRV6_A1Z651_3mutA
EAAAKGGS	16,998	XMRV6_A1Z651_3mutA
EAAAKGSSPAP	16,999	XMRV6_A1Z651_3mut
PAPEAAAK	17,000	MLVAV_P03356_3mutA
GGSGGSGGSGGS	17,001	MLVMS_P03355_3mutA_WS
GGGPAPGGS	17,002	MLVMS_P03355_PLV919
GSSGSSGSSGSS	17,003	PERV_Q4VFZ2_3mutA_WS
EAAAKPAPGGS	17,004	MLVCB_P08361_3mutA
GSSGSS	17,005	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAK	17,006	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAK	17,007	FLV_P10273_3mutA
GSS	17,008	MLVFF_P26809_3mutA
EAAAKEAAAK	17,009	MLVMS_P03355_3mutA_WS
PAPEAAAKGGG	17,010	MLVAV_P03356_3mutA
GGSGSSEAAAK	17,011	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	17,012	PERV_Q4VFZ2
GSSEAAAKPAP	17,013	AVIRE_P03360_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	17,014	MLVCB_P08361_3mutA
EAAAKGGG	17,015	MLVFF_P26809_3mutA
GSSPAPGGG	17,016	MLVCB_P08361_3mutA
GGGPAPGSS	17,017	MLVMS_P03355_PLV919
GGGGGS	17,018	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,019	PERV_Q4VFZ2_3mut
GGGGSGGGGSGGGGSGGGGSGGGGS	17,020	WMSV_P03359_3mutA
EAAAKEAAAKEAAAK	17,021	PERV_Q4VFZ2_3mut
PAPAPAPAP	17,022	MLVCB_P08361_3mutA
GSSGSSGSSGSSGSS	17,023	PERV_Q4VFZ2_3mut
GGGGSSEAAAK	17,024	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGS	17,025	MLVCB_P08361_3mutA
PAPEAAAKGGS	17,026	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,027	MLVCB_P08361_3mutA
EAAAKGGGGSEAAAK	17,028	MLVMS_P03355_PLV919
EAAAKGGGGSEAAAK	17,029	MLVMS_P03355_3mutA_WS

EAAAKGGGPAP	17,030	XMRV6_A1Z651_3mut
EAAAKEAAAKEAAAKEAAAKEAAAK	17,031	MLVMS_P03355_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,032	FLV_P10273_3mutA
GGSEAAAKGGG	17,033	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,034	KORV_Q9TTC1-Pro_3mutA
GGGPAPGGS	17,035	MLVCB_P08361_3mutA
PAPAPAPAPAPAP	17,036	XMRV6_A1Z651_3mutA
GGSGSSGGG	17,037	XMRV6_A1Z651_3mutA
GGSGSSGGG	17,038	MLVCB_P08361_3mutA
GGGEAAAKGGS	17,039	MLVMS_P03355_3mutA_WS
EAAAK	17,040	MLVCB_P08361_3mutA
GGSPAPGSS	17,041	MLVMS_P03355_3mutA_WS
GGGGSSEAAAK	17,042	PERV_Q4VFZ2_3mut
PAPAPAPAPAP	17,043	MLVBM_Q7SVK7_3mut
EAAAKEAAAKEAAAKEAAAK	17,044	MLVAV_P03356_3mutA
GGGGGSGSS	17,045	MLVCB_P08361_3mutA
EAAAKGSSPAP	17,046	MLVMS_P03355_3mutA_WS
PAPAPAPAPAPAP	17,047	MLVMS_P03355_3mutA_WS
GSSGGGGGS	17,048	MLVMS_P03355_3mutA_WS
PAPGSSGGG	17,049	MLVMS_P03355_PLV919
GGSGGGPAP	17,050	MLVCB_P08361_3mutA
GGGGGGG	17,051	MLVCB_P08361_3mutA
GSSGSSGSSGSSGSSGSS	17,052	MLVCB_P08361_3mutA
GGGPAPGGS	17,053	MLVFF_P26809_3mutA
EAAAKGGSGGG	17,054	PERV_Q4VFZ2_3mut
EAAAKGGGGSS	17,055	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSSGSSGSS	17,056	MLVMS_P03355_3mut
GGGGSGGGGSGGGGSGGGGS	17,057	MLVBM_Q7SVK7_3mutA_WS
PAPAPAPAPAP	17,058	MLVMS_P03355_PLV919
GGGEAAAKGGS	17,059	MLVMS_P03355_PLV919
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,060	MLVMS_P03355_3mut
GSAGSAAGSGEF	17,061	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSSGSS	17,062	MLVFF_P26809_3mutA
EAAAKGGSGSS	17,063	MLVFF_P26809_3mutA
PAPGGG	17,064	MLVFF_P26809_3mutA
GGGPAPGSS	17,065	XMRV6_A1Z651_3mutA
PAPEAAAKGGS	17,066	AVIRE_P03360_3mutA
PAPGGGEAAAK	17,067	MLVFF_P26809_3mut
GGGGSSEAAAK	17,068	MLVCB_P08361_3mutA

EAAAK	17,069	MLVMS_P03355_PLV919
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,070	BAEVM_P10272_3mutA
GGSGGGEAAAK	17,071	MLVMS_P03355_PLV919
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,072	MLVFF_P26809_3mutA
GSSPAPGGS	17,073	XMRV6_A1Z651_3mutA
GGSGGGPAP	17,074	MLVMS_P03355_PLV919
EAAAK	17,075	AVIRE_P03360_3mutA
GSS	17,076	XMRV6_A1Z651_3mutA
GGSGGSGGS	17,077	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAK	17,078	AVIRE_P03360_3mut
PAPEAAAKGGG	17,079	PERV_Q4VFZ2_3mutA_WS
GGGGGSEAAAK	17,080	BAEVM_P10272_3mutA
GGSGSSGGG	17,081	MLVMS_P03355_3mutA_WS
GGGGGGG	17,082	MLVMS_P03355_3mutA_WS
GSSEAAAKPAP	17,083	PERV_Q4VFZ2_3mut
GGGGGSEAAAK	17,084	WMSV_P03359_3mut
GGGGSGGGGSGGGGSGGGGSGGGGS	17,085	MLVFF_P26809_3mut
GGGEAAAKGGS	17,086	AVIRE_P03360_3mutA
GGSPAPGGG	17,087	AVIRE_P03360_3mutA
GSAGSAAGSGEF	17,088	MLVAV_P03356_3mutA
EAAAK	17,089	MLVAV_P03356_3mutA
EAAAKPAPGSS	17,090	WMSV_P03359_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,091	PERV_Q4VFZ2_3mutA_WS
GGSEAAAKPAP	17,092	MLVCB_P08361_3mutA
PAPAPAPAPAPAP	17,093	MLVBM_Q7SVK7_3mutA_WS
GGSPAPGGG	17,094	MLVMS_P03355_3mutA_WS
GGSEAAAKGGG	17,095	MLVMS_P03355_3mut
GGSGGSGGSGGS	17,096	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,097	MLVFF_P26809_3mutA
GGG	17,098	AVIRE_P03360_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,099	PERV_Q4VFZ2_3mut
GGSGGSGGSGGS	17,100	MLVMS_P03355_3mutA_WS
GGGEAAAK	17,101	MLVCB_P08361_3mutA
GSSGSSGSSGSSGSSGSS	17,102	MLVMS_P03355_3mutA_WS
GSSGGGPAP	17,103	MLVMS_P03355_3mutA_WS
GSSEAAAKPAP	17,104	MLVFF_P26809_3mutA
EAAAKEAAAK	17,105	MLVMS_P03355_PLV919
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,106	MLVCB_P08361_3mut
GGGGGG	17,107	MLVMS_P03355_3mutA_WS

GGSGSSGGG	17,108	MLVFF_P26809_3mutA
GSSGGGEAAAK	17,109	PERV_Q4VFZ2_3mutA_WS
PAPAPAPAPAP	17,110	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,111	SFV3L_P27401_2mut
EAAAKGGSGGG	17,112	BAEVM_P10272_3mutA
GGGGSSPAP	17,113	PERV_Q4VFZ2_3mutA_WS
GGGEAAAKPAP	17,114	MLVMS_P03355_PLV919
GGSGGGPAP	17,115	BAEVM_P10272_3mutA
PAPGSSGGS	17,116	MLVMS_P03355_PLV919
GGSGGGPAP	17,117	MLVMS_P03355_3mutA_WS
EAAAKGGSPAP	17,118	PERV_Q4VFZ2_3mutA_WS
EAAAKGGSGGG	17,119	MLVMS_P03355_3mutA_WS
PAPGSSGGG	17,120	MLVFF_P26809_3mutA
GSSEAAAKGGS	17,121	MLVFF_P26809_3mutA
PAPGSSEAAAK	17,122	MLVFF_P26809_3mutA
EAAAKGSSPAP	17,123	KORV_Q9TTC1-Pro_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	17,124	MLVBM_Q7SVK7_3mutA_WS
PAPGSSEAAAK	17,125	MLVMS_P03355_PLV919
EAAAKGSSGGG	17,126	MLVMS_P03355_3mutA_WS
EAAAKGGGGGS	17,127	AVIRE_P03360_3mutA
EAAAKEAAAKEAAAK	17,128	MLVMS_P03355_PLV919
PAPAPAPAPAPAP	17,129	MLVFF_P26809_3mutA
GGGGSGGGGSGGGGS	17,130	MLVCB_P08361_3mutA
PAPGGSEAAAK	17,131	MLVCB_P08361_3mutA
PAPGSSEAAAK	17,132	MLVBM_Q7SVK7_3mutA_WS
PAPEAAAKGSS	17,133	AVIRE_P03360_3mutA
GGSPAPGSS	17,134	WMSV_P03359_3mutA
PAPGGSGGG	17,135	MLVMS_P03355_PLV919
EAAAKGGSGSS	17,136	MLVMS_P03355_3mutA_WS
GGSGGG	17,137	MLVFF_P26809_3mutA
GGSEAAAKGSS	17,138	KORV_Q9TTC1_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,139	MLVCB_P08361_3mutA
PAPAPAPAPAPAP	17,140	PERV_Q4VFZ2_3mutA_WS
PAPEAAAK	17,141	MLVMS_P03355_3mutA_WS
GGSEAAAKGGG	17,142	MLVMS_P03355_PLV919
GSSPAP	17,143	MLVMS_P03355_3mutA_WS
GGGGSS	17,144	MLVMS_P03355_PLV919
GGGEAAAKPAP	17,145	AVIRE_P03360_3mutA
EAAAKPAPGGS	17,146	MLVAV_P03356_3mutA

EAAAKGGGPAP	17,147	MLVAV_P03356_3mutA
PAPGGSEAAAK	17,148	BAEVM_P10272_3mutA
PAPGGSGSS	17,149	MLVMS_P03355_3mutA_WS
PAPGGSGSS	17,150	AVIRE_P03360_3mutA
GGSGGGPAP	17,151	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAK	17,152	BAEVM_P10272_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGS	17,153	MLVMS_P03355_PLV919
GGGGSSPAP	17,154	MLVCB_P08361_3mutA
GSSGGGPAP	17,155	MLVFF_P26809_3mutA
GGGGSSGGS	17,156	MLVMS_P03355_PLV919
GGSGGG	17,157	MLVCB_P08361_3mutA
GSSGGGGGS	17,158	MLVMS_P03355_PLV919
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,159	XMRV6_A1Z651_3mutA
GGGGGSGSS	17,160	KORV_Q9TTC1_3mut
GGGEAAAKGGS	17,161	BAEVM_P10272_3mutA
GGSGGG	17,162	BAEVM_P10272_3mutA
PAPAPAP	17,163	KORV_Q9TTC1-Pro_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,164	SFV3L_P27401_2mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,165	MLVBM_Q7SVK7_3mutA_WS
GSSGSSGSSGSSGSS	17,166	MLVMS_P03355_3mutA_WS
GSSGGGEAAAK	17,167	MLVMS_P03355_3mutA_WS
GSSGGSEAAAK	17,168	MLVFF_P26809_3mutA
PAP	17,169	MLVMS_P03355_PLV919
EAAAKGGGGSEAAAK	17,170	MLVBM_Q7SVK7_3mutA_WS
PAPAP	17,171	AVIRE_P03360_3mutA
PAP	17,172	MLVFF_P26809_3mutA
GSSGGG	17,173	MLVMS_P03355_3mut
GSSPAPGGS	17,174	MLVFF_P26809_3mutA
PAPAPAPAP	17,175	XMRV6_A1Z651_3mutA
EAAAKGSSGGS	17,176	PERV_Q4VFZ2_3mut
PAPEAAAKGGG	17,177	KORV_Q9TTC1-Pro_3mutA
PAPGGS	17,178	MLVCB_P08361_3mutA
EAAAKGGG	17,179	MLVCB_P08361_3mutA
GSSEAAAKPAP	17,180	MLVMS_P03355_PLV919
PAPGGS	17,181	MLVFF_P26809_3mutA
EAAAKGGS	17,182	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	17,183	FLV_P10273_3mutA
PAPGGSEAAAK	17,184	MLVAV_P03356_3mutA
GSS	17,185	MLVCB_P08361_3mutA

GSSGSSGSSGSS	17,186	AVIRE_P03360_3mutA
GSSGSSGSS	17,187	MLVFF_P26809_3mutA
GSSGGG	17,188	MLVMS_P03355_PLV919
EAAAK	17,189	MLVFF_P26809_3mutA
GGSPAPEAAAK	17,190	MLVCB_P08361_3mutA
GGSGSS	17,191	MLVCB_P08361_3mutA
GSSPAPGGG	17,192	MLVMS_P03355_PLV919
EAAAKEAAAKEAAAKEAAAKEAAAK	17,193	MLVAV_P03356_3mutA
EAAAKGSSPAP	17,194	FLV_P10273_3mutA
GGGGSS	17,195	XMRV6_A1Z651_3mutA
GGSPAPGSS	17,196	MLVMS_P03355_PLV919
EAAAKEAAAKEAAAKEAAAKEAAAK	17,197	MLVMS_P03355_3mutA_WS
PAPEAAAKGGG	17,198	FLV_P10273_3mutA
EAAAKPAPGGS	17,199	XMRV6_A1Z651_3mut
PAPAP	17,200	BAEVM_P10272_3mutA
EAAAKEAAAKEAAAKEAAAK	17,201	MLVMS_P03355_PLV919
GSSPAPGGG	17,202	MLVMS_P03355_PLV919
EAAAKGGGPAP	17,203	KORV_Q9TTC1_3mutA
PAPEAAAK	17,204	MLVMS_P03355_PLV919
PAPGGGEAAAK	17,205	PERV_Q4VFZ2_3mutA_WS
EAAAKGSSGGS	17,206	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAK	17,207	MLVMS_P03355_PLV919
GSSEAAAK	17,208	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSS	17,209	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGSGGGGS	17,210	MLVMS_P03355_3mutA_WS
EAAAKGGGGSEAAAK	17,211	MLVMS_P03355_3mut
GGS	17,212	MLVCB_P08361_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,213	XMRV6_A1Z651_3mutA
GGSGSSPAP	17,214	MLVCB_P08361_3mutA
GGGGSGGGGSGGGGS	17,215	XMRV6_A1Z651_3mutA
PAPAPAPAPAP	17,216	BAEVM_P10272_3mutA
PAPAPAPAPAP	17,217	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAK	17,218	MLVBM_Q7SVK7_3mut
GGGGSGGGGSGGGGSGGGGSGGGGS	17,219	BAEVM_P10272_3mutA
GGSGGSGGS	17,220	MLVMS_P03355_3mutA_WS
EAAAKPAPGSS	17,221	MLVMS_P03355_PLV919
GSS	17,222	MLVMS_P03355_3mutA_WS
PAPEAAAKGGS	17,223	MLVMS_P03355_3mutA_WS
GGGPAPGGS	17,224	MLVMS_P03355_3mutA_WS
EAAAKGGGGSS	17,225	MLVAV_P03356_3mutA
GSSGSSGSSGSSGSS	17,226	MLVFF_P26809_3mut

SGSETPGTSESATPES	17,227	PERV_Q4VFZ2_3mut
GGSEAAAKGGG	17,228	MLVMS_P03355_3mut
GSSGSSGSSGSSGSSGSS	17,229	AVIRE_P03360_3mutA
PAPAPAPAPAPAP	17,230	AVIRE_P03360_3mut
GGSGGS	17,231	XMRV6_A1Z651_3mutA
PAPGSSEAAAK	17,232	MLVCB_P08361_3mut
GGSPAPEAAAK	17,233	PERV_Q4VFZ2_3mut
EAAAKGGGGGS	17,234	MLVCB_P08361_3mutA
GGSGGSGGSGGS	17,235	MLVMS_P03355_PLV919
GGGGSSEAAAK	17,236	MLVMS_P03355_PLV919
GSSEAAAKGGG	17,237	MLVFF_P26809_3mutA
PAPGGS	17,238	MLVMS_P03355_3mutA_WS
EAAAKGGSGGG	17,239	MLVCB_P08361_3mutA
EAAAKGGG	17,240	PERV_Q4VFZ2_3mut
PAPGGS	17,241	XMRV6_A1Z651_3mutA
GSSPAPGGG	17,242	XMRV6_A1Z651_3mutA
PAPEAAAKGGG	17,243	MLVMS_P03355_3mutA_WS
GSSEAAAKGGG	17,244	PERV_Q4VFZ2_3mutA_WS
PAPGGSEAAAK	17,245	XMRV6_A1Z651_3mutA
GGGGGS	17,246	MLVMS_P03355_3mutA_WS
GGSPAPEAAAK	17,247	MLVMS_P03355_3mutA_WS
GGGPAP	17,248	MLVFF_P26809_3mutA
PAPGSSGGG	17,249	XMRV6_A1Z651_3mutA
PAPGSSGGG	17,250	MLVBM_Q7SVK7_3mutA_WS
GGGEAAAKGSS	17,251	MLVMS_P03355_3mutA_WS
GSSEAAAKGGS	17,252	MLVCB_P08361_3mutA
PAPGGSGSS	17,253	MLVCB_P08361_3mutA
EAAAKGGGGSEAAAK	17,254	BAEVM_P10272_3mutA
PAPAPAP	17,255	PERV_Q4VFZ2_3mutA_WS
GGGGGG	17,256	MLVAV_P03356_3mutA
GSSPAPEAAAK	17,257	MLVCB_P08361_3mutA
GGSGGSGGS	17,258	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSSGSS	17,259	XMRV6_A1Z651_3mut
GGGPAPGGS	17,260	XMRV6_A1Z651_3mutA
GGGPAPEAAAK	17,261	BAEVM_P10272_3mutA
GGSGGG	17,262	AVIRE_P03360_3mutA
SGSETPGTSESATPES	17,263	PERV_Q4VFZ2_3mutA_WS
EAAAKGSSPAP	17,264	MLVMS_P03355_PLV919
GSSEAAAK	17,265	XMRV6_A1Z651_3mut
GSSGGSGGG	17,266	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	17,267	WMSV_P03359_3mutA

GGGGSEAAAKGGGGS	17,268	MLVMS_P03355_PLV919
PAPGGGGSS	17,269	MLVMS_P03355_3mutA_WS
SGSETPGTSESATPES	17,270	MLVMS_P03355_3mutA_WS
GGSPAPEAAAK	17,271	KORV_Q9TTC1-Pro_3mutA
GSSEAAAKGGG	17,272	MLVMS_P03355_3mutA_WS
GSSEAAAK	17,273	WMSV_P03359_3mutA
GGGGSEAAAKGGGGS	17,274	AVIRE_P03360_3mutA
GSS	17,275	WMSV_P03359_3mutA
PAPGGSEAAAK	17,276	MLVFF_P26809_3mutA
GGGGS	17,277	MLVMS_P03355_3mutA_WS
GGGPAP	17,278	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,279	MLVMS_P03355_3mutA_WS
EAAAKPAPGSS	17,280	PERV_Q4VFZ2_3mut
EAAAKPAPGSS	17,281	MLVCB_P08361_3mutA
GGGGGG	17,282	WMSV_P03359_3mutA
EAAAKPAPGGS	17,283	MLVMS_P03355_PLV919
PAPGGGEAAAK	17,284	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAKEAAAK	17,285	AVIRE_P03360_3mutA
GSSEAAAKPAP	17,286	XMRV6_A1Z651_3mutA
PAPGGSEAAAK	17,287	MLVBM_Q7SVK7_3mutA_WS
PAPGSS	17,288	MLVCB_P08361_3mutA
EAAAKGGG	17,289	MLVMS_P03355_3mutA_WS
EAAAKPAP	17,290	MLVCB_P08361_3mutA
PAPEAAAKGGS	17,291	MLVBM_Q7SVK7_3mutA_WS
GGSPAPGGG	17,292	MLVCB_P08361_3mutA
PAPGGSGSS	17,293	WMSV_P03359_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,294	MLVMS_P03355_PLV919
GGSGGGPAP	17,295	MLVMS_P03355_PLV919
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,296	MLVMS_P03355
PAPEAAAKGSS	17,297	MLVCB_P08361_3mutA
EAAAKGSS	17,298	MLVMS_P03355_3mutA_WS
GGSGGS	17,299	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAK	17,300	BAEVM_P10272_3mutA
GGGGSEAAAKGGGGS	17,301	FLV_P10273_3mutA
GGSEAAAKGGG	17,302	MLVCB_P08361_3mutA
GSSGSSGSSGSSGSS	17,303	BAEVM_P10272_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,304	MLVFF_P26809_3mutA
EAAAKGGG	17,305	PERV_Q4VFZ2_3mut
GGGGGSEAAAK	17,306	MLVCB_P08361_3mutA

EAAAKPAPGGS	17,307	MLVMS_P03355_3mutA_WS
GGGGGSGSS	17,308	XMRV6_A1Z651_3mutA
PAPGSSEAAAK	17,309	MLVMS_P03355_3mutA_WS
GSSEAAAKPAP	17,310	MLVCB_P08361_3mutA
EAAAKGSSPAP	17,311	MLVAV_P03356_3mutA
GGGPAPGGS	17,312	WMSV_P03359_3mutA
GGSPAP	17,313	MLVMS_P03355_3mutA_WS
GGSEAAAKGGG	17,314	MLVMS_P03355_3mutA_WS
GGGGGGGG	17,315	MLVFF_P26809_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,316	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,317	MLVBM_Q7SVK7_3mutA_WS
GSSPAPGGG	17,318	MLVAV_P03356_3mutA
GGGGGG	17,319	AVIRE_P03360_3mutA
GSSGGS	17,320	MLVMS_P03355_3mutA_WS
GGSPAPGSS	17,321	MLVFF_P26809_3mutA
PAPEAAAKGGG	17,322	PERV_Q4VFZ2_3mut
EAAAKGGGPAP	17,323	MLVFF_P26809_3mutA
GGGEAAAKGGS	17,324	MLVMS_P03355_PLV919
GGSGSSPAP	17,325	MLVFF_P26809_3mutA
SGSETPGTSESATPES	17,326	WMSV_P03359_3mutA
PAPGGSEAAAK	17,327	MLVBM_Q7SVK7_3mutA_WS
GGSGGG	17,328	MLVMS_P03355_PLV919
GGGGSSPAP	17,329	PERV_Q4VFZ2_3mut
GGGEAAAKGSS	17,330	MLVAV_P03356_3mutA
PAPAPAPAPAPAP	17,331	MLVMS_P03355_3mutA_WS
EAAAKGGGGSEAAAK	17,332	PERV_Q4VFZ2
EAAAKEAAAKEAAAKEAAAKEAAAK	17,333	MLVMS_P03355_PLV919
GGGGGSEAAAK	17,334	PERV_Q4VFZ2_3mut
PAPGSSEAAAK	17,335	MLVCB_P08361_3mutA
GSAGSAAGSGEF	17,336	PERV_Q4VFZ2_3mutA_WS
EAAAKGGGGSEAAAK	17,337	MLVFF_P26809_3mutA
GGSPAPGGG	17,338	PERV_Q4VFZ2_3mutA_WS
GSSEAAAKGGG	17,339	AVIRE_P03360_3mutA
GGGEAAAKPAP	17,340	MLVMS_P03355_3mutA_WS
GGGPAP	17,341	AVIRE_P03360_3mutA
GGSEAAAK	17,342	MLVCB_P08361_3mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,343	PERV_Q4VFZ2_3mut
EAAAKPAPGGS	17,344	MLVBM_Q7SVK7_3mutA_WS

AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,345	XMRV6_A1Z651_3mut
GGGGGGGG	17,346	MLVCB_P08361_3mutA
PAPGSS	17,347	PERV_Q4VFZ2_3mut
EAAAK	17,348	PERV_Q4VFZ2_3mut
GSAGSAAGSGEF	17,349	MLVMS_P03355_3mutA_WS
PAPGGGEAAAK	17,350	PERV_Q4VFZ2_3mut
EAAAKGSSGGS	17,351	MLVFF_P26809_3mut
GGGGSEAAAKGGGGS	17,352	BAEVM_P10272_3mutA
GGGGSGGGGSGGGGS	17,353	MLVMS_P03355_PLV919
EAAAKGGGGSEAAAK	17,354	BAEVM_P10272_3mut
PAPGGGEAAAK	17,355	MLVMS_P03355_3mutA_WS
GGSEAAAKPAP	17,356	MLVMS_P03355_3mutA_WS
PAPAP	17,357	MLVCB_P08361_3mutA
PAPAP	17,358	MLVFF_P26809_3mutA
GGSPAP	17,359	AVIRE_P03360_3mutA
EAAAKGSSGGS	17,360	MLVCB_P08361_3mutA
PAPGSSGGS	17,361	AVIRE_P03360_3mutA
EAAAKGGGGSEAAAK	17,362	XMRV6_A1Z651_3mutA
PAPAPAP	17,363	BAEVM_P10272_3mutA
GGSGGSGGSGGSGGSGGS	17,364	MLVMS_P03355_PLV919
GGGGGSGSS	17,365	MLVMS_P03355_PLV919
PAPGSSEAAAK	17,366	XMRV6_A1Z651_3mut
GGSEAAAKPAP	17,367	XMRV6_A1Z651_3mutA
EAAAKEAAAKEAAAKEAAAK	17,368	XMRV6_A1Z651_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,369	WMSV_P03359_3mut
GGSGGGEAAAK	17,370	XMRV6_A1Z651_3mutA
GGGEAAAK	17,371	XMRV6_A1Z651_3mutA
GGGGSGGGGSGGGGS	17,372	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGSGGS	17,373	MLVFF_P26809_3mutA
GSSGGGGGS	17,374	MLVMS_P03355_3mut
PAPGGSEAAAK	17,375	MLVMS_P03355_3mutA_WS
GSSGGSPAP	17,376	MLVMS_P03355_3mutA_WS
SGSETPGTSESATPES	17,377	XMRV6_A1Z651_3mutA
GGGGSGGGGS	17,378	MLVMS_P03355_PLV919
PAPAPAPAPAP	17,379	MLVMS_P03355_3mut
GSSGSS	17,380	XMRV6_A1Z651_3mutA
GSSEAAAKPAP	17,381	PERV_Q4VFZ2_3mut
GGSGSSGGG	17,382	MLVMS_P03355_3mutA_WS
EAAAKEAAAK	17,383	MLVCB_P08361_3mutA
GSSGSSGSSGSS	17,384	MLVMS_P03355_3mutA_WS

GSSPAPGGG	17,385	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAK	17,386	MLVMS_P03355_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,387	SFV1_P23074_2mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,388	MLVMS_P03355_PLV919
GSAGSAAGSGEF	17,389	MLVMS_P03355_PLV919
PAPGSSEAAAK	17,390	MLVMS_P03355_3mutA_WS
GGSEAAAK	17,391	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSSGSS	17,392	PERV_Q4VFZ2_3mutA_WS
GGSEAAAKPAP	17,393	PERV_Q4VFZ2_3mutA_WS
GGSGGSGGS	17,394	MLVCB_P08361_3mutA
EAAAKGGSGSS	17,395	MLVCB_P08361_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGS	17,396	FLV_P10273_3mutA
EAAAKEAAAKEAAAKEAAAK	17,397	MLVBM_Q7SVK7_3mutA_WS
GGSGSSPAP	17,398	BAEVM_P10272_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	17,399	XMRV6_A1Z651_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGS	17,400	MLVBM_Q7SVK7_3mutA_WS
GGSGSS	17,401	WMSV_P03359_3mutA
PAPEAAAK	17,402	MLVCB_P08361_3mutA
EAAAKPAP	17,403	BAEVM_P10272_3mutA
GSSPAP	17,404	PERV_Q4VFZ2_3mutA_WS
GGGPAP	17,405	PERV_Q4VFZ2_3mutA_WS
EAAAKGGSGSS	17,406	MLVMS_P03355_3mutA_WS
EAAAKGGGGSEAAAK	17,407	AVIRE_P03360_3mutA
GGSGGG	17,408	KORV_Q9TTC1-Pro_3mutA
GSSPAP	17,409	MLVFF_P26809_3mutA
GGSGSSEAAAK	17,410	BAEVM_P10272_3mutA
PAPGSSGGS	17,411	BAEVM_P10272_3mutA
GGGGGG	17,412	MLVFF_P26809_3mutA
PAPGGSEAAAK	17,413	MLVMS_P03355_PLV919
PAPGGS	17,414	MLVMS_P03355_PLV919
GGSGGSGGSGGS	17,415	BAEVM_P10272_3mutA
GSSPAP	17,416	MLVCB_P08361_3mutA
PAPAPAPAP	17,417	MLVMS_P03355_3mutA_WS
GGGGGG	17,418	MLVCB_P08361_3mutA
GSSGSSGSSGSSGSSGSS	17,419	KORV_Q9TTC1-Pro_3mutA
GSSEAAAKGGS	17,420	BAEVM_P10272_3mutA
GGSEAAAK	17,421	FLV_P10273_3mutA
GGSGGSGGSGGSGGS	17,422	KORV_Q9TTC1-Pro_3mutA
GSSPAPEAAAK	17,423	PERV_Q4VFZ2_3mut

GSSGSSGSSGSSGSS	17,424	XMRV6_A1Z651_3mutA
EAAAKPAPGGS	17,425	MLVMS_P03355_3mut
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,426	FLV_P10273_3mut
GGSPAPEAAAK	17,427	XMRV6_A1Z651_3mut
EAAAKGGSGGG	17,428	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAK	17,429	MLVFF_P26809_3mutA
GSSPAP	17,430	WMSV_P03359_3mutA
PAPAPAPAP	17,431	MLVAV_P03356_3mutA
PAPGGSEAAAK	17,432	KORV_Q9TTC1_3mut
GGSGSSEAAAK	17,433	MLVBM_Q7SVK7_3mutA_WS
GSSGGG	17,434	MLVCB_P08361_3mutA
GGGEAAAKGSS	17,435	PERV_Q4VFZ2_3mut
PAPGGSGGG	17,436	MLVFF_P26809_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,437	FFV_O93209
PAPGGGGSS	17,438	MLVMS_P03355_3mutA_WS
EAAAKGGS	17,439	MLVAV_P03356_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,440	MLVBM_Q7SVK7_3mutA_WS
GGSGGSGGS	17,441	WMSV_P03359_3mutA
PAPAP	17,442	MLVMS_P03355_3mutA_WS
GSSGGGEAAAK	17,443	MLVAV_P03356_3mutA
GGGGSSEAAAK	17,444	MLVFF_P26809_3mutA
EAAAKGSSGGS	17,445	MLVMS_P03355_PLV919
EAAAKGGGGSEAAAK	17,446	MLVMS_P03355_3mutA_WS
GGGGGGGG	17,447	MLVMS_P03355_PLV919
GSSGSSGSS	17,448	MLVMS_P03355_PLV919
GGGEAAAKPAP	17,449	PERV_Q4VFZ2_3mutA_WS
GGGGGSGSS	17,450	MLVMS_P03355_3mutA_WS
GGGGGGG	17,451	MLVMS_P03355_PLV919
GGS	17,452	MLVMS_P03355_PLV919
GSSGGG	17,453	MLVMS_P03355_3mutA_WS
EAAAKGGSGSS	17,454	PERV_Q4VFZ2_3mutA_WS
PAPGSSEAAAK	17,455	MLVMS_P03355_PLV919
GSSEAAAKPAP	17,456	MLVMS_P03355_PLV919
GGSPAPGSS	17,457	BAEVM_P10272_3mutA
GSAGSAAGSGEF	17,458	MLVCB_P08361_3mut
GGSPAPGGG	17,459	PERV_Q4VFZ2_3mut
GGGGSGGGGSGGGGSGGGGS	17,460	MLVMS_P03355_3mut
GSSGSSGSS	17,461	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,462	PERV_Q4VFZ2_3mut

GGGGSEAAAKGGGGS	17,463	MLVCB_P08361_3mutA
GGSEAAAKGSS	17,464	MLVAV_P03356_3mutA
EAAAKGGGGSEAAAK	17,465	MLVCB_P08361_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,466	XMRV6_A1Z651_3mutA
PAPGGGEAAAK	17,467	MLVMS_P03355_3mutA_WS
GSSGGGEAAAK	17,468	PERV_Q4VFZ2_3mutA_WS
GSSGSS	17,469	MLVCB_P08361_3mut
PAPAPAPAPAPAP	17,470	PERV_Q4VFZ2_3mut
GGSPAPGGG	17,471	MLVFF_P26809_3mutA
GGSGGSGGSGGSGGS	17,472	MLVCB_P08361_3mutA
EAAAKEAAAK	17,473	MLVFF_P26809_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,474	GALV_P21414_3mut
PAPAPAPAPAPAP	17,475	WMSV_P03359_3mutA
GGGEAAAKGGS	17,476	KORV_Q9TTC1_3mutA
EAAAKGGGPAP	17,477	KORV_Q9TTC1_3mut
PAPEAAAKGSS	17,478	MLVBM_Q7SVK7_3mutA_WS
PAPEAAAKGSS	17,479	FLV_P10273_3mutA
PAPGGSEAAAK	17,480	MLVMS_P03355_3mut
GSSPAPGGG	17,481	BAEVM_P10272_3mutA
GGGEAAAKPAP	17,482	KORV_Q9TTC1-Pro_3mutA
GGGGSGGGGS	17,483	MLVMS_P03355_PLV919
GGGEAAAKGSS	17,484	MLVFF_P26809_3mutA
PAPGGGGSS	17,485	MLVBM_Q7SVK7_3mutA_WS
GSSEAAAK	17,486	BAEVM_P10272_3mutA
GGGGGGGG	17,487	MLVMS_P03355_PLV919
PAPGSSGGS	17,488	MLVAV_P03356_3mutA
GGGGSGGGGSGGGGSGGGGS	17,489	BAEVM_P10272_3mutA
PAP	17,490	MLVMS_P03355_3mut
EAAAKGSSPAP	17,491	XMRV6_A1Z651_3mutA
PAPEAAAKGGS	17,492	MLVFF_P26809_3mutA
GSSGGGEAAAK	17,493	BAEVM_P10272_3mutA
PAPAPAP	17,494	MLVMS_P03355_3mutA_WS
GGSEAAAKGGG	17,495	MLVMS_P03355_PLV919
GSSEAAAK	17,496	PERV_Q4VFZ2_3mut
GGGG	17,497	MLVMS_P03355_3mutA_WS
GGGGGS	17,498	MLVMS_P03355_3mut
GGGGSSEAAAK	17,499	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,500	SFV3L_P27401-Pro_2mutA
GGSEAAAKGSS	17,501	MLVMS_P03355_3mutA_WS

PAPGSSGGS	17,502	XMRV6_A1Z651_3mutA
GGSPAP	17,503	MLVMS_P03355_3mutA_WS
GGGGSSEAAAK	17,504	BAEVM_P10272_3mut
GGSGGSGGSGGS	17,505	AVIRE_P03360_3mutA
PAPGSSGGS	17,506	MLVFF_P26809_3mutA
GSSPAPGGG	17,507	MLVMS_P03355_3mutA_WS
GGGGGGG	17,508	MLVMS_P03355_3mutA_WS
EAAAKGGGGGS	17,509	MLVMS_P03355_3mutA_WS
EAAAKGGSGGG	17,510	MLVMS_P03355_PLV919
GGGGSSEAAAK	17,511	XMRV6_A1Z651_3mutA
GGGGSEAAAKGGGGS	17,512	MLVBM_Q7SVK7_3mutA_WS
GSSGSS	17,513	MLVMS_P03355_PLV919
GGSGGG	17,514	MLVMS_P03355_PLV919
PAPEAAAKGGG	17,515	AVIRE_P03360_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,516	FOAMV_P14350-Pro_2mutA
GGGGGSGSS	17,517	PERV_Q4VFZ2_3mut
GSSGSSGSSGSSGSS	17,518	KORV_Q9TTC1-Pro_3mut
GGGGSEAAAKGGGGS	17,519	MLVMS_P03355_3mutA_WS
GGGGGSPAP	17,520	FLV_P10273_3mut
GGGEAAAK	17,521	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGS	17,522	FLV_P10273_3mutA
GGG	17,523	MLVMS_P03355_PLV919
GGSPAPEAAAK	17,524	BAEVM_P10272_3mutA
EAAAKEAAAK	17,525	FLV_P10273_3mutA
GGGEAAAKPAP	17,526	BAEVM_P10272_3mutA
GGGEAAAKGGS	17,527	PERV_Q4VFZ2_3mut
GGSGGSGGS	17,528	PERV_Q4VFZ2_3mut
EAAAKGGGPAP	17,529	XMRV6_A1Z651_3mutA
EAAAK	17,530	MLVBM_Q7SVK7_3mutA_WS
PAPEAAAKGGG	17,531	PERV_Q4VFZ2_3mut
EAAAKGSS	17,532	MLVCB_P08361_3mutA
GGSEAAAKGGG	17,533	MLVBM_Q7SVK7_3mutA_WS
GGGGSGGGGSGGGGSGGGGS	17,534	XMRV6_A1Z651_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGS	17,535	BAEVM_P10272_3mut
GGGGSSPAP	17,536	PERV_Q4VFZ2_3mutA_WS
GGSGGSGGSGGSGGSGGS	17,537	PERV_Q4VFZ2_3mut
GGGEAAAKPAP	17,538	PERV_Q4VFZ2_3mut
EAAAKEAAAK	17,539	BAEVM_P10272_3mutA

GGSGSSEAAAK	17,540	XMRV6_A1Z651_3mutA
PAPEAAAKGSS	17,541	WMSV_P03359_3mutA
PAPAPAPAPAP	17,542	XMRV6_A1Z651_3mutA
GSSGGGEAAAK	17,543	MLVMS_P03355_PLV919
GSSPAPGGG	17,544	MLVFF_P26809_3mutA
GGSPAPEAAAK	17,545	MLVFF_P26809_3mut
PAPGGSEAAAK	17,546	PERV_Q4VFZ2_3mut
GGGGSS	17,547	MLVFF_P26809_3mutA
GGSGSSGGG	17,548	BAEVM_P10272_3mutA
GSSGGGEAAAK	17,549	MLVMS_P03355_3mutA_WS
EAAAKGGS	17,550	MLVBM_Q7SVK7_3mutA_WS
GGGPAPGGS	17,551	MLVMS_P03355_PLV919
EAAAKEAAAK	17,552	MLVMS_P03355_PLV919
GSSGSSGSS	17,553	MLVMS_P03355_PLV919
GGGEAAAKPAP	17,554	MLVAV_P03356_3mutA
SGSETPGTSESATPES	17,555	FLV_P10273_3mutA
PAPAPAPAPAP	17,556	KORV_Q9TTC1-Pro_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,557	BAEVM_P10272_3mutA
PAPGSSGGG	17,558	MLVMS_P03355_3mutA_WS
GSSGGGEAAAK	17,559	XMRV6_A1Z651_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGS	17,560	XMRV6_A1Z651_3mutA
GGGGSSPAP	17,561	MLVFF_P26809_3mutA
GGSGGGPAP	17,562	PERV_Q4VFZ2_3mutA_WS
GSS	17,563	PERV_Q4VFZ2_3mut
EAAAKGSSPAP	17,564	MLVMS_P03355_3mut
EAAAKGGG	17,565	XMRV6_A1Z651_3mutA
GSSGSSGSSGSS	17,566	WMSV_P03359_3mutA
PAPEAAAKGSS	17,567	MLVMS_P03355_PLV919
GSSEAAAK	17,568	AVIRE_P03360_3mutA
EAAAKGGSGSS	17,569	AVIRE_P03360_3mutA
GSSEAAAK	17,570	MLVMS_P03355_3mut
GGSGSSEAAAK	17,571	MLVMS_P03355_PLV919
GGSEAAAKGGG	17,572	MLVFF_P26809_3mutA
GGGGSGGGGSGGGGSGGGGS	17,573	MLVAV_P03356_3mutA
PAPAPAPAPAPAP	17,574	MLVFF_P26809_3mut
EAAAKPAPGSS	17,575	KORV_Q9TTC1-Pro_3mut
PAPGSSEAAAK	17,576	MLVAV_P03356_3mutA
GGGGSSPAP	17,577	WMSV_P03359_3mutA
EAAAKGGGGGS	17,578	MLVMS_P03355_3mutA_WS
GGGEAAAKGGS	17,579	MLVMS_P03355_3mut

GGSGSSGGG	17,580	MLVMS_P03355_3mut
GGGPAPGGS	17,581	MLVAV_P03356_3mutA
PAPGGGGGS	17,582	MLVMS_P03355_PLV919
GGGPAPGSS	17,583	PERV_Q4VFZ2_3mut
GGGGGGG	17,584	MLVFF_P26809_3mutA
GGSGGGGSS	17,585	MLVCB_P08361_3mutA
GGGGGG	17,586	FLV_P10273_3mutA
GGSEAAAKGSS	17,587	PERV_Q4VFZ2_3mut
GGSPAPGGG	17,588	BAEVM_P10272_3mutA
GGSPAPGSS	17,589	AVIRE_P03360_3mutA
GGSGGSGGSGGS	17,590	KORV_Q9TTC1_3mut
EAAAKEAAAKEAAAKEAAAKEAAAK	17,591	MLVBM_Q7SVK7_3mut
PAPGSSGGS	17,592	XMRV6_A1Z651_3mut
EAAAKGGGGSS	17,593	PERV_Q4VFZ2_3mutA_WS
GGSGGSGGSGGSGGS	17,594	PERV_Q4VFZ2_3mutA_WS
PAPGGSGGG	17,595	MLVMS_P03355_PLV919
PAPGSSGGG	17,596	PERV_Q4VFZ2_3mutA_WS
GSSGSS	17,597	BAEVM_P10272_3mutA
EAAAKGSS	17,598	MLVFF_P26809_3mutA
GGGPAP	17,599	MLVMS_P03355_PLV919
EAAAKGGGGGS	17,600	MLVFF_P26809_3mutA
EAAAKGGSPAP	17,601	MLVBM_Q7SVK7_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,602	WMSV_P03359_3mutA
GSSPAPGGG	17,603	MLVBM_Q7SVK7_3mutA_WS
GGGEAAAKGSS	17,604	AVIRE_P03360_3mutA
GGGGSSEAAAK	17,605	AVIRE_P03360_3mutA
GGGGGGGG	17,606	PERV_Q4VFZ2_3mutA_WS
PAPGSSEAAAK	17,607	BAEVM_P10272_3mutA
EAAAKGSS	17,608	MLVFF_P26809_3mut
GSSEAAAKGGG	17,609	MLVCB_P08361_3mutA
GGSEAAAK	17,610	MLVBM_Q7SVK7_3mutA_WS
GSSEAAAKGGG	17,611	PERV_Q4VFZ2_3mutA_WS
PAPGGSGGG	17,612	WMSV_P03359_3mutA
GSSGGSGGG	17,613	MLVCB_P08361_3mutA
EAAAKGSSGGG	17,614	FLV_P10273_3mutA
GSSEAAAK	17,615	MLVCB_P08361_3mutA
GSSGGGEAAAK	17,616	MLVMS_P03355_3mut
GGGGSGGGGS	17,617	MLVCB_P08361_3mutA

EAAAKGGGGSEAAAK	17,618	MLVBM_Q7SVK7_3mutA_WS
EAAAKGGG	17,619	PERV_Q4VFZ2_3mutA_WS
EAAAKGGSPAP	17,620	MLVMS_P03355_PLV919
GGGPAPGGS	17,621	AVIRE_P03360_3mutA
GSSEAAAK	17,622	MLVBM_Q7SVK7_3mutA_WS
GSSGGGEAAAK	17,623	PERV_Q4VFZ2_3mut
SGSETPGTSESATPES	17,624	MLVMS_P03355_PLV919
GGSGSSPAP	17,625	MLVMS_P03355_3mut
GGGGGG	17,626	MLVBM_Q7SVK7_3mutA_WS
GGSPAPGGG	17,627	XMRV6_A1Z651_3mutA
GGSGSS	17,628	PERV_Q4VFZ2_3mutA_WS
PAP	17,629	MLVBM_Q7SVK7_3mutA_WS
EAAAKPAPGSS	17,630	MLVMS_P03355_PLV919
EAAAKGGG	17,631	MLVMS_P03355_3mut
GSSEAAAKPAP	17,632	PERV_Q4VFZ2_3mutA_WS
GGGGSS	17,633	MLVMS_P03355_3mutA_WS
GGSGSSEAAAK	17,634	PERV_Q4VFZ2_3mut
GGGGSS	17,635	BAEVM_P10272_3mutA
PAPAP	17,636	MLVFF_P26809_3mut
PAPEAAAKGGG	17,637	BAEVM_P10272_3mutA
EAAAKGGS	17,638	MLVMS_P03355_PLV919
PAPAPAPAPAPAP	17,639	PERV_Q4VFZ2_3mutA_WS
GGGGGSEAAAK	17,640	MLVMS_P03355_3mut
PAPGGS	17,641	PERV_Q4VFZ2_3mut
GGGGSS	17,642	MLVCB_P08361_3mutA
GGGGS	17,643	MLVAV_P03356_3mutA
GSSPAPEAAAK	17,644	MLVMS_P03355_PLV919
GGGGSSGGS	17,645	MLVFF_P26809_3mutA
PAPEAAAKGSS	17,646	MLVMS_P03355_PLV919
GGSGSSEAAAK	17,647	MLVMS_P03355_3mutA_WS
EAAAKGGG	17,648	MLVAV_P03356_3mutA
PAPGSSEAAAK	17,649	FLV_P10273_3mutA
EAAAKGSSGGG	17,650	MLVCB_P08361_3mutA
PAPEAAAK	17,651	KORV_Q9TTC1-Pro_3mutA
GGSPAPEAAAK	17,652	KORV_Q9TTC1-Pro_3mut
GGSGGSGGSGGSGGSGGS	17,653	MLVAV_P03356_3mutA
GSSEAAAKPAP	17,654	MLVBM_Q7SVK7_3mutA_WS

AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,655	KORV_Q9TTC1-Pro_3mutA
GSSGGGEAAAK	17,656	XMRV6_A1Z651_3mut
PAPGGSGGG	17,657	AVIRE_P03360_3mutA
PAPGGSEAAAK	17,658	PERV_Q4VFZ2_3mutA_WS
GGGGS	17,659	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGS	17,660	MLVBM_Q7SVK7_3mutA_WS
PAPAPAPAPAP	17,661	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAK	17,662	MLVMS_P03355_3mut
GSSGGSEAAAK	17,663	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGS	17,664	WMSV_P03359_3mutA
EAAAKGSSGGG	17,665	WMSV_P03359_3mutA
EAAAKGGG	17,666	PERV_Q4VFZ2_3mutA_WS
SGSETPGTSESATPES	17,667	PERV_Q4VFZ2_3mut
PAPGSSGGS	17,668	MLVMS_P03355_3mutA_WS
PAPEAAAKGSS	17,669	PERV_Q4VFZ2_3mut
PAPEAAAK	17,670	AVIRE_P03360_3mutA
GSSEAAAKGGG	17,671	BAEVM_P10272_3mutA
GSSPAP	17,672	MLVAV_P03356_3mutA
EAAAKEAAAKEAAAKEAAAK	17,673	MLVFF_P26809_3mut
PAPGGSGSS	17,674	MLVAV_P03356_3mutA
GGGGSGGGGSGGGGS	17,675	PERV_Q4VFZ2_3mutA_WS
GSSGGSEAAAK	17,676	MLVCB_P08361_3mutA
EAAAKGGS	17,677	KORV_Q9TTC1-Pro_3mutA
EAAAKGGS	17,678	MLVFF_P26809_3mutA
GGSPAP	17,679	MLVMS_P03355_PLV919
GGSGSS	17,680	MLVMS_P03355_PLV919
SGSETPGTSESATPES	17,681	WMSV_P03359_3mut
GGGGGGG	17,682	WMSV_P03359_3mut
GGSPAPGSS	17,683	MLVCB_P08361_3mutA
GGGGSSGGS	17,684	WMSV_P03359_3mut
PAPGGS	17,685	MLVMS_P03355_PLV919
PAPGSSGGS	17,686	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	17,687	MLVFF_P26809_3mut
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,688	PERV_Q4VFZ2_3mut
GGSGGSGGSGGSGGS	17,689	BAEVM_P10272_3mutA
GSSEAAAK	17,690	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAK	17,691	KORV_Q9TTC1-Pro_3mutA
GGSGGSGGSGGSGGS	17,692	MLVMS_P03355_3mut
PAPAPAPAPAPAP	17,693	MLVMS_P03355_3mut
GGSPAPEAAAK	17,694	MLVMS_P03355_PLV919

EAAAK	17,695	WMSV_P03359_3mutA
EAAAKGSSGGS	17,696	MLVBM_Q7SVK7_3mutA_WS
GGSGGGGSS	17,697	MLVMS_P03355_3mutA_WS
GGGEAAAKPAP	17,698	MLVMS_P03355_3mut
EAAAKGGSGGG	17,699	XMRV6_A1Z651_3mutA
GGGGGSEAAAK	17,700	KORV_Q9TTC1-Pro_3mutA
GGGGGG	17,701	BAEVM_P10272_3mutA
GGGGGG	17,702	MLVMS_P03355_3mut
GGGGGGG	17,703	MLVBM_Q7SVK7_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,704	AVIRE_P03360
PAPGSSGGS	17,705	PERV_Q4VFZ2_3mut
GGGGGS	17,706	XMRV6_A1Z651_3mut
EAAAKPAP	17,707	XMRV6_A1Z651_3mutA
GGG	17,708	MLVMS_P03355_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,709	FLV_P10273_3mut
EAAAKGSSPAP	17,710	MLVMS_P03355_3mut
SGSETPGTSESATPES	17,711	BAEVM_P10272_3mutA
GGSPAPEAAAK	17,712	MLVMS_P03355_3mut
GSSGSSGSSGSS	17,713	MLVAV_P03356_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,714	MLVMS_P03355_3mut
GGSPAP	17,715	MLVCB_P08361_3mutA
GGGGGSEAAAK	17,716	MLVMS_P03355_3mutA_WS
GGGGG	17,717	MLVFF_P26809_3mutA
GSSEAAAK	17,718	MLVAV_P03356_3mutA
GGS	17,719	BAEVM_P10272_3mut
EAAAKGGSPAP	17,720	MLVCB_P08361_3mutA
PAPAPAPAP	17,721	FLV_P10273_3mutA
PAPGGGEAAAK	17,722	MLVCB_P08361_3mutA
GGGGSSEAAAK	17,723	MLVMS_P03355_3mutA_WS
GGGGG	17,724	PERV_Q4VFZ2_3mutA_WS
GGSGGSGGSGGSGGSGGS	17,725	PERV_Q4VFZ2_3mut
GGGGG	17,726	MLVMS_P03355_3mut
PAPEAAAKGGG	17,727	MLVBM_Q7SVK7_3mutA_WS
GSSGGGPAP	17,728	XMRV6_A1Z651_3mutA
GSSGSSGSSGSSGSSGSS	17,729	PERV_Q4VFZ2_3mutA_WS
EAAAKGGSPAP	17,730	PERV_Q4VFZ2_3mut
GSSGGSEAAAK	17,731	MLVMS_P03355_PLV919

GSS	17,732	PERV_Q4VFZ2_3mut
EAAAKGGS	17,733	WMSV_P03359_3mutA
GGGGGSPAP	17,734	PERV_Q4VFZ2_3mutA_WS
EAAAKGSS	17,735	MLVMS_P03355_PLV919
EAAAKGGGGSS	17,736	KORV_Q9TTC1-Pro_3mutA
PAPGSSGGG	17,737	PERV_Q4VFZ2_3mut
GGGGSSEAAAK	17,738	MLVFF_P26809_3mut
PAPAPAP	17,739	MLVMS_P03355_3mut
GSSGGSEAAAK	17,740	XMRV6_A1Z651_3mut
PAPEAAAKGSS	17,741	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGSGGS	17,742	MLVMS_P03355_3mutA_WS
GGSGSSPAP	17,743	XMRV6_A1Z651_3mutA
GGGGSSPAP	17,744	MLVMS_P03355_PLV919
GGGGS	17,745	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAK	17,746	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAK	17,747	KORV_Q9TTC1_3mutA
PAPGGGEAAAK	17,748	BAEVM_P10272_3mutA
GSSGGSEAAAK	17,749	XMRV6_A1Z651_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,750	FLV_P10273_3mut
GSSEAAAKPAP	17,751	MLVMS_P03355_3mutA_WS
EAAAKPAPGSS	17,752	PERV_Q4VFZ2_3mutA_WS
GSSGGSPAP	17,753	XMRV6_A1Z651_3mutA
GSSEAAAKGGG	17,754	PERV_Q4VFZ2_3mut
GGGEAAAKGGS	17,755	WMSV_P03359_3mutA
GSSEAAAKGGG	17,756	MLVFF_P26809_3mut
PAPAPAP	17,757	KORV_Q9TTC1-Pro_3mutA
EAAAKGGSPAP	17,758	MLVMS_P03355_3mutA_WS
PAPGGSEAAAK	17,759	PERV_Q4VFZ2_3mut
GGGGS	17,760	MLVBM_Q7SVK7_3mutA_WS
EAAAKGSSGGG	17,761	KORV_Q9TTC1_3mut
EAAAKGGGPAP	17,762	MLVCB_P08361_3mutA
EAAAKGSS	17,763	BAEVM_P10272_3mutA
GGSPAPGGG	17,764	MLVBM_Q7SVK7_3mutA_WS
GGGGSEAAAKGGGGS	17,765	MLVMS_P03355_3mutA_WS
GGGEAAAKGGS	17,766	PERV_Q4VFZ2_3mutA_WS
EAAAKGGGGSS	17,767	MLVMS_P03355_3mutA_WS
EAAAKGGGPAP	17,768	MLVFF_P26809_3mut
GSSPAP	17,769	PERV_Q4VFZ2_3mutA_WS
EAAAKGGS	17,770	MLVMS_P03355_3mut
GGGGSS	17,771	KORV_Q9TTC1-Pro_3mutA

EAAAKGSSPAP	17,772	MLVMS_P03355_3mutA_WS
GGGPAP	17,773	PERV_Q4VFZ2_3mut
EAAAKGSSGGS	17,774	XMRV6_A1Z651_3mutA
PAPGGG	17,775	MLVAV_P03356_3mutA
GSSPAPEAAAK	17,776	BAEVM_P10272_3mutA
GGGPAP	17,777	MLVBM_Q7SVK7_3mutA_WS
GSSGGGGGS	17,778	AVIRE_P03360_3mutA
SGSETPGTSESATPES	17,779	MLVMS_P03355_PLV919
GGGPAP	17,780	MLVFF_P26809_3mut
EAAAKGGGGSS	17,781	XMRV6_A1Z651_3mutA
GGGGSSPAP	17,782	XMRV6_A1Z651_3mut
GGGGSEAAAKGGGGS	17,783	MLVMS_P03355_3mut
GSSPAP	17,784	MLVBM_Q7SVK7_3mutA_WS
GGSGSSEAAAK	17,785	FLV_P10273_3mutA
SGSETPGTSESATPES	17,786	MLVBM_Q7SVK7_3mutA_WS
PAPGGG	17,787	AVIRE_P03360_3mutA
GGGEAAAKPAP	17,788	MLVMS_P03355_3mutA_WS
EAAAKGGSGSS	17,789	PERV_Q4VFZ2_3mut
GGSPAPGGG	17,790	MLVAV_P03356_3mutA
PAPGGSGSS	17,791	BAEVM_P10272_3mutA
GSSGGSPAP	17,792	MLVFF_P26809_3mutA
EAAAKGSSGGG	17,793	PERV_Q4VFZ2_3mut
GGGGSGGGGS	17,794	PERV_Q4VFZ2_3mutA_WS
GSSGGGGGS	17,795	BAEVM_P10272_3mutA
GGGGSSGGS	17,796	MLVBM_Q7SVK7_3mutA_WS
EAAAKGGS	17,797	PERV_Q4VFZ2_3mutA_WS
GSSGSSGSSGSS	17,798	MLVMS_P03355_3mut
GGS	17,799	MLVMS_P03355_3mutA_WS
GSSGGSEAAAK	17,800	MLVBM_Q7SVK7_3mutA_WS
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,801	XMRV6_A1Z651
GGGGG	17,802	FLV_P10273_3mutA
PAPEAAAKGSS	17,803	PERV_Q4VFZ2_3mut
GGGGGG	17,804	WMSV_P03359_3mut
EAAAKGGG	17,805	BAEVM_P10272_3mutA
GGGGSS	17,806	MLVMS_P03355_3mutA_WS
GSSGGGEAAAK	17,807	KORV_Q9TTC1_3mut
GGSGSS	17,808	AVIRE_P03360_3mutA

EAAAKPAP	17,809	MLVMS_P03355_3mut
EAAAKEAAAKEAAAK	17,810	FLV_P10273_3mutA
GGGG	17,811	XMRV6_A1Z651_3mutA
GSSPAPGGS	17,812	BAEVM_P10272_3mutA
GSSGGGGGS	17,813	MLVFF_P26809_3mutA
GGGGSSGGS	17,814	MLVAV_P03356_3mutA
GGS	17,815	PERV_Q4VFZ2_3mut
GGGGG	17,816	WMSV_P03359_3mutA
GSSGSSGSSGSSGSSGSS	17,817	FLV_P10273_3mutA
PAPGGGGSS	17,818	MLVAV_P03356_3mutA
GGGGGGGG	17,819	BAEVM_P10272_3mutA
SGSETPGTSESATPES	17,820	MLVCB_P08361_3mutA
PAPGGG	17,821	BAEVM_P10272_3mutA
GSSGSSGSS	17,822	MLVCB_P08361_3mutA
GGSGSS	17,823	MLVMS_P03355_3mutA_WS
EAAAKGGGGSEAAAK	17,824	WMSV_P03359_3mutA
GGGGGGGG	17,825	FLV_P10273_3mutA
GSSGSS	17,826	MLVMS_P03355_3mutA_WS
PAPEAAAKGGS	17,827	XMRV6_A1Z651_3mutA
EAAAKEAAAK	17,828	MLVMS_P03355_3mut
GGGGSGGGGSGGGGS	17,829	BAEVM_P10272_3mutA
EAAAKGSSPAP	17,830	MLVMS_P03355_PLV919
GGGGSSEAAAK	17,831	MLVMS_P03355_3mut
GGGGSSEAAAK	17,832	BAEVM_P10272_3mutA
PAPGGSGSS	17,833	PERV_Q4VFZ2_3mut
GGSGGGEAAAK	17,834	MLVFF_P26809_3mut
PAPEAAAKGGS	17,835	PERV_Q4VFZ2_3mut
GGGPAPGSS	17,836	AVIRE_P03360_3mut
PAPGGSGGG	17,837	PERV_Q4VFZ2_3mutA_WS
GGGGGGGG	17,838	PERV_Q4VFZ2_3mutA_WS
GSSEAAAK	17,839	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGS	17,840	PERV_Q4VFZ2_3mutA_WS
EAAAKGGS	17,841	MLVMS_P03355_3mut
GGGGGSGSS	17,842	MLVCB_P08361_3mut
GGGPAP	17,843	KORV_Q9TTC1-Pro_3mutA
EAAAKPAPGGG	17,844	MLVCB_P08361_3mut
GSSGGSPAP	17,845	MLVCB_P08361_3mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,846	MLVMS_P03355_3mut
PAPAPAPAP	17,847	MLVMS_P03355_3mut
GSSGGS	17,848	XMRV6_A1Z651_3mutA
GSSEAAAKGGG	17,849	MLVMS_P03355_3mut

GGSGSSPAP	17,850	MLVMS_P03355_3mutA_WS
GSSEAAAKGGS	17,851	MLVMS_P03355_PLV919
EAAAKEAAAKEAAAKEAAAKEAAAK	17,852	BAEVM_P10272_3mut
PAPGGGGSS	17,853	KORV_Q9TTC1_3mutA
EAAAKGSS	17,854	MLVMS_P03355_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,855	FFV_O93209_2mut
GGSGGSGGSGGSGGSGGS	17,856	BAEVM_P10272_3mutA
GGGGGG	17,857	MLVMS_P03355_PLV919
PAPEAAAK	17,858	BAEVM_P10272_3mutA
GGSGSSEAAAK	17,859	MLVAV_P03356_3mutA
GGG	17,860	MLVCB_P08361_3mutA
GGGGG	17,861	MLVCB_P08361_3mutA
GGSGGSGGSGGS	17,862	KORV_Q9TTC1-Pro_3mutA
GSSGSSGSSGSSGSSGSS	17,863	XMRV6_A1Z651_3mutA
GSSEAAAKPAP	17,864	FLV_P10273_3mutA
GGGEAAAKPAP	17,865	MLVCB_P08361_3mutA
GSSGSSGSS	17,866	MLVMS_P03355_3mutA_WS
PAPAPAPAP	17,867	MLVMS_P03355_PLV919
EAAAKGGG	17,868	MLVMS_P03355_PLV919
PAPAPAPAPAPAP	17,869	FLV_P10273_3mutA
EAAAKGGSGSS	17,870	MLVMS_P03355_3mut
GGGGGG	17,871	PERV_Q4VFZ2_3mutA_WS
PAPGGG	17,872	MLVCB_P08361_3mutA
GGGGGSGSS	17,873	KORV_Q9TTC1_3mutA
GGGGSGGGGSGGGGSGGGGS	17,874	XMRV6_A1Z651_3mut
GGSGGSGGS	17,875	KORV_Q9TTC1-Pro_3mutA
EAAAKPAPGGG	17,876	MLVMS_P03355_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,877	XMRV6_A1Z651
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,878	FLV_P10273_3mutA
EAAAKGGGGSEAAAK	17,879	PERV_Q4VFZ2_3mutA_WS
GGGPAPGSS	17,880	AVIRE_P03360_3mutA
GGGGG	17,881	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,882	MLVMS_P03355_3mut
GGGGSGGGGS	17,883	MLVMS_P03355_3mutA_WS
EAAAKGGSPAP	17,884	XMRV6_A1Z651_3mutA
EAAAKGSSPAP	17,885	AVIRE_P03360_3mutA
PAPGGSGSS	17,886	KORV_Q9TTC1-Pro_3mutA
GSS	17,887	MLVBM_Q7SVK7_3mutA_WS
GSS	17,888	WMSV_P03359_3mut

GGGPAPGSS	17,889	MLVFF_P26809_3mutA
EAAAKPAP	17,890	MLVMS_P03355_3mut
GSSPAPEAAAK	17,891	FLV_P10273_3mutA
GGSPAPGSS	17,892	MLVBM_Q7SVK7_3mutA_WS
GGGGGSEAAAK	17,893	XMRV6_A1Z651_3mut
PAPEAAAKGGG	17,894	WMSV_P03359_3mutA
PAPGGG	17,895	PERV_Q4VFZ2_3mut
GGSPAPEAAAK	17,896	WMSV_P03359_3mutA
GGSGGGGSS	17,897	PERV_Q4VFZ2_3mut
EAAAKGGGGSS	17,898	PERV_Q4VFZ2_3mut
EAAAKGGSPAP	17,899	AVIRE_P03360_3mut
GGSGGGGSS	17,900	WMSV_P03359_3mutA
PAPGSSEAAAK	17,901	MLVFF_P26809_3mut
GSSEAAAK	17,902	MLVMS_P03355_PLV919
GSAGSAAGSGEF	17,903	AVIRE_P03360_3mutA
EAAAKGGSGSS	17,904	MLVMS_P03355_3mut
GGSEAAAKPAP	17,905	MLVMS_P03355_PLV919
GGGGSGGGGSGGGGSGGGGSGGGGS	17,906	MLVFF_P26809_3mutA
PAPGSSEAAAK	17,907	PERV_Q4VFZ2_3mutA_WS
GGGGSSPAP	17,908	MLVMS_P03355_3mutA_WS
PAPAPAP	17,909	MLVCB_P08361_3mutA
EAAAKPAPGGG	17,910	MLVBM_Q7SVK7_3mutA_WS
GGGPAPGSS	17,911	BAEVM_P10272_3mutA
PAP	17,912	MLVMS_P03355_3mutA_WS
PAPGGSGGG	17,913	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGSGGS	17,914	MLVBM_Q7SVK7_3mutA_WS
PAPAPAPAP	17,915	XMRV6_A1Z651_3mut
GSSPAPGGG	17,916	MLVMS_P03355_3mutA_WS
GSSPAPGGG	17,917	MLVMS_P03355_3mut
PAPGGG	17,918	MLVMS_P03355_PLV919
GGGEAAAKGSS	17,919	WMSV_P03359_3mut
EAAAKGSS	17,920	KORV_Q9TTC1-Pro_3mutA
EAAAKGGS	17,921	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAKEAAAK	17,922	PERV_Q4VFZ2_3mut
PAPEAAAKGGG	17,923	MLVMS_P03355_PLV919
EAAAKGSSGGG	17,924	MLVFF_P26809_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,925	PERV_Q4VFZ2
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,926	MLVAV_P03356_3mutA

GSSGGSGGG	17,927	MLVFF_P26809_3mut
GSSGSSGSSGSS	17,928	PERV_Q4VFZ2_3mutA_WS
GGSPAPGGG	17,929	MLVMS_P03355_PLV919
GSS	17,930	BAEVM_P10272_3mut
GGGPAPGSS	17,931	MLVMS_P03355_3mutA_WS
GGGGSS	17,932	KORV_Q9TTC1_3mutA
GSSGGSGGG	17,933	BAEVM_P10272_3mutA
EAAAKEAAAKEAAAK	17,934	MLVCB_P08361_3mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,935	FLV_P10273_3mutA
PAPGGGGGS	17,936	PERV_Q4VFZ2_3mut
PAPAPAPAPAP	17,937	KORV_Q9TTC1-Pro_3mutA
EAAAK	17,938	MLVMS_P03355_3mutA_WS
GGG	17,939	MLVCB_P08361_3mut
GGSEAAAKGGG	17,940	BAEVM_P10272_3mutA
GGGGGSGSS	17,941	MLVAV_P03356_3mutA
EAAAKGSSPAP	17,942	MLVBM_Q7SVK7_3mutA_WS
GGSGGSGGS	17,943	XMRV6_A1Z651_3mut
EAAAKPAPGGG	17,944	KORV_Q9TTC1-Pro_3mutA
GGGPAPEAAAK	17,945	FLV_P10273_3mutA
GGSPAPEAAAK	17,946	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGSGGS	17,947	MLVFF_P26809_3mut
EAAAKGGSGSS	17,948	MLVMS_P03355_PLV919
GGGEAAAKGGS	17,949	MLVBM_Q7SVK7_3mutA_WS
PAPAPAPAP	17,950	BAEVM_P10272_3mutA
EAAAKEAAAKEAAAKEAAAK	17,951	MLVMS_P03355_3mut
EAAAKPAP	17,952	XMRV6_A1Z651_3mut
EAAAKEAAAK	17,953	MLVBM_Q7SVK7_3mutA_WS
EAAAKGGG	17,954	BAEVM_P10272_3mut
EAAAKGSS	17,955	MLVAV_P03356_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,956	MLVFF_P26809_3mut
GGGPAPGSS	17,957	PERV_Q4VFZ2_3mutA_WS
GGGG	17,958	PERV_Q4VFZ2_3mut
EAAAKGGSGSS	17,959	MLVMS_P03355_PLV919
GGGGSGGGGSGGGGS	17,960	MLVMS_P03355_3mutA_WS
EAAAK	17,961	MLVMS_P03355_3mutA_WS
GGGGSS	17,962	PERV_Q4VFZ2
PAPEAAAKGGS	17,963	MLVCB_P08361_3mut
GSS	17,964	MLVMS_P03355_3mut
GSAGSAAGSGEF	17,965	MLVFF_P26809_3mutA

EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,966	KORV_Q9TTC1-Pro_3mut
GGGGSGGGGS	17,967	AVIRE_P03360_3mutA
EAAAK	17,968	MLVMS_P03355_3mut
GGGPAPGGS	17,969	PERV_Q4VFZ2_3mut
GGGGSGGGGSGGGGS	17,970	MLVMS_P03355_PLV919
PAPGGG	17,971	MLVMS_P03355_3mutA_WS
GGGEAAAKPAP	17,972	PERV_Q4VFZ2_3mutA_WS
EAAAKPAPGSS	17,973	KORV_Q9TTC1-Pro_3mutA
PAPGSS	17,974	KORV_Q9TTC1_3mutA
GSAGSAAGSGEF	17,975	PERV_Q4VFZ2_3mut
PAPGGGGSS	17,976	KORV_Q9TTC1-Pro_3mutA
GSSGGGEAAAK	17,977	MLVCB_P08361_3mutA
GSS	17,978	AVIRE_P03360_3mutA
GSSGSSGSSGSS	17,979	XMRV6_A1Z651_3mutA
PAPEAAAKGGG	17,980	MLVMS_P03355_PLV919
GGGPAPEAAAK	17,981	MLVCB_P08361_3mutA
PAPGGGGGS	17,982	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAK	17,983	PERV_Q4VFZ2_3mutA_WS
GGGGGSPAP	17,984	MLVFF_P26809_3mutA
GSSGSSGSSGSSGSS	17,985	PERV_Q4VFZ2
GSSPAPEAAAK	17,986	MLVMS_P03355_PLV919
GSSGSSGSSGSSGSSGSS	17,987	MLVBM_Q7SVK7_3mutA_WS
GSSGSSGSSGSSGSSGSS	17,988	MLVMS_P03355_3mutA_WS
GGSPAPEAAAK	17,989	MLVAV_P03356_3mutA
GSSGGG	17,990	BAEVM_P10272_3mut
EAAAKGSSGGS	17,991	KORV_Q9TTC1-Pro_3mutA
GGSGSSEAAAK	17,992	MLVMS_P03355_3mutA_WS
GGGPAPEAAAK	17,993	MLVFF_P26809_3mutA
GGGPAPGGS	17,994	MLVMS_P03355_3mutA_WS
GGGGG	17,995	MLVMS_P03355_PLV919
GGGEAAAKPAP	17,996	MLVBM_Q7SVK7_3mutA_WS
GGGGSGGGGS	17,997	WMSV_P03359_3mut
GGGPAPEAAAK	17,998	PERV_Q4VFZ2_3mut
GGSGSSEAAAK	17,999	MLVMS_P03355_PLV919
EAAAKGGGPAP	18,000	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSSGSS	18,001	KORV_Q9TTC1-Pro_3mutA
PAPAP	18,002	WMSV_P03359_3mutA
GGSPAPGSS	18,003	MLVAV_P03356_3mutA
GGSGGGPAP	18,004	MLVMS_P03355_3mut
GGSPAP	18,005	MLVMS_P03355_PLV919

EAAAKGGSPAP	18,006	PERV_Q4VFZ2_3mut
GSSPAPGGG	18,007	KORV_Q9TTC1-Pro_3mutA
GSAGSAAGSGEF	18,008	MLVMS_P03355_3mut
GGSPAP	18,009	PERV_Q4VFZ2_3mut
GSSGSS	18,010	KORV_Q9TTC1-Pro_3mut
GGGPAPGSS	18,011	MLVMS_P03355_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	18,012	FOAMV_P14350
PAPGSSGGG	18,013	MLVMS_P03355_PLV919
GGSEAAAKPAP	18,014	BAEVM_P10272_3mutA
GGGGGS	18,015	MLVCB_P08361_3mutA
PAPEAAAKGGS	18,016	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	18,017	BAEVM_P10272_3mutA
GGSEAAAK	18,018	BAEVM_P10272_3mutA
GSSPAPEAAAK	18,019	MLVMS_P03355_3mutA_WS
PAPGGG	18,020	WMSV_P03359_3mut
EAAAKPAP	18,021	PERV_Q4VFZ2_3mut
GSSGSSGSSGSSGSS	18,022	WMSV_P03359_3mut
PAPGGG	18,023	MLVBM_Q7SVK7_3mutA_WS
GGSGGGEAAAK	18,024	BAEVM_P10272_3mutA
PAPGGS	18,025	MLVMS_P03355_3mut
GGSGGSGGSGGS	18,026	MLVBM_Q7SVK7_3mutA_WS
EAAAKEAAAKEAAAKEAAAK	18,027	PERV_Q4VFZ2_3mut
GGSEAAAKGGG	18,028	WMSV_P03359_3mutA
GGGPAP	18,029	BAEVM_P10272_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	18,030	XMRV6_A1Z651_3mut
GGSPAPGSS	18,031	KORV_Q9TTC1_3mut
GGGPAPGSS	18,032	MLVMS_P03355_3mut
GGGGSSGGS	18,033	BAEVM_P10272_3mutA
GGGEAAAKGSS	18,034	KORV_Q9TTC1-Pro_3mutA
PAPAP	18,035	MLVBM_Q7SVK7_3mutA_WS
GGSPAPGGG	18,036	PERV_Q4VFZ2_3mut
PAPGSS	18,037	PERV_Q4VFZ2_3mutA_WS
GSSGGSPAP	18,038	MLVBM_Q7SVK7_3mutA_WS
EAAAKGGGGSEAAAK	18,039	PERV_Q4VFZ2_3mut
GSSEAAAKGGS	18,040	KORV_Q9TTC1-Pro_3mut
PAPAPAPAP	18,041	KORV_Q9TTC1-Pro_3mutA
GGSEAAAKPAP	18,042	WMSV_P03359_3mutA

PAPGGS	18,043	FLV_P10273_3mutA
EAAAKGGGPAP	18,044	PERV_Q4VFZ2_3mut
GGSGSSGGG	18,045	AVIRE_P03360_3mutA
EAAAKGGSGSS	18,046	BAEVM_P10272_3mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	18,047	MLVCB_P08361_3mutA
GSSEAAAKGGS	18,048	XMRV6_A1Z651_3mutA
GGGGG	18,049	BAEVM_P10272_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	18,050	SFV3L_P27401_2mutA
GGGEAAAKGSS	18,051	MLVMS_P03355_PLV919
EAAAKGGGGSEAAAK	18,052	KORV_Q9TTC1_3mutA
EAAAKGGG	18,053	AVIRE_P03360_3mut
GGSGGG	18,054	MLVMS_P03355_3mutA_WS
GGSGSSGGG	18,055	MLVMS_P03355_PLV919
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	18,056	KORV_Q9TTC1_3mut
GGGGSEAAAKGGGGS	18,057	KORV_Q9TTC1_3mutA
PAPAPAPAPAP	18,058	FLV_P10273_3mutA
GGS	18,059	MLVBM_Q7SVK7_3mutA_WS
GGGGGSEAAAK	18,060	MLVBM_Q7SVK7_3mutA_WS
GSSGSSGSSGSSGSS	18,061	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAK	18,062	MLVMS_P03355_3mut
GGSGSSGGG	18,063	PERV_Q4VFZ2_3mut
PAP	18,064	MLVFF_P26809_3mut
GSSPAPEAAAK	18,065	MLVAV_P03356_3mutA
EAAAKGGGGSS	18,066	MLVMS_P03355_3mut
GGGEAAAKGGS	18,067	XMRV6_A1Z651_3mut
GGSGGGPAP	18,068	MLVBM_Q7SVK7_3mutA_WS
GSAGSAAGSGEF	18,069	BAEVM_P10272_3mutA
GSSEAAAK	18,070	MLVCB_P08361_3mut
PAPGSS	18,071	MLVMS_P03355_3mut
EAAAKEAAAKEAAAK	18,072	MLVAV_P03356_3mutA
GSAGSAAGSGEF	18,073	XMRV6_A1Z651_3mutA
GSSGSSGSSGSS	18,074	BAEVM_P10272_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	18,075	KORV_Q9TTC1-Pro_3mut
GGGGSSEAAAK	18,076	WMSV_P03359_3mut
GSSGGGEAAAK	18,077	MLVBM_Q7SVK7_3mutA_WS
EAAAKPAP	18,078	MLVFF_P26809_3mutA
GGSPAPGGG	18,079	KORV_Q9TTC1_3mutA

PAPEAAAK	18,080	FLV_P10273_3mutA
GSSGSSGSS	18,081	MLVBM_Q7SVK7_3mutA_WS
GSSGGGEAAAK	18,082	FLV_P10273_3mutA
GGSPAP	18,083	MLVBM_Q7SVK7_3mutA_WS
GSAGSAAGSGEF	18,084	KORV_Q9TTC1-Pro_3mutA
PAPGGSEAAAK	18,085	MLVMS_P03355_PLV919
GGSPAPEAAAK	18,086	MLVBM_Q7SVK7_3mutA_WS
GGGGGSPAP	18,087	MLVBM_Q7SVK7_3mutA_WS
EAAAKGSSPAP	18,088	WMSV_P03359_3mut
EAAAKGGGPAP	18,089	MLVBM_Q7SVK7_3mutA_WS
PAPGSS	18,090	KORV_Q9TTC1-Pro_3mutA
GGSGSSGGG	18,091	BAEVM_P10272_3mut
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	16,901	FFV_O93209-Pro_2mut
GGSGGSGGSGGSGGSGGS	16,902	WMSV_P03359_3mutA
GGSGGSGGS	16,903	PERV_Q4VFZ2_3mutA_WS
GGGGG	16,904	PERV_Q4VFZ2_3mutA_WS
GGGPAP	16,905	FLV_P10273_3mutA
PAPGGSGGG	16,906	XMRV6_A1Z651_3mutA
GGGGSEAAAKGGGGS	16,907	XMRV6_A1Z651_3mut
EAAAKGSSGGG	16,908	KORV_Q9TTC1-Pro_3mutA
GSSGGSEAAAK	16,909	WMSV_P03359_3mut
EAAAKGGSGSS	16,910	PERV_Q4VFZ2_3mut
PAPAPAPAPAP	16,911	PERV_Q4VFZ2_3mut
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	16,912	MLVMS_P03355_3mutA_WS
GGGGGGG	16,913	KORV_Q9TTC1_3mutA
EAAAK	16,914	KORV_Q9TTC1-Pro_3mutA
GGGEAAAKGGS	16,915	KORV_Q9TTC1-Pro_3mutA
GGGEAAAKGGS	16,916	PERV_Q4VFZ2_3mutA_WS
GGGGGSPAP	16,917	XMRV6_A1Z651_3mut
GGGGSGGGGSGGGGSGGGGS	16,918	MLVFF_P26809_3mut
GGGGGGG	16,919	MLVFF_P26809_3mut
PAPAPAPAPAP	16,920	AVIRE_P03360_3mutA
GSSPAPGGG	16,921	FLV_P10273_3mutA
GGGGGSPAP	16,922	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGS	16,923	MLVMS_P03355_3mut
GGGGSGGGGSGGGGS	16,924	KORV_Q9TTC1_3mut
GSSEAAAKGGS	16,925	MLVAV_P03356_3mutA

GSSGSSGSSGSSGSS	16,926	MLVMS_P03355_3mut
EAAAKGGGGGS	16,927	PERV_Q4VFZ2_3mutA_WS
GSSGGGGGS	16,928	PERV_Q4VFZ2_3mut
GGGEAAAKPAP	16,929	MLVMS_P03355_3mut
GSSGGSPAP	16,930	PERV_Q4VFZ2_3mutA_WS
GSSGGGPAP	16,931	BAEVM_P10272_3mutA
GGGGGSGSS	16,932	MLVMS_P03355_PLV919
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	16,933	BAEVM_P10272_3mut
PAPEAAAK	16,934	MLVMS_P03355_3mut
GGGGSGGGGSGGGGS	16,935	FLV_P10273_3mutA
GGSGSSGGG	16,936	WMSV_P03359_3mutA
EAAAKGGS	16,937	PERV_Q4VFZ2_3mut
EAAAKGSSPAP	16,938	MLVCB_P08361_3mut
EAAAKGGSGSS	16,939	WMSV_P03359_3mutA
GSSGSS	16,940	PERV_Q4VFZ2_3mutA_WS
PAPAPAPAP	16,941	MLVMS_P03355_PLV919
GGSGGG	16,942	PERV_Q4VFZ2_3mutA_WS
GSS	16,943	MLVBM_Q7SVK7_3mutA_WS
PAP	16,944	KORV_Q9TTC1-Pro_3mutA
GGSGSSEAAAK	16,945	MLVFF_P26809_3mut
PAPEAAAKGSS	16,946	KORV_Q9TTC1-Pro_3mutA
GGSGGS	16,947	MLVCB_P08361_3mutA
GGGGGGG	16,948	PERV_Q4VFZ2_3mutA_WS
GGSPAPEAAAK	16,949	MLVBM_Q7SVK7_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	16,950	KORV_Q9TTC1_3mutA
GGSPAP	16,951	MLVMS_P03355_3mut
GGSEAAAKGGG	16,952	PERV_Q4VFZ2_3mut
GGGGSGGGGS	16,953	FLV_P10273_3mutA
GGGEAAAK	16,954	BAEVM_P10272_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	16,955	SFV3L_P27401_2mut
GGSEAAAKPAP	16,956	KORV_Q9TTC1-Pro_3mutA
GSSGGGEAAAK	16,957	MLVMS_P03355_PLV919
GGGGGSEAAAK	16,958	MLVMS_P03355_PLV919
EAAAKGGSGGG	16,959	MLVMS_P03355_3mutA_WS
GGGGSSPAP	16,960	MLVAV_P03356_3mutA
EAAAKEAAAK	16,961	MLVMS_P03355_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	16,962	SFV3L_P27401_2mut
GSSGSSGSSGSSGSS	16,963	MLVMS_P03355_PLV919
GSSGGG	16,964	KORV_Q9TTC1-Pro_3mutA

GSSGGS	16,965	MLVFF_P26809_3mutA
GGGGSGGGGS	16,966	XMRV6_A1Z651_3mutA
PAPGSS	16,967	MLVBM_Q7SVK7_3mutA_WS
GGGPAPEAAAK	16,968	XMRV6_A1Z651_3mutA
EAAAKGGS	16,969	MLVFF_P26809_3mut
GSS	16,970	KORV_Q9TTC1_3mutA
GGGG	16,971	PERV_Q4VFZ2_3mut
GGGGGSEAAAK	16,972	AVIRE_P03360_3mutA
GSSGSSGSSGSSGSS	16,973	MLVMS_P03355_PLV919
PAPGGSGGG	16,974	PERV_Q4VFZ2_3mut
GGGPAP	16,975	PERV_Q4VFZ2_3mut
GGGPAPEAAAK	16,976	AVIRE_P03360_3mutA
GGGEAAAK	16,977	MLVCB_P08361_3mut
GGG	16,978	MLVFF_P26809_3mutA
EAAAKPAPGSS	16,979	XMRV6_A1Z651_3mutA
GGSGSSEAAAK	16,980	PERV_Q4VFZ2_3mutA_WS
EAAAKGSS	16,981	MLVMS_P03355_3mut
GGSGSSEAAAK	16,982	BAEVM_P10272_3mut
GGSGGG	16,983	MLVBM_Q7SVK7_3mutA_WS
GGGPAP	16,984	MLVMS_P03355_PLV919
GGSPAPGGG	16,985	PERV_Q4VFZ2_3mutA_WS
GGGGGSEAAAK	16,986	MLVFF_P26809_3mutA
EAAAKGSSGGS	16,987	MLVBM_Q7SVK7_3mut
PAPAP	16,988	XMRV6_A1Z651_3mut
GSSPAPGGS	16,989	MLVBM_Q7SVK7_3mutA_WS
GSSEAAAKGGG	16,990	WMSV_P03359_3mutA
EAAAKGGGGGS	16,991	PERV_Q4VFZ2_3mut
GSSGSSGSSGSSGSS	16,992	MLVCB_P08361_3mutA
EAAAKGGGGSS	16,993	PERV_Q4VFZ2_3mut
EAAAKGSS	16,994	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	16,995	AVIRE_P03360_3mutA
EAAAKGGS	16,996	MLVCB_P08361_3mut
GSSGGSEAAAK	16,997	MLVAV_P03356_3mutA
EAAAKPAPGGS	16,998	PERV_Q4VFZ2_3mut
GGSGGS	16,999	MLVAV_P03356_3mutA
EAAAKGSSGGG	17,000	AVIRE_P03360_3mutA
GGSGGSGGSGGS	17,001	PERV_Q4VFZ2_3mut
GGGGGGGG	17,002	KORV_Q9TTC1_3mutA
GGSGSSEAAAK	17,003	MLVCB_P08361_3mutA

EAAAKGGG	17,004	MLVBM_Q7SVK7_3mutA_WS
GGGGSGGGGSGGGGS	17,005	MLVCB_P08361_3mut
GGSGGSGGSGGS	17,006	PERV_Q4VFZ2_3mutA_WS
PAPAPAPAPAP	17,007	WMSV_P03359_3mut
EAAAKEAAAKEAAAKEAAAK	17,008	PERV_Q4VFZ2_3mut
GGSGGSGGS	17,009	XMRV6_A1Z651_3mutA
PAPGGGGSS	17,010	BAEVM_P10272_3mutA
GSSEAAAKGGS	17,011	MLVCB_P08361_3mut
GSSGGGPAP	17,012	MLVCB_P08361_3mutA
GGSGSS	17,013	MLVBM_Q7SVK7_3mutA_WS
GGGGGSEAAAK	17,014	MLVAV_P03356_3mutA
GSSEAAAK	17,015	PERV_Q4VFZ2_3mutA_WS
GGGGGSGSS	17,016	MLVBM_Q7SVK7_3mutA_WS
EAAAKGGSGSS	17,017	MLVFF_P26809_3mut
PAP	17,018	FLV_P10273_3mutA
GGGGG	17,019	MLVMS_P03355_3mutA_WS
EAAAK	17,020	PERV_Q4VFZ2_3mut
GSS	17,021	FLV_P10273_3mutA
PAPAPAPAPAPAP	17,022	KORV_Q9TTC1-Pro_3mutA
EAAAKEAAAKEAAAKEAAAK	17,023	MLVCB_P08361_3mut
EAAAKGGGGSEAAAK	17,024	XMRV6_A1Z651_3mut
PAPGGSGGG	17,025	MLVBM_Q7SVK7_3mutA_WS
GGSGGGPAP	17,026	WMSV_P03359_3mutA
GGGGSSEAAAK	17,027	MLVBM_Q7SVK7_3mutA_WS
PAPGGGGSS	17,028	MLVCB_P08361_3mut
GGSGGSGGSGGS	17,029	PERV_Q4VFZ2_3mutA_WS
PAPGGSGGG	17,030	MLVMS_P03355_3mutA_WS
GSSPAPGGS	17,031	MLVCB_P08361_3mutA
GSSGSSGSS	17,032	MLVFF_P26809_3mut
PAPGGGGGS	17,033	MLVBM_Q7SVK7_3mutA_WS
GSSPAP	17,034	PERV_Q4VFZ2_3mut
GGSGGG	17,035	KORV_Q9TTC1-Pro_3mut
EAAAKGGGGSEAAAK	17,036	PERV_Q4VFZ2_3mutA_WS
GGSPAPEAAAK	17,037	PERV_Q4VFZ2_3mutA_WS
EAAAKPAP	17,038	BAEVM_P10272_3mut
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,039	MLVMS_P03355_3mut
EAAAKGGGGSS	17,040	MLVFF_P26809_3mut

EAAAKEAAAK	17,041	MLVCB_P08361_3mut
GSSEAAAKGGS	17,042	PERV_Q4VFZ2_3mut
GGSPAP	17,043	KORV_Q9TTC1-Pro_3mutA
EAAAKEAAAKEAAAKEAAAK	17,044	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSSGSS	17,045	BAEVM_P10272_3mut
PAPEAAAK	17,046	MLVMS_P03355_3mut
GSSGGSPAP	17,047	PERV_Q4VFZ2
GGGPAPGGS	17,048	BAEVM_P10272_3mutA
EAAAKPAPGGS	17,049	MLVMS_P03355_PLV919
GGGGSGGGGS	17,050	PERV_Q4VFZ2
GGGEAAAK	17,051	KORV_Q9TTC1-Pro_3mut
EAAAKGGGGGS	17,052	FLV_P10273_3mutA
GGSPAPGSS	17,053	MLVMS_P03355_3mut
GSSPAPEAAAK	17,054	MLVMS_P03355_3mutA_WS
GSAGSAAGSGEF	17,055	MLVBM_Q7SVK7_3mutA_WS
EAAAK	17,056	BAEVM_P10272_3mutA
EAAAKGGGGSS	17,057	BAEVM_P10272_3mutA
GGG	17,058	WMSV_P03359_3mut
GGSGSSPAP	17,059	BAEVM_P10272_3mut
GGSEAAAKPAP	17,060	MLVBM_Q7SVK7_3mutA_WS
EAAAKGGSGSS	17,061	MLVCB_P08361_3mut
PAPGSS	17,062	MLVAV_P03356_3mutA
PAPEAAAKGGG	17,063	MLVCB_P08361_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,064	FOAMV_P14350-Pro_2mut
GSSGSSGSS	17,065	PERV_Q4VFZ2_3mut
PAPGGG	17,066	MLVMS_P03355_3mut
PAPGGS	17,067	PERV_Q4VFZ2_3mut
GSSGGG	17,068	MLVMS_P03355_PLV919
GSSGSSGSSGSSGSSGSS	17,069	WMSV_P03359_3mut
PAP	17,070	AVIRE_P03360_3mutA
EAAAKGSSPAP	17,071	MLVBM_Q7SVK7_3mutA_WS
GSSGSSGSSGSS	17,072	MLVMS_P03355_PLV919
GGGGSGGGGSGGGGSGGGGSGGGGS	17,073	AVIRE_P03360
GGGGS	17,074	PERV_Q4VFZ2_3mut
EAAAKGSSGGG	17,075	MLVBM_Q7SVK7_3mutA_WS
GGGGGG	17,076	KORV_Q9TTC1-Pro_3mut
GGSGSSEAAAK	17,077	PERV_Q4VFZ2_3mut

GSSPAPEAAAK	17,078	MLVBM_Q7SVK7_3mutA_WS
GGGGSGGGGS	17,079	MLVBM_Q7SVK7_3mutA_WS
GSSGGGGGS	17,080	MLVAV_P03356_3mutA
GSAGSAAGSGEF	17,081	WMSV_P03359_3mutA
GGGEAAAKGSS	17,082	BAEVM_P10272_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,083	FFV_O93209-Pro_2mut
PAPGGSGGG	17,084	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	17,085	SFV3L_P27401_2mut
GGSGSSPAP	17,086	MLVMS_P03355_PLV919
GGGGGG	17,087	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAKEAAAK	17,088	PERV_Q4VFZ2_3mut
EAAAKGSSPAP	17,089	MLVFF_P26809_3mut
GGGPAPGGS	17,090	MLVBM_Q7SVK7_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,091	SFV3L_P27401
PAP	17,092	PERV_Q4VFZ2_3mut
EAAAKGGS	17,093	MLVMS_P03355_PLV919
GSSGGSEAAAK	17,094	WMSV_P03359_3mutA
GGSGSSEAAAK	17,095	KORV_Q9TTC1-Pro_3mutA
EAAAKEAAAKEAAAK	17,096	PERV_Q4VFZ2
GGSGGGEAAAK	17,097	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGSGGGGS	17,098	BAEVM_P10272_3mut
EAAAKGSS	17,099	XMRV6_A1Z651_3mutA
GSSGGGGGS	17,100	WMSV_P03359_3mutA
GSSGSSGSSGSSGSSGSS	17,101	MLVFF_P26809_3mutA
GGSGSS	17,102	MLVAV_P03356_3mutA
EAAAKGGGGSEAAAK	17,103	MLVMS_P03355_PLV919
EAAAKGGGPAP	17,104	PERV_Q4VFZ2
GGSEAAAKGGG	17,105	MLVAV_P03356_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,106	MLVBM_Q7SVK7_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,107	KORV_Q9TTC1-Pro_3mutA
GSSPAPEAAAK	17,108	MLVFF_P26809_3mutA
GGGGSEAAAKGGGGS	17,109	PERV_Q4VFZ2_3mut
GSSGSSGSSGSS	17,110	PERV_Q4VFZ2_3mut
GGSEAAAK	17,111	MLVFF_P26809_3mutA
GGGGGGGG	17,112	MLVMS_P03355_3mut
GSSGGG	17,113	XMRV6_A1Z651_3mutA
EAAAKGGS	17,114	BAEVM_P10272_3mutA

GGGGS	17,115	BAEVM_P10272_3mutA
GGSEAAAKGGG	17,116	KORV_Q9TTC1-Pro_3mutA
GGSGSSGGG	17,117	KORV_Q9TTC1_3mutA
GGSGSSEAAAK	17,118	WMSV_P03359_3mut
EAAAKGGSGSS	17,119	MLVBM_Q7SVK7_3mutA_WS
GGS	17,120	BAEVM_P10272_3mutA
GGGPAPGSS	17,121	WMSV_P03359_3mutA
GSSGSSGSSGSSGSS	17,122	AVIRE_P03360_3mut
GGGEAAAKPAP	17,123	XMRV6_A1Z651_3mut
GSSGGG	17,124	MLVFF_P26809_3mutA
GGSPAPGSS	17,125	PERV_Q4VFZ2_3mut
PAPGGS	17,126	MLVCB_P08361_3mut
PAPAPAPAPAP	17,127	KORV_Q9TTC1_3mutA
GSSGGS	17,128	MLVCB_P08361_3mutA
GSSGGSEAAAK	17,129	PERV_Q4VFZ2_3mut
EAAAKGSSGGS	17,130	MLVMS_P03355_PLV919
EAAAKGGG	17,131	WMSV_P03359_3mut
PAPGGGGGS	17,132	BAEVM_P10272_3mutA
GGGGSEAAAKGGGGS	17,133	WMSV_P03359_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,134	MLVMS_P03355_3mutA_WS
GGS	17,135	KORV_Q9TTC1-Pro_3mutA
GSSGGSPAP	17,136	BAEVM_P10272_3mutA
GGG	17,137	MLVMS_P03355_PLV919
PAPGSS	17,138	KORV_Q9TTC1-Pro_3mut
GGSEAAAKGGG	17,139	FLV_P10273_3mutA
GGSEAAAKPAP	17,140	PERV_Q4VFZ2_3mutA_WS
GGGGSSPAP	17,141	XMRV6_A1Z651_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	17,142	PERV_Q4VFZ2_3mutA_WS
GGGG	17,143	PERV_Q4VFZ2_3mutA_WS
GGSEAAAKPAP	17,144	MLVMS_P03355_3mut
PAPGSSGGG	17,145	MLVMS_P03355_3mutA_WS
PAPEAAAKGGS	17,146	AVIRE_P03360_3mut
GGGGSSPAP	17,147	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGSGGGGS	17,148	PERV_Q4VFZ2_3mut
GGGEAAAK	17,149	MLVMS_P03355_3mut
GGGGSS	17,150	MLVFF_P26809_3mut
GGSPAPGSS	17,151	XMRV6_A1Z651_3mut
GGGGS	17,152	KORV_Q9TTC1-Pro_3mutA
EAAAKGSSGGS	17,153	FLV_P10273_3mutA
GSS	17,154	MLVMS_P03355_PLV919
GGGG	17,155	MLVMS_P03355_PLV919

GSSGGS	17,156	MLVMS_P03355_PLV919
GGSGGSGGSGGS	17,157	MLVMS_P03355_3mut
PAPEAAAKGGS	17,158	MLVMS_P03355_3mut
EAAAKGSSGGG	17,159	BAEVM_P10272_3mutA
GSSEAAAK	17,160	KORV_Q9TTC1-Pro_3mutA
GSAGSAAGSGEF	17,161	KORV_Q9TTC1_3mutA
GGGGGSEAAAK	17,162	MLVCB_P08361_3mut
GGGG	17,163	WMSV_P03359_3mut
GGGGSSEAAAK	17,164	MLVMS_P03355_PLV919
PAPGGG	17,165	WMSV_P03359_3mutA
EAAAKGGSGGG	17,166	MLVAV_P03356_3mutA
GGGPAPGGS	17,167	MLVMS_P03355_3mut
EAAAKPAP	17,168	PERV_Q4VFZ2_3mutA_WS
GSSGSSGSS	17,169	KORV_Q9TTC1-Pro_3mutA
GSSPAPGGS	17,170	XMRV6_A1Z651_3mut
GGGGGSPAP	17,171	BAEVM_P10272_3mutA
GGSGSSGGG	17,172	PERV_Q4VFZ2_3mutA_WS
GGGEAAAKGSS	17,173	AVIRE_P03360_3mut
GSSEAAAK	17,174	FLV_P10273_3mutA
EAAAK	17,175	MLVMS_P03355_3mut
EAAAKGGSGSS	17,176	WMSV_P03359_3mut
GSSEAAAKGGG	17,177	PERV_Q4VFZ2_3mut
PAPGSSGGG	17,178	BAEVM_P10272_3mutA
EAAAKGGGGGS	17,179	MLVMS_P03355_3mut
GGSEAAAKPAP	17,180	AVIRE_P03360_3mut
GGGPAPGGS	17,181	XMRV6_A1Z651_3mut
GGGGS	17,182	KORV_Q9TTC1_3mutA
GGSGGSGGSGGSGGS	17,183	XMRV6_A1Z651_3mut
GGGPAP	17,184	KORV_Q9TTC1-Pro_3mut
EAAAKPAP	17,185	MLVBM_Q7SVK7_3mutA_WS
GGSEAAAK	17,186	MLVMS_P03355_PLV919
GSSEAAAKPAP	17,187	KORV_Q9TTC1-Pro_3mutA
GGSGSS	17,188	MLVMS_P03355_3mut
EAAAKPAPGGG	17,189	PERV_Q4VFZ2_3mut
GGSPAPEAAAK	17,190	KORV_Q9TTC1_3mutA
GGSEAAAKGGG	17,191	AVIRE_P03360_3mutA
GGGGSEAAAKGGGGS	17,192	MLVMS_P03355_PLV919
GSSGGGEAAAK	17,193	KORV_Q9TTC1-Pro_3mutA
EAAAKGGGPAP	17,194	WMSV_P03359_3mut
GSSPAP	17,195	XMRV6_A1Z651_3mutA

AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,196	SFV3L_P27401-Pro
GGSEAAAKGSS	17,197	MLVMS_P03355_PLV919
GSSGGSEAAAK	17,198	KORV_Q9TTC1-Pro_3mutA
GGSEAAAKGSS	17,199	KORV_Q9TTC1-Pro_3mutA
EAAAKGGG	17,200	AVIRE_P03360_3mutA
GSSGGSEAAAK	17,201	BAEVM_P10272_3mutA
GGGGSEAAAKGGGGS	17,202	KORV_Q9TTC1-Pro_3mut
PAPGSSEAAAK	17,203	MLVMS_P03355_3mut
PAPEAAAK	17,204	WMSV_P03359_3mut
PAPGGSGSS	17,205	PERV_Q4VFZ2_3mutA_WS
PAPGSS	17,206	BAEVM_P10272_3mut
PAPGGGGGS	17,207	MLVMS_P03355_3mut
EAAAKPAPGSS	17,208	MLVBM_Q7SVK7_3mutA_WS
GSSPAPGGS	17,209	MLVMS_P03355_PLV919
GGSGSSEAAAK	17,210	MLVMS_P03355_3mut
GGGGGG	17,211	KORV_Q9TTC1-Pro_3mutA
EAAAKEAAAKEAAAKEAAAK	17,212	MLVBM_Q7SVK7_3mut
GGSPAPGSS	17,213	MLVMS_P03355_PLV919
PAPAPAPAPAP	17,214	MLVCB_P08361_3mut
GGSGSSPAP	17,215	WMSV_P03359_3mutA
EAAAKGGSGGG	17,216	PERV_Q4VFZ2_3mutA_WS
GSSGSSGSSGSSGSS	17,217	PERV_Q4VFZ2_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,218	KORV_Q9TTC1_3mutA
GSSGGGEAAAK	17,219	WMSV_P03359_3mutA
GSSGGSEAAAK	17,220	FLV_P10273_3mutA
GGGGGGGG	17,221	PERV_Q4VFZ2_3mut
PAPGGSEAAAK	17,222	FLV_P10273_3mutA
GGGGSSPAP	17,223	BAEVM_P10272_3mutA
PAPAPAPAP	17,224	WMSV_P03359_3mut
GGSEAAAKPAP	17,225	PERV_Q4VFZ2_3mut
PAPGGSGGG	17,226	BAEVM_P10272_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,227	MLVMS_P03355_3mut
GGGGSGGGGSGGGGS	17,228	PERV_Q4VFZ2_3mut
GGSGGGPAP	17,229	PERV_Q4VFZ2_3mut
GGGPAPEAAAK	17,230	MLVFF_P26809_3mut
GGGGGSGSS	17,231	MLVMS_P03355_3mutA_WS
GSS	17,232	MLVCB_P08361_3mut
GGGGGSPAP	17,233	MLVMS_P03355_PLV919
GGSPAP	17,234	MLVAV_P03356_3mutA

GGGPAPGGS	17,235	KORV_Q9TTC1-Pro_3mutA
PAPGSSGGG	17,236	FLV_P10273_3mutA
PAPGSSGGG	17,237	WMSV_P03359_3mutA
PAPGGS	17,238	MLVBM_Q7SVK7_3mutA_WS
GGGEAAAKGSS	17,239	PERV_Q4VFZ2_3mutA_WS
GGSEAAAKGSS	17,240	MLVBM_Q7SVK7_3mutA_WS
PAPGGSEAAAK	17,241	MLVCB_P08361_3mut
GGSEAAAKGGG	17,242	XMRV6_A1Z651_3mutA
GGSGGGGSS	17,243	WMSV_P03359_3mut
GGGEAAAKPAP	17,244	KORV_Q9TTC1_3mutA
EAAAKGSS	17,245	KORV_Q9TTC1-Pro_3mut
PAPEAAAKGSS	17,246	MLVFF_P26809_3mut
GSAGSAAGSGEF	17,247	PERV_Q4VFZ2_3mut
EAAAKGGGGGS	17,248	WMSV_P03359_3mut
EAAAKGSSPAP	17,249	WMSV_P03359_3mutA
GGGGSEAAAKGGGGS	17,250	XMRV6_A1Z651_3mutA
GSSEAAAKPAP	17,251	SFV3L_P27401-Pro_2mutA
GGGGGG	17,252	PERV_Q4VFZ2_3mutA_WS
PAPGGS	17,253	BAEVM_P10272_3mut
PAP	17,254	AVIRE_P03360_3mut
PAPAPAP	17,255	MLVBM_Q7SVK7_3mutA_WS
GGGG	17,256	PERV_Q4VFZ2_3mutA_WS
GSSGGSEAAAK	17,257	MLVBM_Q7SVK7_3mut
GGSGGGGSS	17,258	MLVFF_P26809_3mut
GGGGSSGGS	17,259	AVIRE_P03360_3mutA
GSSPAPGGG	17,260	PERV_Q4VFZ2_3mutA_WS
GGSEAAAKPAP	17,261	MLVMS_P03355_PLV919
PAP	17,262	KORV_Q9TTC1-Pro_3mut
GSSGGS	17,263	PERV_Q4VFZ2_3mut
GGGGG	17,264	PERV_Q4VFZ2_3mut
GSSGGGPAP	17,265	FLV_P10273_3mutA
GSSEAAAKGGG	17,266	KORV_Q9TTC1-Pro_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,267	MLVCB_P08361_3mut
GGSEAAAKPAP	17,268	MLVCB_P08361_3mut
PAPAPAPAPAPAP	17,269	BAEVM_P10272_3mutA
GGGGSEAAAKGGGGS	17,270	MLVMS_P03355_3mut
EAAAKPAPGSS	17,271	MLVMS_P03355_3mut
GSSGSSGSSGSSGSS	17,272	MLVBM_Q7SVK7_3mutA_WS

PAPEAAAKGSS	17,273	MLVAV_P03356_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,274	AVIRE_P03360_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,275	PERV_Q4VFZ2_3mut
GGSEAAAKGGG	17,276	PERV_Q4VFZ2_3mutA_WS
GGSGGGGSS	17,277	MLVFF_P26809_3mutA
PAPEAAAKGSS	17,278	MLVCB_P08361_3mut
GGG	17,279	PERV_Q4VFZ2_3mutA_WS
GGSGGGEAAAK	17,280	MLVMS_P03355_3mut
EAAAKGGGGSS	17,281	WMSV_P03359_3mut
GSSPAPGGG	17,282	WMSV_P03359_3mutA
EAAAKGSSGGG	17,283	PERV_Q4VFZ2_3mut
GGSGGGEAAAK	17,284	PERV_Q4VFZ2_3mutA_WS
GGSGGSGGSGGSGGS	17,285	PERV_Q4VFZ2_3mutA_WS
EAAAKPAPGGS	17,286	PERV_Q4VFZ2_3mutA_WS
GGGGGSEAAAK	17,287	PERV_Q4VFZ2_3mutA_WS
GSSPAP	17,288	MLVFF_P26809_3mut
GGGEAAAKPAP	17,289	AVIRE_P03360_3mut
GSSGGSEAAAK	17,290	MLVMS_P03355_PLV919
EAAAKPAPGGS	17,291	WMSV_P03359_3mutA
PAPGGG	17,292	KORV_Q9TTC1_3mutA
EAAAKGSSPAP	17,293	KORV_Q9TTC1-Pro_3mut
GSSPAPEAAAK	17,294	MLVFF_P26809_3mut
GGSGGGEAAAK	17,295	MLVFF_P26809_3mutA
GSSGSSGSS	17,296	WMSV_P03359_3mutA
EAAAKGGS	17,297	BAEVM_P10272_3mut
EAAAKPAPGGS	17,298	KORV_Q9TTC1_3mutA
EAAAKPAPGGS	17,299	BAEVM_P10272_3mutA
GSSGGGGGS	17,300	PERV_Q4VFZ2_3mut
PAPGGGGSS	17,301	PERV_Q4VFZ2_3mut
GSSGSSGSS	17,302	WMSV_P03359_3mut
EAAAKEAAAKEAAAKEAAAK	17,303	WMSV_P03359_3mut
GGS	17,304	AVIRE_P03360_3mut
EAAAKPAPGSS	17,305	MLVFF_P26809_3mut
EAAAKGGG	17,306	KORV_Q9TTC1_3mut
PAPGSSEAAAK	17,307	MLVMS_P03355_3mut
PAPGSSGGS	17,308	MLVMS_P03355_PLV919
GSSPAPEAAAK	17,309	MLVMS_P03355_3mut
GSSGSSGSSGSSGSSGSS	17,310	WMSV_P03359_3mutA
GGGGS	17,311	BAEVM_P10272_3mut
GSSPAP	17,312	MLVMS_P03355_3mut

EAAAKGGGGSEAAAK	17,313	KORV_Q9TTC1-Pro_3mutA
EAAAKEAAAK	17,314	WMSV_P03359_3mutA
GGGGSSGGS	17,315	MLVCB_P08361_3mutA
PAPGGSEAAAK	17,316	BAEVM_P10272_3mut
EAAAKGGSPAP	17,317	MLVFF_P26809_3mut
GSSGGSGGG	17,318	MLVBM_Q7SVK7_3mutA_WS
GSSGGS	17,319	PERV_Q4VFZ2_3mut
PAPGGSGSS	17,320	PERV_Q4VFZ2_3mutA_WS
EAAAKGGSGSS	17,321	KORV_Q9TTC1-Pro_3mutA
PAPAP	17,322	MLVCB_P08361_3mut
EAAAKGSSPAP	17,323	PERV_Q4VFZ2_3mutA_WS
EAAAKPAPGGG	17,324	MLVMS_P03355_PLV919
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,325	MLVBM_Q7SVK7_3mut
EAAAKGGGGSS	17,326	MLVMS_P03355_PLV919
PAPEAAAK	17,327	PERV_Q4VFZ2_3mut
EAAAKPAPGSS	17,328	BAEVM_P10272_3mutA
GGSPAP	17,329	PERV_Q4VFZ2_3mutA_WS
GGSGGS	17,330	BAEVM_P10272_3mutA
PAPEAAAKGSS	17,331	KORV_Q9TTC1_3mut
PAPGSS	17,332	MLVMS_P03355_PLV919
PAPAPAPAPAP	17,333	MLVAV_P03356_3mutA
GGG	17,334	XMRV6_A1Z651_3mutA
GGGPAP	17,335	PERV_Q4VFZ2_3mutA_WS
GSSPAPEAAAK	17,336	KORV_Q9TTC1_3mutA
PAP	17,337	BAEVM_P10272_3mutA
GGSPAP	17,338	BAEVM_P10272_3mutA
PAPEAAAKGGS	17,339	MLVMS_P03355_PLV919
PAPGSSGGS	17,340	PERV_Q4VFZ2_3mutA_WS
PAPAPAPAPAPAP	17,341	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAK	17,342	MLVCB_P08361_3mut
GGSGGSGGSGGSGGS	17,343	MLVMS_P03355_PLV919
EAAAKPAPGGS	17,344	MLVMS_P03355_3mut
GGSGGS	17,345	MLVMS_P03355_PLV919
EAAAKPAP	17,346	MLVMS_P03355_3mutA_WS
GGSEAAAK	17,347	XMRV6_A1Z651_3mutA
GGSGGG	17,348	KORV_Q9TTC1_3mut
GGSGGGEAAAK	17,349	PERV_Q4VFZ2_3mut
PAPEAAAKGGG	17,350	AVIRE_P03360
PAPAP	17,351	PERV_Q4VFZ2_3mut
GSS	17,352	KORV_Q9TTC1-Pro_3mutA
EAAAKGSSGGG	17,353	MLVAV_P03356_3mutA

GGSPAPGSS	17,354	MLVBM_Q7SVK7_3mutA_WS
PAPEAAAK	17,355	MLVAV_P03356_3mut
EAAAKGGSPAP	17,356	BAEVM_P10272_3mutA
PAPAPAPAP	17,357	WMSV_P03359_3mutA
PAPGGSEAAAK	17,358	MLVMS_P03355_3mut
GGSGGSGGSGGS	17,359	WMSV_P03359_3mut
GGGGGSGSS	17,360	XMRV6_A1Z651_3mut
PAPGGSGGG	17,361	KORV_Q9TTC1_3mutA
GGS	17,362	MLVMS_P03355_3mut
EAAAK	17,363	WMSV_P03359_3mut
GGGEAAAKGSS	17,364	MLVBM_Q7SVK7_3mutA_WS
GGSPAPGSS	17,365	MLVCB_P08361_3mut
GGSEAAAKPAP	17,366	PERV_Q4VFZ2_3mut
GGGGSGGGGSGGGGSGGGGSGGGGS	17,367	MLVCB_P08361_3mutA
GGSGSS	17,368	BAEVM_P10272_3mutA
GGGEAAAKGSS	17,369	WMSV_P03359_3mutA
EAAAKGGSPAP	17,370	WMSV_P03359_3mut
GSSPAPEAAAK	17,371	MLVMS_P03355_3mut
GGSGGSGGSGGS	17,372	MLVMS_P03355_PLV919
GSSPAPEAAAK	17,373	WMSV_P03359_3mut
GSSGSSGSSGSS	17,374	PERV_Q4VFZ2
GGSGSSEAAAK	17,375	WMSV_P03359_3mutA
GGSGGG	17,376	MLVFF_P26809_3mut
GGSPAPGGG	17,377	MLVFF_P26809_3mut
GGSGGSGGS	17,378	BAEVM_P10272_3mutA
GGGGSSEAAAK	17,379	MLVBM_Q7SVK7_3mut
GGSPAPGSS	17,380	MLVMS_P03355_3mut
EAAAKPAPGSS	17,381	AVIRE_P03360_3mut
GGGGSSGGS	17,382	FLV_P10273_3mutA
GGSPAPEAAAK	17,383	PERV_Q4VFZ2_3mut
GGSEAAAK	17,384	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSS	17,385	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	17,386	MLVMS_P03355_PLV919
GGGGG	17,387	PERV_Q4VFZ2_3mut
GGSEAAAKGSS	17,388	MLVCB_P08361_3mutA
GSSGGG	17,389	MLVBM_Q7SVK7_3mutA_WS
PAPGSSGGG	17,390	KORV_Q9TTC1-Pro_3mutA
GGSGGS	17,391	BAEVM_P10272_3mut

EAAAKGGGGGS	17,392	MLVBM_Q7SVK7_3mutA_WS
GGSGSSPAP	17,393	MLVCB_P08361_3mut
PAPGSSGGG	17,394	KORV_Q9TTC1
PAPGGSGGG	17,395	MLVMS_P03355_3mut
GGGG	17,396	WMSV_P03359_3mutA
EAAAKGGSPAP	17,397	MLVCB_P08361_3mut
GSSGSS	17,398	FLV_P10273_3mutA
GGSEAAAKPAP	17,399	SFV3L_P27401_2mut
EAAAKGSSGGS	17,400	MLVAV_P03356_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,401	MLVAV_P03356_3mutA
EAAAKGGSGSS	17,402	PERV_Q4VFZ2_3mutA_WS
GGGGG	17,403	MLVCB_P08361_3mut
GGGEAAAK	17,404	BAEVM_P10272_3mut
GGSGGSGGSGGS	17,405	MLVCB_P08361_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,406	PERV_Q4VFZ2
PAPAPAPAPAP	17,407	MLVMS_P03355_3mutA_WS
EAAAKEAAAK	17,408	XMRV6_A1Z651_3mut
GSSGGSEAAAK	17,409	PERV_Q4VFZ2_3mutA_WS
PAPGGSEAAAK	17,410	KORV_Q9TTC1-Pro_3mutA
EAAAKGGGPAP	17,411	MLVBM_Q7SVK7_3mutA_WS
PAPGGSGSS	17,412	PERV_Q4VFZ2
SGSETPGTSESATPES	17,413	MLVMS_P03355_3mut
GGSGGS	17,414	MLVMS_P03355_PLV919
EAAAKGGS	17,415	FLV_P10273_3mut
GGSPAPGSS	17,416	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAK	17,417	FFV_O93209_2mut
GSSGGSGGG	17,418	MLVMS_P03355_3mutA_WS
PAPGSSEAAAK	17,419	WMSV_P03359_3mut
PAPAPAPAPAPAP	17,420	KORV_Q9TTC1_3mutA
GGGGSS	17,421	BAEVM_P10272_3mut
GGGGSEAAAKGGGGS	17,422	AVIRE_P03360_3mut
GSSPAPEAAAK	17,423	KORV_Q9TTC1-Pro_3mutA
PAPEAAAKGGG	17,424	MLVBM_Q7SVK7_3mut
EAAAKEAAAK	17,425	WMSV_P03359_3mut
EAAAK	17,426	SFV3L_P27401-Pro_2mutA
GSSGGSGGG	17,427	XMRV6_A1Z651_3mutA
GGGEAAAKPAP	17,428	WMSV_P03359_3mutA
GGSGGS	17,429	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,430	FOAMV_P14350_2mutA

GGGGG	17,431	MLVAV_P03356_3mutA
GSSGGSEAAAK	17,432	BAEVM_P10272_3mut
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,433	SFV1_P23074
GGSGGGPAP	17,434	MLVCB_P08361_3mut
GGSGSS	17,435	PERV_Q4VFZ2_3mut
SGSETPGTSESATPES	17,436	MLVFF_P26809_3mut
EAAAKGGSPAP	17,437	MLVMS_P03355_3mut
PAPAP	17,438	PERV_Q4VFZ2_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,439	MLVBM_Q7SVK7_3mut
GGGGGS	17,440	BAEVM_P10272_3mutA
EAAAKEAAAK	17,441	AVIRE_P03360_3mut
GSSGGSEAAAK	17,442	PERV_Q4VFZ2_3mut
GGGEAAAK	17,443	WMSV_P03359_3mut
GSSGGGEAAAK	17,444	AVIRE_P03360_3mutA
GGG	17,445	XMRV6_A1Z651_3mut
GGGGSEAAAKGGGGS	17,446	BAEVM_P10272_3mut
GGGG	17,447	MLVMS_P03355_3mut
GGSGGS	17,448	MLVMS_P03355_3mutA_WS
GGSGGGGSS	17,449	MLVBM_Q7SVK7_3mutA_WS
GSSPAPGGS	17,450	PERV_Q4VFZ2_3mut
GSSPAPEAAAK	17,451	PERV_Q4VFZ2_3mutA_WS
EAAAKGGS	17,452	WMSV_P03359_3mut
GGSGGSGGSGGS	17,453	PERV_Q4VFZ2_3mut
GGGGSSEAAAK	17,454	KORV_Q9TTC1-Pro_3mut
PAPAPAPAPAPAP	17,455	MLVAV_P03356_3mut
EAAAKGSSGGG	17,456	MLVMS_P03355_PLV919
GGGGG	17,457	MLVBM_Q7SVK7_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,458	FFV_O93209_2mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,459	KORV_Q9TTC1-Pro_3mut
GGSPAPGGG	17,460	MLVMS_P03355_3mutA_WS
GGGEAAAKGGS	17,461	MLVMS_P03355_3mut
GGGEAAAK	17,462	PERV_Q4VFZ2_3mut
PAPEAAAKGGG	17,463	MLVMS_P03355_3mut
GSSGSSGSSGSSGSSGSS	17,464	BAEVM_P10272_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,465	GALV_P21414_3mutA
EAAAKGGSPAP	17,466	FFV_O93209-Pro
EAAAKEAAAK	17,467	MLVFF_P26809_3mut

GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,468	PERV_Q4VFZ2_3mutA_WS
GGSGGSGGSGGS	17,469	MLVAV_P03356_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	17,470	SFV3L_P27401_2mutA
GSSGSSGSSGSSGSSGSS	17,471	BAEVM_P10272_3mut
GGGGS	17,472	MLVMS_P03355_PLV919
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,473	SFV1_P23074
GGGGSGGGGS	17,474	KORV_Q9TTC1-Pro_3mutA
GGGGSGGGGS	17,475	MLVMS_P03355_3mut
GGSGSS	17,476	KORV_Q9TTC1_3mutA
GSSPAPGGG	17,477	PERV_Q4VFZ2_3mut
GSSGGSPAP	17,478	PERV_Q4VFZ2_3mutA_WS
PAPGGS	17,479	PERV_Q4VFZ2_3mutA_WS
GGSPAPEAAAK	17,480	FOAMV_P14350_2mutA
GGGPAPGGS	17,481	SFV3L_P27401_2mut
PAPGSSGGG	17,482	MLVCB_P08361_3mut
GSSGGGEAAAK	17,483	AVIRE_P03360_3mut
GSSGGG	17,484	XMRV6_A1Z651_3mut
GSSGSS	17,485	PERV_Q4VFZ2_3mut
GSSGGG	17,486	MLVAV_P03356_3mutA
PAPGGGGGS	17,487	PERV_Q4VFZ2_3mut
GSSEAAAK	17,488	MLVMS_P03355_3mut
PAPGGG	17,489	FLV_P10273_3mutA
GGGGSGGGGS	17,490	PERV_Q4VFZ2_3mut
GSSGGS	17,491	MLVMS_P03355_PLV919
GGGGSGGGGS	17,492	SFV3L_P27401_2mut
EAAAKGGSGSS	17,493	FLV_P10273_3mutA
GSSEAAAKGGS	17,494	MLVMS_P03355_3mutA_WS
PAPGSSEAAAK	17,495	SFV3L_P27401_2mutA
GGGGSGGGGS	17,496	SFV3L_P27401-Pro_2mutA
PAPGSSEAAAK	17,497	PERV_Q4VFZ2_3mut
PAPGSSEAAAK	17,498	PERV_Q4VFZ2
GGSPAPGGG	17,499	AVIRE_P03360_3mut
GGGGGS	17,500	PERV_Q4VFZ2_3mutA_WS
GGGGSSGGS	17,501	PERV_Q4VFZ2_3mut
PAPAPAPAP	17,502	AVIRE_P03360_3mutA
GGSGGS	17,503	WMSV_P03359_3mutA
GGGPAPGGS	17,504	PERV_Q4VFZ2_3mut
GGSGGSGGSGGSGGS	17,505	MLVMS_P03355_PLV919
GGSGGG	17,506	PERV_Q4VFZ2_3mut
EAAAKEAAAK	17,507	SFV3L_P27401_2mut
PAPGSS	17,508	XMRV6_A1Z651_3mut

GSSEAAAK	17,509	MLVFF_P26809_3mut
GGSPAPGGG	17,510	MLVMS_P03355_3mut
EAAAKGGG	17,511	WMSV_P03359_3mutA
GSSEAAAKGGS	17,512	PERV_Q4VFZ2_3mutA_WS
GSSGGSPAP	17,513	FFV_O93209
GGGGGS	17,514	KORV_Q9TTC1-Pro_3mut
GSSGGG	17,515	MLVCB_P08361_3mut
GSSGSS	17,516	MLVCB_P08361_3mutA
GGSEAAAKPAP	17,517	BAEVM_P10272_3mut
EAAAKGGGGSS	17,518	MLVCB_P08361_3mut
EAAAKPAPGGS	17,519	KORV_Q9TTC1-Pro_3mutA
GSSGSSGSSGSSGSS	17,520	MLVAV_P03356_3mutA
GGGGSEAAAKGGGGS	17,521	PERV_Q4VFZ2_3mutA_WS
GGSGSS	17,522	KORV_Q9TTC1-Pro_3mut
GSS	17,523	SFV3L_P27401-Pro_2mutA
PAPAP	17,524	BAEVM_P10272_3mut
EAAAKPAP	17,525	BAEVM_P10272
EAAAKEAAAKEAAAKEAAAKEAAAK	17,526	KORV_Q9TTC1-Pro_3mut
GGGGGGG	17,527	PERV_Q4VFZ2_3mutA_WS
GGGGS	17,528	MLVMS_P03355_3mut
GSSGGG	17,529	FLV_P10273_3mutA
PAPAPAPAPAP	17,530	FLV_P10273_3mut
EAAAKEAAAKEAAAK	17,531	WMSV_P03359_3mutA
GSSGGS	17,532	MLVBM_Q7SVK7_3mutA_WS
EAAAKPAPGGG	17,533	MLVMS_P03355_3mut
GSSPAPGGS	17,534	WMSV_P03359_3mut
PAPGSSGGG	17,535	PERV_Q4VFZ2_3mutA_WS
GSSGGG	17,536	AVIRE_P03360_3mutA
PAPGGSGSS	17,537	MLVFF_P26809_3mut
PAPGSS	17,538	PERV_Q4VFZ2_3mut
GGGGGSGSS	17,539	WMSV_P03359_3mutA
EAAAKGGGGSS	17,540	MLVBM_Q7SVK7_3mutA_WS
GGGGGGG	17,541	BAEVM_P10272_3mut
PAPEAAAKGSS	17,542	MLVMS_P03355_3mut
GGSGGGEAAAK	17,543	MLVMS_P03355_PLV919
EAAAKGGGGGS	17,544	MLVCB_P08361_3mut
PAPGGS	17,545	KORV_Q9TTC1-Pro_3mut
GGGG	17,546	FLV_P10273_3mutA
EAAAKGGSGSS	17,547	MLVBM_Q7SVK7_3mutA_WS

GGGGSSGGS	17,548	MLVMS_P03355_3mutA_WS
GGGGGGGG	17,549	WMSV_P03359_3mut
GGSGSSGGG	17,550	MLVMS_P03355_PLV919
GSSEAAAKGGS	17,551	KORV_Q9TTC1-Pro_3mutA
EAAAKPAPGSS	17,552	MLVCB_P08361_3mut
GGSPAPGSS	17,553	KORV_Q9TTC1_3mutA
PAPGSSGGG	17,554	BAEVM_P10272_3mut
EAAAKPAPGSS	17,555	WMSV_P03359_3mut
GGSPAPEAAAK	17,556	XMRV6_A1Z651_3mutA
GSSPAP	17,557	FLV_P10273_3mutA
GSS	17,558	BAEVM_P10272_3mutA
EAAAKPAPGGS	17,559	FLV_P10273_3mutA
GGSGSSPAP	17,560	FLV_P10273_3mutA
PAPGSSGGS	17,561	MLVMS_P03355_3mut
GSAGSAAGSGEF	17,562	PERV_Q4VFZ2_3mutA_WS
GSSGGSEAAAK	17,563	KORV_Q9TTC1_3mutA
GSSGGS	17,564	MLVMS_P03355_3mutA_WS
EAAAKGGGGSEAAAK	17,565	SFV3L_P27401_2mut
GSSGGS	17,566	PERV_Q4VFZ2_3mutA_WS
GGSPAPEAAAK	17,567	FLV_P10273_3mut
GGSEAAAKGSS	17,568	PERV_Q4VFZ2_3mutA_WS
GSSPAPEAAAK	17,569	PERV_Q4VFZ2_3mutA_WS
GGSGSSGGG	17,570	PERV_Q4VFZ2_3mut
GGGG	17,571	AVIRE_P03360_3mutA
GGSEAAAKPAP	17,572	WMSV_P03359_3mut
GSSGGSPAP	17,573	MLVAV_P03356_3mutA
GSSGGSEAAAK	17,574	MLVMS_P03355_3mut
PAPEAAAKGGS	17,575	KORV_Q9TTC1-Pro_3mut
GGSPAP	17,576	PERV_Q4VFZ2_3mutA_WS
GGSEAAAK	17,577	MLVAV_P03356_3mutA
EAAAKGGGGSEAAAK	17,578	KORV_Q9TTC1-Pro_3mut
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,579	MLVMS_P03355_PLV919
GSSEAAAK	17,580	KORV_Q9TTC1_3mutA
GGG	17,581	AVIRE_P03360
GGSEAAAKGSS	17,582	MLVBM_Q7SVK7_3mut
GGSEAAAKGSS	17,583	MLVMS_P03355_3mut
GGSPAPEAAAK	17,584	MLVCB_P08361_3mut
GGSGGGEAAAK	17,585	MLVCB_P08361_3mut
GGSEAAAKPAP	17,586	MLVMS_P03355_3mutA_WS
EAAAKGGSGSS	17,587	KORV_Q9TTC1-Pro_3mut
GGGEAAAKGGS	17,588	MLVCB_P08361_3mut

EAAAKGGGGSEAAAK	17,589	FLV_P10273_3mutA
GGSPAP	17,590	MLVFF_P26809_3mut
GGGGSSGGS	17,591	XMRV6_A1Z651_3mutA
PAP	17,592	MLVCB_P08361_3mut
GGS	17,593	SFV3L_P27401-Pro_2mutA
GGGGSGGGGS	17,594	MLVMS_P03355_3mut
GGGEAAAKGGS	17,595	MLVAV_P03356_3mutA
GSSGSSGSSGSSGSSGSS	17,596	MLVMS_P03355_PLV919
PAPGSS	17,597	MLVCB_P08361_3mut
GGSGGSGGS	17,598	MLVMS_P03355_PLV919
PAPGGSGGG	17,599	FLV_P10273_3mutA
GGGGSGGGGSGGGGS	17,600	FLV_P10273_3mut
GGSGSSGGG	17,601	KORV_Q9TTC1-Pro_3mutA
GGSGGSGGS	17,602	GALV_P21414_3mutA
GGGEAAAKGGS	17,603	WMSV_P03359_3mut
SGSETPGTSESATPES	17,604	KORV_Q9TTC1_3mutA
EAAAKGGGGGS	17,605	KORV_Q9TTC1-Pro_3mut
EAAAKGSSPAP	17,606	BAEVM_P10272_3mut
GGGG	17,607	MLVCB_P08361_3mut
GGGGSGGGGSGGGGSGGGGSGGGGS	17,608	MLVBM_Q7SVK7_3mut
GSSGGSGGG	17,609	MLVMS_P03355_PLV919
GGSGSS	17,610	MLVFF_P26809_3mut
EAAAKGGS	17,611	AVIRE_P03360_3mutA
GSSEAAAKGGS	17,612	MLVBM_Q7SVK7_3mutA_WS
EAAAKPAPGGG	17,613	WMSV_P03359_3mut
PAPGSSGGG	17,614	MLVCB_P08361_3mutA
GGGGSSEAAAK	17,615	KORV_Q9TTC1-Pro_3mutA
GSSEAAAKPAP	17,616	BAEVM_P10272_3mutA
PAPGGGEAAAK	17,617	MLVBM_Q7SVK7_3mutA_WS
GGSGGGEAAAK	17,618	MLVCB_P08361_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,619	FFV_O93209
EAAAKGGGGGS	17,620	GALV_P21414_3mutA
GGSPAPGGG	17,621	MLVMS_P03355_3mut
GSSGSSGSS	17,622	FLV_P10273_3mutA
EAAAK	17,623	MLVBM_Q7SVK7_3mut
GGGGSSGGS	17,624	MLVMS_P03355_3mut
GGSGSSPAP	17,625	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAK	17,626	BAEVM_P10272_3mut
GGGPAPGSS	17,627	MLVMS_P03355_3mut
GSSPAPGGS	17,628	PERV_Q4VFZ2_3mutA_WS

PAPAP	17,629	FLV_P10273_3mutA
PAPAPAPAP	17,630	PERV_Q4VFZ2_3mut
GGGGGSEAAAK	17,631	GALV_P21414_3mutA
GGGGGSGSS	17,632	BAEVM_P10272_3mutA
GGGEAAAKGSS	17,633	KORV_Q9TTC1_3mutA
GGGGGSPAP	17,634	AVIRE_P03360_3mut
GGGGGSEAAAK	17,635	SFV3L_P27401_2mutA
GGS	17,636	KORV_Q9TTC1_3mutA
GGGGGGG	17,637	PERV_Q4VFZ2_3mut
SGSETPGTSESATPES	17,638	SFV3L_P27401_2mutA
EAAAKGGSGGG	17,639	MLVMS_P03355_3mut
GGGGS	17,640	MLVFF_P26809_3mut
EAAAKGSSGGG	17,641	BAEVM_P10272_3mut
EAAAKPAPGGS	17,642	MLVF5_P26810_3mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,643	SFV3L_P27401_2mutA
GGSPAPGGG	17,644	WMSV_P03359_3mutA
GSAGSAAGSGEF	17,645	MLVFF_P26809_3mut
GGGGSSGGS	17,646	MLVMS_P03355_3mutA_WS
GGGGGGG	17,647	MLVCB_P08361_3mut
GSSEAAAK	17,648	WMSV_P03359_3mut
PAPGSS	17,649	FLV_P10273_3mutA
GSSGGG	17,650	PERV_Q4VFZ2_3mutA_WS
PAPGGG	17,651	MLVFF_P26809_3mut
GGGGGSPAP	17,652	MLVMS_P03355_3mut
GGSEAAAK	17,653	XMRV6_A1Z651_3mut
GSSGGG	17,654	PERV_Q4VFZ2_3mut
GGSGGSGGSGGS	17,655	MLVMS_P03355_3mut
PAPAP	17,656	AVIRE_P03360_3mut
GGSEAAAK	17,657	PERV_Q4VFZ2_3mut
GGGGS	17,658	MLVMS_P03355_PLV919
GGGG	17,659	BAEVM_P10272_3mutA
EAAAKGGGGSS	17,660	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAK	17,661	GALV_P21414_3mutA
PAPGGGEAAAK	17,662	KORV_Q9TTC1
EAAAKGGSPAP	17,663	MLVMS_P03355_3mut
GGSGSSEAAAK	17,664	MLVMS_P03355_3mut
GGSPAPEAAAK	17,665	FLV_P10273_3mutA
GGGGGGG	17,666	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,667	SFV1_P23074_2mutA
EAAAKGSSGGS	17,668	MLVMS_P03355_3mut
GSSEAAAKPAP	17,669	MLVFF_P26809_3mut

GGGGSS	17,670	FLV_P10273_3mutA
EAAAKGGSGGG	17,671	AVIRE_P03360_3mutA
GGSGGS	17,672	PERV_Q4VFZ2_3mutA_WS
GGGGGSPAP	17,673	AVIRE_P03360_3mutA
EAAAKEAAAKEAAAK	17,674	XMRV6_A1Z651_3mut
PAPEAAAKGGS	17,675	FLV_P10273_3mutA
GSSGGSEAAAK	17,676	MLVCB_P08361_3mut
EAAAKGGSGGG	17,677	MLVMS_P03355
GGSGGGPAP	17,678	MLVMS_P03355_3mut
GGS	17,679	XMRV6_A1Z651_3mut
GGSEAAAKPAP	17,680	MLVFF_P26809_3mut
EAAAKGGG	17,681	MLVMS_P03355_PLV919
GSSGSSGSSGSS	17,682	WMSV_P03359_3mut
GGSGSSPAP	17,683	PERV_Q4VFZ2_3mut
GGGEAAAK	17,684	MLVMS_P03355_3mutA_WS
GSSPAPGGS	17,685	KORV_Q9TTC1-Pro_3mutA
GSSEAAAKGGG	17,686	SFV3L_P27401_2mut
EAAAKPAPGGS	17,687	MLVCB_P08361_3mut
GGSGGGEAAAK	17,688	PERV_Q4VFZ2
GGSGSS	17,689	MLVCB_P08361_3mut
GGSGGGEAAAK	17,690	MLVBM_Q7SVK7_3mutA_WS
GGSGGSGGSGGSGGSGGS	17,691	FLV_P10273_3mut
PAPEAAAKGSS	17,692	MLVMS_P03355_3mut
EAAAKGSSGGS	17,693	WMSV_P03359_3mutA
GGSGSSEAAAK	17,694	MLVCB_P08361_3mut
GGSGSSEAAAK	17,695	KORV_Q9TTC1_3mutA
GSSGGSGGG	17,696	MLVMS_P03355_PLV919
EAAAKGGSGGG	17,697	SFV3L_P27401-Pro_2mutA
GGSGGS	17,698	AVIRE_P03360_3mutA
GSAGSAAGSGEF	17,699	MLVMS_P03355_PLV919
GGSGSS	17,700	GALV_P21414_3mutA
GGGG	17,701	MLVFF_P26809_3mutA
GGGGSGGGGSGGGGSGGGGS	17,702	WMSV_P03359_3mut
SGSETPGTSESATPES	17,703	BAEVM_P10272_3mut
EAAAKEAAAKEAAAKEAAAK	17,704	FOAMV_P14350_2mutA
GGGEAAAKGGS	17,705	FLV_P10273_3mutA
GSSGGSEAAAK	17,706	MLVFF_P26809_3mut
EAAAKGGGGSS	17,707	MLVAV_P03356_3mut
PAPGGSEAAAK	17,708	KORV_Q9TTC1-Pro_3mut
EAAAK	17,709	XMRV6_A1Z651_3mut
GSSGSSGSSGSSGSSGSS	17,710	PERV_Q4VFZ2_3mut

GGGG	17,711	MLVCB_P08361_3mutA
GSSGSS	17,712	WMSV_P03359_3mutA
GSSGGSPAP	17,713	AVIRE_P03360_3mut
GGSGGSGGS	17,714	MLVCB_P08361_3mut
EAAAKGGGPAP	17,715	FLV_P10273_3mutA
GGGGSGGGGS	17,716	MLVCB_P08361_3mut
GGSEAAAKGSS	17,717	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,718	SFV3L_P27401_2mutA
GGSGSSEAAAK	17,719	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAKEAAAK	17,720	SFV3L_P27401-Pro_2mutA
GSSEAAAKGGS	17,721	FLV_P10273_3mutA
GGSGSS	17,722	PERV_Q4VFZ2
GGSGSSEAAAK	17,723	SFV3L_P27401-Pro_2mutA
GSSGSSGSS	17,724	XMRV6_A1Z651_3mutA
EAAAKGSSPAP	17,725	KORV_Q9TTC1_3mutA
EAAAKPAP	17,726	FLV_P10273_3mutA
GGSGSSEAAAK	17,727	KORV_Q9TTC1-Pro_3mut
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,728	KORV_Q9TTC1_3mutA
GGGGSGGGGSGGGGS	17,729	KORV_Q9TTC1-Pro_3mutA
GGGGGGG	17,730	FLV_P10273_3mut
EAAAKGSS	17,731	WMSV_P03359_3mut
EAAAKGGGPAP	17,732	MLVCB_P08361_3mut
GSSGSS	17,733	MLVBM_Q7SVK7_3mutA_WS
EAAAKGGGGGS	17,734	MLVFF_P26809_3mut
GGSGGGEAAAK	17,735	FLV_P10273_3mutA
PAPGSS	17,736	MLVFF_P26809_3mutA
PAPGSS	17,737	BAEVM_P10272_3mutA
GGSPAPGSS	17,738	AVIRE_P03360_3mut
GGGGSSEAAAK	17,739	MLVMS_P03355_3mut
GSSGGGGGS	17,740	FFV_O93209-Pro
EAAAKGSSPAP	17,741	PERV_Q4VFZ2_3mut
GSSPAPGGS	17,742	PERV_Q4VFZ2_3mut
GGGGGG	17,743	BAEVM_P10272_3mut
EAAAKGGGGSS	17,744	PERV_Q4VFZ2_3mutA_WS
PAPGGSEAAAK	17,745	KORV_Q9TTC1_3mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,746	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSS	17,747	MLVMS_P03355_3mut
EAAAKGSSGGG	17,748	MLVMS_P03355_PLV919
GGSEAAAKPAP	17,749	AVIRE_P03360_3mutA
GSSGSSGSSGSSGSS	17,750	WMSV_P03359_3mutA
GGGEAAAKPAP	17,751	FLV_P10273_3mutA

PAPGSSGGG	17,752	KORV_Q9TTC1_3mutA
GSSGSS	17,753	MLVMS_P03355_3mutA_WS
PAPEAAAK	17,754	BAEVM_P10272_3mut
GGGPAPGSS	17,755	PERV_Q4VFZ2
GSSGGSPAP	17,756	MLVFF_P26809_3mut
GGGGSS	17,757	SFV3L_P27401_2mut
PAPEAAAKGSS	17,758	SFV3L_P27401_2mut
GGSGGGPAP	17,759	XMRV6_A1Z651_3mutA
PAPGGS	17,760	BAEVM_P10272_3mutA
EAAAKGGGGGS	17,761	AVIRE_P03360_3mut
GSSGGSPAP	17,762	KORV_Q9TTC1-Pro_3mutA
GSSGGGGGS	17,763	WMSV_P03359_3mut
GGGEAAAKGGS	17,764	AVIRE_P03360_3mut
GGGEAAAKGSS	17,765	BAEVM_P10272_3mut
PAPEAAAKGSS	17,766	MLVAV_P03356_3mutA
GSSGSSGSSGSSGSS	17,767	MLVCB_P08361_3mut
GGSPAPGSS	17,768	FLV_P10273_3mutA
EAAAKGSSPAP	17,769	BAEVM_P10272_3mutA
GGSGGSGGSGGSGGSGGS	17,770	PERV_Q4VFZ2
GGGGSSEAAAK	17,771	FLV_P10273_3mutA
GGGGSSPAP	17,772	FFV_O93209
GSSGGSPAP	17,773	MLVMS_P03355_3mut
GGGPAPGSS	17,774	MLVMS_P03355_PLV919
PAPGSSGGS	17,775	PERV_Q4VFZ2_3mut
GGGGGSPAP	17,776	MLVFF_P26809_3mut
SGSETPGTSESATPES	17,777	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSSGSS	17,778	KORV_Q9TTC1_3mutA
GSSPAPGGG	17,779	WMSV_P03359_3mut
PAPAPAPAPAPAP	17,780	SFV3L_P27401_2mutA
GGGPAPGGS	17,781	MLVMS_P03355_3mut
PAPGGSEAAAK	17,782	WMSV_P03359_3mut
GGGGSSEAAAK	17,783	FFV_O93209-Pro
GGSPAPGGG	17,784	FLV_P10273_3mutA
GSSPAPEAAAK	17,785	AVIRE_P03360_3mut
GGGEAAAK	17,786	FLV_P10273_3mutA
PAPEAAAKGGG	17,787	MLVCB_P08361_3mut
GGSPAPGGG	17,788	MLVCB_P08361_3mut
GGSGGGGSS	17,789	BAEVM_P10272_3mutA
GSSPAPEAAAK	17,790	MLVCB_P08361_3mut
GGSPAPGGG	17,791	KORV_Q9TTC1-Pro_3mutA
PAPGGSGSS	17,792	KORV_Q9TTC1_3mutA

GSSPAP	17,793	KORV_Q9TTC1-Pro_3mutA
SGSETPGTSESATPES	17,794	MLVMS_P03355
GSSGSSGSS	17,795	MLVAV_P03356_3mutA
PAPGSSGGS	17,796	PERV_Q4VFZ2_3mutA_WS
PAPGGS	17,797	KORV_Q9TTC1-Pro_3mutA
PAPEAAAKGGG	17,798	SFV3L_P27401-Pro_2mutA
GGSGGSGGS	17,799	BAEVM_P10272_3mut
PAPGGS	17,800	MLVFF_P26809_3mut
GSSGGSPAP	17,801	MLVMS_P03355_PLV919
GSSGGGGGS	17,802	FLV_P10273_3mutA
GGGGGSPAP	17,803	KORV_Q9TTC1-Pro_3mut
EAAAKPAPGSS	17,804	SFV3L_P27401-Pro_2mutA
EAAAKGGSPAP	17,805	KORV_Q9TTC1-Pro
GGGPAPEAAAK	17,806	MLVMS_P03355_PLV919
GGSEAAAKGSS	17,807	MLVMS_P03355
PAPEAAAKGSS	17,808	KORV_Q9TTC1_3mutA
PAPEAAAKGGS	17,809	WMSV_P03359_3mutA
GSSGGG	17,810	PERV_Q4VFZ2_3mutA_WS
EAAAKGGGGSS	17,811	MLVMS_P03355_PLV919
EAAAKGGSPAP	17,812	AVIRE_P03360_3mutA
GGGGSSGGS	17,813	MLVMS_P03355_PLV919
PAPEAAAKGSS	17,814	PERV_Q4VFZ2_3mutA_WS
EAAAKGGGGGS	17,815	BAEVM_P10272_3mut
GSSGGGGGS	17,816	MLVMS_P03355_3mut
PAPAPAPAP	17,817	KORV_Q9TTC1_3mutA
GGSGGSGGSGGS	17,818	MLVAV_P03356_3mut
PAPAPAPAP	17,819	SFV3L_P27401_2mut
GSSEAAAKPAP	17,820	MLVMS_P03355_3mut
GGSGGGEAAAK	17,821	SFV3L_P27401_2mutA
GSSGGSGGG	17,822	MLVMS_P03355_3mutA_WS
GGGGGSPAP	17,823	MLVCB_P08361_3mutA
GGGEAAAKGSS	17,824	XMRV6_A1Z651_3mutA
GGGGSSPAP	17,825	BAEVM_P10272_3mut
GGSGGG	17,826	PERV_Q4VFZ2_3mut
GGGGSS	17,827	MLVBM_Q7SVK7_3mutA_WS
EAAAKGSSGGS	17,828	PERV_Q4VFZ2_3mutA_WS
GSSGGGGGS	17,829	PERV_Q4VFZ2
EAAAKGSSGGS	17,830	PERV_Q4VFZ2_3mut
EAAAKEAAAK	17,831	MLVAV_P03356_3mut
GSSGGGEAAAK	17,832	MLVAV_P03356_3mut
GSSPAPGGG	17,833	XMRV6_A1Z651_3mut

GGGGSGGGGSGGGGS	17,834	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAK	17,835	KORV_Q9TTC1_3mutA
EAAAKGGSGSS	17,836	MLVBM_Q7SVK7_3mut
PAPEAAAK	17,837	BLVJ_P03361
GSSGGG	17,838	FFV_O93209-Pro
GGSGGGEAAAK	17,839	KORV_Q9TTC1-Pro_3mutA
EAAAK	17,840	FLV_P10273_3mutA
GGGGSSPAP	17,841	MLVMS_P03355_3mut
GSS	17,842	SFV3L_P27401-Pro_2mut
PAPEAAAKGSS	17,843	BAEVM_P10272_3mut
GGGGGSPAP	17,844	PERV_Q4VFZ2_3mut
GSSGSSGSS	17,845	BAEVM_P10272_3mutA
GGGGSGGGGSGGGGSGGGGS	17,846	SFV1_P23074_2mut
GGGGSSEAAAK	17,847	SFV3L_P27401_2mutA
GGGGSGGGGSGGGGSGGGGS	17,848	FOAMV_P14350-Pro_2mut
PAPGSSEAAAK	17,849	MLVBM_Q7SVK7_3mutA_WS
GGGGGSGSS	17,850	MLVFF_P26809_3mutA
GGSEAAAKGGG	17,851	MLVBM_Q7SVK7_3mut
PAPGSSGGG	17,852	PERV_Q4VFZ2
GGS	17,853	PERV_Q4VFZ2_3mutA_WS
EAAAKGGSGSS	17,854	FLV_P10273_3mut
GGGEAAAK	17,855	WMSV_P03359_3mutA
GGSEAAAKPAP	17,856	MLVBM_Q7SVK7_3mut
SGSETPGTSESATPES	17,857	FOAMV_P14350-Pro_2mutA
EAAAKPAPGGS	17,858	AVIRE_P03360_3mut
EAAAKGGGGGS	17,859	KORV_Q9TTC1-Pro_3mutA
GGGGS	17,860	PERV_Q4VFZ2_3mut
GGSEAAAKGSS	17,861	MLVFF_P26809_3mutA
GGSEAAAKGGG	17,862	AVIRE_P03360
GGSGGSGGSGGSGGSGGS	17,863	SFV3L_P27401_2mut
GGSEAAAKGSS	17,864	SFV3L_P27401-Pro_2mutA
GGGEAAAKPAP	17,865	MLVCB_P08361_3mut
GGSEAAAK	17,866	MLVMS_P03355_PLV919
GGSPAPGSS	17,867	KORV_Q9TTC1-Pro_3mutA
GSSPAPEAAAK	17,868	WMSV_P03359_3mutA
GGSGSS	17,869	KORV_Q9TTC1-Pro_3mutA
PAPGGGGGS	17,870	AVIRE_P03360_3mut
PAPEAAAKGSS	17,871	FFV_O93209-Pro
GGSGGGEAAAK	17,872	WMSV_P03359_3mut
PAPGGG	17,873	MLVMS_P03355_3mut
EAAAKGGG	17,874	FLV_P10273_3mutA

GSSGSSGSSGSS	17,875	MLVCB_P08361_3mut
EAAAKGGSGGG	17,876	FFV_O93209
GSSPAPGGS	17,877	PERV_Q4VFZ2_3mutA_WS
GSSPAPGGS	17,878	MLVCB_P08361_3mut
GGGPAP	17,879	WMSV_P03359_3mutA
GGGPAP	17,880	KORV_Q9TTC1_3mutA
GGSPAPGSS	17,881	KORV_Q9TTC1-Pro_3mut
PAPAP	17,882	MLVMS_P03355_3mut
GGGGGGG	17,883	MLVMS_P03355_3mut
GGGGG	17,884	KORV_Q9TTC1-Pro_3mut
GSAGSAAGSGEF	17,885	FOAMV_P14350_2mutA
PAPAP	17,886	KORV_Q9TTC1-Pro_3mutA
GGSEAAAKGGG	17,887	SFV3L_P27401-Pro_2mutA
PAPAP	17,888	WMSV_P03359_3mut
GGGGSGGGGSGGGGS	17,889	SFV3L_P27401_2mut
PAPGGS	17,890	KORV_Q9TTC1_3mutA
GGGEAAAKPAP	17,891	FLV_P10273_3mut
GGGGGS	17,892	MLVAV_P03356_3mutA
GSSEAAAKGGG	17,893	WMSV_P03359_3mut
EAAAKGGGGSS	17,894	GALV_P21414_3mutA
GSSGGS	17,895	MLVAV_P03356_3mutA
GSSGGG	17,896	MLVBM_Q7SVK7_3mut
PAPAPAP	17,897	SFV3L_P27401-Pro_2mutA
GGGG	17,898	KORV_Q9TTC1_3mutA
EAAAKPAPGGS	17,899	MLVFF_P26809_3mut
GGGGSGGGGS	17,900	XMRV6_A1Z651_3mut
EAAAKGGG	17,901	MLVCB_P08361_3mut
GGGGSSPAP	17,902	KORV_Q9TTC1_3mutA
GSSEAAAKGGG	17,903	KORV_Q9TTC1-Pro_3mutA
GGGGG	17,904	BLVJ_P03361_2mutB
GGGEAAAKGSS	17,905	FFV_O93209-Pro
GSSGSSGSS	17,906	BAEVM_P10272_3mut
GSSGGSPAP	17,907	PERV_Q4VFZ2_3mut
EAAAKGGS	17,908	KORV_Q9TTC1_3mut
GGSPAPEAAAK	17,909	AVIRE_P03360_3mut
GGSEAAAK	17,910	WMSV_P03359_3mut
GSSGGS	17,911	KORV_Q9TTC1-Pro_3mutA
GGGPAPEAAAK	17,912	KORV_Q9TTC1_3mutA
PAPGSS	17,913	WMSV_P03359_3mutA
GGSEAAAKGSS	17,914	FLV_P10273_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	17,915	SFV3L_P27401

GSSEAAAKGGG	17,916	SFV3L_P27401-Pro_2mutA
GGGGSEAAAKGGGGS	17,917	KORV_Q9TTC1-Pro_3mutA
GGSGGSGGS	17,918	WMSV_P03359_3mut
GGGGGSGSS	17,919	KORV_Q9TTC1-Pro
GGGGSGGGGSGGGGSGGGGS	17,920	MLVMS_P03355_3mut
EAAAKGGG	17,921	PERV_Q4VFZ2
GGSEAAAKGGG	17,922	KORV_Q9TTC1-Pro_3mut
GSSGGSGGG	17,923	PERV_Q4VFZ2_3mutA_WS
GGGGGS	17,924	PERV_Q4VFZ2_3mut
GSAGSAAGSGEF	17,925	PERV_Q4VFZ2
PAPEAAAKGSS	17,926	BAEVM_P10272_3mutA
GSSPAPGGG	17,927	MLVCB_P08361_3mut
GGGGSSPAP	17,928	KORV_Q9TTC1-Pro_3mutA
PAPGGSGGG	17,929	MLVFF_P26809_3mut
GSSPAP	17,930	KORV_Q9TTC1_3mutA
PAPGSS	17,931	SFV3L_P27401-Pro_2mut
GGSGGGGSS	17,932	MLVMS_P03355_PLV919
GSSGGS	17,933	WMSV_P03359_3mutA
EAAAKGGGGGS	17,934	PERV_Q4VFZ2
GGGGG	17,935	KORV_Q9TTC1_3mutA
EAAAKGSS	17,936	MLVMS_P03355_PLV919
EAAAKEAAAKEAAAKEAAAKEAAAK	17,937	FLV_P10273_3mut
EAAAKEAAAKEAAAKEAAAK	17,938	SFV3L_P27401-Pro_2mut
GSAGSAAGSGEF	17,939	SFV3L_P27401_2mutA
GGGPAPGGS	17,940	FLV_P10273_3mutA
GGSEAAAKGGG	17,941	MLVCB_P08361_3mut
PAPGGGEAAAK	17,942	BAEVM_P10272_3mut
EAAAKPAPGSS	17,943	FOAMV_P14350_2mut
GGSEAAAK	17,944	KORV_Q9TTC1_3mutA
GGSGSS	17,945	AVIRE_P03360
GGSPAPEAAAK	17,946	MLVMS_P03355_PLV919
GGGGS	17,947	XMRV6_A1Z651_3mut
GGSPAPGGG	17,948	XMRV6_A1Z651_3mut
EAAAKPAPGGS	17,949	PERV_Q4VFZ2
GSSPAP	17,950	BAEVM_P10272_3mut
GGSGSSGGG	17,951	FLV_P10273_3mutA
PAPGGG	17,952	PERV_Q4VFZ2_3mutA_WS
GSSGGSEAAAK	17,953	MLVBM_Q7SVK7_3mut
GGSEAAAK	17,954	MLVMS_P03355_3mut
GGGPAPGGS	17,955	MLVFF_P26809_3mut
GSAGSAAGSGEF	17,956	MLVBM_Q7SVK7_3mutA_WS

EAAAKPAPGGS	17,957	SFVCP_Q87040
PAPGGG	17,958	PERV_Q4VFZ2_3mutA_WS
GSSPAPEAAAK	17,959	MLVBM_Q7SVK7
PAPEAAAK	17,960	MLVBM_Q7SVK7_3mut
PAPGGGGGS	17,961	AVIRE_P03360_3mutA
GGSEAAAKPAP	17,962	MLVBM_Q7SVK7_3mut
EAAAKGSS	17,963	WMSV_P03359_3mutA
GGGEAAAK	17,964	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAK	17,965	MLVMS_P03355_3mut
PAPEAAAKGGG	17,966	BAEVM_P10272_3mut
PAPAPAP	17,967	MLVCB_P08361_3mut
EAAAKPAPGGS	17,968	BAEVM_P10272_3mut
GGGGSGGGGS	17,969	FLV_P10273_3mut
GGGGSEAAAKGGGGS	17,970	KORV_Q9TTC1_3mut
EAAAK	17,971	FLV_P10273_3mut
PAPAPAP	17,972	WMSV_P03359_3mut
GGGGSEAAAKGGGGS	17,973	FFV_O93209-Pro
GGSPAPEAAAK	17,974	MLVMS_P03355_3mut
GGSGSSGGG	17,975	XMRV6_A1Z651_3mut
GGSPAPGSS	17,976	PERV_Q4VFZ2_3mut
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,977	SFV3L_P27401-Pro_2mutA
EAAAKGGGPAP	17,978	BAEVM_P10272_3mutA
GSSGGSEAAAK	17,979	MLVMS_P03355_3mutA_WS
SGSETPGTSESATPES	17,980	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAK	17,981	KORV_Q9TTC1-Pro_3mutA
GSSGSSGSS	17,982	KORV_Q9TTC1_3mutA
GSSPAPGGG	17,983	SFV3L_P27401-Pro_2mutA
GSSGGGEAAAK	17,984	KORV_Q9TTC1_3mutA
GGSGGGGSS	17,985	PERV_Q4VFZ2_3mutA_WS
GSSGGGEAAAK	17,986	MLVCB_P08361_3mut
GSSEAAAKGGG	17,987	MLVCB_P08361_3mut
GGSGGGGSS	17,988	KORV_Q9TTC1_3mutA
GGSGSSPAP	17,989	PERV_Q4VFZ2_3mutA_WS
GSSPAP	17,990	MLVMS_P03355_3mut
GGGGSSEAAAK	17,991	AVIRE_P03360
GGS	17,992	WMSV_P03359_3mut
EAAAKEAAAK	17,993	PERV_Q4VFZ2_3mut
PAPAPAPAP	17,994	MLVAV_P03356_3mut
GGSEAAAKGGG	17,995	KORV_Q9TTC1_3mutA
PAPGGG	17,996	MLVAV_P03356_3mut
EAAAKGSS	17,997	BAEVM_P10272_3mut

GGGGSGGGGS	17,998	WMSV_P03359_3mutA
GGSGGSGGS	17,999	SFV3L_P27401_2mut
EAAAK	18,000	MLVCB_P08361_3mut
GGGGSSGGS	18,001	WMSV_P03359_3mutA
GGGPAPEAAAK	18,002	MLVAV_P03356_3mutA
EAAAKEAAAKEAAAK	18,003	FFV_O93209
GSSEAAAKGGG	18,004	MLVBM_Q7SVK7_3mut
GGGPAPGGS	18,005	FLV_P10273_3mut
GGSEAAAKGGG	18,006	WMSV_P03359_3mut
EAAAKGGGGGS	18,007	XMRV6_A1Z651_3mutA
EAAAKGGSGGG	18,008	FLV_P10273_3mutA
GGSEAAAKGGG	18,009	SFV3L_P27401_2mutA
GGGGS	18,010	PERV_Q4VFZ2_3mutA_WS
GSSGGS	18,011	MLVMS_P03355_3mut
GSSGSS	18,012	MLVAV_P03356_3mutA
GGSPAPGGG	18,013	MLVBM_Q7SVK7_3mutA_WS
GSSGGGGGS	18,014	MLVF5_P26810_3mut
PAPAPAPAP	18,015	MLVCB_P08361_3mut
PAPAP	18,016	PERV_Q4VFZ2_3mutA_WS
PAPGSSGGS	18,017	KORV_Q9TTC1_3mut
PAPGSSGGG	18,018	PERV_Q4VFZ2_3mut
GGGEAAAK	18,019	MLVMS_P03355_PLV919
GGSGGSGGSGGSGGS	18,020	SFV3L_P27401-Pro_2mutA
GGSGGG	18,021	FLV_P10273_3mut
PAPEAAAKGGG	18,022	MLVFF_P26809_3mut
PAP	18,023	PERV_Q4VFZ2_3mutA_WS
PAPGGSGSS	18,024	FFV_O93209_2mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	18,025	FFV_O93209-Pro_2mut
GSSGSSGSSGSS	18,026	FFV_O93209-Pro
GSSGSSGSSGSSGSS	18,027	FLV_P10273_3mutA
GGGEAAAKPAP	18,028	PERV_Q4VFZ2
PAPGSSGGG	18,029	SFV3L_P27401_2mut
PAPGGSGSS	18,030	KORV_Q9TTC1-Pro_3mut
PAPAPAPAPAP	18,031	GALV_P21414_3mutA
GGSGGGEAAAK	18,032	PERV_Q4VFZ2_3mut
GSSPAP	18,033	MLVCB_P08361_3mut
EAAAKPAP	18,034	MLVF5_P26810_3mut
GGGGSGGGGSGGGGSGGGGS	18,035	MLVBM_Q7SVK7_3mut
GGSGGG	18,036	WMSV_P03359_3mut
GGSGGSGGS	18,037	KORV_Q9TTC1_3mut
GGGGGGGG	18,038	MLVFF_P26809_3mut

GGGGSS	18,039	MLVAV_P03356_3mut
GSSGGGGGS	18,040	SFV3L_P27401_2mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	18,041	GALV_P21414_3mutA
GSSGSSGSS	18,042	PERV_Q4VFZ2_3mut
GSSPAPGGS	18,043	MLVFF_P26809_3mut
PAPAPAP	18,044	AVIRE_P03360_3mutA
EAAAKEAAAKEAAAKEAAAK	18,045	WMSV_P03359_3mutA
PAPAPAPAP	18,046	SFV3L_P27401_2mutA
GGGGSS	18,047	MLVAV_P03356_3mutA
GSSGSSGSSGSSGSS	18,048	SFV3L_P27401_2mutA
PAPGGS	18,049	WMSV_P03359_3mutA
GSSEAAAKGGG	18,050	PERV_Q4VFZ2
GSSGGSPAP	18,051	MLVMS_P03355_PLV919
GSSGSSGSSGSSGSSGSS	18,052	SFV3L_P27401_2mutA
GGSGSSGGG	18,053	MLVCB_P08361_3mut
GGGPAPGSS	18,054	SFV3L_P27401-Pro_2mutA
GSSEAAAKGGS	18,055	WMSV_P03359_3mut
GSSEAAAKGGG	18,056	MLVAV_P03356_3mut
GGSGGGPAP	18,057	FFV_O93209-Pro
GSSGSS	18,058	PERV_Q4VFZ2_3mut
PAPGGGGGS	18,059	GALV_P21414_3mutA
EAAAKPAPGGS	18,060	MLVAV_P03356_3mut
GSSGSS	18,061	MLVMS_P03355_3mut
EAAAKPAPGGS	18,062	FFV_O93209-Pro
GGGPAPEAAAK	18,063	MLVMS_P03355_3mutA_WS
GSSEAAAKGGG	18,064	MLVBM_Q7SVK7_3mut
GGGEAAAKGGS	18,065	BAEVM_P10272_3mut
GSSGSS	18,066	KORV_Q9TTC1-Pro_3mutA
EAAAKEAAAKEAAAK	18,067	SFV1_P23074
PAPGSSGGS	18,068	KORV_Q9TTC1-Pro_3mut
PAPAPAPAPAP	18,069	MLVMS_P03355
GSSEAAAK	18,070	SFV3L_P27401_2mut
PAP	18,071	PERV_Q4VFZ2_3mut
GGSEAAAKGGG	18,072	MLVBM_Q7SVK7_3mut
GGSGGGPAP	18,073	MLVBM_Q7SVK7_3mutA_WS
GSSGSS	18,074	MLVMS_P03355_3mut
GGSEAAAK	18,075	MLVMS_P03355
GSSEAAAKGGS	18,076	MLVMS_P03355_PLV919
PAPGGGGGS	18,077	MLVFF_P26809_3mut
GSSGGG	18,078	PERV_Q4VFZ2_3mut
GSSGGS	18,079	PERV_Q4VFZ2_3mutA_WS

PAPGGG	18,080	BAEVM_P10272_3mut
PAPGSSGGG	18,081	MLVBM_Q7SVK7_3mut
GGSEAAAK	18,082	SFV3L_P27401_2mut
GSSPAPEAAAK	18,083	SFV3L_P27401-Pro_2mut
GSSGGSPAP	18,084	BAEVM_P10272_3mut
GGSPAPGSS	18,085	PERV_Q4VFZ2_3mutA_WS
GGSGGSGGS	18,086	PERV_Q4VFZ2
GGSGGGPAP	18,087	FLV_P10273_3mut
GGGPAPEAAAK	18,088	SFV3L_P27401_2mutA
GGGGS	18,089	FLV_P10273_3mutA
GSSGGSGGG	18,090	XMRV6_A1Z651_3mut
EAAAKGGGGSS	18,091	PERV_Q4VFZ2
GGSGSSGGG	16,901	SFV3L_P27401-Pro_2mutA
GGSGGSGGS	16,902	MLVFF_P26809_3mut
GGGPAPEAAAK	16,903	FLV_P10273_3mut
GSSGGGEAAAK	16,904	MLVMS_P03355_3mut
GGG	16,905	SFV3L_P27401_2mut
GSAGSAAGSGEF	16,906	WMSV_P03359_3mut
GSSGGGPAP	16,907	MLVMS_P03355_PLV919
GGGGSS	16,908	KORV_Q9TTC1-Pro_3mut
GGGGSSEAAAK	16,909	KORV_Q9TTC1
PAPGGSGGG	16,910	SFV3L_P27401_2mut
GSSGSSGSSGSSGSS	16,911	FFV_O93209
GSSGGSPAP	16,912	MLVMS_P03355_3mut
GGSEAAAK	16,913	KORV_Q9TTC1-Pro_3mutA
GGGGSGGGGS	16,914	BAEVM_P10272_3mut
GSSEAAAKGGG	16,915	AVIRE_P03360_3mut
EAAAKPAPGGG	16,916	FLV_P10273_3mut
EAAAKGGSPAP	16,917	SFV3L_P27401-Pro_2mutA
GSSEAAAKPAP	16,918	MLVBM_Q7SVK7_3mut
GGGPAPGGS	16,919	MLVCB_P08361_3mut
GGG	16,920	SFV3L_P27401_2mutA
EAAAKGGGGSEAAAK	16,921	SFV3L_P27401_2mutA
GGSGSSGGG	16,922	MLVBM_Q7SVK7_3mut
GSAGSAAGSGEF	16,923	BAEVM_P10272_3mut
GGGEAAAK	16,924	FOAMV_P14350_2mutA
PAPEAAAKGGS	16,925	WMSV_P03359_3mut
PAPAPAPAPAPAP	16,926	MLVF5_P26810_3mutA
GGSGGGGSS	16,927	FLV_P10273_3mutA
PAPGSSGGS	16,928	BAEVM_P10272_3mut
PAPEAAAK	16,929	WMSV_P03359_3mutA

GSSGSSGSSGSSGSSGSS	16,930	FFV_O93209-Pro_2mut
GGGGGSGSS	16,931	FFV_O93209-Pro
GGGGGGGG	16,932	SFV3L_P27401-Pro_2mutA
GGGGGG	16,933	FLV_P10273_3mut
GSSGGSGGG	16,934	MLVAV_P03356_3mutA
GGGGSS	16,935	SFV3L_P27401-Pro_2mutA
GGSGGGPAP	16,936	FOAMV_P14350_2mut
GSSGSS	16,937	AVIRE_P03360_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	16,938	SFV3L_P27401-Pro_2mutA
EAAAKEAAAK	16,939	BAEVM_P10272_3mut
GSSPAPEAAAK	16,940	GALV_P21414_3mutA
GGSEAAAKPAP	16,941	SFV3L_P27401_2mutA
GGSGGGEAAAK	16,942	SFV3L_P27401-Pro_2mutA
EAAAKGSSPAP	16,943	FOAMV_P14350_2mut
GGSGSSEAAAK	16,944	SFV3L_P27401_2mut
GGG	16,945	PERV_Q4VFZ2
GGGGGSGSS	16,946	FOAMV_P14350_2mut
GGSGGGEAAAK	16,947	KORV_Q9TTC1-Pro_3mut
GSSGGSGGG	16,948	AVIRE_P03360_3mutA
EAAAKPAPGGG	16,949	SFV3L_P27401_2mutA
PAPGGSGGG	16,950	KORV_Q9TTC1-Pro_3mut
PAPAPAP	16,951	WMSV_P03359_3mutA
GSSEAAAKPAP	16,952	SFV1_P23074
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	16,953	SRV2_P51517
GSSGGSGGG	16,954	PERV_Q4VFZ2_3mutA_WS
GSSGSSGSSGSSGSSGSS	16,955	FFV_O93209
GSSGGGPAP	16,956	WMSV_P03359_3mut
PAPAPAPAPAPAP	16,957	MLVBM_Q7SVK7_3mut
GGGGGSPAP	16,958	KORV_Q9TTC1-Pro_3mutA
PAPGSS	16,959	MLVBM_Q7SVK7_3mutA_WS
PAPEAAAKGGS	16,960	SFV3L_P27401-Pro_2mut
GGGGSSPAP	16,961	MLVMS_P03355_3mut
GGSEAAAK	16,962	FFV_O93209-Pro
EAAAKPAPGGS	16,963	AVIRE_P03360_3mutA
PAPGSS	16,964	WMSV_P03359_3mut
PAPGSSGGG	16,965	SFV3L_P27401-Pro_2mutA
EAAAKEAAAKEAAAK	16,966	SFV3L_P27401_2mut
GGS	16,967	MLVRD_P11227_3mut
GGGGS	16,968	KORV_Q9TTC1-Pro_3mut
GGSGGGGSS	16,969	KORV_Q9TTC1
GGSGGG	16,970	MLVMS_P03355_3mutA_WS

GGGEAAAKPAP	16,971	BAEVM_P10272_3mut
EAAAKEAAAKEAAAKEAAAKEAAAK	16,972	FLV_P10273
PAPGGSGGG	16,973	KORV_Q9TTC1-Pro_3mutA
GSSGSSGSSGSSGSSGSS	16,974	HTL1L_P0C211
GGGEAAAKPAP	16,975	WMSV_P03359
GSSGGSPAP	16,976	FFV_O93209-Pro
PAPAPAPAPAP	16,977	SFV3L_P27401-Pro_2mutA
GSSGGSEAAAK	16,978	SFV3L_P27401_2mutA
GGSPAPGSS	16,979	SFV3L_P27401_2mut
GGSGGSGGS	16,980	KORV_Q9TTC1-Pro_3mut
PAPEAAAKGSS	16,981	KORV_Q9TTC1-Pro_3mut
EAAAKGGS	16,982	KORV_Q9TTC1_3mutA
EAAAKGGGGSEAAAK	16,983	SFV3L_P27401-Pro_2mut
GGGGSSPAP	16,984	FFV_O93209-Pro
EAAAK	16,985	SFV3L_P27401_2mut
EAAAKGGGGSS	16,986	BAEVM_P10272_3mut
GGGGGSEAAAK	16,987	MLVBM_Q7SVK7_3mut
GGGG	16,988	PERV_Q4VFZ2
GGGGGSEAAAK	16,989	FLV_P10273_3mut
EAAAKGGGPAP	16,990	KORV_Q9TTC1-Pro
GGGGSGGGGSGGGGSGGGGS	16,991	FFV_O93209_2mutA
GSSGGSGGG	16,992	PERV_Q4VFZ2_3mut
GGGGSGGGGSGGGGS	16,993	GALV_P21414_3mutA
GGSGGGEAAAK	16,994	AVIRE_P03360_3mutA
PAPEAAAKGGG	16,995	SFV3L_P27401_2mut
GGGGSGGGGS	16,996	AVIRE_P03360
GSSGGGEAAAK	16,997	SFV3L_P27401_2mutA
GGGGG	16,998	AVIRE_P03360_3mutA
GGSGSS	16,999	KORV_Q9TTC1_3mut
PAPAPAPAPAPAP	17,000	FOAMV_P14350_2mut
GGSEAAAKPAP	17,001	KORV_Q9TTC1-Pro_3mut
GGGGGG	17,002	PERV_Q4VFZ2_3mut
GSSGGGEAAAK	17,003	MLVBM_Q7SVK7
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,004	MLVAV_P03356
GGSPAPGSS	17,005	BAEVM_P10272_3mut
GGGGSSPAP	17,006	BAEVM_P10272
GGGGSEAAAKGGGGS	17,007	SFV3L_P27401_2mut
GGGGGGGG	17,008	GALV_P21414_3mutA
PAPAP	17,009	MLVAV_P03356_3mut
GGGEAAAK	17,010	PERV_Q4VFZ2_3mutA_WS
GSSPAPGGG	17,011	FFV_O93209_2mut

GGSGGSGGSGGSGGS	17,012	BAEVM_P10272
GGGGGS	17,013	MLVF5_P26810_3mutA
PAPGGGGSS	17,014	FLV_P10273_3mutA
GGGEAAAK	17,015	MLVBM_Q7SVK7_3mut
PAPEAAAKGGG	17,016	WMSV_P03359_3mut
GSSEAAAK	17,017	MLVBM_Q7SVK7_3mut
EAAAKEAAAK	17,018	AVIRE_P03360
EAAAKGGGGGS	17,019	MLVBM_Q7SVK7_3mut
GGGEAAAKGGS	17,020	SFV3L_P27401-Pro_2mutA
PAPAPAPAPAP	17,021	MLVF5_P26810_3mut
PAPGSSEAAAK	17,022	SFV3L_P27401-Pro_2mutA
EAAAKEAAAKEAAAK	17,023	BAEVM_P10272_3mutA
GGSPAPGSS	17,024	MLVMS_P03355
PAPGSSGGS	17,025	FLV_P10273_3mutA
EAAAKEAAAKEAAAKEAAAK	17,026	FOAMV_P14350-Pro_2mut
EAAAKGGG	17,027	KORV_Q9TTC1_3mutA
EAAAKGGSGGG	17,028	MLVBM_Q7SVK7_3mut
GGGGGS	17,029	KORV_Q9TTC1-Pro_3mutA
PAPGGSGGG	17,030	WMSV_P03359_3mut
GGGPAPGGS	17,031	KORV_Q9TTC1_3mutA
GSS	17,032	FFV_O93209
GGSGGSGGS	17,033	PERV_Q4VFZ2_3mut
GGGGS	17,034	GALV_P21414_3mutA
GGGG	17,035	MLVF5_P26810_3mut
GGSEAAAKPAP	17,036	FFV_O93209-Pro_2mut
PAPAPAPAP	17,037	FFV_O93209-Pro
PAP	17,038	MLVF5_P26810_3mut
EAAAKEAAAKEAAAK	17,039	FFV_O93209_2mut
EAAAKGSS	17,040	MLVCB_P08361_3mut
EAAAKGGG	17,041	MLVBM_Q7SVK7_3mut
PAPEAAAKGGG	17,042	FFV_O93209_2mut
GSSGGGEAAAK	17,043	SFV1_P23074-Pro_2mut
PAPGGGEAAAK	17,044	GALV_P21414_3mutA
GGGGSGGGGSGGGGSGGGGS	17,045	FOAMV_P14350-Pro_2mutA
GSSGGG	17,046	FOAMV_P14350_2mut
GGGGSGGGGSGGGGSGGGGS	17,047	SFV3L_P27401_2mutA
GGSGSS	17,048	AVIRE_P03360_3mut
GGSGSSEAAAK	17,049	MMTVB_P03365_WS
PAPAPAP	17,050	MLVAV_P03356_3mutA
GSSGGSPAP	17,051	SFV3L_P27401-Pro_2mut
GGSPAP	17,052	AVIRE_P03360

GGSGGGPAP	17,053	FFV_O93209
GSSEAAAK	17,054	PERV_Q4VFZ2
GSSGGGPAP	17,055	PERV_Q4VFZ2_3mutA_WS
GGGGSSEAAAK	17,056	KORV_Q9TTC1_3mutA
GGSEAAAKPAP	17,057	SFVCP_Q87040
GGSGGGPAP	17,058	FOAMV_P14350_2mutA
GGGGSGGGGSGGGGSGGGGS	17,059	BLVJ_P03361_2mutB
GGGGSSPAP	17,060	SFV3L_P27401_2mutA
EAAAKGGS	17,061	MLVF5_P26810_3mut
GGSEAAAKGSS	17,062	MLVCB_P08361_3mut
GGGGSSEAAAK	17,063	SFV3L_P27401_2mut
EAAAKGGSGGG	17,064	FOAMV_P14350_2mut
GGSGGS	17,065	FLV_P10273_3mut
EAAAKGGG	17,066	FFV_O93209-Pro
GSSGSSGSSGSSGSS	17,067	SFV3L_P27401
GSSGGGPAP	17,068	PERV_Q4VFZ2_3mutA_WS
PAPGGSEAAAK	17,069	SFV3L_P27401-Pro_2mutA
GGSPAP	17,070	KORV_Q9TTC1
EAAAKPAPGSS	17,071	KORV_Q9TTC1_3mutA
SGSETPGTSESATPES	17,072	SFV1_P23074
GSSPAP	17,073	SFV3L_P27401-Pro_2mutA
GSSPAPGGG	17,074	SFV3L_P27401_2mut
GGGEAAAKGSS	17,075	SFV1_P23074_2mut
GGGPAPGGS	17,076	BAEVM_P10272_3mut
EAAAKGGG	17,077	KORV_Q9TTC1-Pro_3mutA
GSSGGG	17,078	SFV3L_P27401-Pro_2mut
GGSPAPEAAAK	17,079	BAEVM_P10272_3mut
EAAAKGSSPAP	17,080	FFV_O93209
EAAAKGGGGSEAAAK	17,081	SFV3L_P27401-Pro_2mutA
GSSGSSGSSGSSGSS	17,082	SFV1_P23074_2mut
EAAAKGGSPAP	17,083	FOAMV_P14350_2mut
GGSGGS	17,084	KORV_Q9TTC1-Pro_3mutA
EAAAKGSSGGS	17,085	GALV_P21414
GSSGGGPAP	17,086	MLVAV_P03356
PAPEAAAKGGS	17,087	FOAMV_P14350_2mut
EAAAKPAPGGG	17,088	AVIRE_P03360_3mut
GGSPAP	17,089	SFV3L_P27401_2mutA
GGGGSGGGGS	17,090	SFV3L_P27401_2mutA
GGGGSS	17,091	AVIRE_P03360_3mutA
GGSPAPGGG	17,092	SFV3L_P27401-Pro_2mutA
EAAAKPAPGSS	17,093	SFV3L_P27401

EAAAKPAP	17,094	FOAMV_P14350-Pro_2mut
PAPEAAAKGSS	17,095	PERV_Q4VFZ2_3mutA_WS
EAAAKGGSGSS	17,096	SFV3L_P27401_2mutA
GGGEAAAKGSS	17,097	GALV_P21414_3mutA
GGGGSEAAAKGGGGS	17,098	PERV_Q4VFZ2_3mut
PAPGGSGSS	17,099	FFV_O93209-Pro_2mutA
GGSEAAAKPAP	17,100	GALV_P21414_3mutA
GGSGGSGGSGGSGGS	17,101	FFV_O93209-Pro
GSSGGSEAAAK	17,102	SFV3L_P27401-Pro_2mut
GGS	17,103	GALV_P21414_3mutA
PAPGGSEAAAK	17,104	MLVMS_P03355
PAPEAAAKGGS	17,105	BAEVM_P10272_3mutA
GGSGSSPAP	17,106	SFV3L_P27401-Pro_2mutA
GSSPAP	17,107	WMSV_P03359_3mut
GGGEAAAK	17,108	MMTVB_P03365
GGGGSS	17,109	PERV_Q4VFZ2_3mut
GGSPAPGSS	17,110	SFV3L_P27401-Pro_2mut
PAPGGS	17,111	MLVBM_Q7SVK7_3mut
EAAAKGSSPAP	17,112	MLVBM_Q7SVK7_3mut
GGGGSSGGS	17,113	PERV_Q4VFZ2_3mut
PAPAPAPAPAPAP	17,114	SFV1_P23074
GGSEAAAKGGG	17,115	SFV3L_P27401-Pro_2mut
GGSGGS	17,116	SFV1_P23074_2mut
GSSGGGGGS	17,117	MLVF5_P26810_3mutA
EAAAKGGGPAP	17,118	SFV3L_P27401
EAAAKEAAAKEAAAKEAAAK	17,119	FOAMV_P14350-Pro_2mutA
GGGPAPGSS	17,120	SFV3L_P27401_2mutA
GGGGSGGGGSGGGGSGGGGS	17,121	SFV3L_P27401_2mut
EAAAKEAAAKEAAAKEAAAK	17,122	MMTVB_P03365_WS
PAPGSSGGS	17,123	KORV_Q9TTC1-Pro_3mutA
PAPGSSEAAAK	17,124	FOAMV_P14350-Pro_2mut
GSSPAPEAAAK	17,125	BAEVM_P10272_3mut
EAAAKGGGGSEAAAK	17,126	FFV_O93209-Pro
GGSPAP	17,127	PERV_Q4VFZ2
GGSGSSEAAAK	17,128	XMRV6_A1Z651_3mut
GGSEAAAKGGG	17,129	GALV_P21414_3mutA
PAPGGGGSS	17,130	AVIRE_P03360_3mutA
GGSGGSGGSGGS	17,131	PERV_Q4VFZ2
GGGGSSGGS	17,132	PERV_Q4VFZ2_3mutA_WS
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,133	BAEVM_P10272_3mutA
GGGPAP	17,134	MLVAV_P03356_3mut

GGGGSGGGGSGGGGSGGGGS	17,135	FFV_O93209_2mut
GSSEAAAK	17,136	FFV_O93209
GGSPAPEAAAK	17,137	FOAMV_P14350_2mut
GGGGGSEAAAK	17,138	FOAMV_P14350_2mut
GSSPAPGGS	17,139	MLVBM_Q7SVK7_3mut
GSS	17,140	SFVCP_Q87040_2mut
EAAAKPAP	17,141	FOAMV_P14350-Pro
EAAAKGGG	17,142	SFV3L_P27401_2mut
GGGEAAAK	17,143	AVIRE_P03360_3mutA
PAPGSSGGG	17,144	WMSV_P03359_3mut
EAAAKGGSPAP	17,145	SFV3L_P27401
GSSGGSGGG	17,146	SFV3L_P27401-Pro_2mutA
GSSGGGEAAAK	17,147	GALV_P21414_3mutA
GGGPAPGSS	17,148	MLVBM_Q7SVK7_3mutA_WS
PAPGGGEAAAK	17,149	FFV_O93209-Pro_2mut
GSSGSSGSSGSS	17,150	SFV1_P23074_2mut
GGSEAAAK	17,151	PERV_Q4VFZ2_3mutA_WS
GGGEAAAKPAP	17,152	SFV3L_P27401_2mut
EAAAKGGGPAP	17,153	SFV3L_P27401_2mut
GGGGSSPAP	17,154	FLV_P10273_3mut
EAAAKPAPGSS	17,155	FFV_O93209_2mut
GGGGSSPAP	17,156	SFV3L_P27401_2mut
GSSGSS	17,157	KORV_Q9TTC1_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGS	17,158	BLVJ_P03361_2mut
GGGGSSGGS	17,159	GALV_P21414_3mutA
EAAAKGGSGSS	17,160	FFV_O93209-Pro
EAAAKPAP	17,161	PERV_Q4VFZ2
GSSGGGEAAAK	17,162	MLVBM_Q7SVK7_3mut
PAPGGSGGG	17,163	BAEVM_P10272
EAAAKGGGPAP	17,164	MLVF5_P26810
GSSGSSGSS	17,165	MLVBM_Q7SVK7_3mut
GSSGGS	17,166	AVIRE_P03360_3mutA
GGSEAAAKGGG	17,167	FOAMV_P14350_2mut
EAAAKGGS	17,168	MLVF5_P26810_3mutA
GGSGSSGGG	17,169	WMSV_P03359_3mut
EAAAK	17,170	SFV1_P23074_2mut
GSSGGSPAP	17,171	SFV3L_P27401-Pro_2mutA
GGGGSSGGS	17,172	KORV_Q9TTC1_3mut
PAPGGSGGG	17,173	FFV_O93209-Pro_2mut
GGGPAPGGS	17,174	SFV3L_P27401_2mutA
GSSPAPEAAAK	17,175	FLV_P10273_3mut

GGSGSSPAP	17,176	SFV3L_P27401_2mut
GSSEAAAKGGS	17,177	SFV3L_P27401_2mut
PAPGGG	17,178	SFV3L_P27401_2mutA
SGSETPGTSESATPES	17,179	KORV_Q9TTC1-Pro_3mut
GGGGS	17,180	SFV1_P23074-Pro_2mutA
GSSGGGEAAAK	17,181	WMSV_P03359
EAAAKGGGGSEAAAK	17,182	MLVF5_P26810_3mutA
GSSEAAAKPAP	17,183	FFV_O93209
GGGGGG	17,184	SFV1_P23074_2mutA
EAAAKEAAAKEAAAK	17,185	MMTVB_P03365-Pro
EAAAKPAPGSS	17,186	MLVBM_Q7SVK7_3mut
GGSGSSEAAAK	17,187	SFV3L_P27401_2mutA
GGSEAAAK	17,188	MLVMS_P03355_3mut
GGSPAPEAAAK	17,189	SFV3L_P27401_2mut
GGGPAPGSS	17,190	SFV1_P23074
GGGGGSEAAAK	17,191	MLVBM_Q7SVK7_3mutA_WS
EAAAKPAPGSS	17,192	KORV_Q9TTC1-Pro
GSSGSSGSSGSS	17,193	SFV3L_P27401_2mut
EAAAKPAP	17,194	SFV3L_P27401_2mut
GGGEAAAK	17,195	PERV_Q4VFZ2_3mut
GGSGGS	17,196	SFV3L_P27401_2mutA
EAAAKGSSGGS	17,197	MMTVB_P03365
SGSETPGTSESATPES	17,198	SFV3L_P27401
EAAAKGSSGGG	17,199	PERV_Q4VFZ2
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,200	MMTVB_P03365
GGSGGGPAP	17,201	KORV_Q9TTC1_3mutA
PAPAPAPAP	17,202	SFV3L_P27401
GGGEAAAKGGS	17,203	SFV1_P23074_2mut
GSSGGSGGG	17,204	PERV_Q4VFZ2_3mut
PAPEAAAKGGS	17,205	FOAMV_P14350_2mutA
GGGEAAAKGSS	17,206	SFV3L_P27401_2mut
GGGGSGGGGSGGGGSGGGGS	17,207	MLVBM_Q7SVK7
PAPGSSGGG	17,208	FLV_P10273
GGSGSSGGG	17,209	FFV_O93209
EAAAKPAPGSS	17,210	MLVBM_Q7SVK7
GSSEAAAKGGG	17,211	SFV3L_P27401_2mutA
GGSGGSGGSGGSGGS	17,212	MLVF5_P26810
GGSEAAAKPAP	17,213	SFV3L_P27401-Pro_2mutA
EAAAKGGSPAP	17,214	SFV3L_P27401_2mutA
EAAAKGGGGGS	17,215	SFV3L_P27401_2mut
GSSPAPEAAAK	17,216	SFV3L_P27401_2mutA

PAPAP	17,217	MLVBM_Q7SVK7_3mut
PAPGGSEAAAK	17,218	KORV_Q9TTC1-Pro
GGSGSS	17,219	MLVF5_P26810_3mutA
GGSEAAAKPAP	17,220	FFV_O93209_2mut
GSS	17,221	MLVMS_P03355
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,222	SFV3L_P27401-Pro
PAPGGGEAAAK	17,223	SFV3L_P27401_2mut
PAPGGGGGS	17,224	SFV3L_P27401-Pro_2mut
PAPGGSGSS	17,225	BAEVM_P10272_3mut
GSSGGGEAAAK	17,226	FFV_O93209
GGSEAAAKPAP	17,227	SFV1_P23074_2mut
GGGGG	17,228	FLV_P10273_3mut
GGGEAAAKGSS	17,229	SFV3L_P27401
GSSGSSGSSGSSGSS	17,230	SFV1_P23074-Pro
SGSETPGTSESATPES	17,231	AVIRE_P03360
PAPGSSGGG	17,232	MLVBM_Q7SVK7_3mut
GGGGSSPAP	17,233	HTL3P_Q4U0X6_2mut
GGGEAAAK	17,234	SFV1_P23074
GGSGGG	17,235	AVIRE_P03360
EAAAKGSSGGG	17,236	SFV3L_P27401_2mutA
GSSPAPEAAAK	17,237	FOAMV_P14350-Pro_2mutA
GGGPAPGSS	17,238	WMSV_P03359
EAAAKGSSGGG	17,239	MLVMS_P03355
GGGGGSEAAAK	17,240	MLVMS_P03355
EAAAKPAPGGS	17,241	SFV3L_P27401
EAAAKGSSPAP	17,242	SFV3L_P27401
GGGGGGG	17,243	FOAMV_P14350_2mutA
EAAAKEAAAKEAAAK	17,244	SFV3L_P27401
GSSPAPGGS	17,245	FFV_O93209_2mutA
GGGGSSEAAAK	17,246	SFV3L_P27401-Pro_2mutA
GGSEAAAKGSS	17,247	GALV_P21414_3mutA
GGSEAAAKGSS	17,248	BAEVM_P10272_3mutA
EAAAKPAPGGG	17,249	MLVCB_P08361
GSSGSSGSSGSSGSSGSS	17,250	SFV1_P23074-Pro
GGGGSEAAAKGGGGS	17,251	FOAMV_P14350_2mut
GSSPAPGGS	17,252	MLVMS_P03355_PLV919
GGGGSGGGGS	17,253	FFV_O93209-Pro
GSSGGSPAP	17,254	KORV_Q9TTC1_3mutA
GGSGGS	17,255	GALV_P21414_3mutA
PAPGSSEAAAK	17,256	WMSV_P03359
PAPGGGGSS	17,257	MMTVB_P03365-Pro

GGGGSSGGS	17,258	PERV_Q4VFZ2_3mutA_WS
GGGGSGGGGS	17,259	FFV_O93209_2mut
GGGGSGGGGSGGGGSGGGGS	17,260	XMRV6_A1Z651
GGSGSSEAAAK	17,261	SFV1_P23074_2mut
GGSGGGGSS	17,262	GALV_P21414_3mutA
GGSEAAAKPAP	17,263	MLVBM_Q7SVK7
EAAAKGGSPAP	17,264	SFV1_P23074_2mutA
PAPAPAPAP	17,265	FFV_O93209
GSSGGSPAP	17,266	MMTVB_P03365-Pro
GGGGGSPAP	17,267	KORV_Q9TTC1_3mutA
EAAAKGGGPAP	17,268	PERV_Q4VFZ2
GSSGGSPAP	17,269	BAEVM_P10272
GGGGG	17,270	FFV_O93209
GGGGGS	17,271	FLV_P10273_3mutA
EAAAKEAAAKEAAAK	17,272	FOAMV_P14350
PAPGGG	17,273	MLVCB_P08361_3mut
GSSGGSEAAAK	17,274	FOAMV_P14350_2mutA
GGSPAPGGG	17,275	FLV_P10273_3mut
GSSGSSGSSGSSGSSGSS	17,276	SFV1_P23074-Pro_2mutA
GGSPAPEAAAK	17,277	SFV3L_P27401
PAPGGGGSS	17,278	HTL3P_Q4U0X6_2mutB
GGGGSSEAAAK	17,279	MMTVB_P03365_2mut_WS
PAPGGS	17,280	MLVRD_P11227_3mut
GGSGGSGGSGGSGGS	17,281	MMTVB_P03365
GSAGSAAGSGEF	17,282	AVIRE_P03360
GSSGGS	17,283	BAEVM_P10272_3mutA
GGSGGGGSS	17,284	MMTVB_P03365
GGSGGGGSS	17,285	WMSV_P03359
PAPEAAAKGSS	17,286	SFV1_P23074
GSSGSSGSSGSS	17,287	SFV1_P23074-Pro_2mutA
PAPAPAPAPAPAP	17,288	SFV3L_P27401
PAPGSSGGG	17,289	FLV_P10273_3mut
GGSGSSPAP	17,290	MLVMS_P03355
GGSGGGPAP	17,291	FOAMV_P14350
PAPGGGGGS	17,292	KORV_Q9TTC1_3mutA
EAAAKGSSPAP	17,293	GALV_P21414_3mutA
GGSGSSPAP	17,294	MLVBM_Q7SVK7_3mut
EAAAKGSS	17,295	SFV3L_P27401_2mut
GGGGGSEAAAK	17,296	WMSV_P03359
GGGGGGGG	17,297	SFV1_P23074-Pro
EAAAKEAAAK	17,298	MLVBM_Q7SVK7

GGGEAAAKGGS	17,299	MLVBM_Q7SVK7
EAAAKGGSPAP	17,300	SFV3L_P27401_2mut
GSSEAAAK	17,301	XMRV6_A1Z651
PAPGGGEAAAK	17,302	MMTVB_P03365_WS
GGSPAP	17,303	GALV_P21414_3mutA
GSSPAPGGG	17,304	MLVBM_Q7SVK7_3mutA_WS
GGSGSSPAP	17,305	SFV1_P23074_2mutA
GGS	17,306	HTL32_Q0R5R2_2mut
GGSGGGGSS	17,307	MMTVB_P03365-Pro
GGGGSGGGGSGGGGSGGGGS	17,308	SFVCP_Q87040_2mutA
EAAAKGGGPAP	17,309	FOAMV_P14350_2mut
GSSGGGEAAAK	17,310	MMTVB_P03365
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,311	MLVBM_Q7SVK7_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,312	MMTVB_P03365_WS
EAAAKEAAAK	17,313	FOAMV_P14350-Pro_2mut
GSSPAPEAAAK	17,314	FOAMV_P14350_2mutA
EAAAKPAPGGS	17,315	GALV_P21414_3mutA
GSSGGSPAP	17,316	KORV_Q9TTC1-Pro_3mut
GGGPAPEAAAK	17,317	MLVAV_P03356
GGGEAAAKPAP	17,318	SFV1_P23074-Pro_2mut
GGGGGSEAAAK	17,319	SFV3L_P27401_2mut
GGGPAPGSS	17,320	SFV3L_P27401_2mut
GGSEAAAKPAP	17,321	AVIRE_P03360
GSSGSSGSSGSSGSSGSS	17,322	SFV1_P23074-Pro_2mut
EAAAKGSSGGS	17,323	FOAMV_P14350_2mutA
GGGGGG	17,324	MLVBM_Q7SVK7_3mut
GSSPAPGGS	17,325	PERV_Q4VFZ2
GGSGSSPAP	17,326	GALV_P21414_3mutA
GGGPAPEAAAK	17,327	SFV3L_P27401
GGSGGGEAAAK	17,328	WMSV_P03359
GSAGSAAGSGEF	17,329	SFV1_P23074_2mut
GSSGGGEAAAK	17,330	MLVMS_P03355
GGG	17,331	MMTVB_P03365-Pro
PAPGSSGGS	17,332	FOAMV_P14350_2mut
GGGGSSPAP	17,333	FFV_O93209_2mut
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,334	MMTVB_P03365_WS
GGGGGGG	17,335	XMRV6_A1Z651
PAPAPAPAPAP	17,336	FOAMV_P14350
GGGGSGGGGSGGGGSGGGGS	17,337	MMTVB_P03365_2mut_WS

GGSGGGPAP	17,338	SFV3L_P27401_2mut
GGGGGG	17,339	SFV1_P23074-Pro
EAAAKPAPGSS	17,340	SFV3L_P27401_2mut
GGGGSSGGS	17,341	HTL3P_Q4U0X6_2mut
PAPGSSEAAAK	17,342	MMTVB_P03365-Pro
GGGGSSPAP	17,343	FOAMV_P14350-Pro_2mut
PAPGSSGGS	17,344	MMTVB_P03365
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,345	SRV2_P51517
PAPAPAP	17,346	MMTVB_P03365_2mut_WS
PAPGGGGGS	17,347	MMTVB_P03365_2mutB
GGGGSS	17,348	SFV1_P23074-Pro_2mutA
EAAAKEAAAKEAAAKEAAAK	17,349	SFV3L_P27401-Pro
GGSGGSGGSGGSGGS	17,350	MMTVB_P03365-Pro
GGGGGGG	17,351	SFV3L_P27401_2mut
PAPGGGEAAAK	17,352	SFV3L_P27401
PAPGSS	17,353	FOAMV_P14350_2mutA
GGGGSGGGGS	17,354	SFVCP_Q87040_2mutA
GSSGGSGGG	17,355	XMRV6_A1Z651
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,356	MLVBM_Q7SVK7
GSSEAAAKGGG	17,357	FFV_O93209-Pro_2mut
GGSEAAAKPAP	17,358	SFV3L_P27401-Pro
GSSGGSGGG	17,359	SFV1_P23074_2mut
EAAAKGGGGSS	17,360	FOAMV_P14350_2mutA
GGGGGG	17,361	SFV3L_P27401_2mut
GGGGG	17,362	MLVBM_Q7SVK7_3mut
PAPEAAAKGGG	17,363	SFV3L_P27401
EAAAKGGSPAP	17,364	KORV_Q9TTC1_3mutA
GGGEAAAKPAP	17,365	SFV1_P23074_2mut
GSSGSSGSSGSSGSSGSS	17,366	KORV_Q9TTC1-Pro
EAAAKEAAAKEAAAKEAAAK	17,367	SFVCP_Q87040
PAPGSSEAAAK	17,368	MLVBM_Q7SVK7
GSSGSSGSS	17,369	FFV_O93209-Pro_2mut
GSSGGGPAP	17,370	SFV3L_P27401-Pro_2mut
GGGPAPEAAAK	17,371	WMSV_P03359_3mut
GGGEAAAK	17,372	MMTVB_P03365-Pro
GSSGSSGSSGSS	17,373	SFV3L_P27401-Pro_2mutA
PAPAPAPAPAP	17,374	FFV_O93209-Pro
GGSPAPEAAAK	17,375	FFV_O93209-Pro_2mut
GSSGSSGSSGSSGSSGSS	17,376	GALV_P21414
EAAAKEAAAKEAAAKEAAAKEAAAK	17,377	FOAMV_P14350
GGGPAPEAAAK	17,378	MMTVB_P03365-Pro

PAPGGSGGG	17,379	MLVF5_P26810_3mutA
PAPGGSGGG	17,380	FLV_P10273_3mut
GGGEAAAKGGS	17,381	SFV3L_P27401
GSAGSAAGSGEF	17,382	MLVBM_Q7SVK7_3mut
GSSPAPGGG	17,383	MPMV_P07572_2mutB
GSSGSSGSSGSSGSSGSS	17,384	FOAMV_P14350
GGSGGGGSS	17,385	BLVJ_P03361_2mut
PAPEAAAKGSS	17,386	SFV1_P23074-Pro
GGG	17,387	FFV_O93209
EAAAKGGGGSS	17,388	SFV1_P23074_2mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,389	SRV2_P51517
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,390	MMTVB_P03365
GGGEAAAKGGS	17,391	MMTVB_P03365_WS
GSSGSS	17,392	SFV1_P23074
GSSGGGGGS	17,393	SFV3L_P27401
GGGGSSEAAAK	17,394	SFV1_P23074
EAAAKGSSGGS	17,395	HTL1A_P03362_2mutB
GSSEAAAKGGS	17,396	GALV_P21414_3mutA
EAAAKGSSPAP	17,397	SFV1_P23074
EAAAKPAPGSS	17,398	SFV3L_P27401_2mutA
PAPGSSGGG	17,399	SFV3L_P27401-Pro_2mut
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,400	SFV3L_P27401-Pro
EAAAKEAAAKEAAAKEAAAKEAAAK	17,401	MMTVB_P03365_WS
GGGGSSEAAAK	17,402	MLVF5_P26810_3mutA
EAAAKGGSPAP	17,403	GALV_P21414
PAPEAAAKGSS	17,404	MMTVB_P03365_WS
GSSGGGGGS	17,405	SFVCP_Q87040_2mut
GGGGSSPAP	17,406	SFV1_P23074
EAAAKGGGGSS	17,407	XMRV6_A1Z651
PAPAPAPAP	17,408	MMTVB_P03365
GGSEAAAKGSS	17,409	SFV3L_P27401_2mutA
GSSPAPGGG	17,410	MMTVB_P03365_WS
GGGGGG	17,411	SFV3L_P27401-Pro
GGSGGSGGS	17,412	FOAMV_P14350-Pro_2mut
PAPAPAPAPAPAP	17,413	WMSV_P03359
GSSPAP	17,414	MLVBM_Q7SVK7
GGGGGSGSS	17,415	MMTVB_P03365_2mut_WS
EAAAKGSSGGS	17,416	MMTVB_P03365_2mutB_WS
EAAAK	17,417	FFV_O93209_2mutA
PAPEAAAK	17,418	SFV1_P23074-Pro
EAAAKGGSGSS	17,419	SFV3L_P27401

GGSGGGSGGS	17,420	FFV_O93209-Pro
GSSGGGEAAAK	17,421	MMTVB_P03365
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,422	MLVFF_P26809_3mutA
GGSGGSGGSGGSGGSGGS	17,423	HTL1L_P0C211_2mutB
GGGEAAAK	17,424	SFV3L_P27401-Pro_2mutA
GGGGGSGSS	17,425	MMTVB_P03365
GSSPAPGGS	17,426	FOAMV_P14350_2mutA
EAAAKGSS	17,427	MLVMS_P03355
GSSGGSGGG	17,428	FFV_O93209-Pro
GGSGGGGSS	17,429	MMTVB_P03365-Pro_2mut
GGSPAPGSS	17,430	FOAMV_P14350_2mut
GGSGGSGGSGGSGGSGGS	17,431	SFVCP_Q87040-Pro_2mut
GSSEAAAKGGG	17,432	FOAMV_P14350_2mutA
GGSGGSGGS	17,433	MMTVB_P03365-Pro
GSSGSSGSSGSSGSSGSS	17,434	MMTVB_P03365_2mut_WS
GSSGSSGSSGSSGSS	17,435	MMTVB_P03365-Pro
PAPEAAAK	17,436	WDSV_O92815
GSSGSSGSSGSSGSS	17,437	FFV_O93209-Pro_2mut
EAAAKGGGGSEAAAK	17,438	MMTVB_P03365-Pro
GGSPAPEAAAK	17,439	FOAMV_P14350
GSSGSS	17,440	PERV_Q4VFZ2
GGG	17,441	MMTVB_P03365-Pro
GGGGSGGGGSGGGGS	17,442	FFV_O93209_2mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,443	MMTVB_P03365-Pro
GGSGSSPAP	17,444	WMSV_P03359
GGGGGGGG	17,445	SFV3L_P27401_2mut
PAPGSSEAAAK	17,446	FOAMV_P14350-Pro_2mutA
GGGGSSPAP	17,447	FOAMV_P14350_2mut
GSSGGSPAP	17,448	MLVBM_Q7SVK7_3mut
GSSGGGGGS	17,449	GALV_P21414_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	17,450	MMTVB_P03365
GSSGGGGGS	17,451	SFV1_P23074_2mut
GGGGSEAAAKGGGGS	17,452	SFV1_P23074
GGGEAAAKPAP	17,453	FFV_O93209
PAPGGGEAAAK	17,454	SFV1_P23074
GGSGGGEAAAK	17,455	PERV_Q4VFZ2_3mutA_WS
GSSGGG	17,456	MMTVB_P03365-Pro
EAAAKGSSGGS	17,457	FFV_O93209_2mut
GGGGG	17,458	SFV1_P23074_2mut
GGGPAP	17,459	SFV3L_P27401
GSSGGSEAAAK	17,460	FFV_O93209

SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,461	MMTVB_P03365-Pro
GSSGGGEAAAK	17,462	SFV1_P23074_2mutA
GSSGSSGSSGSSGSS	17,463	SFV3L_P27401_2mut
GGSEAAAKPAP	17,464	FLV_P10273
GGGGSGGGGS	17,465	FOAMV_P14350-Pro_2mutA
GSSEAAAKPAP	17,466	SFV3L_P27401
GGGGSEAAAKGGGGS	17,467	MMTVB_P03365-Pro
PAPGSSEAAAK	17,468	MLVF5_P26810_3mut
EAAAKGGSGGG	17,469	SFV3L_P27401
GGGPAPGGS	17,470	SFV3L_P27401
GSSEAAAKGGS	17,471	FOAMV_P14350_2mutA
EAAAKGGSGGG	17,472	HTL1L_P0C211
GSSGGSPAP	17,473	SFV3L_P27401_2mutA
PAPAP	17,474	FFV_O93209
PAPGGSGSS	17,475	MMTVB_P03365_WS
EAAAKGGGGGS	17,476	FOAMV_P14350_2mut
PAPEAAAKGGS	17,477	SFV3L_P27401_2mut
GSSEAAAKPAP	17,478	MMTVB_P03365-Pro
GGSGGS	17,479	PERV_Q4VFZ2_3mut
GSSEAAAKGGG	17,480	FFV_O93209-Pro_2mutA
EAAAK	17,481	HTL1L_P0C211
GSSPAP	17,482	MLVMS_P03355
EAAAKPAPGGG	17,483	FFV_O93209-Pro_2mut
GGGGSEAAAKGGGGS	17,484	SFV1_P23074-Pro_2mut
EAAAKGSSGGS	17,485	SFV3L_P27401
GSAGSAAGSGEF	17,486	FFV_O93209_2mutA
PAPEAAAKGGS	17,487	MMTVB_P03365_2mutB_WS
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	17,488	MMTVB_P03365
GGS	17,489	MMTVB_P03365
GGSEAAAKPAP	17,490	SFV1_P23074
EAAAKGSSGGG	17,491	HTLV2_P03363_2mut
GGSEAAAKGGG	17,492	MMTVB_P03365_WS
GGSGGS	17,493	FFV_O93209-Pro
GSSEAAAKGGS	17,494	MMTVB_P03365-Pro
PAPAPAPAPAP	17,495	SFV1_P23074_2mutA
GGSEAAAKGGG	17,496	MMTVB_P03365_2mutB_WS
PAPAPAPAP	17,497	MMTVB_P03365_WS
GGGGSGGGGSGGGGSGGGGSGGGGS	17,498	HTL3P_Q4U0X6_2mut
PAPGGSEAAAK	17,499	SFV1_P23074-Pro_2mut
GGSGGGPAP	17,500	MMTVB_P03365
GSSGSSGSSGSSGSSGSS	17,501	MMTVB_P03365-Pro

GGSEAAAKPAP	17,502	SFV1_P23074-Pro
GGGEAAAKGSS	17,503	SFV3L_P27401_2mutA
GGGPAPGGS	17,504	AVIRE_P03360
PAPGGG	17,505	MLVRD_P11227
GGSEAAAKGSS	17,506	SFV3L_P27401_2mut
GGGEAAAKGSS	17,507	FOAMV_P14350_2mut
GGGEAAAKGSS	17,508	SFV1_P23074-Pro
EAAAKEAAAKEAAAKEAAAK	17,509	MLVAV_P03356
EAAAKGGGPAP	17,510	JSRV_P31623_2mutB
EAAAKGGGGSS	17,511	FOAMV_P14350_2mut
EAAAKEAAAKEAAAKEAAAKEAAAK	17,512	SRV2_P51517
GSSGGGGGS	17,513	FFV_O93209
PAPAPAP	17,514	FOAMV_P14350_2mutA
GGSGGSGGSGGS	17,515	FOAMV_P14350
GGGEAAAK	17,516	MMTVB_P03365_WS
GGGGGS	17,517	SFV1_P23074_2mutA
GGSGGS	17,518	WMSV_P03359_3mut
EAAAKGGS	17,519	MMTVB_P03365-Pro
GGGGSS	17,520	BLVJ_P03361_2mut
PAPAP	17,521	MMTVB_P03365-Pro_2mut
PAPGGG	17,522	SMRVH_P03364
EAAAKGGGGSS	17,523	SFV3L_P27401
PAPAPAPAPAP	17,524	MMTVB_P03365
GGGPAP	17,525	MMTVB_P03365-Pro
GSSGGSGGG	17,526	MMTVB_P03365
EAAAKGGGPAP	17,527	FOAMV_P14350_2mutA
GSSGSSGSSGSS	17,528	SFV1_P23074
GGGGSGGGGS	17,529	SFV3L_P27401
GSSGGSGGG	17,530	MLVF5_P26810
GGGEAAAKPAP	17,531	MMTVB_P03365-Pro
PAPEAAAK	17,532	HTLV2_P03363_2mut
GSSGSSGSSGSS	17,533	FOAMV_P14350_2mut
GSSEAAAKPAP	17,534	MMTVB_P03365-Pro
PAPEAAAKGGG	17,535	HTL3P_Q4U0X6_2mut
GGSEAAAKGSS	17,536	MMTVB_P03365-Pro
EAAAKPAPGGS	17,537	MMTVB_P03365_2mut_WS
GSSGGSEAAAK	17,538	MLVF5_P26810_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,539	MLVF5_P26810_3mut
EAAAKGGGGSS	17,540	MMTVB_P03365-Pro
GGGGGSGSS	17,541	HTL1A_P03362_2mutB
PAPAP	17,542	FFV_O93209-Pro_2mut

GGGGGSPAP	17,543	HTL1C_P14078_2mut
GGGPAP	17,544	HTLV2_P03363_2mut
EAAAKGGGGSEAAAK	17,545	SFVCP_Q87040
GGSEAAAKGGG	17,546	FFV_O93209-Pro_2mutA
GSSPAPGGS	17,547	FOAMV_P14350-Pro_2mut
GGGGGGG	17,548	MMTVB_P03365-Pro
EAAAKGSS	17,549	SFV3L_P27401_2mutA
EAAAKGGGGSEAAAK	17,550	MMTVB_P03365-Pro
GGGGSEAAAKGGGGS	17,551	SFV1_P23074-Pro_2mutA
EAAAKGGGGSS	17,552	MMTVB_P03365
GGGEAAAKGGS	17,553	SFV1_P23074
PAPEAAAKGGG	17,554	MLVF5_P26810
GGGGSSGGS	17,555	MMTVB_P03365
GGSGSS	17,556	MMTVB_P03365
PAPAPAPAPAPAP	17,557	KORV_Q9TTC1
EAAAKGGG	17,558	SFV1_P23074-Pro_2mut
PAPAPAPAPAPAP	17,559	SRV2_P51517
GSSGSSGSSGSSGSS	17,560	FFV_O93209-Pro_2mutA
GGGGSS	17,561	FOAMV_P14350_2mut
PAPGGGEAAAK	17,562	MMTVB_P03365_WS
GGSGGGEAAAK	17,563	FFV_O93209-Pro_2mut
PAPAPAPAPAP	17,564	MMTVB_P03365_WS
GGGEAAAKGGS	17,565	MMTVB_P03365-Pro
GGGEAAAKGSS	17,566	MMTVB_P03365_2mutB
GSSPAPEAAAK	17,567	MMTVB_P03365_WS
EAAAKEAAAKEAAAKEAAAKEAAAK	17,568	SFV1_P23074-Pro_2mutA
PAPGGG	17,569	SFV3L_P27401
GSSEAAAKGGG	17,570	MMTVB_P03365_WS
GGGGSSEAAAK	17,571	FOAMV_P14350_2mut
PAPGSSGGS	17,572	SFV1_P23074-Pro_2mut
GSSGSSGSSGSSGSSGSS	17,573	SFV3L_P27401
EAAAKGSSGGG	17,574	MMTVB_P03365
PAPGGGGSS	17,575	WDSV_O92815_2mutA
GGSPAP	17,576	MMTVB_P03365-Pro
GGSGGSGGSGGSGGS	17,577	SFVCP_Q87040-Pro_2mut
PAPAPAPAP	17,578	MMTVB_P03365-Pro
GGGGG	17,579	HTL1A_P03362
GGSGGSGGSGGS	17,580	SFV1_P23074_2mutA
GSSGSSGSSGSSGSS	17,581	FOAMV_P14350-Pro_2mut
PAPGGSEAAAK	17,582	MMTVB_P03365_2mutB_WS
PAPAPAPAP	17,583	SFV1_P23074_2mut

PAPGGGGSS	17,584	MMTVB_P03365
GGSGSS	17,585	SFV3L_P27401_2mut
EAAAKEAAAKEAAAKEAAAK	17,586	MMTVB_P03365_2mut
EAAAKGGSGGG	17,587	HTL3P_Q4U0X6_2mut
PAPGGGGSS	17,588	SFVCP_Q87040-Pro_2mutA
EAAAKGGGGGS	17,589	MLVAV_P03356
GGGGGS	17,590	FOAMV_P14350_2mut
GGGEAAAKGGS	17,591	FFV_O93209-Pro_2mutA
EAAAKPAPGGG	17,592	MMTVB_P03365_2mutB
GGSGGGPAP	17,593	FFV_O93209_2mut
GSSEAAAKPAP	17,594	MMTVB_P03365
PAPAPAPAPAPAP	17,595	SFV1_P23074_2mut
GGSPAPGGG	17,596	MMTVB_P03365-Pro
GGSGGGEAAAK	17,597	MMTVB_P03365
PAPAP	17,598	SFVCP_Q87040
GSSEAAAK	17,599	SFVCP_Q87040
GGGGSGGGGSGGGGS	17,600	MMTVB_P03365-Pro
GSSGSSGSS	17,601	SFV3L_P27401
EAAAKGGSGGG	17,602	MMTVB_P03365-Pro
GSSPAP	17,603	SFV1_P23074_2mut
GGGEAAAK	17,604	SFV1_P23074-Pro
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,605	MMTVB_P03365-Pro
PAPGGS	17,606	HTL1C_P14078_2mut
PAPGSSGGS	17,607	SFV1_P23074_2mut
PAPEAAAK	17,608	MMTVB_P03365_WS
PAPAP	17,609	MMTVB_P03365-Pro
EAAAKGGS	17,610	HTL1A_P03362_2mut
GGGGSEAAAKGGGGS	17,611	HTL1C_P14078
EAAAKGSSGGS	17,612	FOAMV_P14350-Pro
PAPGGSGSS	17,613	MMTVB_P03365-Pro
PAPGGSEAAAK	17,614	SFV1_P23074_2mut
PAPGSSEAAAK	17,615	FFV_O93209-Pro_2mut
PAPGSSGGG	17,616	FOAMV_P14350-Pro_2mutA
GSSGGGEAAAK	17,617	AVIRE_P03360
GGGGGG	17,618	SMRVH_P03364_2mut
PAPEAAAKGGG	17,619	MMTVB_P03365-Pro
GGGEAAAKGGS	17,620	SFVCP_Q87040_2mutA
PAPAPAPAPAP	17,621	SRV2_P51517
GSSGSSGSSGSSGSSGSS	17,622	MMTVB_P03365
EAAAKGGGPAP	17,623	MLVAV_P03356
PAPAPAPAPAP	17,624	FOAMV_P14350-Pro_2mutA

PAPGGSEAAAK	17,625	FOAMV_P14350
GSSGGGPAP	17,626	HTL32_Q0R5R2_2mutB
GGGGGSPAP	17,627	HTL3P_Q4U0X6_2mutB
GSSGGSGGG	17,628	MMTVB_P03365-Pro
PAPAP	17,629	SFVCP_Q87040-Pro
GSSGGGPAP	17,630	MMTVB_P03365-Pro
GGSGSS	17,631	MMTVB_P03365-Pro_2mut
GGSPAPEAAAK	17,632	SFV1_P23074-Pro_2mut
EAAAKGGSGGG	17,633	SFV3L_P27401_2mut
GGGGSSEAAAK	17,634	MMTVB_P03365_WS
GGGGGSGSS	17,635	MMTVB_P03365_2mut
GGGGSSGGS	17,636	SFV1_P23074-Pro_2mutA
EAAAKGGGGSEAAAK	17,637	MMTVB_P03365_WS
PAPGGGEAAAK	17,638	SFV1_P23074-Pro
PAPEAAAKGGG	17,639	MMTVB_P03365
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,640	MMTVB_P03365
GSSGGSEAAAK	17,641	FOAMV_P14350-Pro_2mut
GGSPAP	17,642	MLVBM_Q7SVK7_3mut
GSSEAAAK	17,643	FOAMV_P14350
GSSEAAAK	17,644	MMTVB_P03365-Pro
EAAAKGSSGGS	17,645	HTL1A_P03362_2mut
GGGEAAAKPAP	17,646	FOAMV_P14350-Pro_2mut
EAAAKGGSPAP	17,647	FOAMV_P14350
GSSEAAAKPAP	17,648	MMTVB_P03365_WS
GSSGSSGSS	17,649	FOAMV_P14350_2mut
EAAAKEAAAKEAAAKEAAAK	17,650	MMTVB_P03365_WS
EAAAK	17,651	MMTVB_P03365
PAPGSS	17,652	BAEVM_P10272
PAPGGS	17,653	FFV_O93209-Pro_2mut
GGSGGS	17,654	SFV1_P23074-Pro_2mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,655	HTLV2_P03363_2mut
GGSGGGEAAAK	17,656	MMTVB_P03365_WS
PAPGSSGGG	17,657	HTL1A_P03362
GGSGGS	17,658	SFV3L_P27401-Pro
GSSGSS	17,659	SFV1_P23074-Pro
PAPGGSEAAAK	17,660	MMTVB_P03365
GSAGSAAGSGEF	17,661	MMTVB_P03365-Pro
PAPGGG	17,662	FOAMV_P14350_2mut
EAAAKGGSGSS	17,663	MMTVB_P03365_WS
GSSGGGEAAAK	17,664	SFV3L_P27401-Pro
GGSGGGPAP	17,665	FOAMV_P14350-Pro_2mut

PAPAPAPAPAPAP	17,666	WDSV_O92815
SGSETPGTSESATPES	17,667	SFVCP_Q87040-Pro_2mutA
GGSGGSGGS	17,668	SFV1_P23074
GGGGSS	17,669	SFVCP_Q87040_2mut
GGGGGSEAAAK	17,670	MMTVB_P03365
SGSETPGTSESATPES	17,671	MMTVB_P03365_WS
PAPAPAP	17,672	SFV3L_P27401
PAPEAAAKGSS	17,673	MMTVB_P03365_2mutB_WS
GSSGSSGSSGSSGSS	17,674	SRV2_P51517
GGGPAPGSS	17,675	HTL32_Q0R5R2_2mutB
GGSGGGGSS	17,676	MMTVB_P03365-Pro
SGSETPGTSESATPES	17,677	SRV2_P51517
EAAAKGSSGGS	17,678	MMTVB_P03365-Pro
GSSPAPEAAAK	17,679	MMTVB_P03365-Pro
GSSPAPEAAAK	17,680	SRV2_P51517
GGGGSSPAP	17,681	MMTVB_P03365-Pro
PAPGGGEAAAK	17,682	SFV1_P23074-Pro_2mutA
PAPEAAAKGGS	17,683	MMTVB_P03365
GSSGSSGSSGSSGSSGSS	17,684	FOAMV_P14350-Pro
GGSPAPGSS	17,685	SFV3L_P27401
GGGPAPGGS	17,686	SFV1_P23074-Pro_2mutA
GGGPAPGSS	17,687	MMTVB_P03365-Pro
EAAAKPAP	17,688	MLVBM_Q7SVK7
EAAAKEAAAKEAAAK	17,689	HTL1C_P14078
GSSGGSEAAAK	17,690	SRV2_P51517
PAPGGGGGS	17,691	SRV2_P51517
GGGEAAAK	17,692	FFV_O93209-Pro_2mut
EAAAKGGGPAP	17,693	HTL32_Q0R5R2
GGSGSSGGG	17,694	MMTVB_P03365
PAPEAAAKGSS	17,695	MMTVB_P03365-Pro
PAPGGGGGS	17,696	MMTVB_P03365-Pro
EAAAKGGGGGS	17,697	MMTVB_P03365_WS
GGGGGS	17,698	MMTVB_P03365-Pro
GGGGSGGGGSGGGGSGGGGSGGGGS	17,699	HTL1C_P14078
EAAAKGGSPAP	17,700	MMTVB_P03365
GGGGSSPAP	17,701	FFV_O93209-Pro_2mut
GGGGSSGGS	17,702	MMTVB_P03365-Pro
PAPGSSGGS	17,703	MMTVB_P03365-Pro
GGGGGS	17,704	SRV2_P51517
GGSGSSGGG	17,705	MMTVB_P03365
GSSGGSEAAAK	17,706	MMTVB_P03365-Pro

EAAAKEAAAKEAAAKEAAAK	17,707	GALV_P21414
GGSEAAAKGGG	17,708	MMTVB_P03365-Pro
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	17,709	MMTVB_P03365-Pro
GSSEAAAKGGS	17,710	MMTVB_P03365
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,711	HTL3P_Q4U0X6_2mutB
GGGEAAAK	17,712	MMTVB_P03365-Pro
PAPAPAPAP	17,713	MMTVB_P03365-Pro
PAPGSSGGG	17,714	MMTVB_P03365
GSSGSSGSSGSSGSS	17,715	GALV_P21414
GGSPAP	17,716	MMTVB_P03365_WS
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,717	MMTVB_P03365-Pro
PAPEAAAK	17,718	MMTVB_P03365-Pro
PAPGSSGGG	17,719	SFV1_P23074-Pro_2mutA
GGGGGSEAAAK	17,720	MMTVB_P03365_2mutB_WS
PAPAPAPAPAP	17,721	MMTVB_P03365-Pro
EAAAKGGSGSS	17,722	MMTVB_P03365-Pro
EAAAKEAAAKEAAAKEAAAK	17,723	MLVRD_P11227_3mut
PAPAPAPAP	17,724	FOAMV_P14350_2mutA
GGGPAPGSS	17,725	SFVCP_Q87040_2mut
PAPEAAAKGSS	17,726	SFVCP_Q87040_2mut
GGSPAPGGG	17,727	MMTVB_P03365-Pro
GGGGSGGGGSGGGGSGGGGS	17,728	MMTVB_P03365
EAAAKGGS	17,729	HTL3P_Q4U0X6_2mut
PAPGSSGGS	17,730	MMTVB_P03365_WS
GGGGSGGGGS	17,731	MMTVB_P03365
GGSGGS	17,732	FOAMV_P14350
EAAAKGGGGSEAAAK	17,733	SFVCP_Q87040-Pro_2mut
EAAAKEAAAKEAAAKEAAAK	17,734	MMTVB_P03365-Pro_2mutB
PAPGGGEAAAK	17,735	SFVCP_Q87040-Pro
GSSGSS	17,736	JSRV_P31623_2mutB
EAAAKGGGGGS	17,737	MMTVB_P03365_2mut_WS
GSSPAPEAAAK	17,738	MMTVB_P03365-Pro
GGGEAAAK	17,739	HTL1C_P14078
PAPEAAAKGSS	17,740	HTL32_Q0R5R2_2mutB
GGGGSSEAAAK	17,741	MMTVB_P03365-Pro
PAPGSSGGS	17,742	MMTVB_P03365-Pro
EAAAKGGGGGS	17,743	MMTVB_P03365
GGGGSGGGGSGGGGSGGGGS	17,744	MMTVB_P03365
EAAAKGGGGSS	17,745	HTL3P_Q4U0X6_2mut
GGGEAAAKGGS	17,746	SFVCP_Q87040-Pro
GGGGGSPAP	17,747	MMTVB_P03365-Pro_2mutB

GGSGGGEAAAK	17,748	SFV3L_P27401-Pro
PAPGGGGGS	17,749	SFV3L_P27401-Pro
EAAAKGGGGSEAAAK	17,750	MMTVB_P03365
PAPEAAAKGSS	17,751	MMTVB_P03365-Pro
GGSEAAAKGGG	17,752	MMTVB_P03365-Pro
GGSGGSGGSGGSGGS	17,753	SMRVH_P03364_2mutB
GGSGGSGGSGGSGGS	17,754	HTL1L_P0C211_2mut
GGGGGG	17,755	WDSV_O92815
GGGGGSGSS	17,756	MMTVB_P03365-Pro
GGSEAAAKPAP	17,757	SFV3L_P27401-Pro_2mut
GGGPAPGSS	17,758	MMTVB_P03365_2mut_WS
GGGGGS	17,759	MMTVB_P03365_WS
GGSPAPEAAAK	17,760	MMTVB_P03365
PAPEAAAKGGS	17,761	HTL1A_P03362
EAAAKGGSGSS	17,762	MMTVB_P03365_2mut_WS
GGGPAPEAAAK	17,763	SFV3L_P27401-Pro_2mut
PAPGGGGSS	17,764	HTL32_Q0R5R2_2mut
GSSPAPGGG	17,765	HTL3P_Q4U0X6_2mut
GGGGSSGGS	17,766	BLVAU_P25059_2mut
EAAAKGGGGGS	17,767	HTL1L_P0C211
GGSEAAAKGSS	17,768	JSRV_P31623_2mutB
GSSGGG	17,769	JSRV_P31623
GGSGGSGGSGGS	17,770	MMTVB_P03365-Pro
EAAAKPAP	17,771	SFV1_P23074-Pro_2mutA
GGGGSSGGS	17,772	MMTVB_P03365_WS
GGSGGS	17,773	MMTVB_P03365_WS
EAAAKGGGGGS	17,774	MMTVB_P03365-Pro
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	17,775	MMTVB_P03365
GGSGGSGGS	17,776	MMTVB_P03365
GGGGGSEAAAK	17,777	MLVBM_Q7SVK7
GGSGSSPAP	17,778	MMTVB_P03365_WS
EAAAKEAAAKEAAAK	17,779	JSRV_P31623
PAPEAAAKGGS	17,780	MMTVB_P03365-Pro
GGSGSSEAAAK	17,781	FOAMV_P14350
GGGGGSGSS	17,782	MMTVB_P03365-Pro_2mut
GGGPAPGGS	17,783	MMTVB_P03365
SGSETPGTSESATPES	17,784	SFVCP_Q87040_2mut
GSSPAPGGS	17,785	SFV1_P23074-Pro_2mutA
GSSGSSGSSGSSGSS	17,786	MMTVB_P03365
EAAAKGGGPAP	17,787	MMTVB_P03365
GSSGGG	17,788	MMTVB_P03365_2mut_WS

GGGEAAAKPAP	17,789	MMTVB_P03365
PAPGGSGGG	17,790	MMTVB_P03365-Pro
GSSGGSGGG	17,791	WDSV_O92815_2mut
GGSGGG	17,792	HTL32_Q0R5R2_2mut
EAAAKGGSPAP	17,793	HTLV2_P03363_2mut
GGSPAPEAAAK	17,794	MMTVB_P03365-Pro
GSSGGSEAAAK	17,795	MMTVB_P03365_2mut
GSAGSAAGSGEF	17,796	MMTVB_P03365_WS
PAPGGSGSS	17,797	FFV_O93209
GGSEAAAKGGG	17,798	MMTVB_P03365
GGSPAPGSS	17,799	MMTVB_P03365-Pro
GSSGGSGGG	17,800	SFV3L_P27401
PAPEAAAKGGG	17,801	HTL1A_P03362_2mutB
GGGEAAAKPAP	17,802	MMTVB_P03365-Pro
GGSEAAAK	17,803	HTL32_Q0R5R2_2mutB
GGGEAAAKGSS	17,804	MPMV_P07572
GGGGGSEAAAK	17,805	MMTVB_P03365-Pro
PAPAPAPAPAP	17,806	SFVCP_Q87040-Pro_2mutA
PAPAPAPAPAP	17,807	HTL1L_P0C211_2mut
GGGGSSGGS	17,808	HTL3P_Q4U0X6
PAPGGSEAAAK	17,809	MMTVB_P03365_2mut_WS
PAPAPAPAPAP	17,810	HTL1A_P03362
EAAAKPAPGGG	17,811	MMTVB_P03365_2mut_WS
GGSEAAAK	17,812	MMTVB_P03365_2mut_WS
GGGEAAAKGSS	17,813	SFV1_P23074-Pro_2mutA
GGSPAPGSS	17,814	MMTVB_P03365-Pro
GGSEAAAKPAP	17,815	MLVBM_Q7SVK7
PAPEAAAKGGG	17,816	MMTVB_P03365_2mut_WS
GSSEAAAKPAP	17,817	MMTVB_P03365-Pro_2mutB
GGGGSEAAAKGGGGS	17,818	MMTVB_P03365-Pro_2mut
GSSEAAAKGGS	17,819	MMTVB_P03365-Pro_2mutB
GSSGSSGSSGSSGSS	17,820	SRV2_P51517_2mutB
GGGGGSPAP	17,821	HTL1L_P0C211_2mut
GGSEAAAK	17,822	MMTVB_P03365
GSSPAPEAAAK	17,823	SMRVH_P03364_2mutB
GGGPAPGGS	17,824	HTL1C_P14078_2mut
GGSPAPEAAAK	17,825	MMTVB_P03365_WS
GGSEAAAKPAP	17,826	HTL1A_P03362_2mut
PAPAPAPAP	17,827	HTLV2_P03363_2mut
GSSPAPGGG	17,828	MMTVB_P03365
GSSGSSGSSGSS	17,829	MMTVB_P03365-Pro

GGSEAAAKGSS	17,830	MMTVB_P03365_WS
GGSGSSGGG	17,831	MMTVB_P03365_2mutB
GSSGSSGSSGSSGSSGSS	17,832	JSRV_P31623_2mutB
GGSEAAAKPAP	17,833	MMTVB_P03365-Pro
GSSGGSGGG	17,834	HTLV2_P03363_2mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	17,835	WDSV_O92815_2mut
GGSPAPEAAAK	17,836	MMTVB_P03365
GGGGSSEAAAK	17,837	MMTVB_P03365
GGSGGGEAAAK	17,838	SFV1_P23074-Pro_2mutA
GGGGSEAAAKGGGGS	17,839	WDSV_O92815_2mut
GGSGSSEAAAK	17,840	MMTVB_P03365_2mutB_WS
GGSEAAAKPAP	17,841	MMTVB_P03365_WS
GSSGGGEAAAK	17,842	SFVCP_Q87040-Pro
GSSGGS	17,843	SFVCP_Q87040-Pro_2mut
GGSEAAAKPAP	17,844	SFVCP_Q87040_2mut
GSSGGSEAAAK	17,845	SFVCP_Q87040_2mut
GSSPAPEAAAK	17,846	SRV2_P51517_2mutB
GGSGGSGGSGGSGGSGGS	17,847	BLVAU_P25059
GSSGSSGSSGSSGSS	17,848	HTL1C_P14078_2mut
EAAAKGGGGSS	17,849	MMTVB_P03365_2mutB
GGGEAAAKGSS	17,850	SFVCP_Q87040-Pro

Example 6: Generation of Exemplary Drivers and Trans Templates

This example describes the generation of exemplary drivers and trans templates for use in the three, four, or five component RNA-based gene modifying systems described in Examples 1-5.

1) Gene modifying polypeptide:

A series of 216 exemplary gene modifying polypeptides were generated containing:

1. a Cas-nuclease with one endonuclease domain inactivated (in this example, Spy N863A Cas9),
2. either of two reverse transcriptases (RTs) (the RTs of the gene modifying polypeptide encoded by: PLV10993 or PLV10990/RNAIVT338), and
3. an RNA binding protein (RBP) containing 1, 2, or 4 RBP repeats (in this example, 1, 2, or 4 MCPv2 domains as provided in Table 31), with each RBP repeat connected to the other RBP repeats by one of 4 linker peptides from Table 10.

These 3 domains were connected using peptide linkers, e.g., as found in Table 10.

The sequence of the RT of the gene modifying polypeptide encoded by PLV10990/RNAIVT338 and the sequence of the RT of the gene modifying polypeptide encoded by PLV10993 are provided below.

PLV10990/RNAIVT338 RT:

TAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPIT
LEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNP
YTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTL
FNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLC
QEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATR
PGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPV
AYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARIT
QYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFT
DGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAF
ATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNR
RADEVAREVAIRPLSTQATIS (SEQ ID NO: 16,706)

PLV10993 RT:

TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYP
MSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVP
NPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKN
SPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKA
QICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAP
LYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGP
WRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDR
WLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPL
PDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKK
LNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHS
AEARGNRMADQAARKAAITETPDTSTLL (SEQ ID NO: 16,707)

The series of 216 exemplary gene modifying polypeptides encompassed a subset of the configurations set forth in Examples 1 and 2:

- Configuration 1: Cas9 – linker $(RBP)_n$ – linker – RT
- Configuration 2: RT – linker – $(RBP)_n$ – linker – Cas9
- Configuration 3: Cas9 – linker – RT – linker – $(RBP)_n$

The sequences of the exemplary gene modifying polypeptides are given below:

5

Table S1. Exemplary gene modifying sequences for Configuration 1 (i.e., Cas9 – linker (RBP)n – linker – RT)

Identifier	Full Amino Acid Sequence	SEQ ID NO:
PL11941	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAK ALEAEAAAKEAAAKEAAAKEAAAKATAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRK FRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLT WTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQA ILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKR PVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHH CLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIY RERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	16,001
PL11942	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK	16,0

WO 2023/039441 PCT/US2022/076064

	ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATP ESSGGSSGGSSTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPW NTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTL FNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAA GWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDL TDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIK NAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
PL11943	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSTAPL EEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYR MVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLD HPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAE LAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLT REASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTD GSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPK RVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	16,003
PL11944	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE	16,0

	NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTAPLEEEYRLF LEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDL REVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSL LQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLY AATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKL TFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRD GKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVM HCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
PL11945	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY	16,005

	AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTAL PVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPE SQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEE VTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFV EETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFK QTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKS VNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGS EFESPKKKAKVE	
PL11946	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY SGGSSGGSSGSETPGTSESATPESSGGSSGGSSTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKR SLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEE GESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSL SNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQ ALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPET DDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATL HVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	16,006
PL11947	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL	16,0

	LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY GSSGSSGSSGSSGSSGSSTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRP VHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGF KNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKT KRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKR LDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSL TSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLT AGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
PL11948	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY GSSGGGEAAAKTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPW NTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTL	16,008

WO 2023/039441 PCT/US2022/076064

	FNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAA GWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDL TDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIK NAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
PL11949	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY GGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIF ATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTI KVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKA TAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGT SEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQG FRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIP GFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAA AALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEAT LFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAV WLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	16,009
PL11950	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL	16,0

	LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY GGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIF ATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTI KVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTAPLEEEYRLFLEAPI QNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVN KRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYV DDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAAT RPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFG QDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKR YAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCK GHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
PL11951	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNG	16,011

WO 2023/039441 PCT/US2022/076064

	GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY GGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIF ATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTI KVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSTAPLEEEYRLFLEAPIQNVTLLEQWKREIPK VWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTL LSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSA TRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEE AFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSP PDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWA EPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRR ADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
PL11952	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY GGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIF ATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTI KVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEIN PPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPD RIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLM TLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKL ALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLT NARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTS	16,012

	AQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAR EVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
.11953	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKEAAAKEAAAKE AAAKEAAAKATAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWN TPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLF NEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREF LGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	16,013
PL11954	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY	16,014

WO 2023/039441 PCT/US2022/076064

	EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTAPL EEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYR MVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLD HPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAE LAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLT REASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTD GSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPK RVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
PL11955	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSTAPLEEEYRLFLEAPIQNVT LLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVET IHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIA ADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGND PLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITS SHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAA VVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKD DAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	16,015
PL11956	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG	16,0

	SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTAPLEEEYRLFLEAPIQNVTLLEQWK REIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVP NPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQA ACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWG EKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLES LLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDS VIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTST GNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
PL11957	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFAN GIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAA AKEAAAKALEAEAAAKEAAAKEAAAKEAAAKATAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKR SLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEE	16,017

	GESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSL SNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQ ALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPET DDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATL HVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
PL11958	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFAN GIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPG TSESATPESSGGSSGGSSTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRP VHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGF KNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKT KRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKR LDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSL TSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLT AGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	16,018
PL11959	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV	16,0

WO 2023/039441 PCT/US2022/076064

	DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFAN GIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSS GSSTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRK SGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDL QGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRL FIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAI AAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQA EATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALL TAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
PL11960	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFAN GIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTAP LEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEY RMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRL DHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGF AELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAAL LTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFT	16,020

	DGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWL PKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
.11961	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFAN GIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSG GGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQ GLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLN MELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATAPLEEEYRLFLEAPI QNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVN KRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYV DDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAAT RPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFG QDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKR YAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCK GHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	16,021
PL11962	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV	16,0

WO 2023/039441 PCT/US2022/076064

	DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFAN GIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSG GGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQ GLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLN MELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTAPLEEEYRLFLEAPIQNVTLLEQWKREIPK VWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTL LSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSA TRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEE AFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSP PDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWA EPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRR ADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
PL11963	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFAN GIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSG GGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQ GLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLN	16,023

	MELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQ APIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDL KDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRV SGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALA LPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQV LLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALT KALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQ ATISKRTADGSEFESPKKKAKVE	
PL11964	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFAN GIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSG GGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQ GLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLN MELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQL LSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCI PLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQ LCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKP FQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPR VRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWS KDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRT ADGSEFESPKKKAKVE	16,024
PL11965	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE	16,0

	NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKATAPL EEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYR MVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLD HPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAE LAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLT REASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTD GSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPK RVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
PL11966	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTAPLEEEYRLFLEAPIQNVT LLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVET	16,026

	IHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIA ADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGND PLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITS SHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAA VVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKD DAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
PL11967	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAE INPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPP DRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLL MTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSL KLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKW LTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIG TSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVA REVAIRPLSTQATISKRTADGSEFESPKKKAKVE	16,027
PL11968	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV	16,0

	DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGL ASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWY SVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAE LGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQ PPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARI TQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQK AELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAI RPLSTQATISKRTADGSEFESPKKKAKVE	
PL11969	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAY KVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAA KEAAAKEAAAKEAAAKATAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRP VHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGF KNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKT KRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKR LDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSL	16,029

	TSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLT AGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
.11970	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAY KVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSG GSSTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRK SGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDL QGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRL FIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAI AAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQA EATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALL TAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	16,030
PL11971	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD	16,03

WO 2023/039441 PCT/US2022/076064

	SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAY KVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSTAPLEEEYRLFLE APIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLRE VNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQ YVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYA ATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLT FGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRD GKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVM HCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
PL11972	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAY KVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTAPLEEEYRLFLEAPIQNV TLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVE TIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLI AADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGN DPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEI TSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGA AVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQK DDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	16,032

WO 2023/039441 PCT/US2022/076064

PL11973	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAY KVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAAN SGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELI VKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATAPLEEEYRLFLEAPIQNVTLLEQWKREIPK VWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTL LSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSA TRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEE AFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSP PDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWA EPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRR ADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	16,033
PL11974	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK	16,0

	ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAY KVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAAN SGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELI VKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQA PIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLK DAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVS GKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALAL PSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVL LLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALT KALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQ ATISKRTADGSEFESPKKKAKVE	
PL11975	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAY KVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAAN SGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELI VKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVR QYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFA	16,035

	FEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLG FKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSG AAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAAL NPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYT DSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESP KKKAKVE	
PL11976	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAY KVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAAN SGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELI VKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITL EAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWA DAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHK GSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKG VLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATL LPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYA FATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAK VE	16,036
PL11977	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG	16,0

	SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATAPLEEEYRLF LEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDL REVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSL LQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLY AATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKL TFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRD GKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVM HCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
PL11978	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTAPLEEEYRLFLEAPIQNVTLLEQWK REIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVP NPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQA ACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWG	16,038

	EKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLES LLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDS VIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTST GNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
.11979	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGL ASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWY SVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAE LGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQ PPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARI TQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQK AELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAI RPLSTQATISKRTADGSEFESPKKKAKVE	16,039
PL11980	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK	16,0

WO 2023/039441 PCT/US2022/076064

	ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVVAEINPPGLASTQA PIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLK DAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVS GKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALAL PSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVL LLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALT KALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQ ATISKRTADGSEFESPKKKAKVE	
PL11981	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSV RQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAK EAAAKEAAAKATAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPW NTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTL FNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAA GWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDL TDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIK NAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	16,041

PL11982	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSV RQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTAP LEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEY RMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRL DHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGF AELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAAL LTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFT DGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWL PKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	16,042
PL11983	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH	16,043

	QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSV RQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSTAPLEEEYRLFLEAPIQNV TLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVE TIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLI AADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGN DPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEI TSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGA AVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQK DDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
PL11984	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSV RQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTAPLEEEYRLFLEAPIQNVTLLEQW KREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVP NPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQA ACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWG EKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLES LLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDS VIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTST GNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	16,044
PL11985	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS	16,0

	DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSV RQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGT GDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGG SGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQ GLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEIN PPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPD RIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLM TLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKL ALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLT NARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTS AQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAR EVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
PL11986	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH	16,04

WO 2023/039441 PCT/US2022/076064

	QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSV RQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGT GDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGG SGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQ GLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQL LSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCI PLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQ LCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKP FQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPR VRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWS KDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRT ADGSEFESPKKKAKVE	
PL11987	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSV RQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGT GDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGG SGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQ GLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITL EAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWA DAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHK GSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKG VLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATL	16,047

	LPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYA FATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAK VE	
.11988	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSV RQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGT GDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGG SGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQ GLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSL RETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGE SGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSN SRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALG PWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDT LPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVH GMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISKRTADGSEFESPKKKAKVE	
PL11989	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV	16,0

	DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAK ALEAEAAAKEAAAKEAAAKEAAAKATLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHI QRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGI SGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLT EARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQ KLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPAT LLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTD SRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAK VE	
PL11990	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATP ESSGGSSGGSSTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQS PWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTP KTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLS KKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNC	16,050

WO 2023/039441 PCT/US2022/076064

	LDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEI YRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
.11991	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSTLNIE DEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTN DYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLA DFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFC RLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLR MVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTD QPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIK NKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	16,051
PL11992	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY	16,052

WO 2023/039441 PCT/US2022/076064

	EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTLNIEDEHRL HETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPV QDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQ HPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIP GFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAA IAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPD ADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEIL ALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL11993	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKATLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKAT STPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCL RLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQ ICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDL TKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQ ALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAE LIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPD TSTLLKRTADGSEFESPKKKAKVE	16,053
PL11994	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE	16,0

	NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY SGGSSGGSSGSETPGTSESATPESSGGSSGGSSTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEA RLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEW RDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKE GQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGP VVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEG KKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSE FESPKKKAKVE	
PL11995	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS	16,055

	SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY GSSGSSGSSGSSGSSGSSTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGI LVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWT RLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETV MGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWR RPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEE GLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFAT AHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL11996	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY GSSGGGEAAAKTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQS PWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTP KTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLS KKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNC LDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEI YRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	16,056
PL11997	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEI FSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNL	16,0

	LAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKF IKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKED YFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSR KLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENI VIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSI DNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDS PTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVN FLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVD NGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNP IPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYL NMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAE WISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMAS NFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAM QGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATLNIEDEHRLHETSKEPDVSLGSTWLS DFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNK RVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQ HPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGK AGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKL DPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEG LQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYT DSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGS EFESPKKKAKVE	
PL11998	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV	16,0

	DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY GGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIF ATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTI KVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTLNIEDEHRLHETSKE PDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLRE VNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLIL LQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEM AAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTK DAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHT WYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKA LFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL11999	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY GGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIF ATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTI	

	KVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSTLNIEDEHRLHETSKEPDVSLGSTWLSDFP QAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNP YNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELD CQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNW GPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAP HAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKA GAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKG HSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL12000	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWIS SNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY GGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIF ATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTI KVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAE TGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSG LPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGT RALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQ KAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEA LVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVT TETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEA RGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL12001	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE	16,0

	NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKE AAAKEAAAKATLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSP WNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFK NSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSK KLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCL DILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIY RRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL12002	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTLNIE DEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTN	16,062

	DYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLA DFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFC RLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLR MVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTD QPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIK NKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL12003	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSTLNIEDEHRLHETSKEPDVS LGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKR VEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYV DDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLY PLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKL TMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDG SSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	16,063
PL12004	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV	16,0

	DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTLNIEDEHRLHETSKEPDVSLGSTW LSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHP TVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLA ATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKP GTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMG QPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLL QEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIH CPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL12005	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFAN GIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAA AKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQE ARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFE WRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLL KEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQ GYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQF GPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMA	16,065

	EGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADG SEFESPKKKAKVE	
.12006	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFAN GIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPG TSESATPESSGGSSGGSSTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGI LVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWT RLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETV MGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWR RPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEE GLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFAT AHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	16,066
PL12007	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD	16,067

	SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFAN GIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSS GSSTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNE ALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREF LGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAG WPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGT RPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTS EGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL12008	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFAN GIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTLNI EDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGT NDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDL ADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGF CRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCL RMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLT DQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEI KNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	16,068

PL12009	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFAN GIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSG GGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQ GLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLN MELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKATLNIEDEHRLHETSKE PDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLRE VNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLIL LQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEM AAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTK DAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHT WYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKA LFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	16,069
PL12010	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK	16,0

	ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFAN GIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSG GGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQ GLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLN MELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTLNIEDEHRLHETSKEPDVSLGSTWLSDFP QAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNP YNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELD CQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNW GPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAP HAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKA GAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKG HSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL12011	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFAN GIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSG GGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQ GLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLN MELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAV RQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWY	16,071

WO 2023/039441 PCT/US2022/076064

	TVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGN LGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQA LLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDR WLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAK ALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMAD QAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL12012	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVR QSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFAN GIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSG GGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQ GLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLN MELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLII PLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLK DAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRA SAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAP ALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNA RMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAG TSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARK AAITETPDTSTLLKRTADGSEFESPKKKAKVE	16,072
PL12013	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG	16,0

	SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATLNIE DEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTN DYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLA DFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFC RLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLR MVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTD QPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIK NKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL12014	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTLNIEDEHRLHETSKEPDVS LGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKR VEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYV DDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLY	16,074

WO 2023/039441 PCT/US2022/076064

	PLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKL TMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDG SSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKR LSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
.12015	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAW AETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLL SGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQ GTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPD QQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAV EALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAA VTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSA EARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	16,075
PL12016	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK	16,0

	ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGM GLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSH QWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQ EIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQ PPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEV IWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNR MADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL12017	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAY KVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAA KEAAAKEAAAKEAAAKATLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGI LVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWT RLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETV MGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWR RPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEE GLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFAT AHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	16,077

PL12018	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAY KVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSG GSSTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP VKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNE ALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREF LGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAG WPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGT RPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTS EGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	16,078
PL12019	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH	16,07

WO 2023/039441 PCT/US2022/076064

	QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAY KVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSTLNIEDEHRLHET SKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDL REVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDL ILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAE MAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVL TKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADH TWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLK ALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL12020	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAY KVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTLNIEDEHRLHETSKEPDV SLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNK RVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQY VDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAP LYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAG KLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTD GSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPK RLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL12021	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS	16,0

	DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAY KVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAAN SGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELI VKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKATLNIEDEHRLHETSKEPDVSLGSTWLSDFPQ AWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPY NLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDC QQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWG PDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPH AVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAG AAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGH SAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL12022	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH	16,082

	QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAY KVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAAN SGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELI VKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVR QAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNL GYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQAL LTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRW LSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKA LPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQ AARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL12023	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAY KVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAAN SGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELI VKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSI KQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPT SQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQ VKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFE LFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLD	16,083

	TDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALT QALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL KRTADGSEFESPKKKAKVE	
.12024	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEV PKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAY KVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAAN SGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELI VKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPM SQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFA FEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGY LLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEK QGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQ FGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKM AEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTAD GSEFESPKKKAKVE	16,084
PL12025	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE	16,0

WO 2023/039441 PCT/US2022/076064

	HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATLNIEDEHRL HETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPV QDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQ HPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIP GFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAA IAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPD ADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEIL ALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL12026	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTLNIEDEHRLHETSKEPDVSLGSTW LSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHP TVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLA ATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKP GTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMG QPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLL	

	QEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIH CPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
.12027	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGM GLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSH QWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQ TLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQ EIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQ PPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEV IWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNR MADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	16,087
PL12028	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY	16,088

WO 2023/039441 PCT/US2022/076064

	EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVR QAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNL GYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQAL LTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRW LSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKA LPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQ AARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL12029	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSV RQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAK EAAAKEAAAKATLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQS PWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTP KTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLS KKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNC LDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEI YRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	16,089
PL12030	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS	16,0

	DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSV RQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTLNI EDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGT NDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDL ADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGF CRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCL RMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLT DQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEI KNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL12031	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSV RQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSTLNIEDEHRLHETSKEPDV	

	SLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNK RVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQY VDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAP LYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAG KLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTD GSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPK RLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL12032	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSV RQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTLNIEDEHRLHETSKEPDVSLGST WLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDI HPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLL LAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTK PGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTM GQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSL LQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSII HCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL12033	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV	16,0

	LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSV RQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGT GDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGG SGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQ GLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAET GGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGL PPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTR ALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQK AYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEAL VKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTT ETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEAR GNRMADQAARKAAITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL12034	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSV RQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGT	16,094

	GDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGG SGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQ GLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIP LKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKD AFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRAS AKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPA LGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNAR MTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGT SAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKA AITETPDTSTLLKRTADGSEFESPKKKAKVE	
PL12035	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSV RQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGT GDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGG SGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQ GLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPM SQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFA FEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGY LLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEK QGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQ FGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKM AEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTAD GSEFESPKKKAKVE	16,095

PL12036	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLS DILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNG SIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSL LYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIV LTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHE HIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQI TKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAW RSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSV RQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGT GDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGG SGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQ GLLKDGNPIPSAIAANSGIYGSSGGGEAAAKTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARL GIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRD PEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEG QRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYA KGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPV VALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGK KLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLKRTADGSEFE SPKKKAKVE	16,096

WO 2023/039441 PCT/US2022/076064

Table S2. Exemplary gene modifying sequences for Configuration 2 (i.e., RT – linker – (RBP)n – linker – Cas9)

Identifier	Full Amino Acid Sequence	SEQ ID NO:
.12037	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEA AAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNS DCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKADKKYSIGLDIGTNSVGWAVITDEYKVP SKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTI YHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNL IALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKY KEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRK VTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSR KLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN QTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDN VPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGR DFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDK VLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,100
PL12038	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEA AAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNS DCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSI KKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKA	16,

	DLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAG YIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTR KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKT ILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNY WRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQ VNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKL PKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIRE QAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12039	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLAEAAAKEAAAKEAAAKEAAAKEAAAKALEAEAAAKEA AAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNS DCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFR GHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY DDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIK PILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVV DKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRF NASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYP KLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM LASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,102
PL12040	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF	16,

	KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEA AAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNS DCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGEL QKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKY FDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12041	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEA AAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNS DCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVP KGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKADKKYSI GLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKK HERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSK SRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL	16,104

	FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVD ELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL KDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYL DEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGS EFESPKKKAKVE	
PL12042	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEA AAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNS DCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVP KGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSDKKYSIGLDIGTNSVGWAVIT DEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKA RGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,105
PL12043	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI	16,

	LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEA AAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNS DCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVP KGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFA WMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKK MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH HAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12044	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEA AAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNS DCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVP KGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSI KKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKA DLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAG YIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTR KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKT ILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNY	16,107

	WRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQ VNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKL PKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIRE QAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12045	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEA AAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNS DCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVP KGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEW ISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDN GGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGI YAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGE TAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFI KPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVV DKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRF NASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYP KLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM LASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,108
PL12046	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI	16,

	LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEA AAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNS DCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVP KGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEW ISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDN GGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGI YSGGSSGGSSGSETPGTSESATPESSGGSSGGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIK DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKE DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI ARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12047	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEA AAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNS DCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVP KGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEW ISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDN GGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGI YGSSGSSGSSGSSGSSGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQ LFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI	16,110

	LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFD NGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILE DIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM YVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF DSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12048	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEA AAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNS DCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVP KGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEW ISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDN GGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGI YGSSGGGEAAAKDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPI NASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILR VNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVA QILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY SVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP	16,111

	EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE TRIDLSQLGGDKRTADGSEFESPKKKAKVE	
.12049	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLSGGSSGGSSGSETPGTSESATPESSGGSSGGSS MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLK DGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHS IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKA DLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAG YIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTR KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKT ILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNY WRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQ VNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKL PKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIRE QAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12050	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLSGGSSGGSSGSETPGTSESATPESSGGSSGGSS MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLK DGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFI	

WO 2023/039441 PCT/US2022/076064

	KPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVV DKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRF NASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYP KLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM LASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12051	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLSGGSSGGSSGSETPGTSESATPESSGGSSGGSS MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLK DGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIK DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKE DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI ARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,114
PL12052	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK	16,

	LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLSGGSSGGSSGSETPGTSESATPESSGGSSGGSS MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLK DGNPIPSAIAANSGIYGSSGGGEAAAKDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLF IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFL AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDL LRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNL PNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKM IAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYV NFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK EVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12053	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLSGGSSGGSSGSETPGTSESATPESSGGSSGGSS MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLK DGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELT IPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKADKKYSIGLDIGTNSVGWAVI TDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVA YHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKK NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVR QQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVD LLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIV IEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKA	16,116

	RGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIV WDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEK NPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILA DANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12054	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLSGGSSGGSSGSETPGTSESATPESSGGSSGGSS MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLK DGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELT IPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLG NTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFA WMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKK MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH HAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12055	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLSGGSSGGSSGSETPGTSESATPESSGGSSGGSS MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLK	16,

	DGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELT IPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKL QLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQ EEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPW NFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIS GVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIK ELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKL ITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGT ALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEV QTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFT LTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12056	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLSGGSSGGSSGSETPGTSESATPESSGGSSGGSS MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLK DGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELT IPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFI KPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVV DKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRF NASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYP KLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM	16,119

	LASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
.12057	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLSGGSSGGSSGSETPGTSESATPESSGGSSGGSS MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLK DGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELT IPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCS VRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFA NGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAA AKEAAAKALEAEAAAKEAAAKEAAAKEAAAKADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIK DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKE DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI ARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,120
PL12058	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLSGGSSGGSSGSETPGTSESATPESSGGSSGGSS MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLK	16,

WO 2023/039441 PCT/US2022/076064

	DGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12059	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSSGSSGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL	16,122

	KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQ TVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHI VPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKY DENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKT EITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKG KSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV EQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD KRTADGSEFESPKKKAKVE	
PL12060	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLSGGSSGGSSGSETPGTSESATPESSGGSSGGSS MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLK DGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELT IPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCS VRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFA NGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKDK KYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEE DKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSAR LSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMI KRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQE DFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTY AHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKS KKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKR TADGSEFESPKKKAKVE	
PL12061	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF	16,

	KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFI KPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVV DKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRF NASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYP KLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM LASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12062	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYS GGSSGGSSGSETPGTSESATPESSGGSSGGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR RKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSD VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQ YADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVK LNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTN FDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDK DFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQ KAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKL	16,125

	YLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDK AGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDV RKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKK DWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPS KYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYT STKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12063	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY GSSGSSGSSGSSGSSGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNE MAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQ LFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAI LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFD NGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILE DIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM YVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF DSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,126
PL12064	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY	16,

	GSSGGGEAAAKDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVD DSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPI NASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILR VNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVA QILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYF FYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY SVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSP EDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE TRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12065	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY GGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKA MQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVL GNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK LVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGL TPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQ SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSR FAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQL KEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKG QKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVV KKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEV	

WO 2023/039441 PCT/US2022/076064

	KKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
.12067	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY GGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKA MQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRT ARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDL NPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLL AQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDG TEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQS FIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGEL QKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKY FDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,130
PL12068	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY GGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKA MQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRY TRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDN	16,13

WO 2023/039441 PCT/US2022/076064

	SDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIK DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKE DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI ARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12069	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY GGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKA MQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIK VEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQ AYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEA AAKEAAAKEAAAKEAAAKADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT FDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLP KHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM YVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF DSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH	16,132

WO 2023/039441 PCT/US2022/076064

	YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
.12070	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY GGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKA MQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIK VEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQ AYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSS GGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKA PLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMI EERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKK GILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRM NTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQK QLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ LGGDKRTADGSEFESPKKKAKVE	16,133
PL12071	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY	16,

WO 2023/039441 PCT/US2022/076064

	GGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKA MQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIK VEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQ AYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSDKKYSIGLDIGT NSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLEN LIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQ DLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDN REKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLS GEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVM KQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSID NKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVI TLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPL IETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQI SEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPK KKAKVE	
PL12072	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIY GGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKA MQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIK VEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQ AYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKDKKYSIGLDIGTNSVG WAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKI LTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKR	16,135

WO 2023/039441 PCT/US2022/076064

	RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKP ENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTR SDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLV SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGET GEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSS FEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI LADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12073	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGGGEAAAKMASNFTQFVLVDNGGTGDVT VAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKE AAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR TARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEG DLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMD GTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQ SFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDD SLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVEN TQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKR NSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGEL QKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKY FDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,136
PL12074	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGGGEAAAKMASNFTQFVLVDNGGTGDVT	16,

	VAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGG SSGSETPGTSESATPESSGGSSGGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICY LQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQ LVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLA AKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLL RKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLP NEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNE ENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ NGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIA KSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNF LYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVL DATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12075	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGGGEAAAKMASNFTQFVLVDNGGTGDVT VAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSG SSGSSGSSGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDS FFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPIN ASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRV NTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQI HLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFT VYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFE DREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAG SPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDIN RLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQI LDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED	16,138

WO 2023/039441 PCT/US2022/076064

	NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDKRTADGSEFESPKKKAKVE	
.12076	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGGGEAAAKMASNFTQFVLVDNGGTGDVT VAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGE AAAKDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE ESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDA KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKA PLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMI EERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKK GILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRM NTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQK QLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ LGGDKRTADGSEFESPKKKAKVE	16,139
PL12077	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGGGEAAAKMASNFTQFVLVDNGGTGDVT VAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGS MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLK DGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHS IKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKA DLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNF	16,14

WO 2023/039441 PCT/US2022/076064

	DLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAG YIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTR KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI ECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKT ILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNY WRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHD AYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQ VNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKL PKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIRE QAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12078	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGGGEAAAKMASNFTQFVLVDNGGTGDVT VAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGS MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLK DGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET AEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKF RGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDT YDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFI KPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVV DKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRF NASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQIL KEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYP KLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFS KESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM LASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGA PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12079	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF	16,

	KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGGGEAAAKMASNFTQFVLVDNGGTGDVT VAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGS MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLK DGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYT RRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS DVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGD QYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLV KLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMT NFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIK DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKE DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI ARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12080	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGGGEAAAKMASNFTQFVLVDNGGTGDVT VAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGS MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLK DGNPIPSAIAANSGIYGSSGGGEAAAKDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLF IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFL AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDL LRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNL PNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD	16,143

	NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKM IAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYV NFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK EVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12081	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGGGEAAAKMASNFTQFVLVDNGGTGDVT VAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGS MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLK DGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKG AWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTC SVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAA AKEAAAKEAAAKADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP HQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYE YFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIAN LAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHV AQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKY FFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVA YSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGS PEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLY ETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12082	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF	16,

	KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGGGEAAAKMASNFTQFVLVDNGGTGDVT VAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGGS MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLK DGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKG AWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTC SVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSD KKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSA RLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRR QEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTV KVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV PQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKS KKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKR TADGSEFESPKKKAKVE	
PL12083	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGGGEAAAKMASNFTQFVLVDNGGTGDVT VAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGS MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLK DGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKG AWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTC SVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSDKKYSIGLDIGTNSVG	16,14

	WAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKI LTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKR RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKP ENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTR SDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLV SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGET GEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSS FEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI LADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12084	MPAAKRVKLDGGTLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQ SPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKK LDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDI LAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRR RGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLGSSGGGEAAAKMASNFTQFVLVDNGGTGDVT VAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGS MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLK DGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKG AWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTC SVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKDKKYSIGLDIGTNSVGWAVITDE YKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGL FGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYY VGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMA RENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQ FYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK GRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDF	16,147

	LEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLD KVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
.12109	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAA AKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQG LLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTD RHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDST DKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNF KSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKN GYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFA WMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKK MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH HAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,148
PL12110	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAA AKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQG LLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDS GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHM IKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY	16,14

WO 2023/039441 PCT/US2022/076064

	KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIK KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG GFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12111	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAA AKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQG LLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR YTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIG DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIK DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKE DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI ARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,150
PL12112	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG	16,

	WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISAEAAAKEAAAKEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAA AKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQG LLKDGNPIPSAIAANSGIYGGSSGGGEAAAKDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRK NRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD KLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD KNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKA QVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY YLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSK YVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12113	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISAEAAAKEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAA AKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQG LLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNM ELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKADKKYSIGLDIGTNSVGW AVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLP GEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLK ALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKI LTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKK AIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKR RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKP ENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTR	16,152

	SDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLV SDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGET GEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSS FEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVI LADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12114	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAA AKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQG LLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNM ELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFK VLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLR KKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSL GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARG NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTV KQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN GIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSE EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGY KEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAY NKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,153
PL12115	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAA AKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQG	16,

	LLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNM ELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIG ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYL ALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG ASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVK KTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLF ELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENII HLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12116	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAA AKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQG LLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNM ELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDS GETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHM IKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIK KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG GFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK	

WO 2023/039441 PCT/US2022/076064

	RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
.12117	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAA AKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQG LLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNM ELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKV TCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPS NFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAK EAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEE LLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER MTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK EDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS ELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKL IARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,156
PL12118	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAA AKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQG	16,

WO 2023/039441 PCT/US2022/076064

	LLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNM ELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKV TCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPS NFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGS ETPGTSESATPESSGGSSGGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIF SNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQT YNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR TFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDI LEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDS LHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD MYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT LIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12119	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAA AKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQG LLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNM ELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKV TCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPS NFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGS SGSSGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASG VDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEI TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE	16,158

	MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAI KKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD YDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNE QKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL SQLGGDKRTADGSEFESPKKKAKVE	
PL12120	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISAEAAAKEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAA AKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQG LLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNM ELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKV TCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPS NFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAA KDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL SARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSA SMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRR QEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTV KVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV PQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKS KKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKR TADGSEFESPKKKAKVE	
PL12121	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT	16,

	LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANS GIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSG ETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIK KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG GFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12122	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANS GIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR YTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIG DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIK DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKE DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN	16,161

WO 2023/039441 PCT/US2022/076064

	EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI ARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12123	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISSGGSSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANS GIYGSSGSSGSSGSSGSSGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT FDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLP KHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM YVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF DSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,162
PL12124	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISSGGSSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANS	16,

	GIYGSSGGGEAAAKDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAK VDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSD ILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLT LTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE LDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSIT GLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12125	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANS GIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIV KAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFK VLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLR KKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSL GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARG NSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTV KQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLIN GIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSE EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGY	16,164

	KEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAY NKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
.12126	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANS GIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIV KAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIG ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYL ALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG ASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVK KTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLF ELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENII HLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,165
PL12127	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANS GIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIV KAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE	16,16

WO 2023/039441 PCT/US2022/076064

	GDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLD NLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGA SAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIL PKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAA FKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12128	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANS GIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIV KAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARR RYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNP DNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQI GDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEE LLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER MTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFK EDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQ NEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS ELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKL IARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTID RKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,167

WO 2023/039441 PCT/US2022/076064

PL12129	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANS GIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIV KAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKY TIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNS RSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALE AEAAAKEAAAKEAAAKEAAAKADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQE IFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQ TYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQ RTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEK VLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENE DILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG DSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGR DMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLV ETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSE QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKY GGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYL ASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDA TLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,168
PL12130	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANS GIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIV KAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKY	16,

	TIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNS RSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESS GGSSGGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFF HRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNT EITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHL GELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY NELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFED REMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSP AIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL SDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILD SRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSN IMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLV VAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDN EQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGDKRTADGSEFESPKKKAKVE	
PL12131	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANS GIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIV KAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKY TIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNS RSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSDKKYSIGL DIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDE HHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFL KDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDD KVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELV KVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD	16,170

	DSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12132	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISSGGSSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSGGSGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,171
PL12133	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG	16,

	WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSN FANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKE AAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR YTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIG DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIK DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKE DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI ARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12134	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSN FANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSE TPGTSESATPESSGGSSGGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT FDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLP KHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM YVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG	16,173

	KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF DSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
.12135	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSN FANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSS GSSGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRL EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVD AKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGEL HAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREM IEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKK GILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRM NTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMN FFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAK VEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQK QLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ LGGDKRTADGSEFESPKKKAKVE	16,174
PL12136	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSN FANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAK DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLV EEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS	16,

	ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRR QEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTV KVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV PQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKS KKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKR TADGSEFESPKKKAKVE	
PL12137	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSN FANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNF TQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIP SAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIG ALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYL ALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG ASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETI TPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDS VEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLL NAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNA VVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVK KTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLF ELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENII HLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,176

PL12138	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSN FANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNF TQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIP SAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRL KRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIE GDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLD NLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEK MDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGA SAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL GTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLI HDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHP VENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLES EFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIL PKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASA GELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAA FKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,177
PL12139	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSN FANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNF TQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIP SAIAANSGIYGSSGSSGSSGSSGSSGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRI CYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLF IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFL AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDL	16,17

	LRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNL PNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLD NEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQ VSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY LQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFI KRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKM IAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYV NFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK EVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12140	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSN FANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNF TQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIP SAIAANSGIYGSSGGGEAAAKDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIF SNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQT YNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQR TFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVL PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDI LEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDS LHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD MYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLA SHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDAT LIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12141	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE	16,

	FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSN FANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNF TQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIP SAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYL NMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSS AQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAA KEAAAKADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASG VDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEI TKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLG ELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE MIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAI KKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSD YDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSR MNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNI MNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVV AKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNE QKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDL SQLGGDKRTADGSEFESPKKKAKVE	
PL12142	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSN FANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNF TQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIP SAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYL NMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSS AQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSDKKYSIGL DIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHE	16,18

WO 2023/039441 PCT/US2022/076064

	RHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDE HHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFL KDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKP AFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDD KVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELV KVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKD DSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIRE VKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIR KRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVK ELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSE FESPKKKAKVE	
PL12143	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSN FANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNF TQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIP SAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYL NMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSS AQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSDKKYSIGLDIGTNSVGWAVITDE YKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE KYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGL FGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQL PEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYY VGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFK TNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMA RENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKS DNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQ FYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK GRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDF	16,182

	LEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLD KVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
.12144	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSN FANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNF TQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIP SAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYL NMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSS AQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKDKKYSIGLDIGTNSVGWAVITDEYKVPSK KFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIY HLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLI ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYK EIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLA RGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKV TVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRK LINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQ TTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVP SEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKV REINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAK GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLS AYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,183
PL12145	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIA EWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKE AAAKALEAEAAAKEAAAKEAAAKEAAAKADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRYTRRK	16,

WO 2023/039441 PCT/US2022/076064

	NRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD KLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYA DLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLN REDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD KNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDF LDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKA QVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLY YLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAG FIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKD WDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSK YVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTS TKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12146	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIA EWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGSETPGTSE SATPESSGGSSGGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAK VDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSD ILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLT LTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE LDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSIT GLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,185

PL12147	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIA EWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSS DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLV EEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRR QEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTV KVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV PQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKS KKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKR TADGSEFESPKKKAKVE	16,186
PL12148	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIA EWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKDKKYSIG LDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKH ERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKS RRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFY PFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGM RKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHL	16,187

WO 2023/039441 PCT/US2022/076064

	FDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVD ELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFL KDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLI REVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANG EIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKS VKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYL DEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGS EFESPKKKAKVE	
PL12149	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIA EWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANS GIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKADKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSG ETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMI KFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK DTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFE EVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELG SQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQ RKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIK KYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTG GFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRK RMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN LGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,188
PL12150	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT	16,

	DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIA EWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANS GIYSGGSSGGSSGSETPGTSESATPESSGGSSGGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR YTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPD NSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIG DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERM TNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIK DKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKE DIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSE LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLI ARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12151	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIA EWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANS GIYGSSGSSGSSGSSGSSGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFS NEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTY NQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT FDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLP KHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL HEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDM YVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETR	16,190

	QITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF DSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASH YEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIH QSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12152	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIA EWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANS GIYGSSGGGEAAAKDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAK VDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEE NPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSD ILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSI PHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLY EYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLT LTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIA NLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE LDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH VAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATA KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLK GSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSIT GLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	
PL12153	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIA EWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLV	16,

	DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANS GIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIP IFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYT IKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLV EEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILS ARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSAS MIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRR QEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTV KVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIV PQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDE NDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKS KKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKR TADGSEFESPKKKAKVE	
PL12154	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIA EWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANS GIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIP IFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYT IKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYSGGSSGGSSGGSETPGTSESATPESSGGSSGGSSDKKYSIGLDIGTNSV GWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNI VDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQ LPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTL LKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKI EKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQ KKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQL	16,193

WO 2023/039441 PCT/US2022/076064

	KRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRH KPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVL TRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETN GETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIME RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFS KRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKA KVE	
PL12155	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIA EWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANS GIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIP IFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYT IKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGSSGSSGSSGSSGSSDKKYSIGLDIGTNSVGWAVITDEYKVPSKKF KVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHL RKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIAL SLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEI FFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLAR GNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVT VKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLI NGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSE EVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVRE INNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFAT VRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGY KEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAY NKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	16,194
PL12156	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSP WNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPT LFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVRE	16,

	FLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAG WPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLT DQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKN APEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIA EWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLV DNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANS GIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIP IFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYT IKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGSSGGGEAAAKDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGN TDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTP NFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFA WMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE DYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDK QSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQK NSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKK MKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYH HAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHR DKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDKRTADGSEFESPKKKAKVE	

WO 2023/039441 PCT/US2022/076064

Table S3. Exemplary gene modifying sequences for Configuration 3 (i.e., Cas9 – linker – RT – linker – (RBP)n)

Identifier	Full Amino Acid Sequence	SEQ ID NO:
PL12085	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIH VQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAF FCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKA QLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDK PFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPP RVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEW SKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISAEA AAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNR KYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYKRTADGSEFESPKKKAKVE	16,200
PL12086	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL	16,

	FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIH VQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAF FCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKA QLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDK PFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPP RVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEW SKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISAEA AAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNR KYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSN SRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYKRTADGSEFESPKKKAKVE	
PL12087	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIH VQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAF FCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKA	16,202

	QLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDK PFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPP RVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEW SKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISAEA AAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNR KYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSN SRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFT QFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPS AIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNS DCELIVKAMQGLLKDGNPIPSAIAANSGIYKRTADGSEFESPKKKAKVE	
PL12088	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKATAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIH VQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAF FCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKA QLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDK PFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPP RVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEW SKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISSGG SSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRS YLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYKRTADGSEFESPKKKAKVE	16,203

PL12089	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIH VQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAF FCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKA QLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDK PFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPP RVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEW SKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISSGG SSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRS YLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQS SAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYKRTADGSEFESPKKKAKVE	16,204
PL12090	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY	16,2[illegible]

	SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIH VQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAF FCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKA QLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDK PFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPP RVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEW SKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISSGG SSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRS YLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQS SAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVT VAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGS MASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKD GNPIPSAIAANSGIYKRTADGSEFESPKKKAKVE	
PL12091	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIH VQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAF FCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKA QLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDK	16,206

WO 2023/039441 PCT/US2022/076064

	PFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPP RVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEW SKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSS GSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNS DCELIVKAMQGLLKDGNPIPSAIAANSGIYKRTADGSEFESPKKKAKVE	
PL12092	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIH VQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAF FCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKA QLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDK PFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPP RVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEW SKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSS GSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNS DCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVP KGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYKRTADGSEFESPKKKAKVE	16,207
PL12093	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH	16,

	QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIH VQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAF FCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKA QLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDK PFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPP RVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEW SKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSS GSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNS DCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVP KGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWI SSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNG GTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYK RTADGSEFESPKKKAKVE	
PL12094	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV	16,2(

	LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIH VQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAF FCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKA QLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDK PFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPP RVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEW SKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSS GGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVK AMQGLLKDGNPIPSAIAANSGIYKRTADGSEFESPKKKAKVE	
PL12095	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIH VQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAF FCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKA QLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDK PFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPP RVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEW SKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSS GGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVK	16,210

	AMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRS YLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYKRTADGSEFESPKKKAKVE	
.12096	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIH VQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAF FCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKA QLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDK PFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPP RVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAELIALTKALEW SKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISGSS GGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVK AMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRS YLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRS QAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDV TVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYKRTADG SEFESPKKKAKVE	16,211
PL12097	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH	16,

	QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQA PLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDL KDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRA SAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPA LGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNAR MTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTS AQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAI TETPDTSTLLAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKV TCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYKRTADGSEFESPKKKAKVE	
PL12098	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQA PLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDL	16,213

	KDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRA SAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPA LGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNAR MTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTS AQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAI TETPDTSTLLAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKV TCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPS NFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYKRTADGSEFES PKKKAKVE	
PL12099	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQA PLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDL KDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRA SAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPA LGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNAR MTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTS AQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAI TETPDTSTLLAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKAMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKV TCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPS NFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGG GSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKA	16,214

	MQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSY LNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYKRTADGSEFESPKKKAKVE	
.12100	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQA PLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDL KDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRA SAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPA LGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNAR MTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTS AQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAI TETPDTSTLLSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYT IKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYKRTADGSEFESPKKKAKVE	
PL12101	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI	16,

	NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQA PLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDL KDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRA SAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPA LGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNAR MTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTS AQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAI TETPDTSTLLSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYT IKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRS QAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYKRTADGSEFESPKKKAKVE	
PL12102	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQA PLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDL KDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRA SAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPA LGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNAR	16,217

WO 2023/039441 PCT/US2022/076064

	MTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTS AQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAI TETPDTSTLLSGGSSGGSSGSETPGTSESATPESSGGSSGGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYT IKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRS QAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGGSGGGSGGGSMASNFTQFV LVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAA NSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCEL IVKAMQGLLKDGNPIPSAIAANSGIYKRTADGSEFESPKKKAKVE	
PL12103	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQA PLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDL KDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRA SAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPA LGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNAR MTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTS AQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAI TETPDTSTLLGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYL NMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYKRTADGSEFESPKKKAKVE	
PL12104	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN	16,

	PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQA PLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDL KDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRA SAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPA LGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNAR MTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTS AQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAI TETPDTSTLLGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYL NMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSS AQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYKRTADGSEFESPKKKAKVE	
PL12105	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED	16,22

	NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQA PLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDL KDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRA SAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPA LGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNAR MTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTS AQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAI TETPDTSTLLGSSGSSGSSGSSGSSGSSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYL NMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSS AQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTV APSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSM ASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDG NPIPSAIAANSGIYKRTADGSEFESPKKKAKVE	
PL12106	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQA PLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDL KDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRA SAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPA LGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNAR MTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTS	16,221

	AQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAI TETPDTSTLLGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIP IFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYKRTADGSEFESPKKKAKVE	
.12107	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQA PLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDL KDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRA SAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPA LGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNAR MTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTS AQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAI TETPDTSTLLGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIP IFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTI KVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYKRTADGSEFESPKKKAKVE	16,222
PL12108	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKV DDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDIL RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPH QIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEY FTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTL	16,

	FEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLA GSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDI NRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQ ILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFY SNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPED NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI DLSQLGGDAEAAAKEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKEAAAKATLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQA PLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDL KDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRA SAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPA LGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNAR MTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTS AQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAI TETPDTSTLLGSSGGGEAAAKMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIP IFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTI KVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGGSGGGSGGGSGGGSMASNFTQFVLVDNGGTGDVTVAPSNFAN GIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQF VLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIA ANSGIYKRTADGSEFESPKKKAKVE	

WO 2023/039441 PCT/US2022/076064

2) Guide RNA:

The guide RNA (gRNA) molecule (5'-GCCGAAGCACTGCACGCCGTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC -3'; SEQ ID NO: 16,708) binds Cas9 and contains a pro-spacer sequence complementary to the target locus, resulting in localization of the gene modifying complex to the target genomic locus.

3) Trans-template RNA:

A series of 384 ttRNAs were constructed (288 experimental configurations and 96 negative controls) with each of the following regions:

1. a primer binding site (PBS) that is 8, 13, or 17 nucleotides in length that basepairs with the nicked DNA strand allowing primer extension of the nicked DNA, followed immediately by
2. a homology 1 region with homology to the target BFP/GFP locus
3. the exemplary desired modification: a C to T substitution (the edit) which converts the BFP target to GFP, as well as a PAM-nullifying mutation to prevent multiple turnover cutting at the target site.
4. either a 12 nucleotide or 63 nucleotide homology 2 region with homology to the target BFP/GFP locus.
5. an RBP recruitment site (**RRS**) containing 1, 2, or 4 MS2 sequences (1x, 2x, or 4x MS2) with intervening RNA linker sequences.
6. either a 3' end-protecting RNA secondary structure comprising ePEG with an 8 nt linker sequence or no 3' end-protecting RNA secondary structure.
7. either no 5' end-protecting RNA secondary structure or one of the following 5'end-protecting RNA secondary structures:
    a. ePEG
    b. a Cas9 scaffold lacking a spacer sequence
    c. a Cas9 scaffold with a 16 nucleotide spacer sequence

8. either an 8 or 16 nucleotide linker between the RRS and PBS

These components were assembled in the below configuration (if no 3' or 5' end protecting structure was selected, then no such structure was included in that template RNA):

3'-end block - $(RRS)_n$ – linker – PBS-template – 5' end block

Template sequences are included in Table S4 below.

Table S4. Exemplary template sequences

Identifier	Template sequence	SEQ ID NO:
.13280	GACCACCCTGACGTACGGCGTGCAGACATACAAgcACATGAGGATCACCCATGTgc	16,301
PL13281	GACCACCCTGACGTACGGCGTGCAGACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,302
PL13282	GACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,303
PL13283	GACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,304
PL13284	GACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,305
PL13285	GACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,306
PL13286	GACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgcACATGAGGATCACCCATGTgc	16,307
PL13287	GACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,308
PL13288	GACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,309
PL13289	GACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,310
PL13290	GACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,311
PL13291	GACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,312
PL13292	GACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgcACATGAGGATCACCCATGTgc	16,31
PL13293	GACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,31
PL13294	GACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,31
PL13295	GACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,31
PL13296	GACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,31

PL13297	GACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctct ctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGG ATAACCCTCAAAGTCGGGGGGC	16,318
PL13298	GACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgcACATGAGGATCACCCATGTgc	16,319
PL13299	GACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCT GAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,320
PL13300	GACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtctt ccc	16,321
PL13301	GACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtctt cccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,322
PL13302	GACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtctt ccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,323
PL13303	GACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtctt ccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTG AACCCAGGATAACCCTCAAAGTCGGGGGGC	16,324
PL13304	GACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgcACATGAGGATCACCCATGTgc	16,325
PL13305	GACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAAC CCAGGATAACCCTCAAAGTCGGGGGGC	16,326
PL13306	GACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,327
PL13307	GACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTC TCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,328
PL13308	GACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctct ctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,329
PL13309	GACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctct ctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCC AGGATAACCCTCAAAGTCGGGGGGC	16,330
PL13310	GACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgcACATGAGGATCACCCATGTgc	16,33
PL13311	GACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCC CCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,33
PL13312	GACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGg agtcttccc	16,33
PL13313	GACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGg agtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,33
PL13314	GACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGg agtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,33

PL13315	GACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,336
PL13316	GACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgcACATGAGGATCACCCATGTgc	16,337
PL13317	GACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,338
PL13318	GACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,339
PL13319	GACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,340
PL13320	GACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,341
PL13321	GACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,342
PL13322	GACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgcACATGAGGATCACCCATGTgc	16,343
PL13323	GACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,344
PL13324	GACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcc	16,345
PL13325	GACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,346
PL13326	GACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,347
PL13327	GACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,348
PL13328	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgcACATGAGGATCACCCATGTgc	16,34
PL13329	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,35
PL13330	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,35
PL13331	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,35

PL13332	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGagtcttccc	16,353
PL13333	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,354
PL13334	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgcACATGAGGATCACCCATGTgc	16,355
PL13335	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,356
PL13336	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,357
PL13337	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,358
PL13338	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,359
PL13339	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,360
PL13340	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgcACATGAGGATCACCCATGTgc	16,361
PL13341	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,362
PL13342	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGagtcttccc	16,363
PL13343	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,364
PL13344	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGagtcttccc	16,365

PL13345	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcA CATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGT CACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,366
.13346	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACAT AgcACATGAGGATCACCCATGTgc	16,367
PL13347	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACAT AgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,368
PL13348	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACAT AgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,369
PL13349	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACAT AgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCC TCAAAGTCGGGGGGC	16,370
PL13350	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACAT AgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactc CCACAGTCACTGGGgagtcttccc	16,371
PL13351	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACAT AgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactc CCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,372
PL13352	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgcA CATGAGGATCACCCATGTgc	16,373
PL13353	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgcA CATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,374
PL13354	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAggg agcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,375
PL13355	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAggg agcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAA AGTCGGGGGGC	16,376
PL13356	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAggg agcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCA CAGTCACTGGGgagtcttccc	16,37
PL13357	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAggg agcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCA CAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,37
PL13358	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACT AACATAgcACATGAGGATCACCCATGTgc	16,37

PL13359	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACT AACATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,380
PL13360	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACT AACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,381
PL13361	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACT AACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGAT AACCCTCAAAGTCGGGGGGC	16,382
PL13362	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACT AACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGT gcgactcCCACAGTCACTGGGgagtcttccc	16,383
PL13363	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACT AACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGT gcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,384
PL13364	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgcACATG AGGATCACCCATGTgc	16,385
PL13365	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgcACATG AGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,386
PL13366	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcAC ATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,387
PL13367	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcAC ATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTC GGGGGGC	16,388
PL13368	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcAC ATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTC ACTGGGgagtcttccc	16,389
PL13369	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcAC ATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTC ACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,390
PL13370	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACAT AgcACATGAGGATCACCCATGTgc	16,391
PL13371	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACAT AgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,392
PL13372	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACAT AgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,393

PL13373	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACAT AgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCC TCAAAGTCGGGGGGC	16,394
PL13374	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACAT AgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactc CCACAGTCACTGGGgagtcttccc	16,395
PL13375	GAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACAT AgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactc CCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,396
PL13376	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGACATACAAgcACATGAGGATCACCCATGTgc	16,397
PL13377	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGACATACAAgcACATGAGGATCACCCATGTgcTCT CTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,398
PL13378	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTg cgactcCCACAGTCACTGGGgagtcttccc	16,399
PL13379	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTg cgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,400
PL13380	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTg cgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,401
PL13381	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTg cgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCT CTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,402
PL13382	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgcACATGAGGATCACCC ATGTgc	16,403
PL13383	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgcACATGAGGATCACCC ATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,404
PL13384	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgggagcACATGAGGATCA CCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,40
PL13385	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgggagcACATGAGGATCA CCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,40
PL13386	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgggagcACATGAGGATCA CCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtc ttccc	16,40
PL13387	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgggagcACATGAGGATCA CCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtc ttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,40

PL13388	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgcACATGAGGATCACCCATGTgc	16,409
PL13389	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,410
PL13390	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,411
PL13391	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,412
PL13392	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,413
PL13393	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,414
PL13394	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgcACATGAGGATCACCCATGTgc	16,415
PL13395	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,416
PL13396	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,417
PL13397	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,418
PL13398	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,419
PL13399	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,42(
PL13400	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgcACATGAGGATCACCCATGTgc	16,42
PL13401	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,42
PL13402	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,42

PL13403	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,424
PL13404	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,425
PL13405	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,426
PL13406	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgcACATGAGGATCACCCATGTgc	16,427
PL13407	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,428
PL13408	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,429
PL13409	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,430
PL13410	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,431
PL13411	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,432
PL13412	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgcACATGAGGATCACCCATGTgc	16,433
PL13413	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,434
PL13414	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,435
PL13415	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,436
PL13416	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,437

PL13417	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctcgggtcaggagccccccccctgaacccaggataaccctcaaagtcgggggggc	16,438
PL13418	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgcACATGAGGATCACCCATGTgc	16,439
PL13419	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,440
PL13420	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,441
PL13421	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,442
PL13422	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,443
PL13423	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,444
PL13424	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgcACATGAGGATCACCCATGTgc	16,445
PL13425	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,446
PL13426	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,447
PL13427	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,448
PL13428	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,449
PL13429	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,450

PL13430	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGACATACAACTAACATAgcACATGAGGATCACCCATGTgc	16,451
PL13431	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGACATACAACTAACATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATA ACCCTCAAAGTCGGGGGGC	16,452
PL13432	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,453
PL13433	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGG TCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,454
PL13434	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctc tctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,455
PL13435	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctc tctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAAC CCTCAAAGTCGGGGGGC	16,456
PL13436	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTACATACAAgcACATGAGGATCACCCATGTgc	16,457
PL13437	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCC TCAAAGTCGGGGGGC	16,458
PL13438	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,459
PL13439	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCA GGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,460
PL13440	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctc gggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,46
PL13441	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctc gggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCT CAAAGTCGGGGGGC	16,46

PL13442	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgcACATGAGGATCACCCATGTgc	16,463
PL13443	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAG GATAACCCTCAAAGTCGGGGGGC	16,464
PL13444	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,465
PL13445	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCT CGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,466
PL13446	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctct ctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,467
PL13447	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctct ctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGG ATAACCCTCAAAGTCGGGGGGC	16,468
PL13448	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTCGGCACATACAAgcACATGAGGATCACCCATGTgc	16,469
PL13450	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,471
PL13451	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGG GTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,472
PL13452	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctc tctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,473
PL13453	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctc tctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAA CCCTCAAAGTCGGGGGGC	16,47
PL13454	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgcACATGAGGATCACCCATGTgc	16,47

PL13455	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAAC CCAGGATAACCCTCAAAGTCGGGGGGC	16,476
PL13456	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,477
PL13457	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTC TCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,478
PL13458	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctct ctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,479
PL13459	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctct ctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCC AGGATAACCCTCAAAGTCGGGGGGC	16,480
PL13460	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGTCCTGTGAGCTTACATACAAgcACATGAGGATCACCCATGTgc	16,481
PL13461	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGTCCTGTGAGCTTACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCC TCAAAGTCGGGGGGC	16,482
PL13462	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGTCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,483
PL13463	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGTCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCA GGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,484
PL13464	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGTCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcg ggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,485
PL13465	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGTCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcg ggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTC AAAGTCGGGGGGC	16,486
PL13466	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgcACATGAGGATCACCCATGTgc	16,487

PL13467	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAG GATAACCCTCAAAGTCGGGGGGC	16,488
PL13468	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,489
PL13469	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCT CGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,490
PL13470	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctct ctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,491
PL13471	CGCGGTTCTATCTAGTTACGCGTTAAACCAACTAGAAGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCC TGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctct ctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGG ATAACCCTCAAAGTCGGGGGGC	16,492
PL13472	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGACATACAAgcACATGAGGATCACCCATGTgc	16,493
PL13473	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACC CTCAAAGTCGGGGGGC	16,494
PL13474	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,495
PL13475	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTC AGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,496
PL13476	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctct cgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,497
PL13477	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctct cgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCC TCAAAGTCGGGGGGC	16,498
PL13478	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgcACATGAGGATCACCCATGTgc	16,499

PL13479	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCA GGATAACCCTCAAAGTCGGGGGGC	16,500
PL13480	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,501
PL13481	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCT CTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,502
PL13482	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctct ctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,503
PL13483	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctct ctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAG GATAACCCTCAAAGTCGGGGGGC	16,504
PL13484	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgcACATGAGGATCACCCATGTgc	16,505
PL13485	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGA TAACCCTCAAAGTCGGGGGGC	16,506
PL13486	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,507
PL13487	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTC GGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,508
PL13488	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctc tctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,509
PL13489	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctc tctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATA ACCCTCAAAGTCGGGGGGC	16,510
PL13490	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgcACATGAGGATCACCCATGTgc	16,511

PL13491	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGA ACCCAGGATAACCCTCAAAGTCGGGGGGC	16,512
PL13492	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,513
PL13493	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc TCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,514
PL13494	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc tctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,515
PL13495	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc tctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAAC CCAGGATAACCCTCAAAGTCGGGGGGC	16,516
PL13496	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgcACATGAGGATCACCCATGTgc	16,517
PL13497	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCC AGGATAACCCTCAAAGTCGGGGGGC	16,518
PL13498	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,519
PL13499	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTC TCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,520
PL13500	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctc tctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,521
PL13501	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctc tctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAG GATAACCCTCAAAGTCGGGGGGC	16,522
PL13502	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgcACATGAGGATCACCCATGTgc	16,523

PL13503	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCC CTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,524
PL13504	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagt cttccc	16,525
PL13505	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagt cttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,526
PL13506	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagt cttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,527
PL13507	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagt cttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCT GAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,528
PL13508	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgcACATGAGGATCACCCATGTgc	16,529
PL13509	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGA TAACCCTCAAAGTCGGGGGGC	16,530
PL13510	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,531
PL13511	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCG GGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,532
PL13512	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctc tctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,53
PL13513	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctc tctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATA ACCCTCAAAGTCGGGGGGC	16,53
PL13514	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgcACATGAGGATCACCCATGTgc	16,53

PL13515	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGA ACCCAGGATAACCCTCAAAGTCGGGGGGC	16,536
PL13516	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,537
PL13517	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc TCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,538
PL13519	CAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA CCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccct ctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAAC CCAGGATAACCCTCAAAGTCGGGGGGC	16,540
PL13520	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgcACATGAGGATC ACCCATGTgc	16,541
PL13521	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgcACATGAGGATC ACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,542
PL13522	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGG ATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,543
PL13523	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGG ATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,544
PL13524	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGG ATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGG gagtcttccc	16,545
PL13525	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGG ATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGG gagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,546
PL13526	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgcACAT GAGGATCACCCATGTgc	16,547

PL13527	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgcACAT GAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,548
13528	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgggagc ACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,549
PL13529	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgggagc ACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAG TCGGGGGGC	16,550
PL13530	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgggagc ACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAG TCACTGGGgagtcttccc	16,551
PL13532	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgcACATGAG GATCACCCATGTgc	16,553
PL13533	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgcACATGAG GATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,554
PL13534	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACAT GAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,555
PL13535	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACAT GAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGG GGGGC	16,556
PL13536	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACAT GAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCAC TGGGgagtcttccc	16,55
PL13537	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACAT GAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCAC TGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,55

PL13538	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAg cACATGAGGATCACCCATGTgc	16,559
PL13539	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAg cACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,560
PL13540	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAg ggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,561
PL13541	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAg ggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTC AAAGTCGGGGGGC	16,562
PL13542	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAg ggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCC ACAGTCACTGGGgagtcttccc	16,563
PL13543	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAg ggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCC ACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,564
PL13544	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgcACA TGAGGATCACCCATGTgc	16,565
PL13545	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgcACA TGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,566
PL13546	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggag cACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,56
PL13547	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggag cACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAA GTCGGGGGGC	16,56
PL13548	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggag	16,56

	cACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACA GTCACTGGGgagtcttccc	
.13549	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggag cACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACA GTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,570
PL13550	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAAC ATAgcACATGAGGATCACCCATGTgc	16,571
PL13551	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAAC ATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,572
PL13552	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAAC ATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,573
PL13553	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAAC ATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAAC CCTCAAAGTCGGGGGGC	16,574
PL13554	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAAC ATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcga ctcCCACAGTCACTGGGgagtcttccc	16,575
PL13555	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAAC ATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcga ctcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,576
PL13556	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgcACATGAG GATCACCCATGTgc	16,57
PL13557	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgcACATGAG GATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,57
PL13558	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcACAT GAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,57

PL13559	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcACAT GAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGG GGGGC	16,580
PL13560	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcACAT GAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCAC TGGGgagtcttccc	16,581
PL13561	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcACAT GAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCAC TGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,582
PL13562	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAg cACATGAGGATCACCCATGTgc	16,583
PL13563	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAg cACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,584
PL13564	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAg ggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,585
PL13565	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAg ggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTC AAAGTCGGGGGGC	16,586
PL13566	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAg ggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCC ACAGTCACTGGGgagtcttccc	16,587
PL13567	AGGGCGATGCCACCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGA AGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAg ggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCC ACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,588
PL13568	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGACATACAAgcACATGAGGATCACCCATGTgc	16,589

PL13569	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,590
PL13570	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,591
PL13571	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGA ACCCAGGATAACCCTCAAAGTCGGGGGGC	16,592
PL13572	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGAT CACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,593
PL13573	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGAT CACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,594
PL13574	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGACATACAACTAACATAgcACATGAGGATCACCCATGTgc	16,595
PL13575	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGACATACAACTAACATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGT CGGGGGGC	16,596
PL13576	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,597
PL13577	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCC CCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,598
PL13578	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcAC ATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,599
PL13579	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcAC ATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTC GGGGGGC	16,60(
PL13580	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTACATACAAgcACATGAGGATCACCCATGTgc	16,60

PL13581	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGG GGGGC	16,602
PL13582	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,603
PL13583	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCC CCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,604
PL13584	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATG AGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,605
PL13585	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATG AGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGG GGGC	16,606
PL13586	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTACATACAACTAACATAgcACATGAGGATCACCCATGTgc	16,607
PL13587	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTACATACAACTAACATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTC AAAGTCGGGGGGC	16,608
PL13588	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,609
PL13589	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGG AGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,610
PL13590	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggg agcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,61[illegible]
PL13591	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggg agcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAA AGTCGGGGGGC	16,61[illegible]
PL13592	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTCGGCACATACAAgcACATGAGGATCACCCATGTgc	16,61[illegible]

PL13593	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTCGGCACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAG TCGGGGGGC	16,614
PL13594	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,615
PL13595	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCC CCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,616
PL13596	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcA CATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,617
PL13597	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcA CATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGT CGGGGGGC	16,618
PL13598	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTCGGCACATACAACTAACATAgcACATGAGGATCACCCATGTgc	16,619
PL13599	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTCGGCACATACAACTAACATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAAC CCTCAAAGTCGGGGGGC	16,620
PL13600	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,621
PL13601	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGT CAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,622
PL13602	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctc tcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,623
PL13603	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG CGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctc tcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCC TCAAAGTCGGGGGGC	16,624
PL13604	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG TCCTGTGAGCTTACATACAAgcACATGAGGATCACCCATGTgc	16,625

PL13605	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG TCCTGTGAGCTTACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGG GGGGC	16,626
.13606	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG TCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,627
PL13607	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG TCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCC CTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,628
PL13608	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG TCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGA GGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,629
PL13609	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG TCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGA GGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGG GGC	16,630
PL13610	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG TCCTGTGAGCTTACATACAACTAACATAgcACATGAGGATCACCCATGTgc	16,631
PL13612	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG TCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,633
PL13613	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG TCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGA GCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,634
PL13614	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG TCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggg agcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,635
PL13615	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGACCACCCTGACGTACGG TCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggg agcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAA AGTCGGGGGGC	16,636
PL13616	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgcACATGAGGATCACCCATGTgc	16,63
PL13617	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTC GGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,63

PL13618	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactc CCACAGTCACTGGGgagtcttccc	16,639
PL13619	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactc CCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,640
PL13620	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactc CCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,641
PL13621	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAAgggagcACATGAGGATCACCCATGTgcgactc CCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCG GGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,642
PL13622	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgcACATGAGGATCACCCATGTgc	16,643
PL13623	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgcACATGAGGATCACCCATGTgc TCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,644
PL13624	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgggagcACATGAGGATCACCCAT GTgcgactcCCACAGTCACTGGGgagtcttccc	16,645
PL13625	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgggagcACATGAGGATCACCCAT GTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,646
PL13626	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgggagcACATGAGGATCACCCAT GTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,647
PL13627	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGACATACAACTAACATAgggagcACATGAGGATCACCCAT GTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccT CTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,648
PL13628	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgcACATGAGGATCACCCATGTgc	16,649
PL13629	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgcACATGAGGATCACCCATGTgcTCT CTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,650

PL13630	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTg cgactcCCACAGTCACTGGGgagtcttccc	16,651
PL13631	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTg cgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,652
PL13632	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTg cgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,653
PL13633	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAAgggagcACATGAGGATCACCCATGTg cgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCT CTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,654
PL13634	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgcACATGAGGATCACCCA TGTgc	16,655
PL13635	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgcACATGAGGATCACCCA TGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,656
PL13636	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCA CCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,657
PL13637	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCA CCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,658
PL13638	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCA CCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtc ttccc	16,659
PL13639	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTACATACAACTAACATAgggagcACATGAGGATCA CCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtc ttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,66
PL13640	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgcACATGAGGATCACCCATGT gc	16,66

PL13641	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgcACATGAGGATCACCCATGT gcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,662
.13642	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCA TGTgcgactcCCACAGTCACTGGGgagtcttccc	16,663
PL13643	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCA TGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,664
PL13644	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCA TGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,665
PL13645	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAAgggagcACATGAGGATCACCCA TGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccT CTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,666
PL13646	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgcACATGAGGATC ACCCATGTgc	16,667
PL13647	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgcACATGAGGATC ACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,668
PL13648	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGG ATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,669
PL13649	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGG ATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,670
PL13650	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGG ATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGG gagtcttccc	16,67
PL13651	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGCGTGCAGTGCTTCGGCACATACAACTAACATAgggagcACATGAGG	16,67

	ATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	
.13652	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgcACATGAGGATCACCCATGTgc	16,673
.13653	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,674
PL13654	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,675
PL13655	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,676
PL13656	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,677
PL13657	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,678
PL13658	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgcACATGAGGATCACCCATGTGc	16,679
PL13659	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgcACATGAGGATCACCCATGTgcTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,680
PL13660	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,68
PL13661	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,68
PL13662	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACCGGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtcttccc	16,68

PL13663	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCGAAGTTCATCTGTACCACC GGCAAACTGCCCGTGCCTTGGCCCACCCTCGTCACCACCCTGACGTACGGTCCTGTGAGCTTACATACAACTAACATAgggagcACATGAGGATCA CCCATGTgcgactcCCACAGTCACTGGGgagtcttccctctctctctctctctctcgggagcACATGAGGATCACCCATGTgcgactcCCACAGTCACTGGGgagtc ttcccTCTCTCTCGGGTCAGGAGCCCCCCCCTGAACCCAGGATAACCCTCAAAGTCGGGGGGC	16,684

Example 7: Evaluation of Rewriting Activity of Exemplary Gene Modifying Polypeptides Using Cis Template RNAs

This example demonstrates the rewrite activity of exemplary gene modifying systems generated in Example 6 using the benchmark cis template RNA and normalized to exemplary gene modifying polypeptide PLV4921.

Before evaluating the rewriting activities of the gene modifying polypeptides of Example 6 using trans templates, 214 of the polypeptides were evaluated using a 'cis' template RNA containing both guide RNA and a 3'-extension comprising the PBS and template sequences, where the template is comprised of homology 1, edit and homology 2. Each exemplary gene modifying system (the combination of exemplary gene modifying polypeptide and the 'cis' template RNA) was tested to determine its genome-editing capacity using mammalian U2OS cells carrying a genomic landing pad that expresses a BFP-GFP SNP reporter. The U2OS cells are BFP-expressing and can be converted to a GFP-expressing cell line via a single base pair substitution programmed by the cis template RNA. The U2OS cells were electroporated with each exemplary gene modifying system, in the form of plasmid encoding an individual exemplary gene modifying polypeptide and a second plasmid encoding the cis template RNA.

Cells were analyzed by flow cytometry 4 days post-electroporation of the gene modifying system. The percentage of GFP cells was used as a measure for rewriting activity and compared to the rewriting activity of a benchmark gene modifying system containing an exemplary gene modifying polypeptide and the cis template RNA (see e.g., workflow shown in FIG. 7A).

The sequence of the benchmark gene modifying polypeptide and cis template RNA are provided here:

PLV4921:

MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHP
IFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDV
DKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALS
LGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVN
TEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY

KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNR
EKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKN
LPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLK
EDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDR
EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNF
MQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPE
NIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRD
MYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQ
LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDY
KVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKG
RDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAY
SVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFE
LENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEI
IEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDR
KRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDSGGSSGGSSGSETPGTSESATPESSGGSS
GGSSTLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVS
IKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVED
IHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRL
PQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGY
RASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPG
FAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGV
LTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALV
KQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPD
LTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKM
AEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCP
GHQKGHSAEARGNRMADQAARKAAITETPDTSTLLIENSSPSGGSKRTADGSEFEKRTADGSEF
ESPKKKAKVE (SEQ ID NO: 16,709)

PLV4114

5'-
AGAAGTCGTGCTGCTTCATGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG
TTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCACCCTGACCTACGGCGTGCAGTGCTTCG
GCCGCTACCCCGATCACATGAAGCAGCACGAC -3' (SEQ ID NO: 16,710)

FIG. 7B shows a graph of the rewrite activity (normalized to PLV4921) for select exemplary gene modifying polypeptides having the selected polypeptide configurations (see FIG. 6A) and comprising exemplary RBP domains comprising 1, 2, or 4 MCP repeats. Exemplary gene modifying polypeptides comprised either a first exemplary RT domain (left shaded bars) or a second exemplary RT domain (right lighter shaded bars).

Exemplary gene modifying polypeptides shown in FIG. 7B:

Driver containing a first exemplary RT (ID of plasmid encoding)	**Driver containing a second exemplary RT (ID of plasmid encoding)**
PL11941	PL11989
PL11944	PL11992
PL11981	PL12018
PL11945	PL11993
PL11949	PL12034
PL11973	PL11998
PL12109	PL12037
PL12113	PL12053
PL12115	
PL12117	PL12045
PL12120	PL12047
PL12085	PL12097
PL12088	
PL12086	PL12101
PL12095	PL12104
PL12087	PL12108
PL12096	PL12099

The results show that many exemplary gene modifying polypeptides, comprising all configurations tested, number of MCP repeats, and both RT domains, exhibit BFP to GFP rewriting activity in U2OS cells and that their rewrite activities are comparable to that of benchmark gene modifying polypeptide PLV4921 when combined with a cis template RNA. Additionally, the experiment identified 36 exemplary gene modifying polypeptides that were more than 75% as active as the benchmark gene modifying polypeptide when combined with a cis template RNA, and of these 12

exemplary gene modifying polypeptides that were more than 90% as active as the benchmark; sequence exemplary gene modifying polypeptides are provided in Table X0 below. Activity with the cis template RNA demonstrates that an exemplary gene modifying polypeptide is active and potentially capable of rewriting using a trans template RNA.

The exemplary gene modifying polypeptides that had >75% Rewrite activity relative to PLV4921 are listed in Table X0.

Table X0: Exemplary gene modifying polypeptides having >75% rewriting activity relative to PLV4921

Driver (ID of plasmid encoding)	**Rewrite Activity (normalized to PLV4921)**
PL12086	1.057
PL11941	1.039
PL11957	1.035
PL11981	1.026
PL12089	0.956
PL11949	0.941
PL12095	0.921
PL12092	0.917
PL12018	0.889
PL11945	0.885
PL11993	0.885
PL11973	0.865
PL11994	0.849
PL12088	0.847
PL11996	0.842
PL12053	0.838
PL11995	0.827
PL11982	0.820
PL12034	0.820
PL11958	0.813
PL11947	0.811
PL12085	0.801
PL12101	0.801
PL12029	0.800
PL11984	0.789
PL11983	0.773
PL11950	0.771
PL12006	0.771
PL11960	0.762

PL12020	0.758
PL11944	0.758
PL12104	0.757
PL12113	0.755
PL11969	0.753
PL12017	0.753
PL11985	0.746

Example 9: Characterizing Combinations of Pools of Exemplary Trans Template with Exemplary Gene Modifying Polypeptides

This example characterizes the rewriting activity and trans template preferences of approximately 18 of the exemplary gene modifying polypeptides generated in Example 6 and determined to be capable of rewriting when combined with a cis template RNA in Example 7. The 18 exemplary gene modifying polypeptides (amino acid sequences provided in Example 6) listed below each comprise an RT domain comprising the amino acid sequence of SEQ ID NO: 16,711 and a Cas9 endonuclease domain comprising the amino acid sequence of SEQ ID NO: 16,712.

TAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEA
KRSLRETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLP
PDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGF
RLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSL
SNSRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKL
ALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALL
TREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFKQTAALNPATLLPETDD
TLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSA
QKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPK
RVAVMHCKGHQKDDAPTSTGNRRADEVAREVAIRPLSTQATIS （SEQ ID NO: 16,711）

DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRR
YTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKK
LVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAI
LSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQ
IGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIF
FDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHA
ILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSF

IERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTV
KQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREM
IEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDS
LTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQ
KGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHI
VPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEL
DKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEIT
LANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVK
KDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQ
HKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTI
DRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (SEQ ID NO: 16,712)

The exemplary gene modifying polypeptides are:

Configuration details	Driver (ID of plasmid encoding)
Config1_1MCP A	PL11941
Config1_1MCP B	PL11944
Config1_2MCP A	PL11981
Config1_2MCP B	PL11945
Config1_4MCP A	PL11949
Config1_4MCP B	PL11973
Config2_1MCP A	PL12109
Config2_2MCP A	PL12113
Config2_2MCP B	PL12115
Config2_4MCP A	PL12117
Config2_4MCP B	PL12120
Config3_1MCP A	PL12085
Config3_1MCP B	PL12088
Config3_2MCP A	PL12086
Config3_2MCP B	PL12095
Config3_4MCP A	PL12087
Config3_4MCP B	PL12096
Config2_1MCP A	PL12037

The exemplary gene modifying polypeptides included 6 from each of the configurations selected in Example 7 and encompassed exemplars having 1, 2, or 4 RBP repeats and two different linker combinations connecting each of the three protein domains.

The exemplary trans template RNAs of the pools encompassed exemplars having: linkers of either 8 or 16 nucleotides; 1, 2, or 4 RRS repeats (MS2 sequences); PBSs of 8 or 13 nucleotides (or a scrambled random PBS for negative controls); and 5' end-protecting Cas9 scaffolds with 16 nucleotide spacers or no 5’ end protecting feature.

Pool number	Plasmid (registry ID)
1xMS2_PBS8	PL13280
1xMS2_PBS8	PL13472
1xMS2_PBS8	PL13286
1xMS2_PBS8	PL13478
2xMS2_PBS8	PL13282
2xMS2_PBS8	PL13474
2xMS2_PBS8	PL13288
2xMS2_PBS8	PL13480
4xMS2_PBS8	PL13284
4xMS2_PBS8	PL13476
4xMS2_PBS8	PL13290
4xMS2_PBS8	PL13578
1xMS2_PBS13	PL13292
1xMS2_PBS13	PL13484
1xMS2_PBS13	PL13298
1xMS2_PBS13	PL13490
2xMS2_PBS13	PL13294
2xMS2_PBS13	PL13582
2xMS2_PBS13	PL13300
2xMS2_PBS13	PL13492
4xMS2_PBS13	PL13296
4xMS2_PBS13	PL13488
4xMS2_PBS13	PL13302
4xMS2_PBS13	PL13494
1xMS2_scrPBS	PL13316
1xMS2_scrPBS	PL13604

1xMS2_scrPBS	PL13322
1xMS2_scrPBS	PL13610
2xMS2_scrPBS	PL13318
2xMS2_scrPBS	PL13606
2xMS2_scrPBS	PL13324
2xMS2_scrPBS	PL13612
4xMS2_scrPBS	PL13320
4xMS2_scrPBS	PL13512
4xMS2_scrPBS	PL13326

The rewriting activity of exemplary gene modifying systems described in this example were normalized to a benchmark gene modifying system containing an exemplary gene modifying polypeptide (PLV4921) and a cis template RNA encoding the same edit (PLV4106). The sequence of PLV4921 was provided in Example 7, and the sequence of the cis-template used for normalization in this example is provided below:

PLV4106

GCCGAAGCACTGCACGCCGTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
TTGAAAAAGTGGCACCGAGTCGGTGCACCCTGACGTACGGCGTGCAGTGCTT (SEQ ID NO: 16,713)

The sequence of the gRNA that comprises the 3rd component of the gene modifying system is as below:

PLV4165

GCCGAAGCACTGCACGCCGTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC
TTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 16,714)

Each exemplary trans-gene modifying system (the combination of exemplary gene modifying polypeptide, exemplary trans template RNA, and exemplary gRNA) was tested to determine its genome-editing capacity using mammalian U2OS cells carrying a genomic landing pad that expresses a BFP-GFP SNP reporter. The U2OS cells are BFP-expressing and can be converted to a GFP-expressing cell line via a single base pair substitution programmed by the trans template RNAs. The U2OS cells were electroporated with each exemplary trans-gene modifying system, in the form of a plasmid encoding an individual exemplary gene modifying polypeptide, a second plasmid encoding the exemplary gRNA, and

a pool of plasmids encoding a subset of exemplary trans template RNAs. The trans template RNAs of a given pool were characterized by having similar primer binding site (PBS) length (8 or 13 nucleotides) and an RBP recruitment site (RRS) containing the same number of MS2 sequences (1, 2, or 4). Negative control pools contained trans template RNAs containing scrambled sequence PBSs.

Cells were analyzed by flow cytometry 4 days post-electroporation of the gene modifying system. The percentage of GFP cells was used as a measure for rewriting activity and compared to the rewriting activity of a benchmark gene modifying system containing an exemplary gene modifying polypeptide (PLV4921) and a cis template RNA encoding the same edit.

FIG. 8 shows a heat map of the rewriting activity of the exemplary trans-gene modifying systems (normalized to PLV4921 with cis template RNA) in the BFP/GFP SNP reporter assay. The results show that several pools of trans template RNAs provide comparable or better rewriting activity than a cis template RNA. These results further confirm the rewriting activity of the 18 exemplary gene modifying polypeptides examined across trans templates of differing characteristics, and also demonstrate that rewriting activity can be improved by pairing exemplary gene modifying polypeptides with a trans template. The results also suggested that a trans template RNA containing a PBS length of 13 nucleotides on average provides better rewriting activity across the evaluated gene modifying polypeptides than a trans template RNA containing an 8 nucleotide PBS. The results also suggested that gene modifying polypeptides containing 2 or 4 RBP repeats (in this Example, 2 or 4 MCP domains) generally had higher rewriting activity than gene modifying polypeptides containing 1 RBP (1 MCP domain). The results also suggested that trans template RNAs containing 1 or 4 RRS repeats (in this example, 1 or 4 MS2 sequences) generally provided better rewriting activity across the evaluated gene modifying polypeptides than trans template RNAs containing 2 RRS repeats (2 MS2 sequences). The data reiterates that each tested configuration of the gene modifying polypeptide has rewriting activity.

Example 10: Characterizing Combinations of Individual Exemplary Trans Templates with Exemplary Gene Modifying Polypeptides

This example characterizes the rewriting activity and trans template preferences of 6 of the exemplary gene modifying polypeptides generated in Example 6 and determined to be capable of rewriting and trans recruitment in Examples 7 and 9. The 6 exemplary gene modifying polypeptides (amino acid sequences provided in Example 6) listed below each comprise an RT domain comprising the amino acid sequence of SEQ ID NO: 16,711 and a Cas9 endonuclease domain comprising the amino acid

sequence of SEQ ID NO: 16,712. 29 exemplary trans template RNAs of the 384 generated in Example 6 were tested using flow cytometry (assessing BFP to GFP conversion as described herein).

The exemplary gene modifying polypeptides are: PL11945, PL12095, PL12096, PL12109, PL12113, and PL12115.

The exemplary gene modifying polypeptides included at least one exemplar from each of the configurations selected in Examples 7 and 9 and encompassed exemplars having 1, 2, or 4 RBPs and two different linker combinations connecting each of the three protein domains.

The exemplary trans template RNAs of the pools encompassed exemplars having: linkers of either 8 or 16 nucleotides; 1 or 4 RRS repeats (MS2 sequences); PBSs of 13 nucleotides (or a scrambled random PBS for negative controls); and 5' end-protecting Cas9 scaffolds with 16 nucleotide spacers, 5' end-protecting Cas9 scaffolds without spacers, a control 5' end-protecting ePEG, or no 5' end protecting feature.

Exemplary Trans Templates Used in this Example

PL13292
PL13293
PL13296
PL13297
PL13298
PL13299
PL13302
PL13303
PL13388
PL13389
PL13392
PL13393
PL13394
PL13395
PL13398
PL13399
PL13484
PL13488
PL13489
PL13490
PL13491

PL13494
PL13580
PL13581
PL13584
PL13585
PL13586
PL13590
PL13591

The sequences of PLV4921 and the cis template RNA were provided in Example 7, and the sequences of exemplary gene modifying polypeptides and trans template RNAs were provided in Example 6.

Each exemplary trans-gene modifying system (the combination of exemplary gene modifying polypeptide, exemplary trans template RNA and exemplary gRNA) was tested to determine its genome-editing capacity using mammalian U2OS cells carrying a genomic landing pad that expresses a BFP-GFP SNP reporter. The U2OS cells are BFP-expressing and can be converted to a GFP-expressing cell line via a single base pair substitution programmed by the trans template RNAs. The U2OS cells were electroporated with each exemplary trans-gene modifying system, in the form of a plasmid encoding an individual exemplary gene modifying polypeptide, a second plasmid encoding the exemplary gRNA, and a third plasmid encoding an individual exemplary trans template RNA.

Cells were analyzed by flow cytometry 4 days post-electroporation of the gene modifying system. The percentage of GFP cells was used as a measure for rewriting activity and normalized to the rewriting activity of a gene modifying system containing the same gene modifying polypeptide and a cis template RNA.

FIG. 9 shows a heat map of the rewriting activity of the exemplary trans-gene modifying systems in the BFP/GFP SNP reporter assay. The results show a number of trans template RNAs provide comparable or better rewriting activity than a gene modifying system containing the same gene modifying polypeptide with a cis template RNA. These results also show that a trans template RNA containing a 5' end-block comprising a Cas9 scaffold and a 16 nucleotide spacer provides better rewriting activity across most evaluated gene modifying polypeptides than a trans template RNA containing 5' end-block Cas9 scaffold without a spacer. The results further indicate that the 5'end protecting Cas9 scaffold with 16 nt spacer provides better rewriting activity than a control 5' end-protecting ePEG, or no 5' end protecting feature.

Example 11: Evaluating the Contributions of Trans Template RNA Components to Rewriting Activity

This Example demonstrates the importance of the RRS-RBP interaction to rewriting activity of gene modifying systems comprising a trans template RNA and exemplary RBP-containing gene modifying polypeptide.

Exemplary gene modifying systems comprised an RBP-containing gene modifying polypeptide E1 or an exemplary gene modifying polypeptide lacking an RBP G1 (each having previously demonstrated rewriting activity using cis templates, data not shown) and various trans template RNAs comprising a 5' end block comprising a Cas9 scaffold and no spacer and either 4 RRSs (in this Example 4 MS2 sequences) or 1 RRS (1 MS2 sequence). Sequences for the exemplary gene modifying polypeptides tested in this Example are provided in Table X1 below.

Table X1. Exemplary gene modifying polypeptides

Gene modifying polypeptide	Plasmid (registry ID)	Amino acid sequence	SEQ ID NO:
E1*	PL12113	MPAAKRVKLDGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPK VWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLR ETIRKFRAAGILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREV NKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPE SQLIFAFEWADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRD LQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAE LGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPV PKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLV WGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSGAAK GVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLT REASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLL DPPRVRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPD LTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEP LPIGTSAQKAELIALTKALEWSKDKSVNIYTDSRYAFATLHVH GMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCK GHQKDDAPTSTGNRRADEVAREVAIRPLSTQATISAEAAAKEA AAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKAMA SNFTQFVLVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKV TCSVRQSSAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSD CELIVKAMQGLLKDGNPIPSAIAANSGIYGGSGGSMASNFTQFV LVDNGGTGDVTVAPSNFANGIAEWISSNSRSQAYKVTCSVRQS SAQNRKYTIKVEVPKGAWRSYLNMELTIPIFATNSDCELIVKA MQGLLKDGNPIPSAIAANSGIYAEAAAKEAAAKEAAAKEAAA KALEAEAAAKEAAAKEAAAKEAAAKADKKYSIGLDIGTNSVG WAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAE ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEE SFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTD KADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQT	16,801

		YNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKN GLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNL LAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIK RYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDG GASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPL ARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNF DKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDR EMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIR DKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVS GQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPE NIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPV ENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVP QSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQ LLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITK HVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQF YKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYK VYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIR KRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQ TGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSK YVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLG APAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ LGGDKRTADGSEFESPKKKAKVE	
G2	PLV10990	MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKK FKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTR RKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPI FGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHM IKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASG VDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLT PNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFL AAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLL KALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPI LEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAIL RRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTR KSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK HSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLF KTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHD LLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAH LFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLK SDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTT QKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYL YYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNK VLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKF DNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMN TKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH	16,802

		AHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG EIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSK KLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPK YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHY EKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAN LDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTT IDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGGAEAA AKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAA KAGGTAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPG LASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAA GILRPVHSPWNTPLLPVRKSGTSEYRMVQDLREVNKRVETIHP TVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEW ADAEEGESGQLTWTRLPQGFKNSPTLFNEALNRDLQGFRLDHP SVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGK KAQLCQEEVTYLGFKIHKGSRSLSNSRTQAILQIPVPKTKRQVR EFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAF QSLKLALTQPPALALPSLDKPFQLFVEETSGAAKGVLTQALGP WKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFG QDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPRVRFK QTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQ AEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQ KAELIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERG WLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDA PTSTGNRRADEVAREVAIRPLSTQATISAGKRTADGSEFEKRTA DGSEFESPKKKAKVE	
N863A Cas9	PLV1425	MGKPIPNPLLGLDSTAPKKKRKVGIHGVPAADKKYSIGLDIGT NSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSG ETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQ LVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPG EKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDD LDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRT FDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPY YVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIE RMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMR KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSV EISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTL TLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRK LINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQ KAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVM GRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD VDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMK NYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVE TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFV	16,803

		YGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITL ANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIV KKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPT VAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFL EAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLD EIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLF TLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYE TRIDLSQLGGDSRADPKKKRKVEFSR	

* Sequence is included in the driver tables.

The exemplary cis template RNA is PLV4106 (sequence described earlier in Example 9). The exemplary trans template RNAs are as listed in Table X2 below.

Table X2. Exemplary trans template RNAs

Trans template RNA	**Plasmid (registry ID)**	**Plasmid name (see Table S4)**
4x MS2 no spacer	PL13584	305 scaffold_RT12_P13_L8_4xMS2_
1x MS2 no spacer	PL13580	301 scaffold_RT12_P13_L8_1xMS2_

Negative control polypeptides included N863A Cas9 (lacking any RT or RBP). Exemplary gene modifying systems were tested to determine their genome-editing capacity using mammalian U2OS cells carrying a genomic landing pad that expresses a BFP-GFP SNP reporter. The U2OS cells are BFP-expressing and can be converted to a GFP-expressing cell line via a single base pair substitution programmed by the trans template RNAs. The U2OS cells were electroporated with varying doses (37.5 ng, 75 ng, or 150 ng) of a plasmid encoding exemplary trans template RNA, 300 ng plasmid encoding exemplary gene modifying polypeptide, and 200 ng of a second plasmid encoding the exemplary gRNA (i.e., gRNA PLV4165, as described in Example 9).

Cells were analyzed by flow cytometry 4 days post-electroporation of the gene modifying system. The percentage of GFP-expressing cells was used as a measure for rewriting activity.

FIG. 10 shows graphs of rewriting activity for the exemplary gene modifying systems. The results show that negative control N863A Cas9 showed minimal rewriting activity. The RBP-less gene modifying polypeptide G2 showed no rewriting activity when combined with a trans template RNA comprising a 4x or 1x RRS and a 5' end block comprising a Cas9 scaffold and lacking a spacer. The results show that the exemplary RBP-containing gene modifying polypeptide E1 shows rewriting activity when combined with a trans template RNA comprising a 4x RRS and a 5' end block Cas9 scaffold lacking a spacer, and that rewriting activity increases with increasing dose of trans template RNA. E1 shows low levels of rewriting activity when combined with a trans template RNA comprising a 1x RRS

and a 5' end block Cas9 scaffold lacking a spacer, but that rewriting activity increases with increasing dose of trans template RNA. The results show that the RRS-RBP interaction can be important for assembly of the trans template RNA:gene modifying polypeptide:target DNA complex when using trans template RNAs comprising a 5' end block Cas9 scaffold and lacking a spacer. Without wishing to be bound by theory, even a weaker RBP:RRS interaction as may be provided by the RBP of E1 interacting with a 1x RRS appears to provide sufficient interaction as the dose of trans template RNA increases to provide observable rewriting activity.

Taken together, these results show that the RRS:RBP interaction between a trans template RNA and an exemplary gene modifying polypeptide is critical to recruitment and rewriting activity of gene modifying systems comprising said components when the trans template RNA contains a 5' end block comprising a Cas9 scaffold and lacking a spacer.

Example 12: Evaluating the Contribution of the 5' End Block Spacer to Rewriting Activity

This Example demonstrates the importance and role of the 5' end block to rewriting activity of gene modifying systems comprising a trans template RNA and exemplary RBP-containing gene modifying polypeptide. The presence of a 5' end block with Cas9 scaffold and spacer increased rewriting activity of gene modifying systems in Example 10 relative to similar systems where the 5' end block contained a Cas9 scaffold and no spacer. To further investigate the role of the spacer in the 5' end block, exemplary gene modifying systems were assembled comprising an exemplary RBP-containing gene modifying polypeptide E1 and various trans template RNAs comprising a 5' end block comprising a Cas9 scaffold and no spacer (no spacer) or a 16 nucleotide spacer (spacer), and either 4 RRSs (in this Example 4 MS2 sequences) or 1 RRS (1 MS2 sequence). Without wishing to be bound by theory, a 16 nucleotide spacer is thought to enable only limited nicking activity by the *S. pyogenes* Cas9 domain of E1, but sufficient to enable binding of the E1:spacer to the spacer's complementary sequence in the target DNA. A control trans template RNA comprising a 5' end block not comprising a Cas9 scaffold (pseudoknot) was also included. A gRNA targeting the site where the 16 nucleotide spacer would have (with a full spacer length of 20 nt) targeted a nick was included in some samples (+2nd nick). N863A Cas9 (lacking any RT or RBP) was used as a control polypeptide. The exemplary sgRNA used in this Example was PLV4165 (sequence described in Example 9). Exemplary gene modifying polypeptides used in this Example were those described in Example 11. The exemplary cis template RNA used in this Example was PLV4106 (sequence described in Example 9). The exemplary trans template RNAs used in this Example were as listed in Table X3 below.

Table X3. Exemplary trans template RNAs

Trans template RNA	Plasmid (registry ID)	Plasmid name (see Table S4)
4xMS2 pseudoknot	PLV13392	113 pseudoknot _RT12_P13_L8_4xMS2_
4xMS2 no spacer	PLV13584	305 scaffold_RT12_P13_L8_4xMS2_
4xMS2 + spacer	PLV13488	209 g16A_RT12_P13_L8_4xMS2_
1xMS2 pseudoknot	PLV13388	109 pseudoknot _RT12_P13_L8_1xMS2_
1xMS2 no spacer	PLV13580	301 scaffold_RT12_P13_L8_1xMS2_
1xMS2 + spacer	PLV13484	205 g16A_RT12_P13_L8_1xMS2_

The exemplary 2nd nick gRNA that binds/nicks the same target as the 5' spacer in PL13488 and PL13484 was PLV4163, having the nucleic acid sequence:

CCGGCAAGCTGCCCGTGCCCGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC TTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 16,715).

Exemplary gene modifying systems were tested to determine their genome-editing capacity using mammalian U2OS cells carrying a genomic landing pad that expresses a BFP-GFP SNP reporter. The U2OS cells are BFP-expressing and can be converted to a GFP-expressing cell line via a single base pair substitution programmed by the trans template RNAs. The U2OS cells were electroporated with each exemplary trans-gene modifying system, in the form of a plasmid encoding an individual exemplary gene modifying polypeptide, a second plasmid encoding the exemplary gRNA, a third plasmid encoding an individual exemplary trans template RNA, and a fourth plasmid (if needed for that sample) encoding the 2nd nick gRNA.

Cells were analyzed by flow cytometry 4 days post-electroporation of the gene modifying system. The percentage of GFP-expressing cells was used as a measure for rewriting activity.

FIGS. 11A-11B show graphs of rewriting activity for the exemplary gene modifying systems. The results show that negative control N863A Cas9 showed minimal rewriting activity with trans template RNA containing a 4x RRS and a 5' end block with Cas9 scaffold and spacer. A positive control cis template RNA showed significant rewriting activity when combined with the gene modifying polypeptide (FIGs. 11A-11B).

The results show that the addition of a gRNA directing a nick to the same target site DNA as the spacer of the trans template 5' end block (striped bars) did not increase rewriting activity (open bars;

compare pseudoknot, no spacer, and + spacer with and without 2nd nick). The lack of any increase in rewriting with the gRNA suggests that for these exemplary gene modifying systems nicking at the site targeted by the 5' end block spacer is not important for rewriting activity.

In contrast, the results show that the presence of the spacer in the 5' end block of the trans template RNA significantly increased rewriting activity (FIGs. 11A-11B). The presence of the spacer increased rewriting activity in both a 4x RRS trans template RNA system as well as a 1x RRS trans template RNA system, suggesting that the contributions of the RRS:RBP interaction and the 5' end block spacer:additional gene modifying polypeptide interaction to rewriting activity are additive. This result further suggests that the 5' end block spacer may rescue an otherwise weak RRS:RBP interaction (compare FIG. 11A no spacer to +spacer versus FIG. 11B no spacer to +spacer). Rewriting activity using systems comprising a trans template RNA comprising a 4x RRS and a 5' end block with Cas9 scaffold and spacer was unaffected by addition of the gRNA providing nicking (compare +spacer with and without 2nd nick). This result suggests that the primary role of the 5' end block spacer is providing Cas9-mediated binding of the trans template RNAs to the target locus in these exemplary gene modifying systems, and that nicking at the site targeted by the spacer is unimportant to rewriting activity for these exemplary systems.

Taken together, these results suggest that the role of the 5' end block spacer in these exemplary systems in this BFP to GFP rewriting assay is for a second mode of recruitment of the trans template to the target site.

Without wishing to be bound by theory, the Examples and disclosure are directed, in part, to the discovery that anchoring a trans template RNA to a gene modifying polypeptide:sgRNA:target genomic DNA complex by two or more interactions is important for achieving high rewriting activity, e.g., for achieving single or several nucleotide long edits. This Example suggests that, e.g., 1) an RRS:RBP interaction and 2) a 5' end block Cas9 scaffold and spacer to target DNA interaction (mediated via an additional gene modifying polypeptide) represent exemplary interactions that together anchor a trans template RNA to a gene modifying polypeptide:sgRNA:target genomic DNA complex to enable rewriting. While the RRS:RBP interaction is critical in the absence of the 5' end block spacer, the presence of both provides high rewriting activity and the presence of the 5' end block spacer in combination with a weaker RRS:RBP interaction rescues rewriting activity.

It should be understood that for all numerical bounds describing some parameter in this application, such as "about," "at least," "less than," and "more than," the description also necessarily

encompasses any range bounded by the recited values. Accordingly, for example, the description "at least 1, 2, 3, 4, or 5" also describes, inter alia, the ranges 1-2, 1-3, 1-4, 1-5, 2-3, 2-4, 2-5, 3-4, 3-5, and 4-5, et cetera.

For all patents, applications, or other reference cited herein, such as non-patent literature and reference sequence information, it should be understood that they are incorporated by reference in their entirety for all purposes as well as for the proposition that is recited. Where any conflict exists between a document incorporated by reference and the present application, this application will control. All information associated with reference gene sequences disclosed in this application, such as GeneIDs or accession numbers (typically referencing NCBI accession numbers), including, for example, genomic loci, genomic sequences, functional annotations, allelic variants, and reference mRNA (including, e.g., exon boundaries or response elements) and protein sequences (such as conserved domain structures), as well as chemical references (e.g., PubChem compound, PubChem substance, or PubChem Bioassay entries, including the annotations therein, such as structures and assays, et cetera), are hereby incorporated by reference in their entirety.

Headings used in this application are for convenience only and do not affect the interpretation of this application.

CLAIMS

1. A template RNA comprising:

a) a heterologous object sequence comprising a mutation region to introduce a mutation into a target nucleic acid sequence (wherein optionally the heterologous object sequence comprises, from 5' to 3', a post-edit homology region, the mutation region, and a pre-edit homology region), and

b) a primer binding site sequence (PBS sequence) that binds a first portion of the target nucleic acid sequence, wherein first portion is in the first strand of the target nucleic acid sequence, and wherein the PBS sequence is 3' of the heterologous object sequence, and

c) an RBD recruitment site (RRS), wherein the RRS is 3' of the PBS sequence or 5' of the heterologous object sequence.

2. The template RNA of claim 1, which further comprises an end block sequence, e.g., an end block sequence of Table 41 or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto.

3. The template RNA of claim 2, wherein the end block sequence is 5' of the heterologous object sequence and the RRS is 3' of the PBS sequence.

4. The template RNA of claim 2, wherein the end block sequence is 3' of the PBS sequence and the RRS is 5' of the heterologous object sequence.

5. The template RNA of any of the preceding claims, wherein the RRS has a sequence according to Table 40 or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto.

6. The template RNA of any of the preceding claims, which comprises a plurality of RRSs, e.g., a tandem array of 2, 3, 4, 5, or 10 RRSs.

7. The template RNA of any if the preceding claims, wherein the PBS sequence comprises 8-17 nucleotides, e.g., 8-17 nucleotides of 100% identity to the target nucleic acid sequence.

8. The template RNA of any of the preceding claims wherein the pre-edit homology region comprises up to 20 nucleotides, e.g., up to 20 nucleotides of 100% identity to the target nucleic acid sequence.

9. The template RNA of any of the preceding claims wherein the post-edit homology region comprises 5-500 nucleotides, e.g., 5-500 nucleotides of 100% identity to the target nucleic acid sequence.

10. The template RNA of any of the preceding claims, wherein the mutation region is configured to produce an insertion, a deletion, or a substitution in the target nucleic acid.

11. The template RNA of any of the preceding claims, which further comprises:

a gRNA spacer that is complementary to a different portion (e.g., a third portion) of the target nucleic acid sequence, e.g., wherein the different portion (e.g., third portion) is on the first strand of the target nucleic acid sequence; and

a gRNA scaffold.

12. The template RNA of claim 11, wherein the gRNA spacer is 5' of the heterologous object sequence.

13. The template RNA of claim 11 or 12, wherein the gRNA scaffold is situated between the gRNA spacer and the heterologous object sequence.

14. The template RNA of any of claims 11-13 wherein the gRNA spacer and the PBS sequence bind the same strand of the target nucleic acid sequence.

15. The template RNA of any of claims 11-14 wherein the gRNA spacer, the heterologous object sequence, and the PBS sequence bind the same strand of the target nucleic acid sequence.

16. The template RNA of any of claims 1-4, which does not comprise a gRNA spacer or a gRNA scaffold.

17. The template RNA of any of the preceding claims, which comprises a linker of up to 20 nucleotides between the RRS and the PBS sequence.

18. A gene modifying polypeptide comprising:

a reverse transcriptase (RT) domain; and

a DNA binding domain (DBD) that binds to a target nucleic acid sequence and is heterologous to the RT domain (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain); and

a RNA-binding domain (RBD) that is heterologous to the DBD and the RT domain, wherein the domains are arranged, in an N-terminal to C-terminal direction:

m) DBD, RT domain, RBD;

n) RT domain, DBD, RBD;

o) RBD, DBD, RT domain;

p) RBD, RT domain, DBD;

q) DBD, RBD, RT domain; or

r) RT domain, RBD, DBD.

19. A gene modifying polypeptide comprising:

a reverse transcriptase (RT) domain; and

a DNA binding domain (DBD) that binds to a target nucleic acid sequence and is heterologous to the RT domain (e.g., a Cas nickase domain, e.g., a Cas9 nickase domain); and

a plurality (e.g., 2, 3, 4, or 5) RNA-binding domains (RBD) that are heterologous to the DBD and the RT domain.

20. The gene modifying polypeptide of claim 6, wherein the RBD has an amino acid sequence according to Table 31, or at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

21. The gene modifying polypeptide of any of the preceding claims, wherein the plurality of RBDs have the same amino acid sequence as each other.

22. The gene modifying polypeptide of any of the preceding claims, wherein the plurality of RBDs have different amino acid sequences from each other.

23. The gene modifying polypeptide of any of the preceding claims, wherein the DBD has an amino acid sequence according to Table 7 or 8, or at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

24. The gene modifying polypeptide of any of any of the preceding claims, wherein the RT domain is from a retrovirus, or a polypeptide domain having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acids sequence identity thereto.

25. The gene modifying polypeptide of any of the preceding claims ,wherein the RT domain has an amino acid sequence according to Table 6, or at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

26. The gene modifying polypeptide of any of the preceding claims ,wherein the gene modifying polypeptide comprises a linker.

27. The gene modifying polypeptide of any of the preceding claims , wherein the linker comprises a sequence according to Table 10, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

28. The gene modifying polypeptide of claim 26 or 27, wherein the linker is disposed between the DBD and the RT domain, the RT domain and the RBD, or between the RBD and the DBD.

29. The gene modifying polypeptide of any of the preceding claims, wherein the gene modifying polypeptide comprises, in an N-terminal to C-terminal direction:

m) the DBD, a first linker, the RT domain, a second linker, the RBD;
n) the RT domain, a first linker, the DBD, a second linker, the RBD;
o) the RBD, a first linker, the DBD, a second linker, the RT domain;
p) RBD, a first linker, RT domain, a second linker, DBD;
q) the DBD, a first linker, the RBD, a second linker, the RT domain; or
r) the RT domain, a first linker, the RBD, a second linker, the DBD.

30. The gene modifying polypeptide of any of the preceding claims , which was produced by intein-mediated fusion of an N-terminal portion comprising an intein-N domain and a C-terminal portion comprising an intein-C domain.

31. A polypeptide system (e.g., a polypeptide complex) comprising:

a) a reverse transcriptase (RT) domain; and

b) a DNA binding domain (DBD) that binds to a target nucleic acid sequence and is heterologous to the RT domain (e.g., a Cas domain, e.g., a Cas9 domain, e.g., a Cas9 nickase domain); and

c) a RNA-binding domain (RBD) that is heterologous to the DBD and the RT domain,

wherein at least 2 of (e.g., all of) (a), (b), and (c) are in separate polypeptides, e.g., separate polypeptides that noncovalently form a complex.

32. The polypeptide system of claim 31, wherein complex formation is mediated by a first dimerization domain that binds a second, compatible dimerization domain.

33. The polypeptide system of claim 32, wherein complex formation is mediated by a third dimerization domain that binds a fourth, compatible dimerization domain.

34. The polypeptide system of any of claims 31-33, wherein:

the RBD is operably linked (e.g., via a linker) to a first dimerization domain;

the DBD is operably linked (e.g., via a linker) to a second dimerization domain that binds the first dimerization domain;

the DBD is operably linked (e.g., via a linker) to a third dimerization domain; and

the RT domain is operably linked (e.g., via a linker) to a fourth dimerization domain that binds the third dimerization domain.

35. The polypeptide system of any of claims 31-34 wherein the first and second dimerization domains are: chemical- induced dimerization domains, light-induced dimerization domains, antibody-peptide dimerization domains, or coiled coil dimerization domains.

36. The polypeptide system of any of claims 31-35, wherein the third and fourth dimerization domains are: chemical- induced dimerization domains, light-induced dimerization domains, antibody-peptide dimerization domains, or coiled coil dimerization domains.

37. The polypeptide system of any of claims 31-36, wherein the first dimerization domain and the second dimerization domain are each present in a plurality of copies, e.g., 2, 3, 4, 5, 10, 15, 20, or 30 copies.

38. The polypeptide system of any of claims 31-37, wherein the third dimerization domain and the fourth dimerization domain are each present in a plurality of copies, e.g., 2, 3, 4, 5, 10, 15, 20, or 30 copies.

39. The polypeptide system of any of claims 31-38, wherein the first dimerization domain and the second dimerization domain have the same sequence (e.g., wherein the first dimerization domain and the second dimerization domain form a homodimer).

40. The polypeptide system of any of claims 31-39, wherein the third dimerization domain and the fourth dimerization domain have the same sequence (e.g., wherein the third dimerization domain and the fourth dimerization domain form a homodimer).

41. The polypeptide system of any of claims 31-38, wherein the first dimerization domain and the second dimerization domain have different sequences (e.g., wherein the first dimerization domain and the second dimerization domain form a heterodimer).

42. The polypeptide system of any of claims 31-41, wherein the third dimerization domain and the fourth dimerization domain have different sequences (e.g., wherein the third dimerization domain and the fourth dimerization domain form a hetero dimer).

43. The polypeptide system of any of claims 31-42, wherein the DBD is operably linked to one or more additional DBDs, wherein optionally the additional DBDs have the same sequence as the DBD.

44. The polypeptide system of any of claims 31-43, wherein the RBD has an amino acid sequence according to Table 31, or at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

45. The polypeptide system of any of claims 31-44, wherein the plurality of RBDs have the same amino acid sequence as each other.

46. The polypeptide system of any of claims 31-45, wherein the plurality of RBDs have different amino acid sequences from each other.

47. The polypeptide system of any of claims 31-46, wherein the DBD has an amino acid sequence according to Table 31, or at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

48. The polypeptide system of any of claims 31-47, wherein the RT domain is from a retrovirus, or a polypeptide domain having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% amino acids sequence identity thereto.

49. The polypeptide system of any of claims 31-48, wherein the RT domain has an amino acid sequence according to Table 6, or at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

50. The polypeptide system of any of claims 31-49, wherein each linker independently comprises a sequence according to Table 10, or a sequence having at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity thereto.

51. A nucleic acid or a plurality of nucleic acids encoding the polypeptides of any of the systems of claim 31-50.

52. A system comprising:

a template RNA of any of claims 1-17;

a gene modifying polypeptide of any of claims 18-30 or the polypeptide system of any of claims 31-50; and

a first gRNA comprising:

a gRNA spacer that binds a second portion of the target nucleic acid sequence, wherein the second portion is one the second strand of the target nucleic acid sequence; and

a gRNA scaffold that binds the DBD of the gene modifying polypeptide or the polypeptide system.

53. The system of claim 52, wherein the template RNA does not comprise a gRNA spacer or a gRNA scaffold.

54. The system of claim 52 or 53, wherein the gRNA spacer binds to a region of the target nucleic acid sequence that is within about 5, 10, 15, 20, 25, 30, or 40 nucleotides of the region of the target nucleic acid sequence bound by the PBS sequence.

55. The system of any of claims 52-54, which further comprises:

a second Cas protein (e.g., a dead Cas protein) and

a second gRNA comprising:

a gRNA spacer that binds the first strand of the target nucleic acid at a location 3' of the location bound by the PBS sequence, and

a gRNA scaffold that binds the second Cas protein.

56. The system of claim 55, wherein the second Cas protein is a dead Cas protein (e.g., a dead Cas9 protein) or a Cas nickase protein (e.g., a Cas9 nickase protein).

57. The system of claim 55, wherein the gRNA spacer of the second gRNA has a length of at least 18 nucleotides (e.g., 18-28 nucleotides, e.g., 18-21 nucleotides) and the second Cas protein is a dead Cas protein.

58. The system of claim 55, wherein the gRNA spacer of the second gRNA has a length of 17 nucleotides or less (e.g., 14-17 nucleotides), wherein optionally the second Cas protein is a Cas nickase protein.

59. The system of claim 52, wherein the template RNA further comprises:

a gRNA spacer that is complementary to a third portion of the target nucleic acid sequence wherein the third portion is on the first strand of the target nucleic acid sequence; and

a gRNA scaffold.

60. The system of claim 59, wherein the gRNA scaffold binds the DBD of the gene modifying polypeptide or the polypeptide system.

61. The system of claim 59 or 60, wherein the gRNA spacer has a length of 17 nucleotides or less.

62. The system of any of claims 52-61, wherein the gRNA spacer of the template RNA induces nicking of the template nucleic acid, e.g., at the second strand of the target nucleic acid sequence.

63. The system of any of claims 52-61, wherein the gRNA spacer of the template RNA does not induce nicking of the template nucleic acid.

64. A system comprising:

i) a template RNA of any of claims 1-17 (e.g., a template RNA of claim 16);

ii) a first polypeptide comprising:

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain); and

a RNA-binding domain (RBD) that is heterologous to the DBD, wherein the RBD binds the RRS of the template RNA;

iii) a first gRNA comprising:

a gRNA spacer that directs the DBD of the first polypeptide to a second portion of the target nucleic acid sequence, wherein the second portion of the target nucleic acid sequence is on the second strand of the nucleic acid sequence; and

a gRNA scaffold that binds the DBD of the first polypeptide;

iv) a second polypeptide comprising:

an RT domain, and

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain), that is heterologous to the RT domain, and wherein the DBD of the second polypeptide has a different sequence from the DBD of the first polypeptide; and

v) a second gRNA comprising:

a gRNA spacer that directs the DBD of the second polypeptide to a third portion of the target nucleic acid sequence, wherein the third portion is on the first strand of the target nucleic acid, and

a gRNA scaffold that binds the DBD of the second polypeptide.

65. The system of claim 64, wherein the DBD of the second polypeptide comprises a Cas nickase domain or a dead Cas domain.

66. The system of claim 64, wherein the gRNA spacer of the second RNA induces nicking of the template nucleic acid, e.g., at the second strand of the target nucleic acid sequence.

67. The system of claim 64, wherein the gRNA spacer of the second RNA does not induce nicking of the template nucleic acid.

68. The system of claim 64, wherein the first gRNA does not detectably bind to the DBD of the second polypeptide.

69. The system of claim 64, wherein the second gRNA does not detectably bind to the DBD of the first polypeptide.

70. A system comprising:

i) a template RNA of any of the preceding claims, wherein the template RNA comprises:

a gRNA spacer that is complementary to a third portion of the target nucleic acid sequence wherein the third portion is on the first strand of the target nucleic acid sequence; and

a gRNA scaffold;

ii) a first polypeptide comprising:

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain); and

a RNA-binding domain (RBD) that is heterologous to the DBD, wherein the RBD binds the RRS of the template RNA;

iii) a first gRNA comprising:

a gRNA spacer that directs the DBD of the first polypeptide to a second portion of the target nucleic acid sequence, wherein the second portion of the target nucleic acid sequence is on the second strand of the nucleic acid sequence; and

a gRNA scaffold that binds the DBD of the first polypeptide; and

iv) a second polypeptide comprising:

an RT domain, and

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain), that is heterologous to the RT domain, and wherein the DBD of the second polypeptide has a different sequence from the DBD of the first polypeptide,

and wherein the gRNA scaffold of the template RNA binds the DBD of the second polypeptide.

71. The system of claim 70, wherein the DBD of the second polypeptide comprises a Cas nickase domain or a dead Cas domain.

72. The system of claim 70, wherein the gRNA spacer of the template RNA induces nicking of the template nucleic acid, e.g., at the second strand of the target nucleic acid sequence.

73. The system of claim 70, wherein the gRNA spacer of the template RNA does not induce nicking of the template nucleic acid.

74. The system of any of claims 70-73, wherein the first gRNA does not detectably bind to the DBD of the second polypeptide.

75. The system of any of claims 70-74, wherein the gRNA of the template RNA does not detectably bind to the DBD of the first polypeptide.

76. A polypeptide system comprising:

a first polypeptide comprising:

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain);

a RNA-binding domain (RBD) that is heterologous to the DBD; and

optionally, a linker disposed between the DBD and the RBD; and

a second polypeptide comprising:

an RT domain, and

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain), that is heterologous to the RT domain; and

optionally, a linker disposed between the RT domain and the DBD.

77. The template RNA or system of any of the preceding claims, wherein the target nucleic acid sequence is a target gene, enhancer, or promoter.

78. The template RNA of system any of the preceding claims, wherein the target nucleic acid sequence is a human target gene, human enhancer, or human promoter.

79. The system or polypeptide system of any of the preceding claims, wherein the RBD has a sequence of Table 31, or a sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99% identity thereto.

80. A method for modifying a target nucleic acid in a cell (e.g., a human cell), the method comprising contacting the cell with the system of any one of the preceding claims, or nucleic acid encoding the same, thereby modifying the target nucleic acid.

81. The method of claim 80, wherein presence of the second polypeptide, compared to an otherwise similar system lacking the second polypeptide, results in one or more of:

increased unwinding of the target nucleic acid;

increased number of target nucleic acids that are modified;

increased length of insertion into the target nucleic acid; or

reduced MMR activity at the target nucleic acid.

82. The method of any of claims 80 and 81, wherein the cell is in vivo or ex vivo.

83. A template RNA comprising:

a) a heterologous object sequence comprising a mutation region to introduce a mutation into a target nucleic acid sequence (wherein optionally the heterologous object sequence comprises, from 5' to 3', a post-edit homology region, the mutation region, and a pre-edit homology region), and

b) a primer binding site sequence (PBS sequence) that binds a first portion of the target nucleic acid sequence, wherein first portion is in the first strand of the target nucleic acid sequence, and wherein the PBS sequence is 3' of the heterologous object sequence, and

c) an RBD recruitment site (RRS), wherein the RRS is 3' of the PBS sequence or 5' of the heterologous object sequence.

84. A gene modifying polypeptide comprising:

a reverse transcriptase (RT) domain; and

a DNA binding domain (DBD) that binds to a target nucleic acid sequence and is heterologous to the RT domain (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain); and

a RNA-binding domain (RBD) that is heterologous to the DBD and the RT domain,

wherein the domains are arranged, in an N-terminal to C-terminal direction:

s) DBD, RT domain, RBD;

t) RT domain, DBD, RBD;

u) RBD, DBD, RT domain;

v) RBD, RT domain, DBD;

w) DBD, RBD, RT domain; or

x) RT domain, RBD, DBD.

85. A polypeptide system (e.g., a polypeptide complex) comprising:

a) a reverse transcriptase (RT) domain; and

b) a DNA binding domain (DBD) that binds to a target nucleic acid sequence and is heterologous to the RT domain (e.g., a Cas domain, e.g., a Cas9 domain, e.g., a Cas9 nickase domain); and

c) a RNA-binding domain (RBD) that is heterologous to the DBD and the RT domain,

wherein at least 2 of (e.g., all of) (a), (b), and (c) are in separate polypeptides, e.g., separate polypeptides that noncovalently form a complex.

86. A nucleic acid or a plurality of nucleic acids encoding the polypeptides of the system claim 85.

87. A system comprising:

a template RNA of claim 83;

a gene modifying polypeptide, e.g., a gene modifying polypeptide of claim 84, or a polypeptide system, e.g., a polypeptide system of claim 85; and

a first gRNA comprising:

a gRNA spacer that binds a second portion of the target nucleic acid sequence, wherein the second portion is one the second strand of the target nucleic acid sequence; and

a gRNA scaffold that binds the DBD of the gene modifying polypeptide or the polypeptide system.

88. A system comprising:

i) a template RNA of claim 83;

ii) a first polypeptide comprising:

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain); and

a RNA-binding domain (RBD) that is heterologous to the DBD, wherein the RBD binds the RRS of the template RNA;

iii) a first gRNA comprising:

a gRNA spacer that directs the DBD of the first polypeptide to a second portion of the target nucleic acid sequence, wherein the second portion of the target nucleic acid sequence is on the second strand of the nucleic acid sequence; and

a gRNA scaffold that binds the DBD of the first polypeptide;

iv) a second polypeptide comprising:

an RT domain, and

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain), that is heterologous to the RT domain, and wherein the DBD of the second polypeptide has a different sequence from the DBD of the first polypeptide; and

v) a second gRNA comprising:

a gRNA spacer that directs the DBD of the second polypeptide to a third portion of the target nucleic acid sequence, wherein the third portion is on the first strand of the target nucleic acid, and

a gRNA scaffold that binds the DBD of the second polypeptide.

89. A system comprising:

i) a template RNA of any of the preceding claims, wherein the template RNA comprises:

a gRNA spacer that is complementary to a third portion of the target nucleic acid sequence wherein the third portion is on the first strand of the target nucleic acid sequence; and

a gRNA scaffold;

ii) a first polypeptide comprising:

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain); and

a RNA-binding domain (RBD) that is heterologous to the DBD, wherein the RBD binds the RRS of the template RNA;

iii) a first gRNA comprising:

a gRNA spacer that directs the DBD of the first polypeptide to a second portion of the target nucleic acid sequence, wherein the second portion of the target nucleic acid sequence is on the second strand of the nucleic acid sequence; and

a gRNA scaffold that binds the DBD of the first polypeptide; and

iv) a second polypeptide comprising:

an RT domain, and

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain), that is heterologous to the RT domain, and wherein the DBD of the second polypeptide has a different sequence from the DBD of the first polypeptide, and wherein the gRNA scaffold of the template RNA binds the DBD of the second polypeptide.

90. A polypeptide system comprising:

a first polypeptide comprising:

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain);

a RNA-binding domain (RBD) that is heterologous to the DBD; and

optionally, a linker disposed between the DBD and the RBD; and

a second polypeptide comprising:

an RT domain, and

a DNA binding domain (DBD) (e.g., a Cas domain, e.g., a Cas nickase domain, e.g., a Cas9 nickase domain), that is heterologous to the RT domain; and

optionally, a linker disposed between the RT domain and the DBD.

91. A method for modifying a target nucleic acid in a cell (e.g., a human cell), the method comprising contacting the cell with the system of any one of the preceding claims, or nucleic acid encoding the same, thereby modifying the target nucleic acid.