

ENGINEERED PROTEINS WITH A PROTEASE CLEAVAGE SITE

RELATED APPLICATIONS

[0001] This application claims priority under 35 U.S.C. § 119(e) to U.S. provisional patent applications, U.S.S. N. 61/862,363, filed August 5, 2013, and U.S.S.N. 61/987,518, filed May 2, 2014, each of which is incorporated herein by reference.

BACKGROUND OF INVENTION

[0002] Metabolic engineering permits production of compounds through manipulation of biochemical reactions (*e.g.*, biosynthetic pathways) in a cell. Nonetheless, production of certain compounds may conflict with essential cellular goals. For example, diversion of nutrients and energy for the production of a compound may result in a shortage of those substrates and cofactors for production of biomass. The engineered organism may either evolve away from producing the compound of interest or grow sub-optimally. To address this issue, cell-free systems have been developed for the *in vitro* production of compounds through coordinated expression of proteins in a biosynthetic pathway. One caveat to both *in vivo* and *in vitro* bioproduction systems is that many key proteins that divert flux from a biosynthetic pathway are also important or even essential for cell growth. Deletion or inactivation of these proteins is often difficult or impossible because doing so results in reduced cell growth or viability. One way to inactivate proteins is through protease-mediated inactivation. Protease-mediated inactivation of a target protein can be achieved through the incorporation of a protease recognition site in the primary amino acid sequence of the target protein. The protease recognition site can be incorporated into the primary sequence such that the resulting protein is active in the absence of a protease that cleaves the recognition site and inactive in the presence of the protease. Such engineered or recombinant target proteins are particularly useful for the cell-free synthesis of compounds of interest.

SUMMARY OF INVENTION

[0003] Provided herein are recombinant enzymes that can be inactivated selectively during *in vitro* cell-free production of a compound. Selective inactivation of recombinant enzymes is achieved by introducing, between two codons (*e.g.*, two adjacent codons) in a gene that codes for the recombinant enzyme, at least one nucleotide sequence that codes for a

protease recognition sequence. Prior to *in vitro* cell-free production of a compound, the cognate protease (*i.e.*, a protease that specifically recognizes and cleaves the protease recognition sequence in the recombinant protein) is introduced to, or activated in, the cell-free system such that it can cleave the recombinant enzyme, thereby inactivating it. Also provided herein are methods of producing and screening for recombinant enzymes that can be inactivated selectively and that retain an activity level comparable to that of the wild-type enzymes.

[0004] In some aspects of the invention, provided herein are recombinant phosphoglucose isomerase (Pgi) proteins with at least one (or one) protease recognition sequence located between two amino acids (*e.g.*, two adjacent amino acids) of the native protein. In other aspects, nucleic acids encoding the recombinant phosphoglucose isomerase proteins are provided.

[0005] In some embodiments, the Pgi protein comprises the sequence of SEQ ID NO:17. In certain embodiments, the Pgi protein may comprise a sequence that is 90%, 95%, 98%, or 99% homologous to SEQ ID NO:25. In some embodiments, the nucleic acid encoding the Pgi protein comprises the sequence of SEQ ID NO:1. In certain embodiments, the nucleic acid encoding the Pgi protein may comprise a sequence that is 90%, 95%, 98%, or 99% homologous to SEQ ID NO:9.

[0006] In some aspects of the invention, provided herein are recombinant phosphotransacetylase (Pta) proteins with at least one (or one) protease recognition sequence located between two amino acids (*e.g.*, two adjacent amino acids) of the native protein. In other aspects, nucleic acids encoding the recombinant phosphotransacetylase proteins are provided.

[0007] In some embodiments, the Pta protein comprises the sequence of SEQ ID NO:48. In certain embodiments, the Pta protein may comprise a sequence that is 90%, 95%, 98%, or 99% homologous to SEQ ID NO:48. In some embodiments, the nucleic acid encoding the Pta protein comprises the sequence of SEQ ID NO:47. In certain embodiments, the nucleic acid encoding the Pta protein may comprise a sequence that is 90%, 95%, 98%, or 99% homologous to SEQ ID NO:47.

[0008] In some aspects of the invention, provided herein are recombinant transketolase A (TktA) proteins with at least one (or one) protease recognition sequence located between two amino acids (*e.g.*, two adjacent amino acids) of the native protein. In other aspects, nucleic acids encoding the recombinant transketolase A proteins are provided.

[0009] In some embodiments, the TktA protein comprises the sequence of SEQ ID NO:63. In certain embodiments, the TktA protein may comprise a sequence that is 90%, 95%, 98%, or 99% homologous to SEQ ID NO:63. In some embodiments, the nucleic acid encoding the TktA protein comprises the sequence of SEQ ID NO:57. In certain embodiments, the nucleic acid encoding the TktA protein may comprise a sequence that is 90%, 95%, 98%, or 99% homologous to SEQ ID NO:57.

[00010] In some embodiments, at least one protease recognition sequence is a protease recognition sequence recognized by a protease selected from the group consisting of alanine carboxypeptidase, *Armillaria mellea* astacin, bacterial leucyl aminopeptidase, cancer procoagulant, cathepsin B, clostripain, cytosol alanyl aminopeptidase, elastase, endoproteinase Arg-C, enterokinase, gastricsin, gelatinase, Gly-X carboxypeptidase, glycyl endopeptidase, human rhinovirus 3C protease, hypodermin C, Iga-specific serine endopeptidase, leucyl aminopeptidase, leucyl endopeptidase, lysC, lysosomal pro-X carboxypeptidase, lysyl aminopeptidase, methionyl aminopeptidase, myxobacter, nardilysin, pancreatic endopeptidase E, picornain 2A, picornain 3C, proendopeptidase, prolyl aminopeptidase, proprotein convertase I, proprotein convertase II, russellysin, saccharopepsin, semenogelase, T-plasminogen activator, thrombin, tissue kallikrein, tobacco etch virus (TEV), togavirin, tryptophanyl aminopeptidase, U-plasminogen activator, V8, venombin A, venombin AB, and Xaa-pro aminopeptidase.

[00011] In some embodiments, at least one protease recognition sequence is a protease recognition sequence recognized by human rhinovirus 3C protease. In some embodiments, the amino acid sequence of the protease recognition sequence recognized by human rhinovirus 3C protease comprises the sequence of SEQ ID NO:38. In some embodiments, the nucleic acid sequence of the protease recognition sequence recognized by human rhinovirus 3C protease comprises the sequence of SEQ ID NO:37.

[00012] In some embodiments, at least one protease recognition sequence of the Pgi protein is located after amino acid 410, 526, 527, 528, 529, 530, 531 or 532 of the sequence of SEQ ID NO:17. In some embodiments, at least one protease recognition sequence of the Pgi protein is located after amino acid 526 of the sequence of SEQ ID NO:17. In some embodiments, at least one protease recognition sequence of the nucleic acid encoding the Pgi protein is located after codons 410, 526, 527, 528, 529, 530, 531 or 532 of the sequence of SEQ ID NO:1. In some embodiments, the protease recognition sequence of the nucleic acid encoding the Pgi protein is located after codon 526 of the sequence of SEQ ID NO:1.

[00013] In some embodiments, at least one protease recognition sequence of the Pta protein is located after amino acid 381, 382, 387, or 409 of the sequence of SEQ ID NO:48. In some embodiments, at least one protease recognition sequence of the Pta protein is located after amino acid 381 of the sequence of SEQ ID NO:48. In some embodiments, at least one protease recognition sequence of the nucleic acid encoding the Pta protein is located after codons 381, 382, 387, or 409 of the sequence of SEQ ID NO:47. In some embodiments, the protease recognition sequence of the nucleic acid encoding the Pta protein is located after codon 381 of the sequence of SEQ ID NO:47.

[00014] In some embodiments, at least one protease recognition sequence of the TktA protein is located after amino acid 635, 636, 637, 638, or 640 of the sequence of SEQ ID NO:63. In some embodiments, at least one protease recognition sequence of the TktA protein is located after amino acid 637 of the sequence of SEQ ID NO:63. In some embodiments, at least one protease recognition sequence of the nucleic acid encoding the TktA protein is located after codons 635, 636, 637, 638, or 640 of the sequence of SEQ ID NO:57. In some embodiments, the protease recognition sequence of the nucleic acid encoding the TktA protein is located after codon 637 of the sequence of SEQ ID NO:57.

[00015] In some embodiments, at least one protease recognition sequence of the Pgi protein may be located in a C terminal region of the protein. In some embodiments, at least one protease recognition sequence of the Pta protein may be located in a central region or a C terminal region of the protein. In some embodiments, at least one protease recognition sequence of the TktA protein may be located in a central region or a C terminal region of the protein. In some embodiments, at least one protease recognition sequence of the Pgi protein may be located in a solvent-exposed loop region of the protein. In some embodiments, at least one protease recognition sequence of the Pta protein may be located in a solvent-exposed loop region of the protein. In some embodiments, at least one protease recognition sequence of the TktA protein may be located in a solvent-exposed loop region of the protein.

[00016] Some aspects of the invention provide vectors comprising a nucleic acid encoding a recombinant Pgi protein. Some aspects of the invention provide vectors comprising a nucleic acid encoding a recombinant Pta protein. Some aspects of the invention provide vectors comprising a nucleic acid encoding a recombinant TktA protein. In some embodiments, a vector may be a cloning vector or an expression vector. In some embodiments, a vector may be a plasmid, a fosmid, a phagemid, a virus genome or an artificial chromosome. In certain embodiments, a vector is a plasmid.

[00017] Other aspects of the invention provide cells that comprise any one of the proteins, nucleic acids, or vectors described herein. In some embodiments, the cell is a bacterial cell, a fungal cell, a mammalian cell or a plant cell. In some embodiments, the cell is a bacterial cell. In some embodiments, the bacterial cell is an *Escherichia coli* cell.

[00018] Further aspects of the invention provide cells that express recombinant proteins with a protease cleave site as described herein. In certain embodiments, the recombinant protein is a recombinant Pgi protein. In certain other embodiments, the recombinant protein is a recombinant Pta protein. In certain other embodiments, the recombinant protein is a recombinant TktA protein. In some embodiments, the cell is a bacterial cell, a fungal cell, a mammalian cell, or a plant cell. In certain embodiments, the cell is a bacterial cell such as, for example, an *Escherichia coli* cell.

[00019] Also provided are lysates of any of the cells described herein.

[00020] In various aspects of the invention, provided herein are recombinant proteins with at least one protease recognition sequence located between two amino acids (*e.g.*, two adjacent amino acids) of the native protein in a solvent-exposed loop region of the protein, wherein at least one protease recognition sequence is cleaved by a cognate protease with single recognition sequence specificity, and wherein the recombinant protein activity in the presence of the cognate protease is reduced by at least 30%, at least 40%, at least 50%, at least 75%, at least 90%, at least 99%, or at least 99.9% (or is about 70%, about 60%, about 50%, about 25%, about 10%, about 1%, or about 0.1%) of the recombinant protein activity in the absence of the cognate protease. In some embodiments, the activity of the recombinant protein in the absence of the cognate protease is sufficient to permit wild-type growth of the cell. In some embodiments, the activity of the recombinant protein in the presence of the cognate protease is reduced by at least 50% in comparison to a wild-type control protein, and wherein the activity of the recombinant protein in the absence of the cognate protease is at least 80% in comparison to a wild-type control protein. In some embodiments, the recombinant protein, in the absence of the cognate protease, maintains a cellular growth rate of at least 75% of a wild-type cellular growth rate (*e.g.*, growth rate of a cell without the recombinant protein).

[00021] Methods of engineering recombinant proteins (*e.g.*, recombinant Pgi proteins and/or recombinant Pta proteins and/or recombinant TktA proteins) with a protease recognition sequence are also provided. The methods may comprise the steps of: (a) transforming cells with a plurality of nucleic acid variants, wherein each nucleic acid variant

contains a nucleotide sequence that encodes a recombinant protein with at least one protease recognition sequence located between two amino acids (*e.g.*, two adjacent amino acids) of the native protein, and a nucleic acid with an inducible promoter sequence operably linked to a nucleotide sequence encoding a cognate protease; (b) culturing the cells under non-inducing conditions on media that prohibits cell growth of inactive recombinant proteins and collecting cells with a growth rate comparable to wild-type control cells; and (c) culturing the cells under conditions that induce expression of the cognate protease and collecting cells that do not grow or that have a reduced growth rate in comparison to wild-type control cells. In some embodiments, the cells are transformed with the nucleic acid with an inducible promoter sequence operably linked to a nucleotide sequence encoding a cognate protease after step (b) and before step (c). In some embodiments, the methods further comprise sequencing the nucleic acid variants of the cells collected in step (c) to identify the location of at least one protease recognition sequence.

[00022] In some embodiments, the wild-type protein (*e.g.*, Pgi protein and/or Pta protein and/or TktA protein) is required for cell growth and the genome of the cell lacks a chromosomal copy of the wild-type gene encoding the wild-type protein.

[00023] In some aspects, provided herein are methods of engineering a plurality of nucleic acid variants encoding recombinant proteins. The methods may comprise inserting at least one sequence that encodes at least one protease recognition sequence after each codon of a nucleic acid encoding a target protein to produce a plurality of nucleic acid variants encoding recombinant proteins, wherein each recombinant protein has a protease recognition sequence at a unique location in its primary amino acid sequence. In some embodiments, the methods further comprise (a) transforming cells with the plurality of nucleic acid variants, and a nucleic acid with an inducible promoter sequence operably linked to a nucleotide sequence encoding a cognate protease; (b) culturing the cells under non-inducing conditions on media that prohibits cell growth of inactive recombinant proteins and collecting cells with a normal growth rate; and (c) culturing the cells under conditions that induce expression of the cognate protease and collecting cells that do not grow or that have a reduced growth rate. In some embodiments, the cells are transformed with the nucleic acid with an inducible promoter sequence operably linked to a nucleotide sequence encoding a cognate protease after step (b) and before step (c). In some embodiments, the methods further comprise sequencing the nucleic acid variants of the cells collected in step (c) to identify the location of the protease recognition sequence.

[00024] In other aspects of the invention, provided herein are heterogeneous pluralities of nucleic acid variants, wherein each nucleic acid variant encodes a recombinant protein that is modified to include at least one protease recognition sequence located between two amino acids (*e.g.*, two adjacent amino acids) of the native sequence. In some embodiments, a single protease recognition sequence is located between two amino acids of the native protein.

[00025] Also provided herein are heterogeneous cell populations, wherein each cell of the population comprises a nucleic acid variant, and wherein each nucleic acid variant encodes a recombinant protein that is modified to include at least one protease recognition sequence located between two amino acids (*e.g.*, two adjacent amino acids) of the native sequence; and a nucleic acid encoding a cognate protease operably linked to an inducible promoter. In some embodiments, a single protease recognition sequence is located between two amino acids of the native protein.

BRIEF DESCRIPTION OF DRAWINGS

[00026] The accompanying drawings are not intended to be drawn to scale. For purposes of clarity, not every component may be labeled in every drawing.

[00027] Figure 1A shows a schematic of one exemplary method for preparing a recombinant protein with a protease recognition sequence by inserting the protease recognition sequence between two amino acids of the recombinant protein.

[00028] Figure 1B shows a schematic of another exemplary method for preparing a recombinant protein with a protease recognition sequence by replacing native amino acids of the recombinant protein with a protease recognition sequence.

[00029] Figure 2 shows a schematic of a modification to one of the methods of the invention, where a partial protease recognition sequence is inserted between two amino acids of the wild-type protein or codons such that the full recognition sequence is reconstituted in the final product.

[00030] Figure 3 is a diagram of plasmid pGLA042.

[00031] Figure 4 is a diagram of plasmid pGLC217.

[00032] Figure 5 shows a graph of phosphoglucose isomerase (Pgi) protein activity in cell lysates as a function of time.

[00033] Figure 6 shows a crystal structure of Pgi indicating a location of protease site insertion.

[00034] Figure 7 is a diagram of plasmid pGLC089.

- [00035] Figure 8 is a diagram of plasmid pGLC221.
- [00036] Figure 9 is a table summarizing strain and growth rate data (left) and a graph of Pgi activity with and without protease induction (right).
- [00037] Figure 10A shows a graph of acetate excretion in Pta variants with and without protease induction.
- [00038] Figure 10B shows a graph of lactate excretion in Pta variants with and without protease induction.
- [00039] Figure 11 shows a graph of phosphotransacetylase (Pta) protein activity in cell lysates with and without protease induction.
- [00040] Figure 12 shows a crystal structure of a transketolase A (TktA) protein dimer. The loops highlighted by a white dotted line (top-center and left-center) illustrate the location where the human rhinovirus (HRV) protease recognition sequence motif was inserted to enable protease-mediated deactivation of the TktA protein.

BRIEF DESCRIPTION OF SEQUENCE LISTING

- [00041] SEQ ID NO:1 is a nucleotide sequence of a wild-type *pgi* gene.
- [00042] SEQ ID NO:2 is a nucleotide sequence of a *pgi* gene variant with a human rhinovirus (HRV) 3C recognition sequence inserted after codon 108.
- [00043] SEQ ID NO:3 is a nucleotide sequence of a *pgi* gene variant with an HRV 3C recognition sequence inserted after codon 109.
- [00044] SEQ ID NO:4 is a nucleotide sequence of a *pgi* gene variant with an HRV 3C recognition sequence inserted after codon 110.
- [00045] SEQ ID NO:5 is a nucleotide sequence of a *pgi* gene variant with an HRV 3C recognition sequence inserted after codon 138.
- [00046] SEQ ID NO:6 is a nucleotide sequence of a *pgi* gene variant with an HRV 3C recognition sequence inserted after codon 410.
- [00047] SEQ ID NO:7 is a nucleotide sequence of a *pgi* gene variant with an HRV 3C recognition sequence inserted after codon 524.
- [00048] SEQ ID NO:8 is a nucleotide sequence of a *pgi* gene variant with an HRV 3C recognition sequence inserted after codon 525.
- [00049] SEQ ID NO:9 is a nucleotide sequence of a *pgi* gene variant with an HRV 3C recognition sequence inserted after codon 526.

- [00050] SEQ ID NO:10 is a nucleotide sequence of a *pgi* gene variant with an HRV 3C recognition sequence inserted after codon 527.
- [00051] SEQ ID NO:11 is a nucleotide sequence of a *pgi* gene variant with an HRV 3C recognition sequence inserted after codon 528.
- [00052] SEQ ID NO:12 is a nucleotide sequence of a *pgi* gene variant with an HRV 3C recognition sequence inserted after codon 529.
- [00053] SEQ ID NO:13 is a nucleotide sequence of a *pgi* gene variant with an HRV 3C recognition sequence inserted after codon 530.
- [00054] SEQ ID NO:14 is a nucleotide sequence of a *pgi* gene variant with an HRV 3C recognition sequence inserted after codon 531.
- [00055] SEQ ID NO:15 is a nucleotide sequence of a *pgi* gene variant with an HRV 3C recognition sequence inserted after codon 532.
- [00056] SEQ ID NO:16 is a nucleotide sequence of a *pgi* gene variant with an HRV 3C recognition sequence inserted after codon 545.
- [00057] SEQ ID NO:17 is an amino acid sequence of a wild-type Pgi protein.
- [00058] SEQ ID NO:18 is an amino acid sequence of a Pgi protein encoded by SEQ ID NO:2.
- [00059] SEQ ID NO:19 is an amino acid sequence of a Pgi protein encoded by SEQ ID NO:3.
- [00060] SEQ ID NO:20 is an amino acid sequence of a Pgi protein encoded by SEQ ID NO:4.
- [00061] SEQ ID NO:21 is an amino acid sequence of a Pgi protein encoded by SEQ ID NO:5.
- [00062] SEQ ID NO:22 is an amino acid sequence of a Pgi protein encoded by SEQ ID NO:6.
- [00063] SEQ ID NO:23 is an amino acid sequence of a Pgi protein encoded by SEQ ID NO:7.
- [00064] SEQ ID NO:24 is an amino acid sequence of a Pgi protein encoded by SEQ ID NO:8.
- [00065] SEQ ID NO:25 is an amino acid sequence of a Pgi protein encoded by SEQ ID NO:9.
- [00066] SEQ ID NO:26 is an amino acid sequence of a Pgi protein encoded by SEQ ID NO:10.

- [00067] SEQ ID NO:27 is an amino acid sequence of a Pgi protein encoded by SEQ ID NO:11.
- [00068] SEQ ID NO:28 is an amino acid sequence of a Pgi protein encoded by SEQ ID NO:12.
- [00069] SEQ ID NO:29 is an amino acid sequence of a Pgi protein encoded by SEQ ID NO:13.
- [00070] SEQ ID NO:30 is an amino acid sequence of a Pgi protein encoded by SEQ ID NO:14.
- [00071] SEQ ID NO:31 is an amino acid sequence of a Pgi protein encoded by SEQ ID NO:15.
- [00072] SEQ ID NO:32 is an amino acid sequence of a Pgi protein encoded by SEQ ID NO:16.
- [00073] SEQ ID NO:33 is a nucleotide sequence of a codon-optimized HRV 3C protease.
- [00074] SEQ ID NO:34 is an amino acid sequence of a codon-optimized HRV 3C protease.
- [00075] SEQ ID NO:35 is a nucleotide sequence of a codon-optimized HRV 3C protease with an OmpA leader sequence.
- [00076] SEQ ID NO:36 is an amino acid sequence of a codon-optimized HRV 3C protease with an OmpA leader sequence.
- [00077] SEQ ID NO:37 is a nucleotide sequence of a HRV 3C protease recognition sequence.
- [00078] SEQ ID NO:38 is an amino acid sequence of a HRV 3C protease recognition sequence.
- [00079] SEQ ID NO:39 is an amino acid sequence of a partial HRV 3C protease recognition sequence.
- [00080] SEQ ID NO:40 is an amino acid sequence of a partial HRV 3C protease recognition sequence.
- [00081] SEQ ID NO:41 is an amino acid sequence of a partial HRV 3C protease recognition sequence.
- [00082] SEQ ID NO:42 is an amino acid sequence of a partial HRV 3C protease recognition sequence.

- [00083] SEQ ID NO:43 is an amino acid sequence of a partial HRV 3C protease recognition sequence.
- [00084] SEQ ID NO:44 is an amino acid sequence of a partial HRV 3C protease recognition sequence.
- [00085] SEQ ID NO:45 is an amino acid sequence of a partial HRV 3C protease recognition sequence.
- [00086] SEQ ID NO:46 is an amino acid sequence of an N-terminal OmpA leader sequence.
- [00087] SEQ ID NO:47 is a nucleotide sequence of a wild-type *pta* gene.
- [00088] SEQ ID NO:48 is an amino acid sequence of a wild-type Pta protein.
- [00089] SEQ ID NO:49 is a nucleotide sequence of a *pta* gene variant with an HRV 3C recognition sequence inserted after codon 381.
- [00090] SEQ ID NO:50 is an amino acid sequence of a Pta protein encoded by SEQ ID NO:49.
- [00091] SEQ ID NO:51 is a nucleotide sequence of a *pta* gene variant with an HRV 3C recognition sequence inserted after codon 382.
- [00092] SEQ ID NO:52 is an amino acid sequence of a Pta protein encoded by SEQ ID NO:51.
- [00093] SEQ ID NO:53 is a nucleotide sequence of a *pta* gene variant with an HRV 3C recognition sequence inserted after codon 387.
- [00094] SEQ ID NO:54 is an amino acid sequence of a Pta protein encoded by SEQ ID NO:53.
- [00095] SEQ ID NO:55 is a nucleotide sequence of a *pta* gene variant with an HRV 3C recognition sequence inserted after codon 409.
- [00096] SEQ ID NO:56 is an amino acid sequence of a Pta protein encoded by SEQ ID NO:55.
- [00097] SEQ ID NO:57 is a nucleotide sequence of a wild-type *tktA* gene.
- [00098] SEQ ID NO:58 is a nucleotide sequence of a *tktA* gene variant with an HRV 3C recognition sequence inserted after codon 635.
- [00099] SEQ ID NO:59 is a nucleotide sequence of a *tktA* gene variant with an HRV 3C recognition sequence inserted after codon 636.
- [00100] SEQ ID NO:60 is a nucleotide sequence of a *tktA* gene variant with an HRV 3C recognition sequence inserted after codon 637.

- [000101] SEQ ID NO:61 is a nucleotide sequence of a *tktA* gene variant with an HRV 3C recognition sequence inserted after codon 638.
- [000102] SEQ ID NO:62 is a nucleotide sequence of a *tktA* gene variant with an HRV 3C recognition sequence inserted after codon 640.
- [000103] SEQ ID NO:63 is an amino acid sequence of a wild-type TktA protein.
- [000104] SEQ ID NO:64 is an amino acid sequence of a TktA protein encoded by SEQ ID NO:58.
- [000105] SEQ ID NO:65 is an amino acid sequence of a TktA protein encoded by SEQ ID NO:59.
- [000106] SEQ ID NO:66 is an amino acid sequence of a TktA protein encoded by SEQ ID NO:60.
- [000107] SEQ ID NO:67 is an amino acid sequence of a TktA protein encoded by SEQ ID NO:61.
- [000108] SEQ ID NO:68 is an amino acid sequence of a TktA protein encoded by SEQ ID NO:62.
- [000109] SEQ ID NO:69 is a nucleotide sequence of a HRV protease recognition sequence.
- [000110] SEQ ID NO:70 is an amino acid sequence of a Pgi protein encoded by SEQ ID NO:71.
- [000111] SEQ ID NO:71 is a nucleotide sequence of a Pgi protein.
- [000112] SEQ ID NO:72 is an amino acid sequence of a Pgi-HRV-I001 protein encoded by SEQ ID NO:73.
- [000113] SEQ ID NO:73 is a nucleotide sequence of a *pgi-hrv-I001* gene variant with an HRV protease recognition sequence inserted after the first illustrated codon in Figure 1A.
- [000114] SEQ ID NO:74 is an amino acid sequence of a Pgi-HRV-I002 protein encoded by SEQ ID NO:75.
- [000115] SEQ ID NO:75 is a nucleotide sequence of a *pgi-hrv-I002* gene variant with an HRV protease recognition sequence inserted after the second illustrated codon in Figure 1A.
- [000116] SEQ ID NO:76 is an amino acid sequence of a Pgi-HRV-I003 protein encoded by SEQ ID NO:77.
- [000117] SEQ ID NO:77 is a nucleotide sequence of a *pgi-hrv-I003* gene variant with an HRV protease recognition sequence inserted after the third illustrated codon in Figure 1A.

[000118] SEQ ID NO:78 is an amino acid sequence of a Pgi-HRV-R001 protein encoded by SEQ ID NO:79.

[000119] SEQ ID NO:79 is a nucleotide sequence of a *pgi-hrv-R001* gene variant with an HRV protease recognition sequence substitution after the first illustrated codon in Figure 2A.

[000120] SEQ ID NO:80 is an amino acid sequence of a Pgi-HRV-R002 protein encoded by SEQ ID NO:81.

[000121] SEQ ID NO:81 is a nucleotide sequence of a *pgi-hrv-R002* gene variant with an HRV protease recognition sequence substitution after the second illustrated codon in Figure 2A.

[000122] SEQ ID NO:82 is an amino acid sequence of a Pgi-HRV-R003 protein encoded by SEQ ID NO:83.

[000123] SEQ ID NO:83 is a nucleotide sequence of a *pgi-hrv-R003* gene variant with an HRV protease recognition sequence substitution after the third illustrated codon in Figure 2A.

[000124] SEQ ID NO:84 is an amino acid sequence of a Pgi-HRV-I005 protein encoded by SEQ ID NO:85.

[000125] SEQ ID NO:85 is a nucleotide sequence of a *pgi-hrv-I005* gene variant with an HRV protease recognition sequence insertion before a proline.

[000126] SEQ ID NO:86 is an amino acid sequence of a Pgi-HRV-I015 protein encoded by SEQ ID NO:87.

[000127] SEQ ID NO:87 is a nucleotide sequence of a *pgi-prv-I015* gene variant with an HRV protease recognition sequence insertion after a leucine.

DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS OF THE INVENTION

[000128] Many key proteins in biosynthetic pathways are important for cell growth. Deletion or inactivation of these proteins is often difficult or impossible because doing so results in reduced cell growth or viability, rendering the cells unsatisfactory for producing a compound of interest. The invention addresses this problem of cell growth inhibition by providing recombinant proteins (*e.g.*, enzymes) that are active during cell growth and inactive during *in vitro* cell-free production of a compound of interest. The recombinant proteins engineered by the methods provided herein have a protease recognition sequence selectively located in their primary amino acid sequence such that, despite the presence of the

recognition sequence, the activity of the recombinant protein is sufficient to enable wild-type growth of the cell. The recombinant proteins can be selectively inactivated by the introduction, expression, and/or activation of a cognate protease, which cleaves the recombinant target protein specifically at the protease recognition sequence, thereby rendering the recombinant target protein inactive (or with reduced activity). Thus, the recombinant proteins of the invention are useful for engineering and/or altering biosynthetic pathways to produce a compound of interest.

[000129] The recombinant proteins and methods provided herein are useful for engineering and altering metabolic pathways in a cell and a cell-free system. For example, protease targeting of the recombinant Pgi enzymes of the invention permits control of carbon flux between glycolysis and the pentose phosphate pathway in cell-free reactions without altering the function of this key enzyme during cell growth. Thus, the invention provides a way to control metabolic flux through the manipulation of key pathway enzymes, including manipulation to maintain or alter the cellular activity of key pathway enzymes or competitive enzymes.

[000130] To produce recombinant proteins with a protease recognition sequence that have an activity level that enables wild-type growth of the cell, a library of nucleic acid variants may be generated, wherein each nucleic acid variant encodes a recombinant protein with at least one protease recognition sequence located between two amino acids (*e.g.*, two adjacent amino acids) of the wild-type primary sequence of the protein. From this library of variants, recombinant proteins are produced and screened for protein activity in the absence of protease and protein inactivity in the presence of protease. A population of cells is first transformed with the library of nucleic acid variants. To select for active expressed recombinant proteins, the cells are grown on media that prohibits growth of inactive recombinant proteins. Cells that do not display growth defects are presumed to contain an active recombinant protein, despite introduction of the protease recognition sequence, and are selected for further characterization. To select for cells that contain a recombinant protein that can be inactivated, a nucleic acid with an inducible promoter sequence operably linked to a nucleotide sequence encoding a cognate protease may be introduced into the cells. The nucleic acid encoding the inducible cognate protease may be introduced either before or after selection of active recombinant protein. The cells presumably containing active recombinant protein are then grown under conditions that induce expression of a cognate protease. Cells that do not grow, or that grow poorly (*e.g.*, display growth defects), are presumed to contain

the desired inactive recombinant protein and are selected for further characterization and/or use. The nucleic acid variants of the selected cells may be amplified and sequenced to identify the location of the protease recognition sequence.

Proteins and Metabolic Pathways

[000131] A “protein” or “wild-type protein,” as used herein, refers to a molecule composed of amino acids joined together by peptide bonds in a linear chain. A “native” amino acid, as used herein, refers to an amino acid in the primary amino acid sequence of a wild-type protein (*i.e.*, not a modified or mutated amino acid). A “target protein,” as used herein, refers to a wild-type protein of interest (*i.e.*, not a recombinant protein) or a protein to be engineered with a protease recognition sequence as described herein. A “recombinant protein,” as used herein, refers to a protein derived from recombinant nucleic acid, which is formed artificially by combining nucleic acid from different sources. In some embodiments, recombinant proteins of the invention differ from each other in that the location of the single protease recognition sequence is unique to each recombinant protein. For example, one recombinant protein may have a protease recognition sequence located after the first amino acid of the primary amino acid sequence, another recombinant protein may have a protease recognition sequence located after the second amino acid of the primary amino acid sequence, yet another recombinant protein may have a protease recognition sequence located after the third amino acid of the primary amino acid sequence, and so on. Thus, a plurality of recombinant proteins is typically a heterogeneous plurality.

[000132] The recombinant proteins of the invention may be used to engineer metabolic pathways, or a sequence of biochemical reactions catalyzed by enzymes. Examples of metabolic pathways that may be engineered in accordance with the invention include, without limitation, those involved in carbohydrate metabolism, lipid metabolism, amino acid metabolism, and energy metabolism. In some embodiments, the metabolic pathway is glycolysis. In some embodiments, the metabolic pathway is acetate overflow metabolism. In some embodiments, the metabolic pathway is the pentose phosphate pathway.

Phosphoglucose isomerase (Pgi)

[000133] In some embodiments, the target protein is a phosphoglucose isomerase (Pgi) enzyme, for example, a Pgi enzyme from *Escherichia coli* (*E. coli*). This enzyme catalyzes the inter-conversion of glucose-6-phosphate and fructose-6-phosphate and is the first

committed step in glycolysis. Inactivation of Pgi inhibits cell growth; however, Pgi activity results in the diversion of glucose to the glycolysis pathway, which in turn results in a shortage of glucose for cell-free production of compounds of interest that are derived from ribose. A nucleic acid containing the *pgi* gene that encodes Pgi enzyme may be modified by any of the methods provided herein or known in the art to generate a variant comprising a protease recognition sequence. In some embodiments, the protease recognition sequence used is a human rhinovirus (HRV) 3C protease recognition sequence (*e.g.*, SEQ ID NO:37, SEQ ID NO:38), though the invention is not so limited. In some embodiments, the HRV 3C recognition sequence is inserted in-frame after each codon of the *pgi* gene. In some embodiments, the HRV 3C recognition sequence is inserted after each codon of the *pgi* gene, excluding the first and/or last codon. In some embodiments, the HRV 3C recognition sequence is inserted after amino acids 2-5, 9, 24-25, 33-36, 58-59, 85-96, 105-111, 113-115, 137-141, 143-144, 146, 173-176, 196, 250-251, 254, 366-370, 398-399, 410-414, 447-451, 477, 526-532 or 545. In some embodiments, the HRV 3C recognition sequence is inserted after at least one, or each, codon of the solvent-exposed loop regions of the Pgi protein.

[000134] In some embodiments, the recombinant Pgi proteins of the invention contain an HRV 3C recognition sequence located after amino acid 108, 109, 110, 138, 410, 524, 525, 526, 527, 528, 529, 530, 531, 532 or 545. In some embodiments, the Pgi variants of the invention comprise an amino acid sequence selected from SEQ ID NO:18-32.

[000135] In some embodiments, the nucleic acid *pgi* variants (*e.g.*, genes) of the invention contain an HRV 3C recognition sequence located after codon 108, 109, 110, 138, 410, 524, 525, 526, 527, 528, 529, 530, 531, 532 or 545. In some embodiments, the *pgi* variants of the invention comprise a nucleotide sequence selected from SEQ ID NO:2-16.

[000136] In some embodiments, the HRV 3C recognition sequence is inserted between non-adjacent codons of the *pgi* gene. In some embodiments, the HRV 3C recognition sequence replaces native codons of the *pgi* gene. For example, in some embodiments, the eight codons of HRV 3C recognition sequence (*i.e.*, CTG GAA GTG CTG TTT CAG GGT CCG; SEQ ID NO:37) may replace eight contiguous codons of the *pgi* gene.

Phosphotransacetylase (Pta)

[000137] In some embodiments, the target protein is a phosphotransacetylase (Pta) enzyme, for example, a Pta enzyme from *Escherichia coli* (*E. coli*). This enzyme catalyzes the reversible interconversion of acetyl-CoA and acetyl phosphate. A nucleic acid containing

the *pta* gene that encodes Pta enzyme may be modified by any of the methods provided herein or known in the art to generate a variant comprising a protease recognition sequence. In some embodiments, the protease recognition sequence used is a human rhinovirus (HRV) 3C protease recognition sequence (*e.g.*, SEQ ID NO:37, SEQ ID NO:38), though the invention is not so limited. In some embodiments, the HRV 3C recognition sequence is inserted in-frame after each codon of the *pta* gene. In some embodiments, the HRV 3C recognition sequence is inserted after each codon of the *pta* gene, excluding the first and/or last codon. In some embodiments, the HRV 3C recognition sequence is inserted after at least one, or each, codon of the solvent-exposed loop regions of the Pta protein.

[000138] In some embodiments, the recombinant Pta proteins of the invention contain an HRV 3C recognition sequence located after amino acid 381, 382, 387, or 409. In some embodiments, the Pta variants of the invention comprise an amino acid sequence selected from SEQ ID NO:50, SEQ ID NO:52, SEQ ID NO:54, and SEQ ID NO:56.

[000139] In some embodiments, the nucleic acid *pta* variants (*e.g.*, genes) of the invention contain an HRV 3C recognition sequence located after codon 381, 382, 387, or 409. In some embodiments, the *pta* variants of the invention comprise a nucleotide sequence selected from SEQ ID NO:49, SEQ ID NO:51, SEQ ID NO:53, and SEQ ID NO:55.

[000140] In some embodiments, the HRV 3C recognition sequence is inserted between non-adjacent codons of the *pta* gene. In some embodiments, the HRV 3C recognition sequence replaces native codons of the *pta* gene. For example, in some embodiments, the eight codons of HRV 3C recognition sequence (*i.e.*, CTG GAA GTG CTG TTT CAG GGT CCG; SEQ ID NO:37) may replace eight contiguous codons of the *pta* gene.

Transketolase A (TktA)

[000141] In some embodiments, the target protein is a transketolase A (TktA) enzyme, for example, a TktA enzyme from *Escherichia coli* (*E. coli*). TktA, together with transketolase B (TktB) catalyze two reversible ketol transfer reactions in the pentose phosphate pathway. A nucleic acid containing the *tktA* gene that encodes TktA enzyme may be modified by any of the methods provided herein or known in the art to generate a variant comprising a protease recognition sequence. In some embodiments, the protease recognition sequence used is a human rhinovirus (HRV) 3C protease recognition sequence (*e.g.*, SEQ ID NO:37, SEQ ID NO:38), though the invention is not so limited. In some embodiments, the HRV 3C recognition sequence is inserted in-frame after each codon of the *tktA* gene. In

some embodiments, the HRV 3C recognition sequence is inserted after each codon of the *tktA* gene, excluding the first and/or last codon. In some embodiments, the HRV 3C recognition sequence is inserted after at least one, or each, codon of the solvent-exposed loop regions of the TktA protein.

[000142] In some embodiments, the recombinant TktA proteins of the invention contain an HRV 3C recognition sequence located after amino acid 635, 636, 637, 638, or 640. In some embodiments, the TktA variants of the invention comprise an amino acid sequence selected from SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO: 67 and SEQ ID NO:68.

[000143] In some embodiments, the nucleic acid *tktA* variants (*e.g.*, genes) of the invention contain an HRV 3C recognition sequence located after codon 635, 636, 637, 638, or 640. In some embodiments, the *tktA* variants of the invention comprise a nucleotide sequence selected from SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61 and SEQ ID NO:62.

[000144] In some embodiments, the HRV 3C recognition sequence is inserted between non-adjacent codons of the *tktA* gene. In some embodiments, the HRV 3C recognition sequence replaces native codons of the *tktA* gene. For example, in some embodiments, the eight codons of HRV 3C recognition sequence (*i.e.*, CTG GAA GTG CTG TTT CAG GGT CCG; SEQ ID NO:37) may replace eight contiguous codons of the *tktA* gene.

Proteases and Cognate Recognition Sequences

[000145] The proteins of the invention may be inactivated by any one of a variety of proteases that cleave at specific recognition sequences. As used herein, “protease recognition sequence,” in the context of a protein, refers to an amino acid sequence that is recognized and cleaved by a cognate protease. In the context of a nucleic acid that encodes a protein, a “protease recognition sequence” refers to a sequence that encodes the amino acid sequence recognized and cleaved by a cognate protease. As used herein, “cognate protease” refers to a protease that cleaves and thereby inactivates a recombinant target protein (*e.g.*, enzyme). Cognate proteases that may be used herein include those with single, specific recognition sequence, meaning the proteases cleave within or adjacent to a specific sequence of one or more amino acids. For example, human rhinovirus 3C protease is highly specific for the recognition sequence Leu-Glu-Val-Leu-Phe-Gln-Gly-Pro (SEQ ID NO:38). The protease recognizes this sequence and cleaves after the glutamine residue. The human rhinovirus 3C

protease does not typically recognize and cleave other recognition sequences although all proteases are somewhat promiscuous and may recognize and cleave other sites but at a much reduced rate. In some embodiments, the proteins of the invention are prepared with an engineered human rhinovirus 3C protease recognition sequence.

[000146] Other examples of proteases that may be used in accordance with the invention include, without limitation, alanine carboxypeptidase, *Armillaria mellea* astacin, bacterial leucyl aminopeptidase, cancer procoagulant, cathepsin B, clostripain, cytosol alanyl aminopeptidase, elastase, endoproteinase Arg-C, enterokinase, gastricsin, gelatinase, Gly-X carboxypeptidase, glycyl endopeptidase, human rhinovirus 3C protease, hypodermin C, Iga-specific serine endopeptidase, leucyl aminopeptidase, leucyl endopeptidase, lysC, lysosomal pro-X carboxypeptidase, lysyl aminopeptidase, methionyl aminopeptidase, myxobacter, nardilysin, pancreatic endopeptidase E, picornain 2A, picornain 3C, proendopeptidase, prolyl aminopeptidase, proprotein convertase I, proprotein convertase II, russellysin, saccharopepsin, semenogelase, T-plasminogen activator, thrombin, tissue kallikrein, tobacco etch virus (TEV), togavirin, tryptophanyl aminopeptidase, U-plasminogen activator, V8, venombin A, venombin AB and Xaa-pro aminopeptidase (*see* Rawlings, S. D., *et al.*, *Handbook of Proteolytic Enzymes*, Academic Press, 2013, *Science*, Elsevier Ltd., 4094 pages, incorporated herein by reference for its teachings relating to the structural chemistry and biological aspects of the proteases described therein). Other proteases may be used in accordance with the invention.

Nucleic Acids

[000147] The invention encompasses nucleic acids encoding the recombinant proteins (*e.g.*, recombinant Pgi proteins and/or recombinant Pta proteins and/or recombinant TktA proteins) described herein. A “nucleic acid,” as used herein, refers to at least two nucleotides (*e.g.*, adenine, thymine, cytosine, guanine, uracil) covalently linked together. A nucleic acid of the invention will generally contain phosphodiester bonds. A nucleic acid may be single-stranded (ss) or double-stranded (ds), DNA or RNA. In some embodiments, the nucleic acid is in the form of cDNA. In some embodiments, the nucleic acid is in the form of genomic DNA. A “codon,” as used herein, refers to a set of three adjacent nucleotides that encode an amino acid. The codons of the invention are defined and numbered by the initial nucleotide from which translation starts.

[000148] In some embodiments, linear double-stranded nucleic acid (*e.g.*, DNA) variants are prepared in accordance with the invention. In some instances, the linear double-stranded nucleic acid variants comprise a variant gene sequence encoding a recombinant protein with a protease recognition sequence as well as at least 30 nucleotide base pairs (bp) of additional sequence upstream of the start codon and at least 30 nucleotide base pairs of additional sequence downstream of the stop codon of the gene, wherein each additional sequence is homologous to the wild-type gene locus of the genome of the cell into which the nucleic acid will be transformed. As used herein, “wild-type gene” refers to the wild-type gene encoding the wild-type protein that corresponds to the recombinant protein with at least one (or one) protease recognition site. For example, if the target protein is Pgi and the cell being transformed is *E. coli*, the nucleic acid will contain a gene variant encoding Pgi with at least one protease recognition sequence, at least 30 bp of additional sequence upstream of the start codon of the gene variant and homologous to the *pgi* locus of the *E. coli* genome, and at least 30 bp of additional sequence downstream of the start codon of the gene variant and homologous to the *pgi* locus of the *E. coli* genome. The additional sequence, in some instances, facilitates recombination of the gene variant with the chromosomal wild-type copy of the gene.

[000149] The invention encompasses vectors comprising a nucleic acid variant provided herein. A “vector,” as used herein, may be any of a number of nucleic acids into which a desired sequence or sequences may be inserted by restriction and ligation for transport between different genetic environments or for expression in a cell. Vectors are typically composed of DNA although RNA vectors are also available. Examples of vectors in accordance with the invention include, without limitation, plasmids, fosmids, phagemids, virus genomes, and artificial chromosomes. In some embodiments, a nucleic acid variant of the invention is provided in a recombinant cloning vector. In some embodiments, a nucleic acid variant of the invention is expressed in a recombinant expression vector.

[000150] A cloning vector of the invention is able to replicate autonomously or integrated in the genome of a cell. A cloning vector has an endonuclease restriction sequence at which the vector may be cut in a determinable fashion and into which a desired DNA sequence may be ligated such that the new recombinant vector retains its ability to replicate in a cell. In the case of plasmids, replication of the desired sequence may occur many times as the plasmid increases in copy number within a cell such as a bacterium or just a single time

per cell before the cell reproduces by mitosis. In the case of phage, replication may occur actively during a lytic phase or passively during a lysogenic phase.

[000151] An expression vector of the invention is one into which a desired DNA coding sequence may be inserted by restriction and ligation such that it is operably linked to regulatory sequences and may be expressed as an RNA transcript.

[000152] As used herein, a coding sequence and regulatory sequences (*e.g.*, promoter sequences) are said to be “operably” linked when they are covalently linked in such a way as to place the expression or transcription of the coding sequence under the influence or control of the regulatory sequences (*e.g.*, such that the regulatory sequence “drives” transcriptional initiation and/or expression of the coding sequence). If the coding sequences are to be translated into a functional protein, two DNA sequences are considered operably linked if induction of a promoter in the 5’ regulatory sequences results in the transcription of the coding sequence and if the nature of the linkage between the two DNA sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter region to direct the transcription of the coding sequences, or (3) interfere with the ability of the corresponding RNA transcript to be translated into a protein. Thus, a promoter region would be operably linked to a coding sequence if the promoter region can effect transcription of that DNA sequence such that the resulting transcript can be translated into the desired protein or polypeptide.

[000153] Vectors of the invention may further comprise a marker sequence for use in the identification of cells that have or have not been transformed or transfected with the vector. Markers include, for example, genes encoding proteins that increase or decrease either resistance or sensitivity to antibiotics (*e.g.*, ampicillin resistance genes, kanamycin resistance genes, neomycin resistance genes, tetracycline resistance genes and chloramphenicol resistance genes) or other compounds, genes encoding enzymes with activities detectable by standard assays known in the art (*e.g.*, β -galactosidase, luciferase or alkaline phosphatase), and genes that visibly affect the phenotype of transformed or transfected cells, hosts, colonies or plaques (*e.g.*, green fluorescent protein). In some embodiments, the vectors used herein are capable of autonomous replication and expression of the structural gene products present in the DNA segments to which they are operably linked.

[000154] When a nucleic acid that encodes any of the proteins of the invention is expressed in a cell, a variety of transcription control sequences may be used to direct its

expression. For example, a nucleic acid of the invention may contain a promoter, an enhancer, and/or a terminator. Alternatively, the vector into which the nucleic acid is inserted may contain such regulatory sequences.

[000155] A “promoter,” as used herein, refers to a control region of a nucleic acid sequence at which initiation and rate of transcription of the remainder of a nucleic acid sequence are controlled. A promoter may also contain subregions at which regulatory proteins and molecules may bind, such as RNA polymerase and other transcription factors. Promoters may be constitutive, inducible, activatable, repressible, tissue-specific or any combination thereof. A promoter drives expression or drives transcription of the nucleic acid sequence that it regulates. A promoter may be one naturally associated with a gene or sequence, as may be obtained by isolating the 5′-non-coding sequences located upstream of the coding segment and/or exon of a given gene or sequence. Such a promoter can be referred to as “endogenous.”

[000156] In some embodiments, a coding nucleic acid segment may be positioned under the control of a recombinant or heterologous promoter, which refers to a promoter that is not normally associated with the encoded nucleic acid sequence in its natural environment. A recombinant or heterologous enhancer refers to an enhancer not normally associated with a nucleic acid sequence in its natural environment. Such promoters or enhancers may include promoters or enhancers of other genes; promoters or enhancers isolated from any other prokaryotic, viral or eukaryotic cell; and synthetic promoters or enhancers that are not “naturally occurring” such as, for example, those that contain different elements of different transcriptional regulatory regions and/or mutations that alter expression through methods of genetic engineering that are known in the art. In addition to producing nucleic acid sequences of promoters and enhancers synthetically, sequences may be produced using recombinant cloning and/or nucleic acid amplification technology, including polymerase chain reaction (PCR). Furthermore, control sequences that direct transcription and/or expression of sequences within non-nuclear organelles such as mitochondria, chloroplasts and the like, may be used in accordance with the invention.

[000157] An “inducible promoter,” as used herein, is one that is characterized by initiating or enhancing transcriptional activity when in the presence of, influenced by or contacted by an inducer or inducing agent. An “inducer” or “inducing agent” may be endogenous or a normally exogenous compound or protein that is administered in such a way as to be active in inducing transcriptional activity from the inducible promoter.

[000158] Inducible promoters for use in accordance with the invention include any inducible promoter described herein or known to one of ordinary skill in the art. Examples of inducible promoters include, without limitation, chemically/biochemically-regulated and physically-regulated promoters such as isopropyl β -D-1-thiogalactopyranoside (IPTG)-regulated promoters, alcohol-regulated promoters, tetracycline-regulated promoters (*e.g.*, anhydrotetracycline (aTc)-responsive promoters and other tetracycline-responsive promoter systems, which include a tetracycline repressor protein (tetR), a tetracycline operator sequence (tetO) and a tetracycline transactivator fusion protein (tTA)), steroid-regulated promoters (*e.g.*, promoters based on the rat glucocorticoid receptor, human estrogen receptor, moth ecdysone receptors, and promoters from the steroid/retinoid/thyroid receptor superfamily), metal-regulated promoters (*e.g.*, promoters derived from metallothionein (proteins that bind and sequester metal ions) genes from yeast, mouse and human), pathogenesis-regulated promoters (*e.g.*, induced by salicylic acid, ethylene or benzothiadiazole (BTH)), temperature/heat-inducible promoters (*e.g.*, heat shock promoters), and light-regulated promoters (*e.g.*, light responsive promoters from plant cells).

[000159] Inducible promoters for use in accordance with the invention may function in both prokaryotic and eukaryotic host organisms. In some embodiments, mammalian inducible promoters are used. Examples of mammalian inducible promoters for use herein include, without limitation, promoter type PAct:PAIR, PART, PBIT, PCR5, PCTA, PETR, PNIC, PPIP, PROP, PSPA /PSCA, PTET, PTtgR, promoter type PRep:PCuO, PETR ON8, PNIC, PPIR ON, PSCA ON8, PTetO, PUREX8, promoter type PHyb:tetO7 -ETR8 - PhCMVmin, tetO7 -PIR3 -ETR8 -PhCMVmin, and scbR8 -PIR3 - PhCMVmin. In some embodiments, inducible promoters from other organisms, as well as synthetic promoters designed to function in a prokaryotic or eukaryotic host may be used. Examples of non-mammalian inducible promoters for use herein include, without limitation, Lentivirus promoters (*e.g.*, EF α , CMV, Human SynapsinI (hSynI), CaMKII α , hGFAP and TPH-2) and Adeno-Associated Virus promoters (*e.g.*, CaMKII α (AAV5), hSynI (AAV2), hThy1 (AAV5), fSST (AAV1), hGFAP (AAV5, AAV8), MBP (AAV8), SST (AAV2)). One important functional characteristic of the inducible promoters of the present invention is their inducibility by exposure to an externally applied inducer.

[000160] An inducible promoter for use in accordance with the invention may be induced by (or repressed by) one or more physiological condition(s), such as changes in pH, temperature, radiation, osmotic pressure, saline gradients, cell surface binding, and the

concentration of one or more extrinsic or intrinsic inducing agent(s). The extrinsic inducer or inducing agent may comprise, without limitation, amino acids and amino acid analogs, saccharides and polysaccharides, nucleic acids, protein transcriptional activators and repressors, cytokines, toxins, petroleum-based compounds, metal containing compounds, salts, ions, enzyme substrate analogs, hormones, or combinations thereof. Other inducible promoters may be used in accordance with the invention.

[000161] In some embodiments of the invention, a promoter may or may not be used in conjunction with an “enhancer,” which refers to a *cis*-acting regulatory sequence involved in the transcriptional activation of a nucleic acid sequence downstream of the promoter. An enhancer may be one naturally associated with a nucleic acid sequence, located either downstream or upstream of that sequence. The enhancer may be located at any functional location before or after the promoter and/or the encoded nucleic acid.

[000162] A “terminator” or “terminator sequence,” as used herein, is a nucleic acid sequence that causes transcription to stop. A terminator may be unidirectional or bidirectional. It is comprised of a DNA sequence involved in specific termination of an RNA transcript by an RNA polymerase. A terminator sequence prevents transcriptional activation of downstream nucleic acid sequences by upstream promoters. Thus, in certain embodiments, a terminator that ends the production of an RNA transcript is contemplated.

[000163] The most commonly used type of terminator is a forward terminator. When placed downstream of a nucleic acid sequence that is usually transcribed, a forward transcriptional terminator will cause transcription to abort. In some embodiments, bidirectional transcriptional terminators may be used, which usually cause transcription to terminate on both the forward and reverse strand. In some embodiments, reverse transcriptional terminators may be used, which usually terminate transcription on the reverse strand only.

[000164] In prokaryotic systems, terminators usually fall into two categories (1) rho-independent terminators and (2) rho-dependent terminators. Rho-independent terminators are generally composed of palindromic sequence that forms a stem loop rich in G-C base pairs followed by several T bases. Terminators for use in accordance with the invention include any terminator of transcription described herein or known to one of ordinary skill in the art. Examples of terminators include, without limitation, the termination sequences of genes such as, for example, the bovine growth hormone terminator, and viral termination sequences such as, for example, the SV40 terminator, *spy*, *yejM*, *secG-leuU*, *thrLABC*, *rrnB T1*,

hisLGDCBHAFI, *metZWV*, *rrnC*, *xapR*, *aspA*, and *arcA* terminator. In some embodiments, the termination signal may be a sequence that cannot be transcribed or translated, such as those resulting from a sequence truncation. Other terminators may be used in accordance with the invention.

[000165] In some embodiments, the nucleic acids are codon-optimized for improved expression of the recombinant proteins of the invention. Codon optimization, also referred to as biased codon usage, refers to differences in the frequency of occurrence of synonymous codons in coding DNA.

Cells

[000166] The invention encompasses any type of cell, including a prokaryotic and a eukaryotic cell, that recombinantly expresses the proteins provided herein. In some embodiments, the cell is a bacterial cell. In some embodiments, the bacterial cell is a cell of a bacterium from the genus *Escherichia*. In some embodiments, the bacterial cell is an *Escherichia coli* (*E. coli*) cell. In some embodiments, the cell is a fungal cell, such as, for example, a yeast cell (*e.g.*, a *Saccharomyces cerevisiae* cell). In some embodiments, the cell is a mammalian cell or a plant cell. It should be appreciated that some cells for use in accordance with the invention do not contain the wild-type chromosomal copy of a gene encoding the wild-type protein (*e.g.*, the wild-type protein corresponding to the recombinant protein with a protease recognition sequence).

[000167] The cells provided herein, in some embodiments, are prokaryotic cells that may be transformed with any of the nucleic acids of the invention. Transformation and transfection are processes by which exogenous genetic material is introduced into a prokaryotic cell and into a eukaryotic cell, respectively. Transformation can be achieved by electroporation or by chemical means. The cells to be transformed are typically in a state of competence. Thus, in some embodiments, the cells provided herein are electrocompetent or chemically competent cells. A variety of electrocompetent and chemically competent cells are known in the art and may be used in accordance with the invention.

[000168] In some embodiments, the cells are *Escherichia coli* (*E. coli*) cells such as, for example, JW3985-1 *E. coli* cells (Coli Genetic Shock Center; CHSC #10867; *Mol. Sys. Biol.* 2:2006-08, 2006, incorporated by reference herein). Other commercially available and non-commercially available cell lines may be used in accordance with the invention.

[000169] The cells of the invention may comprise selectable markers. Selectable markers include, without limitation, genes encoding proteins that increase or decrease either resistance or sensitivity to antibiotics (*e.g.*, ampicillin resistance genes, kanamycin resistance genes, neomycin resistance genes, tetracyclin resistance genes and chloramphenicol resistance genes) or other compounds, genes encoding enzymes with activities detectable by standard assays known in the art (*e.g.*, β -galactosidase, luciferase or alkaline phosphatase), and genes that visibly affect the phenotype of transformed or transfected cells, hosts, colonies, or plaques (*e.g.*, green fluorescent protein). Other selectable markers may be used in accordance with the invention.

Library Construction

[000170] The methods of the invention may be used to construct a library of the nucleic acid variants provided herein. Library design may utilize two nucleotide sequences – one that codes for the primary amino acid sequence of the target protein, and one that codes for the protease recognition sequence of the protease that will be used for inactivation of the recombinant protein of the invention. The protease recognition sequence may be “walked” along the former sequence in one of two methods (FIG. 1A and 1B).

[000171] In one method, the protease recognition sequence may be inserted after multiple codons of the nucleotide sequence that encodes the target protein, thereby producing a plurality of nucleic acid variants, wherein each nucleic acid variant contains the protease recognition sequence located at a unique position between two native codons (FIG. 1A). In another method, the protease recognition sequence may replace an equivalent number of nucleotides in the sequence that encodes the target protein, thereby producing a plurality of nucleic acid variants, wherein each nucleic acid variant contains the protease recognition sequence in place of an equivalent stretch of native nucleotides (FIG. 1B).

[000172] In some embodiments, the protease recognition sequence may be inserted after every codon of the nucleic acid sequence that encodes the target protein, thereby producing a plurality of nucleic acid variants, wherein each nucleic acid variant contains the protease recognition sequence located at a unique position between two native codons (*e.g.*, two adjacent native codons). In some embodiments, the protease recognition sequence may be inserted after every codon of the nucleic acid sequence excluding the first and/or last codon. Alternatively, in some embodiments, the protease recognition sequence may be inserted after every other codon, after every third codon, after every fourth codon, after every fifth codon,

after every tenth codon or after every twentieth codon. In some embodiments, the protease recognition sequence may be inserted randomly. In some embodiments, the protease recognition sequence may be inserted in a particular region of the nucleic acid, such as, for example, the N terminal region or the C terminal region. In some embodiments, the protease recognition sequence may replace contiguous codons of the nucleic acid sequence that encodes the target protein. The “N terminal region” of a protein, as used herein, may refer to the stretch of 100, 90, 80, 70, 60, 50, 40, 30, 20 or 10 amino acids located adjacent to the 5' terminal amino acid. The “C terminal region” of a protein, as used herein, may refer to the stretch of 100, 90, 80, 70, 60, 50, 40, 30, 20 or 10 amino acids located adjacent to the 3' terminal amino acid. In each embodiment, it is to be understood that each nucleic acid variant contains at least one (or one) protease recognition sequence.

[000173] In some embodiments, wherein the structure of a target protein is known or can be predicted, the protease recognition sequence may be inserted in regions corresponding to solvent-exposed loop regions of the protein. It has been discovered that, in some instances, these solvent-exposed loop regions are tolerant to insertion of a protease recognition sequence that is readily cleavable. Thus, in some embodiments, provided herein are methods of constructing a library of nucleic acids containing a protease recognition sequence in regions corresponding to solvent-exposed loop regions of the protein. Such methods save time and the cost of preparing proteins with a protease recognition sequence. In some embodiments, it may be advantageous to use polymerase chain reaction (PCR)-based library construction strategies that prohibit synthesis of the wild-type nucleic acid sequences.

[000174] It is to be understood that, in some embodiments, it may be advantageous to incorporate as much of the protease recognition sequence as is necessary to produce a full-length recognition sequence. For example, if the protease recognition sequence begins with a leucine and the sequence is being inserted after a leucine, only a partial recognition sequence may be inserted such that the leucine is not repeated (FIG. 2). Likewise, if the last amino acid in the protease recognition sequence is a proline, and the protease recognition sequence is being inserted before a proline, only a partial protease recognition sequence may be inserted such that the proline is not doubled. Thus, a protein or nucleic acid with a protease recognition sequence located between two native amino acids or codons (*e.g.*, two adjacent native amino acids or codons), respectively, encompasses proteins and nucleic acids with partial protease recognition sequences inserted between two native amino acids or codons such that the full recognition sequence is reconstituted in the final product.

Strain Construction

[000175] The nucleic acid variants of the invention may be transformed into recombinant cells (*e.g.*, bacterial cells) to screen for optimal (*e.g.*, active and inactivatable) recombinant proteins. It is to be understood that the cells used for screening are not necessarily the cells used to express an optimal recombinant protein for the purpose of engineering, for example, a metabolic pathway of interest.

[000176] In some embodiments, the genome of the cells may be modified to (a) delete or mutate the chromosomal wild-type (or endogenous) copy of a gene encoding the target protein and/or (b) include a means of inducing cytoplasmic cognate protease expression. The latter may be accomplished by adding a gene encoding the cognate protease with an inducible promoter to the cell genome, or by providing a gene encoding the cognate protease with an inducible promoter on a vector, such as, for example, a plasmid. Alternatively, in some embodiments, the cells may altogether lack the cognate protease, which can then be added at a later screening/selection step. In some embodiments, the cognate protease is added in purified form.

[000177] In some embodiments, the recombinant cell is modified to lack a functional chromosomal copy of the wild-type gene (*i.e.*, the wild-type gene encoding the wild-type protein that corresponds to the recombinant protein with the protease recognition site) and is transformed with a plasmid containing nucleic acid variants of the invention. Without being bound by theory, deletion of the chromosomal wild-type copy of the gene from the cell permits complementation and aids in minimizing background when the nucleic acid is inserted through low efficiency recombination methods (*e.g.*, when cell growth due to the presence of a compensatory wild-type gene represents a false positive). In some embodiments, inclusion of a selectable marker (*e.g.*, an antibiotic resistance marker) in the cells, for example, in an episomal vector containing a nucleic acid variant, may reduce the rate of false positives.

[000178] In some embodiments, the cells may be modified to delete wild-type genes encoding proteins with functions similar to those of the target protein. For example, in some embodiments, chromosomal copies of genes encoding isozymes of a target enzyme (*i.e.*, enzymes that provide similar function) are deleted from the cells in order to minimize background in the screening/selection step(s).

[000179] In some embodiments, the recombinant cells are modified to contain an inducible recombinase system such as, for example, at least one nucleic acid containing the lambda phage (λ) recombinase system genes *gamma* (γ), *beta* (β), and *exo*. Thus, in some embodiments, recombineering (or recombination-mediated genetic engineering) methods are used to modify the recombinant cells of the invention. Such homologous recombination systems may be used to introduce or delete chromosomal copies of wild-type genes from the cell genome. Other recombineering methods are also contemplated and may be used herein. The invention also contemplates the use of restriction enzymes and ligases to combine nucleic acid sequences in a specified order (Strachan, T., *et al.*, Human Molecular Genetics, Chapter 4, Garland Science, New York, 1999).

Selection for Protein Activity

[000180] The recombinant cells expressing the nucleic acid variants of the invention may be grown in selective media in the absence of a functional protease to permit recovery of recombinant proteins encoded by the nucleic acid variants. For example, in some embodiments, the activity of the target protein may be required for cell growth. If insertion of a cognate protease recognition sequence adversely affects the activity of the recombinant protein, then presumably, the cells will display growth defects such as, for example, a reduced growth rate. Accordingly, at this screening/selection step, only those cells with a normal growth rate (or without growth defects) are selected for further characterization. A “normal growth rate,” as used herein, refers to a growth rate that is comparable to control wild-type cells. In some embodiments, a cell is considered to have a “normal growth rate” if its growth rate is within about 15% of the growth rate of a wild-type control cell (*e.g.*, cell without a nucleic acid variant/recombinant protein of the invention). For example, a cell may be considered to have a normal growth rate if its growth rate is within 50%, 40%, 30%, 25%, 20%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2% or 1% of the growth rate of a wild-type control cell. A cell with a “growth defect,” as used herein, may refer to a cell that fails to grow or that has a reduced growth rate of greater than 10%, greater than 15%, greater than 20%, or greater than 25% in comparison to the growth rate of a wild-type control cell.

[000181] The selective growth media used in accordance with the invention, in some embodiments, may depend on particular characteristics of the recombinant proteins, such as the specific function of the active recombinant protein. For example, when recombinant Pgi

protein is grown on minimal medium where the only available carbon source is glucose, cells containing an active recombinant Pgi variant grow well, whereas cells containing an inactive recombinant Pgi variant grow poorly. In some embodiments, the selective media used may depend on the substrate of the recombinant Pgi variant. In some embodiments, a “rescue” approach is used to produce selective growth conditions, where the activity of the recombinant protein, which is required for cell growth, is deleted from the genome of the cell (*e.g.*, gene(s) encoding the wild-type protein(s) is/are deleted or mutated), and then nucleic acid *pgi* variants are introduced into the cell. Those nucleic acid variants that are active should rescue cell growth (*e.g.*, cells grows), and those that are inactive should not rescue cell growth (*e.g.*, cells do not grow).

Selection for Protein Inactivation

[000182] Cells that do not display growth defects are then grown under selective conditions that induce cognate protease expression. This step permits recovery of cells that display growth defects.

[000183] The cells that display growth defects presumably carry recombinant proteins that are inactivated in the presence of a functional cognate protease (and are active in the absence of a functional cognate protease). These growth defective cells are then recovered, and the nucleic acid variants contained therein are sequenced for further characterization.

[000184] Further characterization may involve the expression of selected nucleic acid variants in cells that lack functional cognate protease. Growth of these cells may then be characterized and lysates produced and collected. The lysates may then be tested *in vitro* for loss of recombinant protein activity. Such testing may utilize protein activity assays upon incubation with or without exogenous, purified cognate protease. Various protein activity assays are known in the art, any of which may be used in accordance with the invention. The protein activity assay selected will depend on the type of protein. In some embodiments, the recombinant protein that inactivates the most completely and rapidly when exposed to protease may be selected for further use in, for example, engineering metabolic pathways of interest.

[000185] Cognate protease induction conditions will depend on the type of inducible promoter system chosen to drive the expression of the cognate protease and are known in the art. For example, isopropyl β -D-1-thiogalactopyranoside (IPTG) may be added to an *in vitro*

cell-free system to activate an IPTG-responsive promoter operably linked to the cognate protease.

[000186] These and other aspects are illustrated by the following non-limiting examples.

[000187] Having thus described several aspects of at least one embodiment of this invention, it is to be appreciated various alterations, modifications, and improvements will readily occur to those skilled in the art. Such alterations, modifications, and improvements are intended to be part of this disclosure, and are intended to be within the spirit and scope of the invention. Accordingly, the foregoing description and drawings are by way of example only.

EXAMPLES

Example 1 – Phosphoglucose isomerase enzymes of *Escherichia coli*

[000188] The phosphoglucose isomerase (Pgi) enzyme of *Escherichia coli* (*E. coli*) catalyzes the inter-conversion of glucose-6-phosphate and fructose-6-phosphate and is the first committed step in glycolysis. Protease targeting of this enzyme, without altering the function/activity of this key enzyme during cell growth, enables control of carbon flux between glycolysis and the pentose phosphate pathway in cell-free reactions.

Pgi variant library construction

[000189] A 562-member linear double-stranded DNA library was designed and constructed by polymerase chain reaction (PCR), where the native *E. coli pgi* gene sequence (SEQ ID NO:1) was modified to include a nucleotide sequence (SEQ ID NO:37) encoding the eight amino acid protease recognition sequence (SEQ ID NO:38) of the human rhinovirus 3C (HRV) protease (FIGs. 1A and 1B). 547 members of the library contained mutant *pgi* genes with nucleotides encoding the protease recognition sequence inserted after each of the 549 codons in the wild-type *pgi* gene (excluding the first and last codons). Additional library members were created by replacing wild-type *pgi* gene sequence with a nucleotide sequence encoding the eight amino acids of the protease recognition sequence. These members contained replacements in the wild-type gene at 15 different locations starting with codon numbers 244, 245, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 461, 462 (where the codon numbers correspond to the first codon of the replacement sequence). In addition to the gene's coding sequence, each library member also contained 50 bp homology arms (*e.g.*,

50 bp additional sequence upstream of the gene's start codon and 50 bp downstream of the gene's stop codon that are homologous to the wild-type *pgi* locus of the *E. coli* genome). If the LEVLFQGP (SEQ ID NO:38) sequence was to be inserted after amino acids L, LE or LEV, only EVLFQGP (SEQ ID NO:39), VLFQGP (SEQ ID NO:40) or LFQGP (SEQ ID NO:41) were inserted, respectively. Similarly, if the sequence was to be inserted before amino acids P, GP or QGP, only LEVLFQG (SEQ ID NO:42), LEVLFQ (SEQ ID NO:43) or LEVLF (SEQ ID NO:44) were inserted, respectively. In addition, if the insertion (or replacement) was between amino acids LP, for example, only EVLFQG (SEQ ID NO:45) was inserted (or replaced).

Strain Design

[000190] *E. coli* JW3985-1 (Coli Genetic Stock Center; CGSC #10867) from the Keio collection (*Mol. Syst. Biol.* 2006;2:2006-08) was selected as the strain for the Pgi library screen. This strain contains a kanamycin-resistance marker (KanR) in place of the *pgi* gene. To prepare the strain for use in the screen, several modifications were made. First, KanR was removed using pCP20, which was obtained from *E. coli* BT340 (CGSC # 7629), by employing the method described by Datsenko & Wanner (*Proc Natl Acad Sci USA*. 2000 Jun 6;97(12):6640-45, incorporated by reference herein). The resulting *pgi* locus of the strain contained the first three bases and the last 21 bases of the *pgi* gene, with a short scar sequence in between. Two plasmids (pGLA042 and pGLC217; FIGs. 3 and 4) were co-transformed into this strain to create the final strain that was used in the Pgi screen (GL12-085) through the pooled high-throughput selection approach (described below). pGLA042 was created from pKD46 (obtained from *E. coli* BW25113; GCSC # 7739) by replacing pKD46's arabinose-inducible expression system with the temperature-inducible λ cI857-pR expression system. This change enables the temperature-inducible expression of the phage λ recombinase system genes γ , β , and *exo* from pGLA042. pGLC217 is a low-copy plasmid that provides for arabinose-inducible expression of the HRV 3C protease (codon-optimized for expression in *E. coli*; SEQ ID NO:34), the translation of which is facilitate by a strong ribosome binding site. A strain lacking pGLC217 was also produced (GL12-052) for use in the individual selection and assay approach (see below).

Individual Selection and Assay Approach

[000191] The chromosomal locus of *pgi* in GL12-052 was recombined with a 76-member subset of the linear, double-stranded DNA library described above. This subset contained the protease recognition sequence in solvent-accessible loop regions of Pgi, as predicted by its crystal structure (Protein Data Bank ID: 3NBU). Resulting Pgi library members carried protease recognition sequence insertions after the following positions in the wild-type Pgi primary amino acid sequence: 2-5, 9, 24-25, 33- 36, 58-59, 85- 96, 105-111, 113-115, 137-141, 143-144, 146, 173-176, 196, 250-251, 254, 366-370, 398-399, 410-414, 447-451, 477, 526-532.

[000192] GL12-052 was grown at 30 °C in low-salt-LB (lysogeny broth) (0.5X NaCl) to an optical density (OD) of 0.5. The culture was transferred to a 42 °C water bath and shaken for 15 minutes to induce the recombinase system from pGLA042. Induced cells were made electrocompetent following standard methods and transformed with the library of nucleic acid variants. Each library member (or nucleic acid variant) was transformed individually (25 μ L cells and 250 ng library member), or in three-member subsets, and recovered in 1 mL low-salt-LB for greater than 1 h at 30 °C. Recovered transformations were plated on M9-agar medium supplemented with 1% glucose (M9G). Plates were incubated at 30 °C for 1.5-2 days. The resulting colonies represented library members who contained active Pgi molecules despite the inclusion of the protease recognition sequence. The Pgi region of these library members' genomes were PCR-amplified and sequenced. Sequence-verified strains were then grown in small-scale shake flask cultures with M9G medium at 37 °C in order to determine their growth rates, thereby providing an *in vivo* method for assessing the impact of protease recognition sequence insertion on Pgi activity. Those strains with growth rates within 15% of the wild-type growth rate were advanced for a second round of screening to determine susceptibility to protease inactivation; 41 of the original 76-member subset were advanced.

[000193] To assess protease susceptibility, clarified lysates were created and assayed *in vitro* for Pgi activity in the presence or absence of exogenous HRV 3C protease. M9G cultures were grown to an OD of 2, pelleted (8000xg, 8 min., 4 °C), washed (10mL 1X PBS, 4 °C), resuspended (12mL 100 mM Tris-HCl, pH 7.5, 4 °C), lysed (AVESTIN[®] Emulsiflex C3 homogenizer at 15,000 psi), and clarified (22,000xg, 15 min, 4 °C). Clarified lysates (100 μ L) were treated with $\pm$ 10 units of exogenous HRV 3C protease (ACCELAGEN[™] H0101S) for 4 hours at 37 °C and assayed for Pgi activity. Pgi activity was assayed by coupling to

glucose-6-phosphate dehydrogenase (G6PDH) and following a reduction of nicotinamide adenine dinucleotide phosphate (NADP⁺) at an absorbance of 340 nm for five minutes at 37 °C. Reactions contained 100 mM Tris-HCl (pH 7.5), 8 mM MgSO₄, 5 mM fructose-6-phosphate, 1 mM NADP⁺, 0.25 mg/mL bovine serum albumin, 2.5 units of purified G6PDH from *Leuconostoc mesenteroides* (MEGAZYME[®] E-GPDH5), and 30 volume-percent lysate/protease samples.

[000194] Ultimately, 12 unique *pgi* gene sequences were selected, which contained the HRV 3C recognition sequence inserted after the following codons: 108, 109, 110, 138, 410, 526, 527, 528, 529, 530, 531, and 532. These 12 members of the original 76-member subset showed growth rates in M9G medium within 15% of wild-type and were significantly inactivated upon exposure to exogenous protease (Table 1). FIG. 5 shows a longer time-course treatment with protease for a subset of the library. Based on data obtained from the experiments described above, the optimal Pgi variant contained the HRV 3C recognition sequence after amino acid 526 (aka, Pgi-HRV-I526 from gene *pgi*-HRV-I526).

Table 1: Comparison of cell growth rates and Pgi activity.

Protease Recognition Sequence Inserted after Amino Acid #	$\mu(\text{h}^{-1})$	μ / wild-type	Pgi Activity ($\mu\text{mol}/\text{min}/\text{mg}$ total protein)		Activity / Activity ^{wild type}	
			- protease	+ protease	- protease	+ protease
Δpgi	0.15	0.22	0	0		
wild-type	0.68	1	2.31	2.37	1	1
108	0.60	0.88	2.16	2.03	0.94	0.86
109	0.71	1.04	2.54	1.86	1.10	0.78
110	0.70	1.03	2.13	1.99	0.92	0.84
138	0.58	0.86	0.62	0.37	0.27	0.16
410	0.59	0.87	1.16	0.96	0.50	0.41
526	0.66	0.97	1.92	0.44	0.83	0.19
527	0.70	1.03	1.93	1.42	0.84	0.60
528	0.65	0.96	1.04	0.63	0.45	0.26
529	0.66	0.97	1.48	1.19	0.64	0.50
530	0.68	1.00	0.77	0.29	0.33	0.12
531	0.62	0.91	1.29	0.67	0.56	0.28
532	0.69	1.01	1.15	0.52	0.50	0.22

Note: Small Δpgi control activity assay value (0.089) subtracted from all reported Pgi activities

Pooled High-Throughput Selection Approach

[000195] The chromosomal locus of *pgi* in GL12-085 was recombined with the 562-member pooled library (at equimolar concentrations), using the methods described above. The resulting cell library was plated onto M9-agar medium lacking arabinose and supplemented with 34 $\mu\text{g}/\text{mL}$ of chloramphenicol and 1% glucose (M9CG). Cells were plated to obtain 5X coverage of the library, which amounted to 11 plates with approximately 250 cells/plate, thereby permitting easier replica-plating in the next step of the method.

These pooled-library plates were incubated at 37 °C for 1.5-2 days. Colonies representing library members that provide active Pgi were subsequently replica-plated onto both M9GC media as well as M9-agar medium supplemented with 34 µg/mL chloramphenicol, 1% glucose and 2% arabinose (M9CGA). Replica plates were incubated at 37 °C for 1.5-2 days. All colonies that were present on M9CG plates but not present on M9CGA plates were recovered for further analysis.

[000196] The Pgi region of these colonies' genomes were PCR-amplified and sequenced. Seven unique *pgi* gene sequences were identified, which contained the HRV 3C recognition sequence inserted after the following codons: 524, 525, 526, 528, 529, 531 and 545. Members 526, 528, 529, and 531 were also identified from the 76-member-subset using the individual screening assay, described above.

Protein crystal structure

[000197] Nearly all 15 Pgi variants obtained from the individual and pooled screens map onto solvent-exposed loop regions of the published crystal structure for Pgi (Protein Data Bank ID: 3NBU). Further, variants 526-532 correspond to a loop region that precedes the C-terminal helix of Pgi, and variants 524-525 correspond to the C-terminal end of another helix, which possesses catalytic activity (FIG. 6). Protease-mediated cleavage of a recognition sequence inserted in the 524-532 region would thus cleave off the C-terminal helix and may result in the malformation of the preceding catalytic helix. Removal of the C-terminal helix is likely detrimental because Pgi is a dimer, and this helix helps to "latch" the dimer together.

Demonstration of effective protease-mediated inactivation of targeted Pgi

[000198] The *pgi* gene of *E. coli* BL21(DE3) was replaced with *pgi-HRV-I526* (SEQ ID NO:9). This strain (GL12-116) was transformed individually with three plasmids: pACYCDuet-1, pGLC089 (FIG. 7), and pGLC221 (FIG. 8). pACYCDuet-1 is a low-copy empty vector control plasmid, whereas pGLC089 & pGLC221 are capable of expressing HRV 3C protease (codon-optimized for expression in *E. coli*) from a T7 promoter when induced with isopropyl β-D-1 thiogalactopyranoside (IPTG). The protease gene of pGLC089 contains additional sequence such that the resulting HRV 3C protease possesses an N-terminal OmpA leader (MKKTAIAIAVALAGFATVAQA) (SEQ ID NO:46) that sequesters the protease in the periplasm, whereas the protease gene of pGLC221 lacks such a leader and

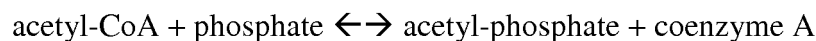
expresses in the cytoplasm. These strains were grown in a defined glucose medium at 37 °C to mid-log phase and induced with 0.8 mM IPTG for 2 hours. Clarified lysates were created and subsequently assayed for Pgi activity (as described above). Table 2 shows the growth rate of each strain prior to induction and the Pgi activity measured in the clarified lysate. When HRV 3C protease is expressed cytoplasmically, the growth rate falls by 40% as compared to the strain lacking protease, presumably due to leaky expression of the protease prior to IPTG-induced over-expression.

Table 2: Comparison

OmpA Leader Sequence for Periplasmic Sequestration	Pre-Induction $\mu(\text{h}^{-1})$	Pgi Activity ($\mu\text{mol}/\text{min}/\text{mg}$ total protein)
no protease	0.62	1.3
No	0.38	-0.01
Yes	0.55	0.01

Example 2 – Phosphotransacetylase enzymes of *Escherichia coli*

[000199] The phosphotransacetylase (Pta) enzyme (also referred to as phosphate acetyltransferase) of *Escherichia coli* (*E. coli*) catalyzes the first committed reaction of acetate overflow metabolism:



Acetate overflow occurs in rapidly growing glucose-fed aerobic cultures of *E. coli*.

Accumulation of acetate excreted into production media limits growth rate, growth density, and recombinant protein production, which is a typical problem in industrial fermentation.

Strains deleted for Pta activity routinely grow 15-20% slower than their wild-type counterparts, and while they diminish acetate excretion, they do not solve the overflow problem, as strains instead excrete lactate and pyruvate with similar deleterious effects.

Protease targeting of Pta in cell-free production processes prevents acetate accumulation and shifts carbon flux into the tricarboxylic acid cycle, while allowing strains to grow at faster maximum growth rates than their *pta*-deleted counterparts.

Pta Variant Library Construction

[000200] A 200-member linear double-stranded DNA library was designed and constructed by polymerase chain reaction (PCR), where the native *E. coli pta* gene sequence (SEQ ID NO: 47) was modified to include a nucleotide sequence (SEQ ID NO:37) encoding

the eight amino acid protease recognition sequence (SEQ ID NO:38) of the human rhinovirus 3C (HRV) protease. The nucleotide sequence encoding the protease recognition motif was inserted after the following codons in the wild-type *pta* gene: 350, 380-388, 401-405, 409-415, 426-431, 434-438, 446-465, 475-483, 490-495, 502-508, 511-518, 526-538, 544-549, 552-563, 577-586, 589-603, 615-620, 626-627, 629-632, 639-650, 653-660, 669-674, 681-687, 689-698, 709-713. In addition to the gene's coding sequence, each library member also contained 50 base pair (bp) homology arms (*e.g.*, 50 bp additional sequence upstream of the gene's start codon and 50 bp downstream of the gene's stop codon that are homologous to the wild-type *pta* locus of the *E. coli* genome). If the LEVLFQGP (SEQ ID NO:38) sequence was to be inserted after amino acids L, LE or LEV, only EVLFQGP (SEQ ID NO:39), VLFQGP (SEQ ID NO:40) or LFQGP (SEQ ID NO:41) were inserted, respectively. Similarly, if the sequence was to be inserted before amino acids P, GP or QGP, only LEVLFQG (SEQ ID NO:42), LEVLFQ (SEQ ID NO:43) or LEVLF (SEQ ID NO:44) were inserted, respectively. In addition, if the insertion was between amino acids LP, for example, only EVLFQG (SEQ ID NO:45) was inserted.

Strain Design

[000201] For the Pta library screen, a modified version of *E. coli* JW2294-1 (Coli Genetic Stock Center; CGSC #9844) from the Keio collection (*Mol. Syst. Biol.* 2006; 2:2006-08) was created. To prepare the strain for use in the screen, the genome of JW2294-1 was modified in two ways, using methods described previously in Example 1 for phosphoglucose isomerase. First, the KanR marker, located in place of the *pta* gene, was removed, leaving a *pta* locus that contained the first three bases and the last 21 bases of the *pta* gene, with a short scar sequence in between. Second, the gene encoding the acetyl-CoA synthetase (*i.e.*, *acs*) was replaced with KanR, thereby creating a strain that lacks the ability to grow on acetate as a sole carbon source and restores resistance to kanamycin. The recombinase plasmid described previously (pGLA042; Figure 3) was transformed into this strain to create the final screen strain (GL13-052), which was used in an individual selection and assay approach.

Individual Selection and Assay Approach

[000202] The chromosomal locus of *pta* in GL13-052 was recombined individually with the 200-member linear, double-stranded DNA library described above, which places the HRV protease recognition sequence in predicted solvent-accessible loop regions of the C-

terminal catalytic domain of Pta. As the crystal structure of the *E. coli* Pta has not yet been determined, the aforementioned loop regions were predicted by performing an amino acid sequence alignment of the C-terminal catalytic domain of the *E. coli* Pta to that of heterologous enzymes with published crystal structures (*i.e.*, Protein Data Bank IDs: 1R5J & 2AF3).

[000203] A 2 liter (L) culture of GL13-052 was grown to an optical density at 600 nm of 0.5 at 30 °C in low-salt LB (5 g/L yeast extract, 10 g/L tryptone, 5 g/L NaCl) containing 10 µg/mL kanamycin and 50 µg/mL carbenicillin. The culture was transferred to a 42 °C water bath and shaken for 30 minutes to induce the recombinase system from pGLA042. Induced cells were made electrocompetent following standard methods, and 50 µL aliquots were flash frozen in liquid nitrogen and subsequently stored at -80 °C prior to recombination with library members.

[000204] Each library member DNA was transformed individually (25 µL cells and 50 ng library member DNA) and recovered in 1 mL low-salt LB for greater than 3 hours at 30 °C. Recovered transformations were plated on M9-agar medium supplemented with 1% acetate and 10 µg/mL kanamycin, and plates were incubated at 37 °C for 2-3 days. Recombination events that yielded colonies were representative of library members that contained active Pta enzymes despite the inclusion of the protease recognition sequence motif. The Pta region of these library members' genomes were PCR-amplified and sequenced. Sequence-verified strains were then made competent and transformed with pGLC217 to enable the arabinose-inducible expression of HRV protease. Strains were also transformed with a control plasmid (pGLC219), which is identical to pGLC217 except that the gene encoding the HRV protease was replaced with a gene encoding the tobacco etch virus (TEV) protease. This protease will not cleave Pta variants.

[000205] To assess protease susceptibility *in vivo*, growth rates and acetate excretion were measured for strains containing viable Pta variants bearing either pGLC217 or pGLC219 that were grown at 37 °C in M9-minimal medium containing 0.5% glucose, 10 µg/mL kanamycin, and 34 µg/mL chloramphenicol. These cultures were grown $\pm$ 2% arabinose to overexpress the protease. Thus, the growth rate of each viable Pta variant was examined under four conditions: (1) TEV protease expression not induced, (2) arabinose-induced TEV protease expression, (3) HRV protease expression not induced, and (4) arabinose-induced HRV protease expression. Conditions (3) and (4) enable HRV protease-mediated Pta deactivation to be assessed via growth rate, while conditions (1) and (2) provide

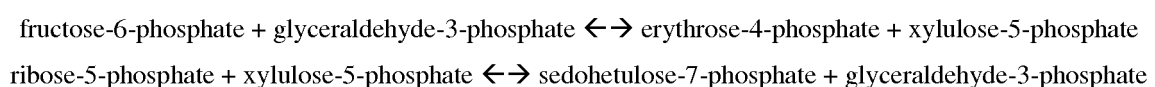
a control for the metabolic burden brought about by plasmid maintenance and arabinose-induced protein expression. As a positive control, a strain with wild-type *pta* was included in the study, whereas a strain deleted for *pta* served as the negative control.

[000206] Based on growth and acetate excretion phenotypes in comparison to the control strains, four unique Pta variants were ultimately selected from the 200-member library. These variants contained the HRV protease cleavage motif inserted after amino acid 381, 382, 387, and 409 (SEQ ID NO: 50, 52, 54, 56, encoded respectively by SEQ ID NO: 49, 51, 53, 55). Figure 10A shows the acetate excretion data, and Figure 10B shows the lactate excretion data for the selected variants in comparison to the wild-type Pta and *pta*-controls. The wild-type Pta control exhibits a phenotype of high acetate and low lactate, whereas the *pta*- control exhibits the opposite phenotype. Thus, if a particular Pta variant is susceptible to HRV protease-mediated deactivation, it should show high acetate and low lactate in the absence of HRV protease induction, while showing the opposite when the HRV protease is induced. In the TEV protease controls, acetate should be high and lactate should be low regardless of induction, as the TEV protease recognition sequence is not present in the Pta variants. Clearly, variants 381, 382, 387, and 409 meet these criteria.

[000207] To further demonstrate that these Pta variants were susceptible to HRV protease, *in vitro* activity assays were performed. Figure 11 shows the activity of each Pta variant in the presence or absence of exogenously supplied HRV protease after incubation for 30 minutes at 37 °C. In all cases, Pta activity was dramatically reduced upon incubation with HRV protease. Variant 381 (SEQ ID NO: 50) exhibited the best balance between having both near-wild-type activity and being particularly susceptible to protease-mediated deactivation.

Example 3— Transketolase enzymes of *Escherichia coli*

[000208] The major and minor transketolase isozymes (TktA and TktB, respectively) of *Escherichia coli* (*E. coli*) catalyze two reversible ketol transfer reactions in the pentose phosphate pathway:



Transketolase activity ensures an adequate supply of erythrose-4-phosphate, a key central carbon metabolite necessary for the production of the three aromatic amino acids as well as

several vitamins. Strains lacking transketolase activity require supplementation with erythrose-4-phosphate-derived compounds, such as aromatic amino acids and vitamins. Even with such supplementation, which can be costly, growth to high cell densities in fermentations is challenging. Transketolase also connects the pentose phosphate pathway to glycolysis, siphoning pentose phosphates away from excessive nucleotide production when the flux of glucose is high. Protease targeting of transketolase, without significantly altering its critical function during cell growth, benefits the production of pentose-derived molecules in cell-free reactions, as high pentose phosphate pathway flux is prevented from siphoning away to glycolysis. The crystal structure of transketolase is shown in Figure 12.

Transketolase A Variant Library Construction

[000209] A 200-member linear double-stranded DNA library was designed and constructed by polymerase chain reaction (PCR), where the native *E. coli transketolase A* (*tktA*) gene sequence (SEQ ID NO:57) was modified to include a nucleotide sequence (SEQ ID NO:37) encoding the eight amino acid protease recognition sequence (SEQ ID NO:38) of the HRV protease. The nucleotide sequence encoding the protease recognition motif was inserted after the following codons in the wild-type *tktA* gene: 22-24, 43-55, 78-83, 88-110, 138-148, 172-175, 185-192, 196-200, 208-210, 218-220, 233-238, 245-257, 261-287, 294-296, 331-336, 350-354, 371-372, 388-403, 484-487, 508-511, 523-529, 544-551, 573-577, 591-593, 601-607, 624-627, 633-640, 648-651. In addition to the gene's coding sequence, each library member also contained 50 base pair (bp) homology arms (*e.g.*, 50 bp of additional sequence upstream of the gene's start codon and 50 bp of additional sequence downstream of the gene's stop codon, each 50 bp sequence homologous to the wild-type *tktA* locus of the *E. coli* genome). For insertion of LEVLFQGP (SEQ ID NO:38) sequence after amino acids L, LE or LEV, only EVLFQGP (SEQ ID NO:39), VLFQGP (SEQ ID NO:40) or LFQGP (SEQ ID NO:41) were inserted, respectively. Similarly, for insertion of LEVLFQGP (SEQ ID NO:38) before amino acids P, GP or QGP, only LEVLFQG (SEQ ID NO:42), LEVLFQ (SEQ ID NO:43) or LEVLF (SEQ ID NO:44) were inserted, respectively. In addition, for insertion of LEVLFQGP (SEQ ID NO:38) between amino acids LP, for example, only EVLFQG (SEQ ID NO:45) was inserted.

Strain Design

[000210] For the TktA library screen, a modified version of *E. coli* JW5478-1 (Coli Genetic Stock Center; CGSC #11606) from the Keio collection (*Mol. Syst. Biol.* 2006; 2:2006-08) was created. To prepare the strain for use in the screen, the genome of JW5478-1 was modified in two ways using methods described in Example 1. First, the KanR marker, located in place of the *tktA* gene, was removed, leaving a *tktA* locus that contained the first 3 bases and the last 21 bases of the *tktA* gene, with a short scar sequence in between. Second, the gene encoding the minor transketolase (*tktB*) was replaced with KanR, thereby creating a strain that lacks any transketolase activity and restores resistance to kanamycin. The recombinase plasmid, pGLA042, (FIG. 3) was transformed into this strain to create the final screen strain, GL13-050, which was used in an individual selection and assay approach.

Individual Selection and Assay Approach

[000211] The chromosomal locus of *tktA* in GL13-050 was recombined individually with the 200-member linear, double-stranded DNA library described above, which places the HRV protease recognition sequence in solvent-accessible loop regions of TktA, as predicted by its crystal structure (Protein Data Bank ID: 1QGD) (*see* FIG. 11).

[000212] A 2 liter (L) culture of GL13-050 was grown to an optical density at 600 nm of 0.5 at 30 °C in a modified Vogel Bonner E (mVBE) medium (1X VBE minimal salts, 0.4% glucose, 2% casamino acids, 1 mM tryptophan, 0.25 mM 2,3-dihydroxybenzoate, 30 µM *p*-aminobenzoate, 30 µM *p*-hydroxybenzoate, 5 µM pyridoxine-HCl, 10 µg/mL kanamycin). The culture was transferred to a 42 °C water bath and shaken for 30 minutes to induce the recombinase system from pGLA042. Induced cells were made electrocompetent following standard methods, and 50 µL aliquots were flash frozen in liquid nitrogen and subsequently stored at -80 °C prior to recombination with library members.

[000213] Each library member DNA was transformed individually (25 µL cells and 50 ng library member DNA) and recovered in 1 mL low-salt-LB for greater than 3 hours at 30 °C. Recovered transformants were plated on M9-agar medium supplemented with 1% glucose and 10 µg/mL kanamycin, and plates were incubated at 37 °C for 2-3 days. Transformations/recombinations that yielded colonies were representative of library members that contained active TktA molecules despite the inclusion of the protease recognition sequence motif. The TktA region of the genomes of these library members were PCR-amplified and sequenced. Sequence-verified strains were then made competent and

transformed with pGLC217 to enable the arabinose-inducible expression of HRV protease. Strains were also transformed with a control plasmid, pGLC219, which is nearly identical to pGLC217, with the difference that the gene encoding the HRV protease was replaced with a gene encoding the tobacco etch virus (TEV) protease. This protease will not cleave TktA variants.

[000214] To assess protease susceptibility *in vivo*, growth rates were measured for strains containing viable TktA variants bearing either pGLC217 or pGLC219 that were grown at 37 °C in M9-minimal medium containing 0.5% glucose, 10 µg/mL kanamycin, and 34 µg/mL chloramphenicol. These cultures were grown $\pm$ 2% arabinose to overexpress the protease. Thus, the growth rate of each viable TktA variant was examined under four conditions: (1) TEV protease expression not induced, (2) arabinose-induced TEV protease expression, (3) HRV protease expression not induced, and (4) arabinose-induced HRV protease expression. Conditions (3) and (4) enabled HRV protease-mediated TktA deactivation to be assessed via growth rate, while conditions (1) and (2) provided a control for the metabolic burden brought about by plasmid maintenance and arabinose-induced protein expression.

[000215] Five unique TktA variants were selected from the 200-member library. The selected variants contained the HRV protease cleavage motif inserted after amino acid 635, 636, 637, 638, and 640 (SEQ ID NO:64-SEQ ID NO:68, encoded respectively by SEQ ID NO:58-SEQ ID NO:62). As shown in Table 3, the selected variants were able to support growth in minimal glucose medium when HRV protease expression was not induced but were significantly impaired for growth when HRV protease was induced. As the TEV protease controls show, the growth impairment brought about via protease induction was not simply the result of the burdensome nature of protein over-expression. Thus, these TktA variants were susceptible to HRV protease-mediated deactivation *in vivo*.

Table 3. Growth rates of protease-targetable TktA variants in minimal glucose medium

Protease Recognition Sequence Inserted after Amino Acid #	TEV protease		HRV protease	
	- inducer	+ inducer	- inducer	+ inducer
wild-type	0.55	0.53	0.60	0.57
635	0.52	0.52	0.58	0.47
636	0.39	0.42	0.43	0.17
637	0.46	0.46	0.40	0.11
638	0.18	0.18	0.13	0.08
640	0.48	0.52	0.55	0.34

Protein Crystal Structure

[000216] According to the published crystal structure for TktA (Protein Data Bank ID: 1QGD), the five TktA variants obtained from the screen (Table 3) map to the loop that immediately precedes the C-terminal helix (FIG. 11). TktA is active as a dimer, and this loop occurs at a dimerization interface. Without being bound by theory, cleavage of this loop by HRV protease likely disrupts the ability of the TktA to dimerize. Indeed, dimerization of transketolase is rate-limiting in formation of the active enzyme (*J. Biol. Chem.* 1981; 256:4877-83).

OTHER EMBODIMENTS

[000217] All of the features disclosed in this specification may be combined in any combination. Each feature disclosed in this specification may be replaced by an alternative feature serving the same, equivalent, or similar purpose. Thus, unless expressly stated otherwise, each feature disclosed is only an example of a generic series of equivalent or similar features.

[000218] From the above description, one skilled in the art can easily ascertain the essential characteristics of the present invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions. Thus, other embodiments are also within the claims.

EQUIVALENTS

[000219] While several inventive embodiments have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the function and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the inventive embodiments described herein. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the inventive teachings is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific inventive embodiments described herein. It is, therefore, to be

understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, inventive embodiments may be practiced otherwise than as specifically described and claimed. Inventive embodiments of the present disclosure are directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles, materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the inventive scope of the present disclosure.

[000220] All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

[000221] The indefinite articles “a” and “an,” as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean “at least one.”

[000222] The phrase “and/or,” as used herein in the specification and in the claims, should be understood to mean “either or both” of the elements so conjoined, *i.e.*, elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with “and/or” should be construed in the same fashion, *i.e.*, “one or more” of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the “and/or” clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to “A and/or B”, when used in conjunction with open-ended language such as “comprising” can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); *etc.*

[000223] As used herein in the specification and in the claims, “or” should be understood to have the same meaning as “and/or” as defined above. For example, when separating items in a list, “or” or “and/or” shall be interpreted as being inclusive, *i.e.*, the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as “only one of” or “exactly one of,” or, when used in the claims, “consisting of,” will refer to the inclusion of exactly one element of a number or list of elements. In general, the term “or” as used herein shall only be interpreted as indicating exclusive alternatives (*i.e.*, “one or the

other but not both”) when preceded by terms of exclusivity, such as “either,” “one of,” “only one of,” or “exactly one of.” “Consisting essentially of,” when used in the claims, shall have its ordinary meaning as used in the field of patent law.

[000224] As used herein in the specification and in the claims, the phrase “at least one,” in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase “at least one” refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, “at least one of A and B” (or, equivalently, “at least one of A or B,” or, equivalently “at least one of A and/or B”) can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); *etc.*

[000225] It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

[000226] All references, patents and patent applications disclosed herein are incorporated by reference with respect to the subject matter for which each is cited, which in some cases may encompass the entirety of the document.

[000227] In the claims, as well as in the specification above, all transitional phrases such as “comprising,” “including,” “carrying,” “having,” “containing,” “involving,” “holding,” “composed of,” and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases “consisting of” and “consisting essentially of” shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

SEQUENCE LISTING

SEQ ID NO:1

ATGAAAAACATCAATCCAACGCAGACCGCTGCCTGGCAGGCACTACAGAAACACTTCGATGA
AATGAAAGACGTTACGATCGCCGATCTTTTTGCTAAAGACGGCGATCGTTTTTCTAAGTTCT
CCGCAACCTTCGACGATCAGATGCTGGTGGATTACTCCAAAAACCGCATCACTGAAGAGACG
CTGGCGAAATTACAGGATCTGGCGAAAGAGTGCGATCTGGCGGGCGCGATTAAAGTCGATGTT
CTCTGGCGAGAAGATCAACCGCACTGAAAACCGCGCCGTGCTGCACGTAGCGCTGCGTAACC
GTAGCAATACCCCGATTTTGGTTGATGGCAAAGACGTAATGCCGGAAGTCAACGCGGTGCTG
GAGAAGATGAAAACCTTCTCAGAAGCGATTATTTCCGGTGAGTGGAAGGTTATACCGGCAA
AGCAATCACTGACGTAGTGAACATCGGGATCGGCGGTTCTGACCTCGGCCCATACATGGTGA
CCGAAGCTCTGCGTCCGTACAAAAACCACCTGAACATGCACTTTGTTTTCTAACGTCGATGGG
ACTCACATCGCGGAAGTGCTGAAAAAAGTAAACCCGGAACACGCTGTTCTTGGTAGCATC
TAAAACCTTCACCACTCAGGAACTATGACCAACGCCCATAGCGCGCGTGACTGGTTCCTGA
AAGCGGCAGGTGATGAAAAACACGTTGCAAAACACTTTGCGGCGCTTTCCACCAATGCCAAA
GCCGTTGGCGAGTTTGGTATTGATACTGCCAACATGTTTCGAGTTCTGGGACTGGGTTGGCGG
CCGTTACTCTTTGTGGTCAGCGATTGGCCTGTCGATTGTTCTCTCCATCGGCTTTGATAACT
TCGTTGAACTGCTTTCCGGCGCACACGCGATGGACAAGCATTTCTCCACCACGCCTGCCGAG
AAAAACCTGCCTGTACTGCTGGCGCTGATTGGCATCTGGTACAACAATTTCTTTGGTGCGGA
AACTGAAGCGATTCTGCCGTATGACCAGTATATGCACCGTTTCGCGGCGTACTTCCAGCAGG
GCAATATGGAGTCCAACGGTAAGTATGTTGACCGTAACGGTAACGTTGTGGATTACCAGACT
GGCCCGATTATCTGGGGTGAACCAGGCACTAACGGTCAGCACGCGTTCTACCAGCTGATCCA
CCAGGGAACCAAAATGGTACCGTGCGATTTTCATCGCTCCGGCTATCACCCATAACCCGCTCT
CTGATCATCACCCAGAACTGCTGTCTAACTTCTTCGCCCAGACCGAAGCGCTGGCGTTTGGT
AAATCCCGCGAAGTGGTTGAGCAGGAATATCGTGATCAGGGTAAAGATCCGGCAACGCTTGA
CTACGTGGTGCCGTTCAAAGTATTCGAAGGTAACCGCCGACCAACTCCATCCTGCTGCGTG
AAATCACTCCGTTACGCCTGGGTGCGTTGATTGCGCTGTATGAGCACAAAATCTTTACTCAG
GGCGTGATCCTGAACATCTTCACCTTCGACCAGTGGGGCGTGGAAGTGGGTAAACAGCTGGC
GAACCGTATTCTGCCAGAGCTGAAAGATGATAAAGAAATCAGCAGCCACGATAGCTCGACCA
ATGGTCTGATTAAACCGCTATAAAGCGTGGCGCGGTTAA

SEQ ID NO:2

ATGAAAAACATCAATCCAACGCAGACCGCTGCCTGGCAGGCACTACAGAAACACTTCGATGAAATGAA
AGACGTTACGATCGCCGATCTTTTTGCTAAAGATGGTGATCGTTTTTCTAAGTTCTCCGCAACCTTCG
ACGATCAGATGCTGGTGGATTACTCCAAAAACCGCATCACTGAAGAGACGCTGGCGAAATTACAGGAT
CTGGCGAAAGAGTGCGATCTGGCGGGCGCGATTAAAGTCGATGTTCTCTGGCGAGAAGATCAACCGCAC
TGAAAACCGCGCCGTGCTGCACGTAGCGCTGCGTAACCGTAGCAATACCCCGCTGGAAGTGCTGTTTT
AGGGTCCGATTTTGGTTGATGGCAAAGACGTAATGCCGGAAGTCAACGCGGTGCTGGAGAAGATGAAA
ACCTTCTCAGAAGCGATTATTTCCGGTGAGTGGAAGGTTATACCGGCAAAGCAATCACTGACGTAGT
GAACATCGGGATCGGCGGTTCTGACCTCGGCCCATACATGGTGACCGAAGCTCTGCGTCCGTACAAAA
ACCACCTGAACATGCACTTTGTTTCTAACGTCGATGGGACTCACATCGCGGAAGTGCTGAAAAAAGTA
AACCCGGAACACGCTGTTCTTGGTAGCATCTAAAACCTTCACCACTCAGGAACTATGACCAACGC
CCATAGCGCGCGTGACTGGTTCCTGAAAGCGGCAGGTGATGAGAAGCACGTTGCAAAACACTTTGCGG
CGTTTTCCACCAATGCCAAAGCCGTTGGCGAGTTTGGTATTGATACTGCCAACATGTTTCGAGTTCTGG
GACTGGGTTGGCGGCCGTTACTCTTTGTGGTCAGCGATTGGCCTGTCGATTGTTCTCTCCATCGGCTT
TGATAACTTCGTTGAACTGCTTTCCGGCGCACACGCGATGGACAAGCATTTCTCCACCACGCCTGCCG
AGAAAAACCTGCCTGTACTGCTGGCGCTGATTGGCATCTGGTACAACAATTTCTTTGGTGCGGAAACT
GAAGCGATTCTGCCGTATGACCAGTATATGCACCGTTTCGCGGCGTACTTCCAGCAGGGCAATATGGA
GTCCAACGGTAAGTATGTTGACCGTAACGGTAACGTTGTGGATTACCAGACTGGCCCGATTATCTGGG
GTGAACCAGGCACTAACGGTCAGCACGCGTTCTACCAGCTGATCCACCAGGGAACCAAAATGGTACCG

TGCGATTTTCATCGCTCCGGCTATCACCCATAACCCGCTCTCTGATCATCACCAGAACTGCTGTCTAA
CTTCTTCGCCCAGACCGAAGCGCTGGCGTTTGGTAAATCCCGCGAAGTGGTTGAGCAGGAATATCGTG
ATCAGGGTAAAGATCCGGCAACGCTTGACTACGTGGTGCCGTTCAAAGTATTCTGAAGGTAACCGCCCC
ACCAACTCCATCCTGCTGCGTGAAATCACTCCGTTTACGCTGGGTGCGTTGATTGCGCTGTATGAGCA
CAAAATCTTTACTCAGGGCGTGATCCTGAACATCTTCACCTTCGACCAGTGGGGCGTGGAAGTGGGTA
AACAGCTGGCGAACCGTATTCTGCCAGAGCTGAAAGATGATAAAGAAATCAGCAGCCACGATAGCTCG
ACCAATGGTCTGATTAACCGCTATAAAGCGTGGCGCGGTtaa

SEQ ID NO:3

ATGAAAAACATCAATCCAACGCAGACCGCTGCCTGGCAGGCACTACAGAAACACTTCGATGA
AATGAAAGACGTTACGATCGCCGATCTTTTTGCTAAAGATGGTGATCGTTTTTCTAAGTTCT
CCGCAACCTTCGACGATCAGATGCTGGTGGATTACTCCAAAACCGCATCACTGAAGAGACG
CTGGCGAAATTACAGGATCTGGCGAAAGAGTGCGATCTGGCGGGCGCGATTAAAGTCGATGTT
CTCTGGCGAGAAGATCAACCGCACTGAAAACCGCGCCGTGCTGCACGTAGCGCTGCGTAACC
GTAGCAATACCCCGATTCTGGAAGTGCTGTTTCAGGGTCCGTTGGTTGATGGCAAAGACGTA
ATGCCGGAAGTCAACGCGGTGCTGGAGAAGATGAAAACCTTCTCAGAAGCGATTATTTCCGG
TGAGTGGAAGGTTATACCGGCAAAGCAATCACTGACGTAGTGAACATCGGGATCGGCGGTT
CTGACCTCGGCCCATACATGGTGACCGAAGCTCTGCGTCCGTACAAAACACCTGAACATG
CACTTTGTTTCTAACGTCGATGGGACTCACATCGCGGAAGTGCTGAAAAAGTAAACCCGGA
AACCACGCTGTTCTTGGTAGCATCTAAAACCTTCACCACTCAGGAACTATGACCAACGCCC
ATAGCGCGCGTGACTGGTTTCTGAAAGCGGCAGGTGATGAGAAGCACGTTGCAAAACACTTT
GCGGCGCTTTCCACCAATGCCAAAGCCGTTGGCGAGTTTGGTATTGATACTGCCAACATGTT
CGAGTTCTGGGACTGGGTTGGCGGCCGTTACTCTTTGTGGTCAGCGATTGGCCTGTCGATTG
TTCTCTCCATCGGCTTTGATAACTTCGTTGAACTGCTTTCCGGCGCACACGCGATGGACAAG
CATTTCTCCACCACGCCTGCCGAGAAAAACCTGCCTGTACTGCTGGCGCTGATTGGCATCTG
GTACAACAATTTCTTTGGTGCGGAACTGAAGCGATTCTGCCGTATGACCAGTATATGCACC
GTTTCGCGGCGTACTTCCAGCAGGGCAATATGGAGTCCAACGGTAAGTATGTTGACCGTAAC
GGTAACGTTGTGGATTACCAGACTGGCCCGATTATCTGGGGTGAACCAGGCACTAACGGTCA
GCACGCGTTCTACCAGCTGATCCACCAGGGAACCAAAATGGTACCGTGCGATTTTCATCGCTC
CGGCTATCACCCATAACCCGCTCTCTGATCATCACCAGAACTGCTGTCTAACTTCTTCGCC
CAGACCGAAGCGCTGGCGTTTGGTAAATCCCGCGAAGTGGTTGAGCAGGAATATCGTGATCA
GGGTAAAGATCCGGCAACGCTTGACTACGTGGTGCCGTTCAAAGTATTCTGAAGGTAACCGCC
CGACCAACTCCATCCTGCTGCGTGAAATCACTCCGTTTACGCTGGGTGCGTTGATTGCGCTG
TATGAGCACAAAATCTTTACTCAGGGCGTGATCCTGAACATCTTCACCTTCGACCAGTGGGG
CGTGGAAGTGGGTAAACAGCTGGCGAACCGTATTCTGCCAGAGCTGAAAGATGATAAAGAAA
TCAGCAGCCACGATAGCTCGACCAATGGTCTGATTAACCGCTATAAAGCGTGGCGCGGTtaa

SEQ ID NO:4

ATGAAAAACATCAATCCAACGCAGACCGCTGCCTGGCAGGCACTACAGAAACACTTCGATGA
AATGAAAGACGTTACGATCGCCGATCTTTTTGCTAAAGATGGTGATCGTTTTTCTAAGTTCT
CCGCAACCTTCGACGATCAGATGCTGGTGGATTACTCCAAAACCGCATCACTGAAGAGACG
CTGGCGAAATTACAGGATCTGGCGAAAGAGTGCGATCTGGCGGGCGCGATTAAAGTCGATGTT
CTCTGGCGAGAAGATCAACCGCACTGAAAACCGCGCCGTGCTGCACGTAGCGCTGCGTAACC
GTAGCAATACCCCGATTTTGGGAAGTGCTGTTTCAGGGTCCGTTGATGGCAAAGACGTAATG
CCGGAAGTCAACGCGGTGCTGGAGAAGATGAAAACCTTCTCAGAAGCGATTATTTCCGGTGA
GTGGAAAGGTTATACCGGCAAAGCAATCACTGACGTAGTGAACATCGGGATCGGCGGTTCTG
ACCTCGGCCCATACATGGTGACCGAAGCTCTGCGTCCGTACAAAACACCTGAACATGCAC
TTTGTTTCTAACGTCGATGGGACTCACATCGCGGAAGTGCTGAAAAAGTAAACCCGGAAAC
CACGCTGTTCTTGGTAGCATCTAAAACCTTCACCACTCAGGAACTATGACCAACGCCCCATA

GCGCGCGTGA CTGGTTCC TGAAAGCG GCAGGTGA TGAGAAGC ACGTTGCA AAAACACT TTGCGG
GCGCTTTCC ACCAATGCC AAAGCCGT TGGCGAG TTTGGTAT TGATACTG CCAACATG TTCGA
GTTCTGGG ACTGGGTT GGCGGCCG TTA CTCTTTGT GGTGAG CATTGGC CTGTCG ATTGTT C
TCTCCATCG GCTTTGATA AACTTCGT TGAAGTGC TTTCCGG CGCACAC GCGATGG ACAAGCAT
TTCTCCACC ACGCCTGCC GAGAAAAA CCTGCCTG TACTGCT GGCCTG ATTGGC ATCTGGTA
CAACAATTT CTTTGGT GCGGAACT GAAGCGAT TCTGCCG TATGACC AGTATATG CACCGTT
TCGCGGCGT ACTTCCAG CAGGGCAAT ATGGAGT CCAACGGT AAGTATG TTGACCGT AACGGT
AACGTTGTG GATTACCAG ACTGGCCC GATTATCT GGGGTGA ACCAGGCA CTAACGGT CAGCA
CGCGTTCTA CCAGCTGA TCCACCAG GGAACCAA AATGGTACC GTGCGATT TCATCGCT CCGG
CTATCACCC ATAACCCG CTCTCTG ATCATCACC AGAAACTG CTGTCTA ACTTCTTC GCCCAG
ACCGAAGCG CTGGCGTT TGGTAAAT CCCGCGA AGTGTTG AGCAGGA ATATCGT GATCAGGG
TAAAGATCC GGCACGCT TGACTACG TGCTGCC GTTCAAAGT ATTCGA AGGTAACC GCGCCGA
CCAACTCCA TCCTGCTG CGTGAAAT CACTCCGT TCAGCCTG GGTGCGTT GATTGCG CTGTAT
GAGCACAAA ATCTTTACT CAGGGCGT GATCCTGA ACATCTTC ACCTTCG ACCAGTGG GGGCGT
GGAAGTGGG TAAACAG CTGGCGA ACCGTATT CTGCCAG AGCTGAA AGATGATA AAAGAAAT CA
GCAGCCACG ATAGCTCG ACCAATGG TCTGATTA ACCGCTATA AAAGCGTGG CGCGGTtaa

SEQ ID NO:5

ATGAAAAAC ATCAATCCA ACGCAGAC CGCTGCCT GGCAGGCA CTACAGAA ACACTTCG ATGA
AATGAAAGAC GTTACGAT CGCCGATC TTTTTT GCTAAAG ATGGTGAT CGTTTTT CTAAAGTT CT
CCGCAACCTTC GACGATC AGATGCTG GTGGATT ACTCCAAA AACCGCAT CACTGAAG AGACG
CTGGCGAAATT ACAGGATCT GGCAGAAAG AGTGCGAT CTGGCGGG CGCGATT AAGTCGAT GTT
CTCTGGCGAGA AGATCAACC GCACTGAAA ACCGCGCC GTGCTGCAC GTAGCGCT GCGTAACC
GTAGCAATAAC CCCGATTTT TGGTTGAT GGCAGAAAG ACCTAATGCC CGGAAGTCA ACGCGGTG CTG
GAGAAGATGAAA ACCTTCTC AGAAGCGATT ATTTCCGGT GAGCTGGA AGTGCTGTTT CAGGG
TCCGTGGAAAG GTTATACCG GCAAAGCA ATCACTGAC GTAGTGA ACATCGGG ATCGGCG GTT
CTGACCTCGG CCCATACAT GGTGACCGA AGCTCTG CGTCCGTAC AAAAAACC ACCTGAAC ATG
CACTTTGTTTCT AACGTCGAT GGGACTC ACATCGCG GAAGTGCTG AAAAAAGTAA ACCCGGA
AACCACGCTGT TCTTGGT AGCATCTA AAACCTTC ACCACTC AGGAACTATG ACCAACGCCC
ATAGCGCGCGT GACTGGTT CCTGAAAG CGGCAGGTG ATGAGAAG CACGTTG CAAAACACT TT
GCGGCGCTTTCC ACCAATGCC AAAGCCGT TGGCGAG TTTGGTATT GATACTG CCAACATG TT
CGAGTTCTGGG ACTGGGTT GGCAGGCC GTTACTCT TTTGTGGT CAGCGATT GGCCTGT CATTG
TTCTCTCCATC GGCTTTGATA AACTTCGT TGAAGTGC TTTCCGG CGCACAC GCGATGG ACAAG
CATTTCTCCACC ACGCCTGCC GAGAAAAA CCTGCCTG TACTGCT GGCCTG ATTGGC ATCTG
GTACAACAATTT CTTTGGT GCGGAACT GAAGCGAT TCTGCCG TATGACC AGTATATG CACC
GTTTCGCGGCGT ACTTCCAG CAGGGCAAT ATGGAGT CCAACGGT AAGTATG TTGACCGT AAC
GGTAACGTTGTG GATTACCAG ACTGGCCC GATTATCT GGGGTGA ACCAGGCA CTAACGGT CA
GCACGCGTTCTA CCAGCTGA TCCACCAG GGAACCAA AATGGTACC GTGCGATT TCATCGCT C
CGGCTATCACCC ATAACCCG CTCTCTG ATCATCACC AGAAACTG CTGTCTA ACTTCTTC GCC
CAGACCGAAG CGCTGGCG TTTGGTAA ATCCCGCGA AGTGTTG AGCAGGA ATATCGT GATCA
GGGTAAAGATCC GGCACGCT TGACTACG TGCTGCC GTTCAAAGT ATTCGA AGGTAACC GCC
CGACCAACTCCA TCCTGCTG CGTGAAAT CACTCCGT TCAGCCTG GGTGCGTT GATTGCG CTG
TATGAGCACAAA ATCTTTACT CAGGGCGT GATCCTGA ACATCTTC ACCTTCG ACCAGTGG GG
CGTGGAAGTGGG TAAACAG CTGGCGA ACCGTATT CTGCCAG AGCTGAA AGATGATA AAAGAAA
TCAGCAGCCAC GATAGCTCG ACCAATGG TCTGATTA ACCGCTATA AAAGCGTGG CGCGGTtaa

SEQ ID NO:6

ATGAAAAACATCAATCCAACGCAGACCGCTGCCTGGCAGGCACTACAGAAACACTTCGATGA
AATGAAAGACGTTACGATCGCCGATCTTTTTGCTAAAGATGGTGATCGTTTTTCTAAGTTCT
CCGCAACCTTCGACGATCAGATGCTGGTGGATTACTCCAAAAACCGCATCACTGAAGAGACG
CTGGCGAAATTACAGGATCTGGCGAAAGAGTGCGATCTGGCGGGCGCGATTAAAGTCGATGTT
CTCTGGCGAGAAGATCAACCGCACTGAAAACCGCGCCGTGCTGCACGTAGCGCTGCGTAACC
GTAGCAATACCCCGATTTTGGTTGATGGCAAAGACGTAATGCCGGAAGTCAACGCGGTGCTG
GAGAAGATGAAAACCTTCTCAGAAGCGATTATTTCCGGTGAGTGGAAGGTTATACCGGCAA
AGCAATCACTGACGTAGTGAACATCGGGATCGGCGGTTCTGACCTCGGCCCATACATGGTGA
CCGAAGCTCTGCGTCCGTACAAAAACCACCTGAACATGCACTTTGTTTCTAACGTCGATGGG
ACTCACATCGCGGAAGTGCTGAAAAAAGTAAACCCGGAAACCACGCTGTTCTTGGTAGCATC
TAAACCTTTCACCACTCAGGAAACTATGACCAACGCCCATAGCGCGCGTGACTGGTTCCTGA
AAGCGGCAGGTGATGAGAAGCACGTTGCAAAACACTTTGCGGCGCTTTCCACCAATGCCAAA
GCCGTTGGCGAGTTTGGTATTGATACTGCCAACATGTTTCGAGTTCTGGGACTGGGTTGGCGG
CCGTTACTCTTTGTGGTCAGCGATTGGCCTGTTCGATTGTTCTCTCCATCGGCTTTGATAACT
TCGTTGAACTGCTTTCCGGCGCACACGCGATGGACAAGCATTCTCCACCACGCCTGCCGAG
AAAAACCTGCCTGTACTGCTGGCGCTGATTGGCATCTGGTACAACAATTTCTTTGGTGCGGA
AACTGAAGCGATTCTGCCGTATGACCAGTATATGCACCGTTTCGCGGCGTACTTCCAGCAGG
GCAATATGGAGTCCAACGGTAAGTATGTTGACCGTAACGGTAACGTTGTGGATTACCAGACT
GGCCCGATTATCTGGGGTGAACCAGGCACCTAACGGTCAGCACGCGTTCTACCAGCTGATCCA
CCAGGGAACCAAAATGGTACCGTGCGATTTTCATCGCTCCGGCTATCACCCATCTGGAAGTGC
TGTTTCAGGGTCCGAACCCGCTCTCTGATCATCACAGAAACTGCTGTCTAACTTCTTCGCC
CAGACCGAAGCGCTGGCGTTTGGTAAATCCCGCGAAGTGGTTGAGCAGGAATATCGTGATCA
GGGTAAAGATCCGGCAACGCTTGACTACGTGGTGCCGTTCAAAGTATTCTGAAGGTAACCGCC
CGACCAACTCCATCCTGCTGCGTGAAATCACTCCGTTACGCCTGGGTGCGTTGATTGCGCTG
TATGAGCACAAAATCTTTACTCAGGGCGTGATCCTGAACATCTTCACCTTCGACCAGTGGGG
CGTGGAAGTGGGTAAACAGCTGGCGAACCGTATTCTGCCAGAGCTGAAAGATGATAAAGAAA
TCAGCAGCCACGATAGCTCGACCAATGGTCTGATTAACCGCTATAAAGCGTGGCGCGGTtaa

SEQ ID NO:7

ATGAAAAACATCAATCCAACGCAGACCGCTGCCTGGCAGGCACTACAGAAACACTTCGATGA
AATGAAAGACGTTACGATCGCCGATCTTTTTGCTAAAGATGGTGATCGTTTTTCTAAGTTCT
CCGCAACCTTCGACGATCAGATGCTGGTGGATTACTCCAAAAACCGCATCACTGAAGAGACG
CTGGCGAAATTACAGGATCTGGCGAAAGAGTGCGATCTGGCGGGCGCGATTAAAGTCGATGTT
CTCTGGCGAGAAGATCAACCGCACTGAAAACCGCGCCGTGCTGCACGTAGCGCTGCGTAACC
GTAGCAATACCCCGATTTTGGTTGATGGCAAAGACGTAATGCCGGAAGTCAACGCGGTGCTG
GAGAAGATGAAAACCTTCTCAGAAGCGATTATTTCCGGTGAGTGGAAGGTTATACCGGCAA
AGCAATCACTGACGTAGTGAACATCGGGATCGGCGGTTCTGACCTCGGCCCATACATGGTGA
CCGAAGCTCTGCGTCCGTACAAAAACCACCTGAACATGCACTTTGTTTCTAACGTCGATGGG
ACTCACATCGCGGAAGTGCTGAAAAAAGTAAACCCGGAAACCACGCTGTTCTTGGTAGCATC
TAAACCTTTCACCACTCAGGAAACTATGACCAACGCCCATAGCGCGCGTGACTGGTTCCTGA
AAGCGGCAGGTGATGAGAAGCACGTTGCAAAACACTTTGCGGCGCTTTCCACCAATGCCAAA
GCCGTTGGCGAGTTTGGTATTGATACTGCCAACATGTTTCGAGTTCTGGGACTGGGTTGGCGG
CCGTTACTCTTTGTGGTCAGCGATTGGCCTGTTCGATTGTTCTCTCCATCGGCTTTGATAACT
TCGTTGAACTGCTTTCCGGCGCACACGCGATGGACAAGCATTCTCCACCACGCCTGCCGAG
AAAAACCTGCCTGTACTGCTGGCGCTGATTGGCATCTGGTACAACAATTTCTTTGGTGCGGA
AACTGAAGCGATTCTGCCGTATGACCAGTATATGCACCGTTTCGCGGCGTACTTCCAGCAGG
GCAATATGGAGTCCAACGGTAAGTATGTTGACCGTAACGGTAACGTTGTGGATTACCAGACT

GGCCCGATTATCTGGGGTGAACCAGGCACTAACGGTCAGCACGCGTTCTACCAGCTGATCCA
CCAGGGAACCAAAATGGTACCGTGCGATTTTCATCGCTCCGGCTATCACCCATAACCCGCTCT
CTGATCATCACCAGAACTGCTGTCTAACTTCTTCGCCCAGACCGAAGCGCTGGCGTTTGGT
AAATCCCGCGAAGTGGTTGAGCAGGAATATCGTGATCAGGGTAAAGATCCGGCAACGCTTGA
CTACGTGGTGCCGTTCAAAGTATTCGAAGGTAACCGCCGACCAACTCCATCCTGCTGCGTG
AAATCACTCCGTTACGCCTGGGTGCGTTGATTGCGCTGTATGAGCACAAAATCTTTACTCAG
GGCGTGATCCTGAACATCTTCACCTTCGACCAGTGGGGCGTGGAAGTGGGTAAACAGCTGGC
GAACCGTATTCTGCCAGAGCTGGAAGTGCTGTTTCAGGGTCCGAAAGATGATAAAGAAATCA
GCAGCCACGATAGCTCGACCAATGGTCTGATTAACCGCTATAAAGCGTGGCGCGGTtaa

SEQ ID NO:8

ATGAAAAACATCAATCCAACGCAGACCGCTGCCTGGCAGGCACTACAGAAACACTTCGATGA
AATGAAAGACGTTACGATCGCCGATCTTTTTGCTAAAGATGGTGATCGTTTTTCTAAGTTCT
CCGCAACCTTCGACGATCAGATGCTGGTGGATTACTCCAAAACCGCATCACTGAAGAGACG
CTGGCGAAATTACAGGATCTGGCGAAAGAGTGCGATCTGGCGGGCGCGATTAAAGTCGATGTT
CTCTGGCGAGAAGATCAACCGCACTGAAAACCGCGCCGTGCTGCACGTAGCGCTGCGTAACC
GTAGCAATACCCCGATTTTGGTTGATGGCAAAGACGTAATGCCGGAAGTCAACGCGGTGCTG
GAGAAGATGAAAACCTTCTCAGAAGCGATTATTTCCGGTGAGTGGAAGGTTATACCGGCAA
AGCAATCACTGACGTAGTGAACATCGGGATCGGCGGTTCTGACCTCGGCCCATACATGGTGA
CCGAAGCTCTGCGTCCGTACAAAACACCTGAACATGCACTTTGTTTTCTAACGTCGATGGG
ACTCACATCGCGGAAGTGCTGAAAAAGTAAACCCGGAAACACGCTGTTCTTGGTAGCATC
TAAAACCTTCACCACTCAGGAACTATGACCAACGCCCATAGCGCGCGTGACTGGTTCCCTGA
AAGCGGCAGGTGATGAGAAGCACGTTGCAAAACACTTTGCGGCGCTTTCCACCAATGCCAAA
GCCGTTGGCGAGTTTGGTATTGATACTGCCAACATGTTTCGAGTTCTGGGACTGGGTTGGCGG
CCGTTACTCTTTGTGGTCAGCGATTGGCCTGTGCGATTGTTCTCTCCATCGGCTTTGATAACT
TCGTTGAACTGCTTTCCGGCGCACACGCGATGGACAAGCATTTCTCCACCACGCCTGCCGAG
AAAAACCTGCCTGTACTGCTGGCGCTGATTGGCATCTGGTACAACAATTTCTTTGGTGCGGA
AACTGAAGCGATTCTGCCGTATGACCAGTATATGCACCGTTTCGCGGCGTACTTCCAGCAGG
GCAATATGGAGTCCAACGGTAAGTATGTTGACCGTAACGGTAACGTTGTGGATTACCAGACT
GGCCCGATTATCTGGGGTGAACCAGGCACTAACGGTCAGCACGCGTTCTACCAGCTGATCCA
CCAGGGAACCAAAATGGTACCGTGCGATTTTCATCGCTCCGGCTATCACCCATAACCCGCTCT
CTGATCATCACCAGAACTGCTGTCTAACTTCTTCGCCCAGACCGAAGCGCTGGCGTTTGGT
AAATCCCGCGAAGTGGTTGAGCAGGAATATCGTGATCAGGGTAAAGATCCGGCAACGCTTGA
CTACGTGGTGCCGTTCAAAGTATTCGAAGGTAACCGCCGACCAACTCCATCCTGCTGCGTG
AAATCACTCCGTTACGCCTGGGTGCGTTGATTGCGCTGTATGAGCACAAAATCTTTACTCAG
GGCGTGATCCTGAACATCTTCACCTTCGACCAGTGGGGCGTGGAAGTGGGTAAACAGCTGGC
GAACCGTATTCTGCCAGAGCTGAAACTGGAAGTGCTGTTTCAGGGTCCGGATGATAAAGAAA
TCAGCAGCCACGATAGCTCGACCAATGGTCTGATTAACCGCTATAAAGCGTGGCGCGGTtaa

SEQ ID NO:9

ATGAAAAACATCAATCCAACGCAGACCGCTGCCTGGCAGGCACTACAGAAACACTTCGATGA
AATGAAAGACGTTACGATCGCCGATCTTTTTGCTAAAGATGGTGATCGTTTTTCTAAGTTCT
CCGCAACCTTCGACGATCAGATGCTGGTGGATTACTCCAAAACCGCATCACTGAAGAGACG
CTGGCGAAATTACAGGATCTGGCGAAAGAGTGCGATCTGGCGGGCGCGATTAAAGTCGATGTT
CTCTGGCGAGAAGATCAACCGCACTGAAAACCGCGCCGTGCTGCACGTAGCGCTGCGTAACC
GTAGCAATACCCCGATTTTGGTTGATGGCAAAGACGTAATGCCGGAAGTCAACGCGGTGCTG
GAGAAGATGAAAACCTTCTCAGAAGCGATTATTTCCGGTGAGTGGAAGGTTATACCGGCAA
AGCAATCACTGACGTAGTGAACATCGGGATCGGCGGTTCTGACCTCGGCCCATACATGGTGA

CCGAAGCTCTGCGTCCGTACAAAAACCACCTGAACATGCACTTTGTTTCTAACGTCGATGGG
ACTCACATCGCGGAAGTGCTGAAAAAAGTAAACCCGGAAACCACGCTGTTCTTGGTAGCATC
TAAAACCTTCACCACTCAGGAAACTATGACCAACGCCCATAGCGCGCGTGACTGGTTCCTGA
AAGCGGCAGGTGATGAGAAGCACGTTGCAAAACACTTTGCGGCGCTTTCCACCAATGCCAAA
GCCGTTGGCGAGTTTGGTATTGATACTGCCAACATGTTTCGAGTTCTGGGACTGGGTTGGCGG
CCGTTACTCTTTGTGGTCAGCGATTGGCCTGTTCGATTGTTCTCTCCATCGGCTTTGATAACT
TCGTTGAACTGCTTTCCGGCGCACACGCGATGGACAAGCATTCTCCACCACGCCTGCCGAG
AAAAACCTGCCTGTACTGCTGGCGCTGATTGGCATCTGGTACAACAATTTCTTTGGTGCGGA
AACTGAAGCGATTCTGCCGTATGACCAGTATATGCACCGTTTCGCGGCGTACTTCCAGCAGG
GCAATATGGAGTCCAACGGTAAGTATGTTGACCGTAACGGTAACGTTGTGGATTACCAGACT
GGCCCGATTATCTGGGGTGAACCAGGCACTAACGGTCAGCACGCGTTCTACCAGCTGATCCA
CCAGGGAACCAAAATGGTACCGTGCGATTTTCATCGCTCCGGCTATCACCCATAACCCGCTCT
CTGATCATCACCAAGAACTGCTGTCTAACTTCTTCGCCCAGACCGAAGCGCTGGCGTTTGGT
AAATCCCGCGAAGTGTTGAGCAGGAATATCGTGATCAGGGTAAAGATCCGGCAACGCTTGA
CTACGTGGTGCCGTTCAAAGTATTCTGAAGGTAAACGCCCGACCAACTCCATCCTGCTGCGTG
AAATCACTCCGTTTCAGCCTGGGTGCGTTGATTGCGCTGTATGAGCACAAAATCTTTACTCAG
GGCGTGATCCTGAACATCTTCACCTTCGACCAGTGGGGCGTGGAAGTGGGTAAACAGCTGGC
GAACCGTATTCTGCCAGAGCTGAAAGATCTGGAAGTGCTGTTTCAGGGTCCGGATAAAGAAA
TCAGCAGCCACGATAGCTCGACCAATGGTCTGATTAACCGCTATAAAGCGTGGCGCGGTtaa

SEQ ID NO:10

ATGAAAAACATCAATCCAACGCAGACCGCTGCCTGGCAGGCACTACAGAAACACTTCGATGA
AATGAAAAGACGTTACGATCGCCGATCTTTTTTGCTAAAGATGGTGATCGTTTTTCTAAGTTCT
CCGCAACCTTCGACGATCAGATGCTGGTGGATTACTCCAAAACCGCATCACTGAAGAGACG
CTGGCGAAATTACAGGATCTGGCGAAAGAGTGCGATCTGGCGGGCGCGATTAAAGTCGATGTT
CTCTGGCGAGAAGATCAACCGCACTGAAAACCGCGCCGTGCTGCACGTAGCGCTGCGTAACC
GTAGCAATACCCCGATTTTGGTTGATGGCAAAGACGTAATGCCGGAAGTCAACGCGGTGCTG
GAGAAGATGAAAACCTTCTCAGAAGCGATTATTTCCGGTGAGTGAAAGGTTATACCGGCAA
AGCAATCACTGACGTAGTGAACATCGGGATCGGCGGTTCTGACCTCGGCCCATACATGGTGA
CCGAAGCTCTGCGTCCGTACAAAAACCACCTGAACATGCACTTTGTTTCTAACGTCGATGGG
ACTCACATCGCGGAAGTGCTGAAAAAAGTAAACCCGGAAACCACGCTGTTCTTGGTAGCATC
TAAAACCTTCACCACTCAGGAAACTATGACCAACGCCCATAGCGCGCGTGACTGGTTCCTGA
AAGCGGCAGGTGATGAGAAGCACGTTGCAAAACACTTTGCGGCGCTTTCCACCAATGCCAAA
GCCGTTGGCGAGTTTGGTATTGATACTGCCAACATGTTTCGAGTTCTGGGACTGGGTTGGCGG
CCGTTACTCTTTGTGGTCAGCGATTGGCCTGTTCGATTGTTCTCTCCATCGGCTTTGATAACT
TCGTTGAACTGCTTTCCGGCGCACACGCGATGGACAAGCATTCTCCACCACGCCTGCCGAG
AAAAACCTGCCTGTACTGCTGGCGCTGATTGGCATCTGGTACAACAATTTCTTTGGTGCGGA
AACTGAAGCGATTCTGCCGTATGACCAGTATATGCACCGTTTCGCGGCGTACTTCCAGCAGG
GCAATATGGAGTCCAACGGTAAGTATGTTGACCGTAACGGTAACGTTGTGGATTACCAGACT
GGCCCGATTATCTGGGGTGAACCAGGCACTAACGGTCAGCACGCGTTCTACCAGCTGATCCA
CCAGGGAACCAAAATGGTACCGTGCGATTTTCATCGCTCCGGCTATCACCCATAACCCGCTCT
CTGATCATCACCAAGAACTGCTGTCTAACTTCTTCGCCCAGACCGAAGCGCTGGCGTTTGGT
AAATCCCGCGAAGTGTTGAGCAGGAATATCGTGATCAGGGTAAAGATCCGGCAACGCTTGA
CTACGTGGTGCCGTTCAAAGTATTCTGAAGGTAAACGCCCGACCAACTCCATCCTGCTGCGTG
AAATCACTCCGTTTCAGCCTGGGTGCGTTGATTGCGCTGTATGAGCACAAAATCTTTACTCAG
GGCGTGATCCTGAACATCTTCACCTTCGACCAGTGGGGCGTGGAAGTGGGTAAACAGCTGGC
GAACCGTATTCTGCCAGAGCTGAAAGATGATCTGGAAGTGCTGTTTCAGGGTCCGAAAGAAA
TCAGCAGCCACGATAGCTCGACCAATGGTCTGATTAACCGCTATAAAGCGTGGCGCGGTtaa

SEQ ID NO:11

ATGAAAAACATCAATCCAACGCAGACCGCTGCCTGGCAGGCACTACAGAAACACTTCGATGA
AATGAAAGACGTTACGATCGCCGATCTTTTTGCTAAAGATGGTGATCGTTTTTCTAAGTTCT
CCGCAACCTTCGACGATCAGATGCTGGTGGATTACTCCAAAAACCGCATCACTGAAGAGACG
CTGGCGAAATTACAGGATCTGGCGAAAGAGTGCGATCTGGCGGGCGCGATTAAAGTCGATGTT
CTCTGGCGAGAAGATCAACCGCACTGAAAACCGCGCCGTGCTGCACGTAGCGCTGCGTAACC
GTAGCAATACCCCGATTTTGGTTGATGGCAAAGACGTAATGCCGGAAGTCAACGCGGTGCTG
GAGAAGATGAAAACCTTCTCAGAAGCGATTATTTCCGGTGAGTGGAAGGTTATACCGGCAA
AGCAATCACTGACGTAGTGAACATCGGGATCGGCGGTTCTGACCTCGGCCCATACATGGTGA
CCGAAGCTCTGCGTCCGTACAAAAACCACCTGAACATGCACTTTGTTTCTAACGTCGATGGG
ACTCACATCGCGGAAGTGCTGAAAAAAGTAAACCCGGAAACCACGCTGTTCTTGGTAGCATC
TAAACCTTTCACCACTCAGGAAACTATGACCAACGCCCATAGCGCGCGTGACTGGTTCCTGA
AAGCGGCAGGTGATGAGAAGCACGTTGCAAAACACTTTGCGGCGCTTTCACCAATGCCAAA
GCCGTTGGCGAGTTTGGTATTGATACTGCCAACATGTTTCGAGTTCTGGGACTGGGTTGGCGG
CCGTTACTCTTTGTGGTCAGCGATTGGCCTGTTCGATTGTTCTCTCCATCGGCTTTGATAACT
TCGTTGAACTGCTTTCCGGCGCACACGCGATGGACAAGCATTCTCTCCACCACGCCTGCCGAG
AAAAACCTGCCTGTACTGCTGGCGCTGATTGGCATCTGGTACAACAATTTCTTTGGTGCGGA
AACTGAAGCGATTCTGCCGTATGACCAGTATATGCACCGTTTCGCGGCGTACTTCCAGCAGG
GCAATATGGAGTCCAACGGTAAGTATGTTGACCGTAACGGTAACGTTGTGGATTACCAGACT
GGCCCGATTATCTGGGGTGAACCAGGCACCTAACGGTCAGCACGCGTTCTACCAGCTGATCCA
CCAGGGAACCAAAATGGTACCGTGCGATTTTCATCGCTCCGGCTATCACCCATAACCCGCTCT
CTGATCATCACCAAGAACTGCTGTCTAACTTCTTCGCCCAGACCGAAGCGCTGGCGTTTGGT
AAATCCCGCGAAGTGGTTGAGCAGGAATATCGTGATCAGGGTAAAGATCCGGCAACGCTTGA
CTACGTGGTGCCGTTCAAAGTATTCGAAGGTAACCGCCCGACCAACTCCATCCTGCTGCGTG
AAATCACTCCGTTTCAGCCTGGGTGCGTTGATTGCGCTGTATGAGCACAAAATCTTTACTCAG
GGCGTGATCCTGAACATCTTCACCTTCGACCAGTGGGGCGTGGAAGTGGGTAAACAGCTGGC
GAACCGTATTCTGCCAGAGCTGAAAGATGATAAACTGGAAGTGCTGTTTCAGGGTCCGGAAA
TCAGCAGCCACGATAGCTCGACCAATGGTCTGATTAACCGCTATAAAGCGTGGCGCGGTtaa

SEQ ID NO:12

ATGAAAAACATCAATCCAACGCAGACCGCTGCCTGGCAGGCACTACAGAAACACTTCGATGA
AATGAAAGACGTTACGATCGCCGATCTTTTTGCTAAAGATGGTGATCGTTTTTCTAAGTTCT
CCGCAACCTTCGACGATCAGATGCTGGTGGATTACTCCAAAAACCGCATCACTGAAGAGACG
CTGGCGAAATTACAGGATCTGGCGAAAGAGTGCGATCTGGCGGGCGCGATTAAAGTCGATGTT
CTCTGGCGAGAAGATCAACCGCACTGAAAACCGCGCCGTGCTGCACGTAGCGCTGCGTAACC
GTAGCAATACCCCGATTTTGGTTGATGGCAAAGACGTAATGCCGGAAGTCAACGCGGTGCTG
GAGAAGATGAAAACCTTCTCAGAAGCGATTATTTCCGGTGAGTGGAAGGTTATACCGGCAA
AGCAATCACTGACGTAGTGAACATCGGGATCGGCGGTTCTGACCTCGGCCCATACATGGTGA
CCGAAGCTCTGCGTCCGTACAAAAACCACCTGAACATGCACTTTGTTTCTAACGTCGATGGG
ACTCACATCGCGGAAGTGCTGAAAAAAGTAAACCCGGAAACCACGCTGTTCTTGGTAGCATC
TAAACCTTTCACCACTCAGGAAACTATGACCAACGCCCATAGCGCGCGTGACTGGTTCCTGA
AAGCGGCAGGTGATGAGAAGCACGTTGCAAAACACTTTGCGGCGCTTTCACCAATGCCAAA
GCCGTTGGCGAGTTTGGTATTGATACTGCCAACATGTTTCGAGTTCTGGGACTGGGTTGGCGG
CCGTTACTCTTTGTGGTCAGCGATTGGCCTGTTCGATTGTTCTCTCCATCGGCTTTGATAACT
TCGTTGAACTGCTTTCCGGCGCACACGCGATGGACAAGCATTCTCTCCACCACGCCTGCCGAG
AAAAACCTGCCTGTACTGCTGGCGCTGATTGGCATCTGGTACAACAATTTCTTTGGTGCGGA
AACTGAAGCGATTCTGCCGTATGACCAGTATATGCACCGTTTCGCGGCGTACTTCCAGCAGG
GCAATATGGAGTCCAACGGTAAGTATGTTGACCGTAACGGTAACGTTGTGGATTACCAGACT

GGCCCGATTATCTGGGGTGAACCAGGCACTAACGGTCAGCACGCGTTCTACCAGCTGATCCA
CCAGGGAACCAAAATGGTACCGTGCGATTTTCATCGCTCCGGCTATCACCCATAACCCGCTCT
CTGATCATCACCAGAACTGCTGTCTAACTTCTTCGCCCAGACCGAAGCGCTGGCGTTTGGT
AAATCCCGCGAAGTGGTTGAGCAGGAATATCGTGATCAGGGTAAAGATCCGGCAACGCTTGA
CTACGTGGTGCCGTTCAAAGTATTCTGAAGGTAACCGCCGACCAACTCCATCCTGCTGCGTG
AAATCACTCCGTTACGCCTGGGTGCGTTGATTGCGCTGTATGAGCACAAAATCTTTACTCAG
GGCGTGATCCTGAACATCTTCACCTTCGACCAGTGGGGCGTGGAAGTGGGTAAACAGCTGGC
GAACCGTATTCTGCCAGAGCTGAAAGATGATAAAGAACTGGAAGTGCTGTTTCAGGGTCCGA
TCAGCAGCCACGATAGCTCGACCAATGGTCTGATTAACCGCTATAAAGCGTGGCGCGGTtaa

SEQ ID NO:13

ATGAAAAACATCAATCCAACGCAGACCGCTGCCTGGCAGGCACTACAGAAACACTTCGATGA
AATGAAAGACGTTACGATCGCCGATCTTTTTGCTAAAGATGGTGATCGTTTTTCTAAGTTCT
CCGCAACCTTCGACGATCAGATGCTGGTGGATTACTCCAAAACCGCATCACTGAAGAGACG
CTGGCGAAATTACAGGATCTGGCGAAAGAGTGCGATCTGGCGGGCGCGATTAAAGTCGATGTT
CTCTGGCGAGAAGATCAACCGCACTGAAAACCGCGCCGTGCTGCACGTAGCGCTGCGTAACC
GTAGCAATACCCCGATTTTGGTTGATGGCAAAGACGTAATGCCGGAAGTCAACGCGGTGCTG
GAGAAGATGAAAACCTTCTCAGAAGCGATTATTTCCGGTGAGTGGAAGGTTATACCGGCAA
AGCAATCACTGACGTAGTGAACATCGGGATCGGCGGTTCTGACCTCGGCCCATACATGGTGA
CCGAAGCTCTGCGTCCGTACAAAAACCACTGAACATGCACTTTGTTTTCTAACGTCGATGGG
ACTCACATCGCGGAAGTGCTGAAAAAGTAAACCCGGAACCAACGCTGTTCTTGGTAGCATC
TAAAACCTTCACCACTCAGGAACTATGACCAACGCCCATAGCGCGCGTGACTGGTTCCCTGA
AAGCGGCAGGTGATGAGAAGCACGTTGCAAAACACTTTGCGGCGCTTTCCACCAATGCCAAA
GCCGTTGGCGAGTTTGGTATTGATACTGCCAACATGTTTCGAGTTCTGGGACTGGGTTGGCGG
CCGTTACTCTTTGTGGTCAGCGATTGGCCTGTGATTGTTCTCTCCATCGGCTTTGATAACT
TCGTTGAACTGCTTTCCGGCGCACACGCGATGGACAAGCATTCTCCACCACGCCTGCCGAG
AAAAACCTGCCTGTACTGCTGGCGCTGATTGGCATCTGGTACAACAATTTCTTTGGTGCGGA
AACTGAAGCGATTCTGCCGTATGACCAGTATATGCACCGTTTCGCGGCGTACTTCCAGCAGG
GCAATATGGAGTCCAACGGTAAGTATGTTGACCGTAACGGTAACGTTGTGGATTACCAGACT
GGCCCGATTATCTGGGGTGAACCAGGCACTAACGGTCAGCACGCGTTCTACCAGCTGATCCA
CCAGGGAACCAAAATGGTACCGTGCGATTTTCATCGCTCCGGCTATCACCCATAACCCGCTCT
CTGATCATCACCAGAACTGCTGTCTAACTTCTTCGCCCAGACCGAAGCGCTGGCGTTTGGT
AAATCCCGCGAAGTGGTTGAGCAGGAATATCGTGATCAGGGTAAAGATCCGGCAACGCTTGA
CTACGTGGTGCCGTTCAAAGTATTCTGAAGGTAACCGCCGACCAACTCCATCCTGCTGCGTG
AAATCACTCCGTTACGCCTGGGTGCGTTGATTGCGCTGTATGAGCACAAAATCTTTACTCAG
GGCGTGATCCTGAACATCTTCACCTTCGACCAGTGGGGCGTGGAAGTGGGTAAACAGCTGGC
GAACCGTATTCTGCCAGAGCTGAAAGATGATAAAGAAATCCTGGAAGTGCTGTTTCAGGGTC
CGAGCAGCCACGATAGCTCGACCAATGGTCTGATTAACCGCTATAAAGCGTGGCGCGGTtaa

SEQ ID NO:14

ATGAAAAACATCAATCCAACGCAGACCGCTGCCTGGCAGGCACTACAGAAACACTTCGATGA
AATGAAAGACGTTACGATCGCCGATCTTTTTGCTAAAGATGGTGATCGTTTTTCTAAGTTCT
CCGCAACCTTCGACGATCAGATGCTGGTGGATTACTCCAAAACCGCATCACTGAAGAGACG
CTGGCGAAATTACAGGATCTGGCGAAAGAGTGCGATCTGGCGGGCGCGATTAAAGTCGATGTT
CTCTGGCGAGAAGATCAACCGCACTGAAAACCGCGCCGTGCTGCACGTAGCGCTGCGTAACC
GTAGCAATACCCCGATTTTGGTTGATGGCAAAGACGTAATGCCGGAAGTCAACGCGGTGCTG
GAGAAGATGAAAACCTTCTCAGAAGCGATTATTTCCGGTGAGTGGAAGGTTATACCGGCAA
AGCAATCACTGACGTAGTGAACATCGGGATCGGCGGTTCTGACCTCGGCCCATACATGGTGA

CCGAAGCTCTGCGTCCGTACAAAAACCACCTGAACATGCACTTTGTTTCTAACGTCGATGGG
ACTCACATCGCGGAAGTGCTGAAAAAAGTAAACCCGGAAACCACGCTGTTCTTGGTAGCATC
TAAAACCTTCACCACTCAGGAAACTATGACCAACGCCCATAGCGCGCGTGACTGGTTCCTGA
AAGCGGCAGGTGATGAGAAGCACGTTGCAAAACACTTTGCGGCGCTTTCCACCAATGCCAAA
GCCGTTGGCGAGTTTGGTATTGATACTGCCAACATGTTTCGAGTTCTGGGACTGGGTTGGCGG
CCGTTACTCTTTGTGGTCAGCGATTGGCCTGTTCGATTGTTCTCTCCATCGGCTTTGATAACT
TCGTTGAACTGCTTTCCGGCGCACACGCGATGGACAAGCATTCTCCACCACGCCTGCCGAG
AAAAACCTGCCTGTACTGCTGGCGCTGATTGGCATCTGGTACAACAATTTCTTTGGTGCGGA
AACTGAAGCGATTCTGCCGTATGACCAGTATATGCACCGTTTCGCGGCGTACTTCCAGCAGG
GCAATATGGAGTCCAACGGTAAGTATGTTGACCGTAACGGTAACGTTGTGGATTACCAGACT
GGCCCGATTATCTGGGGTGAACCAGGCACTAACGGTCAGCACGCGTTCTACCAGCTGATCCA
CCAGGGAACCAAAATGGTACCGTGCGATTTTCATCGCTCCGGCTATCACCCATAACCCGCTCT
CTGATCATCACCAGAACTGCTGTCTAACTTCTTCGCCCAGACCGAAGCGCTGGCGTTTGGT
AAATCCCGCGAAGTGTTGAGCAGGAATATCGTGATCAGGGTAAAGATCCGGCAACGCTTGA
CTACGTGGTGCCGTTCAAAGTATTCTGAAGGTAAACGCCCGACCAACTCCATCCTGCTGCGTG
AAATCACTCCGTTTCAGCCTGGGTGCGTTGATTGCGCTGTATGAGCACAAAATCTTTACTCAG
GGCGTGATCCTGAACATCTTCACCTTCGACCAGTGGGGCGTGGAAGTGGGTAAACAGCTGGC
GAACCGTATTCTGCCAGAGCTGAAAGATGATAAAGAAATCAGCCTGGAAGTGCTGTTTCAGG
GTCCGAGCCACGATAGCTCGACCAATGGTCTGATTAACCGCTATAAAGCGTGGCGCGGTtaa

SEQ ID NO:15

ATGAAAAACATCAATCCAACGCAGACCGCTGCCTGGCAGGCACTACAGAAACACTTCGATGA
AATGAAAAGACGTTACGATCGCCGATCTTTTTTGCTAAAGATGGTGATCGTTTTTCTAAGTTCT
CCGCAACCTTCGACGATCAGATGCTGGTGGATTACTCCAAAAACCGCATCACTGAAGAGACG
CTGGCGAAATTACAGGATCTGGCGAAAGAGTGCGATCTGGCGGGCGCGATTAAAGTCGATGTT
CTCTGGCGAGAAGATCAACCGCACTGAAAACCGCGCCGTGCTGCACGTAGCGCTGCGTAACC
GTAGCAATACCCCGATTTTGGTTGATGGCAAAGACGTAATGCCGGAAGTCAACGCGGTGCTG
GAGAAGATGAAAACCTTCTCAGAAGCGATTATTTCCGGTGAGTGAAAGGTTATACCGGCAA
AGCAATCACTGACGTAGTGAACATCGGGATCGGCGGTTCTGACCTCGGCCCATACATGGTGA
CCGAAGCTCTGCGTCCGTACAAAAACCACCTGAACATGCACTTTGTTTCTAACGTCGATGGG
ACTCACATCGCGGAAGTGCTGAAAAAAGTAAACCCGGAAACCACGCTGTTCTTGGTAGCATC
TAAAACCTTCACCACTCAGGAAACTATGACCAACGCCCATAGCGCGCGTGACTGGTTCCTGA
AAGCGGCAGGTGATGAGAAGCACGTTGCAAAACACTTTGCGGCGCTTTCCACCAATGCCAAA
GCCGTTGGCGAGTTTGGTATTGATACTGCCAACATGTTTCGAGTTCTGGGACTGGGTTGGCGG
CCGTTACTCTTTGTGGTCAGCGATTGGCCTGTTCGATTGTTCTCTCCATCGGCTTTGATAACT
TCGTTGAACTGCTTTCCGGCGCACACGCGATGGACAAGCATTCTCCACCACGCCTGCCGAG
AAAAACCTGCCTGTACTGCTGGCGCTGATTGGCATCTGGTACAACAATTTCTTTGGTGCGGA
AACTGAAGCGATTCTGCCGTATGACCAGTATATGCACCGTTTCGCGGCGTACTTCCAGCAGG
GCAATATGGAGTCCAACGGTAAGTATGTTGACCGTAACGGTAACGTTGTGGATTACCAGACT
GGCCCGATTATCTGGGGTGAACCAGGCACTAACGGTCAGCACGCGTTCTACCAGCTGATCCA
CCAGGGAACCAAAATGGTACCGTGCGATTTTCATCGCTCCGGCTATCACCCATAACCCGCTCT
CTGATCATCACCAGAACTGCTGTCTAACTTCTTCGCCCAGACCGAAGCGCTGGCGTTTGGT
AAATCCCGCGAAGTGTTGAGCAGGAATATCGTGATCAGGGTAAAGATCCGGCAACGCTTGA
CTACGTGGTGCCGTTCAAAGTATTCTGAAGGTAAACGCCCGACCAACTCCATCCTGCTGCGTG
AAATCACTCCGTTTCAGCCTGGGTGCGTTGATTGCGCTGTATGAGCACAAAATCTTTACTCAG
GGCGTGATCCTGAACATCTTCACCTTCGACCAGTGGGGCGTGGAAGTGGGTAAACAGCTGGC
GAACCGTATTCTGCCAGAGCTGAAAGATGATAAAGAAATCAGCAGCCTGGAAGTGCTGTTTC
AGGGTCCGCACGATAGCTCGACCAATGGTCTGATTAACCGCTATAAAGCGTGGCGCGGTtaa

SEQ ID NO:16

ATGAAAAACATCAATCCAACGCAGACCGCTGCCTGGCAGGCACTACAGAAACACTTCGATGA
AATGAAAGACGTTACGATCGCCGATCTTTTTGCTAAAGATGGTGATCGTTTTTCTAAGTTCT
CCGCAACCTTCGACGATCAGATGCTGGTGGATTACTCCAAAAACCGCATCACTGAAGAGACG
CTGGCGAAATTACAGGATCTGGCGAAAGAGTGCGATCTGGCGGGCGCGATTAAAGTCGATGTT
CTCTGGCGAGAAGATCAACCGCACTGAAAACCGCGCCGTGCTGCACGTAGCGCTGCGTAACC
GTAGCAATACCCCGATTTTGGTTGATGGCAAAGACGTAATGCCGGAAGTCAACGCGGTGCTG
GAGAAGATGAAAACCTTCTCAGAAGCGATTATTTCCGGTGAGTGGAAGGTTATACCGGCAA
AGCAATCACTGACGTAGTGAACATCGGGATCGGCGGTTCTGACCTCGGCCCATACATGGTGA
CCGAAGCTCTGCGTCCGTACAAAAACCACCTGAACATGCACTTTGTTTCTAACGTCGATGGG
ACTCACATCGCGGAAGTGCTGAAAAAAGTAAACCCGGAAACCACGCTGTTCTTGGTAGCATC
TAAACCTTTCACCACTCAGGAACTATGACCAACGCCCATAGCGCGCGTGACTGGTTCCTGA
AAGCGGCAGGTGATGAGAAGCACGTTGCAAAACACTTTGCGGCGCTTTCACCAATGCCAAA
GCCGTTGGCGAGTTTGGTATTGATACTGCCAACATGTTTCGAGTTCTGGGACTGGGTTGGCGG
CCGTTACTCTTTGTGGTCAGCGATTGGCCTGTGCGATTGTTCTCTCCATCGGCTTTGATAACT
TCGTTGAACTGCTTTCCGGCGCACACGCGATGGACAAGCATTCTCTCCACCACGCCTGCCGAG
AAAAACCTGCCTGTACTGCTGGCGCTGATTGGCATCTGGTACAACAATTTCTTTGGTGCGGA
AACTGAAGCGATTCTGCCGTATGACCAGTATATGCACCGTTTCGCGGCGTACTTCCAGCAGG
GCAATATGGAGTCCAACGGTAAGTATGTTGACCGTAACGGTAACGTTGTGGATTACCAGACT
GGCCCGATTATCTGGGGTGAACCAGGCACTAACGGTCAGCACGCGTTCTACCAGCTGATCCA
CCAGGGAACCAAAATGGTACCGTGCGATTTTCATCGCTCCGGCTATCACCCATAACCCGCTCT
CTGATCATCACCAAGAACTGCTGTCTAACTTCTTCGCCCAGACCGAAGCGCTGGCGTTTGGT
AAATCCCGCGAAGTGGTTGAGCAGGAATATCGTGATCAGGGTAAAGATCCGGCAACGCTTGA
CTACGTGGTGCCGTTCAAAGTATTCGAAGGTAACCGCCGACCAACTCCATCCTGCTGCGTG
AAATCACTCCGTTTCAGCCTGGGTGCGTTGATTGCGCTGTATGAGCACAAAATCTTTACTCAG
GGCGTGATCCTGAACATCTTCACCTTCGACCAGTGGGGCGTGGAAGTGGGTAAACAGCTGGC
GAACCGTATTCTGCCAGAGCTGAAAGATGATAAAGAAATCAGCAGCCACGATAGCTCGACCA
ATGGTCTGATTAACCGCTATAAACTGGAAGTGCTGTTTCAGGGTCCGGCGTGGCGCGGTtaa

SEQ ID NO:17

MKNINPTQTAAWQALQKHFDKMDVTIADLFAKDGRFSKFSATFDDQMLVDYSKNRITEET
LAKLQDLAKECDLAGAIKSMFSGEKINRTENRAVLHVALRNRSNTPILVDGKDVMPEVNAVL
EKMKTFSIAIISGEWKGYTGKAITDVVNIGIGGSDLGPYMVTEALRPYKNHLNMHFVSNVDG
THIAEVLKKNVPETTLFLVASKTFTTQETMTNAHSARDWFLKAAGDEKHHVAKHFAALSTNAK
AVGEFGIDTANMFEFWDWVGGRYSLWSAIGLSIVLSIGFDNFVELLSGAHAMDKHFSTTPAE
KNLPVLLALIGIWYNNFFGAETEAILPYDQYMHRFAAYFQQGNMESNGKYVDRNGNVVDYQT
GPIIWGEPGTNGQHAFYQLIHQGTKMVPCDFIAPAITHNPLSDHHQKLLSNFFAQTEALAFG
KSREVVEQEYRDQGKDPATLDYVVPFKVFEGNRPTNSILLREITPFSLGALIALYEHKIFTQ
GVILNIFTFDQWGVELGKQLANRILPELKDDKEISSHDSSTNGLINRYKAWRG

SEQ ID NO:18

MKNINPTQTAAWQALQKHFDKMDVTIADLFAKDGRFSKFSATFDDQMLVDYSKNRITEET
LAKLQDLAKECDLAGAIKSMFSGEKINRTENRAVLHVALRNRSNTPLEVLFGPIILVDGKDV
MPEVNAVLEKMKTFSIAIISGEWKGYTGKAITDVVNIGIGGSDLGPYMVTEALRPYKNHLNM
HFVSNVDGTHIAEVLKKNVPETTLFLVASKTFTTQETMTNAHSARDWFLKAAGDEKHHVAKHF
AALSTNAKAVGEFGIDTANMFEFWDWVGGRYSLWSAIGLSIVLSIGFDNFVELLSGAHAMDK
HFSTTPAEKNLPVLLALIGIWYNNFFGAETEAILPYDQYMHRFAAYFQQGNMESNGKYVDRN

GNVVDYQTGP I IWGEPGTNGQHAFYQLIHQGTKMVPCDFIAPAI THNPLSDHHQKLLSNFFA
QTEALAFGKSREVVEQEYRDQGKDPATLDYVVPFKVFEGNRPTNSILLREITPFSLGALIAL
YEHKIFTQGVILNIFTFDQWGVELGKQLANRILPELKDDKEISSHDSSTNGLINRYKAWRG

SEQ ID NO:19

MKNINPTQTAAWQALQKHFDKMDVTIADLFAKDGDRFSKFSATFDDQMLVDYSKNRITEET
LAKLQDLAKECDLAGAIKSMFSGEKINRTENRAVLHVALRNRSNTPILEVLFQGPLVLDGKDV
MPEVNAVLEKMKTFSEAIISGEWKGYTGKAITDVVNIGIGGSDLGPYMVTEALRPYKNHLNM
HFVSNVDGTHIAEVLKKNVPETTFLVASKTFTTQETMTNAHSARDWFLKAAGDEKHHVAKHF
AALSTNAKAVGEFGIDTANMFEFWDWVGGRYSLWSAIGLSIVLSIGFDNFVELLSGAHAMDK
HFSTTPAEKNLPVLLALIGIWYNNFFGAETEAILPYDQYMHRFAAYFQQGNMESNGKYVDRN
GNVVDYQTGP I IWGEPGTNGQHAFYQLIHQGTKMVPCDFIAPAI THNPLSDHHQKLLSNFFA
QTEALAFGKSREVVEQEYRDQGKDPATLDYVVPFKVFEGNRPTNSILLREITPFSLGALIAL
YEHKIFTQGVILNIFTFDQWGVELGKQLANRILPELKDDKEISSHDSSTNGLINRYKAWRG

SEQ ID NO:20

MKNINPTQTAAWQALQKHFDKMDVTIADLFAKDGDRFSKFSATFDDQMLVDYSKNRITEET
LAKLQDLAKECDLAGAIKSMFSGEKINRTENRAVLHVALRNRSNTPILEVLFQGPVDGKDVM
PEVNAVLEKMKTFSEAIISGEWKGYTGKAITDVVNIGIGGSDLGPYMVTEALRPYKNHLNMH
FVSNVDGTHIAEVLKKNVPETTFLVASKTFTTQETMTNAHSARDWFLKAAGDEKHHVAKHFA
ALSTNAKAVGEFGIDTANMFEFWDWVGGRYSLWSAIGLSIVLSIGFDNFVELLSGAHAMDKH
FSTTPAEKNLPVLLALIGIWYNNFFGAETEAILPYDQYMHRFAAYFQQGNMESNGKYVDRNG
NVVDYQTGP I IWGEPGTNGQHAFYQLIHQGTKMVPCDFIAPAI THNPLSDHHQKLLSNFFAQ
TEALAFGKSREVVEQEYRDQGKDPATLDYVVPFKVFEGNRPTNSILLREITPFSLGALIALY
EHKIFTQGVILNIFTFDQWGVELGKQLANRILPELKDDKEISSHDSSTNGLINRYKAWRG

SEQ ID NO:21

MKNINPTQTAAWQALQKHFDKMDVTIADLFAKDGDRFSKFSATFDDQMLVDYSKNRITEET
LAKLQDLAKECDLAGAIKSMFSGEKINRTENRAVLHVALRNRSNTPILVDGKDVMPEVNAV
LEKMKTFSEAIISGELEVLFQGPWKGYTGKAITDVVNIGIGGSDLGPYMVTEALRPYKNHLNM
HFVSNVDGTHIAEVLKKNVPETTFLVASKTFTTQETMTNAHSARDWFLKAAGDEKHHVAKHF
AALSTNAKAVGEFGIDTANMFEFWDWVGGRYSLWSAIGLSIVLSIGFDNFVELLSGAHAMDK
HFSTTPAEKNLPVLLALIGIWYNNFFGAETEAILPYDQYMHRFAAYFQQGNMESNGKYVDRN
GNVVDYQTGP I IWGEPGTNGQHAFYQLIHQGTKMVPCDFIAPAI THNPLSDHHQKLLSNFFA
QTEALAFGKSREVVEQEYRDQGKDPATLDYVVPFKVFEGNRPTNSILLREITPFSLGALIAL
YEHKIFTQGVILNIFTFDQWGVELGKQLANRILPELKDDKEISSHDSSTNGLINRYKAWRG

SEQ ID NO:22

MKNINPTQTAAWQALQKHFDKMDVTIADLFAKDGDRFSKFSATFDDQMLVDYSKNRITEET
LAKLQDLAKECDLAGAIKSMFSGEKINRTENRAVLHVALRNRSNTPILVDGKDVMPEVNAV
LEKMKTFSEAIISGEWKGYTGKAITDVVNIGIGGSDLGPYMVTEALRPYKNHLNMHFVSNVDG
THIAEVLKKNVPETTFLVASKTFTTQETMTNAHSARDWFLKAAGDEKHHVAKHFAALSTNAK
AVGEFGIDTANMFEFWDWVGGRYSLWSAIGLSIVLSIGFDNFVELLSGAHAMDKHFSTTPAE
KNLPVLLALIGIWYNNFFGAETEAILPYDQYMHRFAAYFQQGNMESNGKYVDRNGNVVDYQT
GP I IWGEPGTNGQHAFYQLIHQGTKMVPCDFIAPAI THLEVLFQGPNPLSDHHQKLLSNFFA

QTEALAFGKSREVVEQEYRDQGKDPATLDYVVPFKVFEGNRPTNSILLREITPFSLGALIAL
YEHKIFTQGVLNIFTFDQWGVELGKQLANRILPELKDDKEISSHDSSTNGLINRYKAWRG

SEQ ID NO:23

MKNINPTQTAAWQALQKHFDKMDVTIADLFAKDGDRFSKFSATFDDQMLVDYSKNRITEET
LAKLQDLAKECDLAGAIKSMFSGEKINRTENRAVLHVALRNRSNTPILVDGKDVMPEVNAVL
EKMKTFSIAISGEWKGYTGKAITDVVNIGIGGSDLGPYMVTEALRPYKNHLNMHFVSNVDG
THIAEVLKKVNPETTFLVASKTFTTQETMTNAHSARDWFLKAAGDEKHHVAKHFAALSTNAK
AVGEFGIDTANMFEFWDWVGGRYSLWSAIGLSIVLSIGFDNFVELLSGAHAMDKHFSTTPAE
KNLPVLLALIGIWYNNFFGAETAILPYDQYMHRFAAYFQQGNMESNGKYVDRNGNVVDYQT
GPIIWGEPGTNGQHAFYQLIHQGTKMVPCDFIAPAITHNPLSDHHQKLLSNFFAQTEALAFG
KSREVVEQEYRDQGKDPATLDYVVPFKVFEGNRPTNSILLREITPFSLGALIALYEHKIFTQ
GVILNIFTFDQWGVELGKQLANRILPELEVLFQGPKDDKEISSHDSSTNGLINRYKAWRG

SEQ ID NO:24

MKNINPTQTAAWQALQKHFDKMDVTIADLFAKDGDRFSKFSATFDDQMLVDYSKNRITEET
LAKLQDLAKECDLAGAIKSMFSGEKINRTENRAVLHVALRNRSNTPILVDGKDVMPEVNAVL
EKMKTFSIAISGEWKGYTGKAITDVVNIGIGGSDLGPYMVTEALRPYKNHLNMHFVSNVDG
THIAEVLKKVNPETTFLVASKTFTTQETMTNAHSARDWFLKAAGDEKHHVAKHFAALSTNAK
AVGEFGIDTANMFEFWDWVGGRYSLWSAIGLSIVLSIGFDNFVELLSGAHAMDKHFSTTPAE
KNLPVLLALIGIWYNNFFGAETAILPYDQYMHRFAAYFQQGNMESNGKYVDRNGNVVDYQT
GPIIWGEPGTNGQHAFYQLIHQGTKMVPCDFIAPAITHNPLSDHHQKLLSNFFAQTEALAFG
KSREVVEQEYRDQGKDPATLDYVVPFKVFEGNRPTNSILLREITPFSLGALIALYEHKIFTQ
GVILNIFTFDQWGVELGKQLANRILPELKLEVLFQGPDDKEISSHDSSTNGLINRYKAWRG

SEQ ID NO:25

MKNINPTQTAAWQALQKHFDKMDVTIADLFAKDGDRFSKFSATFDDQMLVDYSKNRITEET
LAKLQDLAKECDLAGAIKSMFSGEKINRTENRAVLHVALRNRSNTPILVDGKDVMPEVNAVL
EKMKTFSIAISGEWKGYTGKAITDVVNIGIGGSDLGPYMVTEALRPYKNHLNMHFVSNVDG
THIAEVLKKVNPETTFLVASKTFTTQETMTNAHSARDWFLKAAGDEKHHVAKHFAALSTNAK
AVGEFGIDTANMFEFWDWVGGRYSLWSAIGLSIVLSIGFDNFVELLSGAHAMDKHFSTTPAE
KNLPVLLALIGIWYNNFFGAETAILPYDQYMHRFAAYFQQGNMESNGKYVDRNGNVVDYQT
GPIIWGEPGTNGQHAFYQLIHQGTKMVPCDFIAPAITHNPLSDHHQKLLSNFFAQTEALAFG
KSREVVEQEYRDQGKDPATLDYVVPFKVFEGNRPTNSILLREITPFSLGALIALYEHKIFTQ
GVILNIFTFDQWGVELGKQLANRILPELKDLEVLFQGPDKEISSHDSSTNGLINRYKAWRG

SEQ ID NO:26

MKNINPTQTAAWQALQKHFDKMDVTIADLFAKDGDRFSKFSATFDDQMLVDYSKNRITEET
LAKLQDLAKECDLAGAIKSMFSGEKINRTENRAVLHVALRNRSNTPILVDGKDVMPEVNAVL
EKMKTFSIAISGEWKGYTGKAITDVVNIGIGGSDLGPYMVTEALRPYKNHLNMHFVSNVDG
THIAEVLKKVNPETTFLVASKTFTTQETMTNAHSARDWFLKAAGDEKHHVAKHFAALSTNAK
AVGEFGIDTANMFEFWDWVGGRYSLWSAIGLSIVLSIGFDNFVELLSGAHAMDKHFSTTPAE
KNLPVLLALIGIWYNNFFGAETAILPYDQYMHRFAAYFQQGNMESNGKYVDRNGNVVDYQT
GPIIWGEPGTNGQHAFYQLIHQGTKMVPCDFIAPAITHNPLSDHHQKLLSNFFAQTEALAFG
KSREVVEQEYRDQGKDPATLDYVVPFKVFEGNRPTNSILLREITPFSLGALIALYEHKIFTQ
GVILNIFTFDQWGVELGKQLANRILPELKDDLEVLFQGPKKEISSHDSSTNGLINRYKAWRG

SEQ ID NO:27

MKNINPTQTAAWQALQKHFDKMDVTIADLFAKDGDRFSKFSATFDDQMLVDYSKNRITEET
LAKLQDLAKECDLAGAIKSMFSGEKINRTENRAVLHVALRNRSNTPILVDGKDVMPEVNAVL
EKMKTFSIAIISGEWKGYTGKAITDVVNIGIGGSDLGPYMVTEALRPYKNHLNMHFVSNVDG
THIAEVLKKVNPETTLFLVASKTFTTQETMTNAHSARDWFLKAAGDEKHHVAKHFAALSTNAK
AVGEFGIDTANMFEFWDWVGGRYSLWSAIGLSIVLSIGFDNFVELLSGAHAMDKHFSTTPAE
KNLPVLLALIGIWINNFFGAETAILPYDQYMHRFAAYFQQGNMESNGKYVDRNGNVVDYQT
GPIIWGEPGTNGQHAFYQLIHQGTKMVPCDFIAPAITHNPLSDHHQKLLSNFFAQTEALAFG
KSREVVEQEYRDQGKDPATLDYVVPFKVFEGNRPTNSILLREITPFSLGALIALYEHKIFTQ
GVILNIFTFDQWGVELGKQLANRILPELKDDKLEVLFGQPEISSHDSSTNGLINRYKAWRG

SEQ ID NO:28

MKNINPTQTAAWQALQKHFDKMDVTIADLFAKDGDRFSKFSATFDDQMLVDYSKNRITEET
LAKLQDLAKECDLAGAIKSMFSGEKINRTENRAVLHVALRNRSNTPILVDGKDVMPEVNAVL
EKMKTFSIAIISGEWKGYTGKAITDVVNIGIGGSDLGPYMVTEALRPYKNHLNMHFVSNVDG
THIAEVLKKVNPETTLFLVASKTFTTQETMTNAHSARDWFLKAAGDEKHHVAKHFAALSTNAK
AVGEFGIDTANMFEFWDWVGGRYSLWSAIGLSIVLSIGFDNFVELLSGAHAMDKHFSTTPAE
KNLPVLLALIGIWINNFFGAETAILPYDQYMHRFAAYFQQGNMESNGKYVDRNGNVVDYQT
GPIIWGEPGTNGQHAFYQLIHQGTKMVPCDFIAPAITHNPLSDHHQKLLSNFFAQTEALAFG
KSREVVEQEYRDQGKDPATLDYVVPFKVFEGNRPTNSILLREITPFSLGALIALYEHKIFTQ
GVILNIFTFDQWGVELGKQLANRILPELKDDKLEVLFGQPIISSHDSSTNGLINRYKAWRG

SEQ ID NO:29

MKNINPTQTAAWQALQKHFDKMDVTIADLFAKDGDRFSKFSATFDDQMLVDYSKNRITEET
LAKLQDLAKECDLAGAIKSMFSGEKINRTENRAVLHVALRNRSNTPILVDGKDVMPEVNAVL
EKMKTFSIAIISGEWKGYTGKAITDVVNIGIGGSDLGPYMVTEALRPYKNHLNMHFVSNVDG
THIAEVLKKVNPETTLFLVASKTFTTQETMTNAHSARDWFLKAAGDEKHHVAKHFAALSTNAK
AVGEFGIDTANMFEFWDWVGGRYSLWSAIGLSIVLSIGFDNFVELLSGAHAMDKHFSTTPAE
KNLPVLLALIGIWINNFFGAETAILPYDQYMHRFAAYFQQGNMESNGKYVDRNGNVVDYQT
GPIIWGEPGTNGQHAFYQLIHQGTKMVPCDFIAPAITHNPLSDHHQKLLSNFFAQTEALAFG
KSREVVEQEYRDQGKDPATLDYVVPFKVFEGNRPTNSILLREITPFSLGALIALYEHKIFTQ
GVILNIFTFDQWGVELGKQLANRILPELKDDKEILEVLFGQPSHDSSTNGLINRYKAWRG

SEQ ID NO:30

MKNINPTQTAAWQALQKHFDKMDVTIADLFAKDGDRFSKFSATFDDQMLVDYSKNRITEET
LAKLQDLAKECDLAGAIKSMFSGEKINRTENRAVLHVALRNRSNTPILVDGKDVMPEVNAVL
EKMKTFSIAIISGEWKGYTGKAITDVVNIGIGGSDLGPYMVTEALRPYKNHLNMHFVSNVDG
THIAEVLKKVNPETTLFLVASKTFTTQETMTNAHSARDWFLKAAGDEKHHVAKHFAALSTNAK
AVGEFGIDTANMFEFWDWVGGRYSLWSAIGLSIVLSIGFDNFVELLSGAHAMDKHFSTTPAE
KNLPVLLALIGIWINNFFGAETAILPYDQYMHRFAAYFQQGNMESNGKYVDRNGNVVDYQT
GPIIWGEPGTNGQHAFYQLIHQGTKMVPCDFIAPAITHNPLSDHHQKLLSNFFAQTEALAFG
KSREVVEQEYRDQGKDPATLDYVVPFKVFEGNRPTNSILLREITPFSLGALIALYEHKIFTQ
GVILNIFTFDQWGVELGKQLANRILPELKDDKEISLEVLFGQPSHDSSTNGLINRYKAWRG

SEQ ID NO:31

MKNINPTQTAAWQALQKHFDKMDVTIADLFAKDGDRFSKFSATFDDQMLVDYSKNRITEET
LAKLQDLAKECDLAGAIKSMFSGEKINRTENRAVLHVALRNRSNTPILVDGKDVMPEVNAVL
EKMKTFSIAIISGEWKGYTGKAITDVVNIGIGGSDLGPYMVTEALRPYKNHLNMHFVSNVDG
THIAEVLKKVNPETTFLVASKTFTTQETMTNAHSARDWFLKAAGDEKHHVAKHFAALSTNAK
AVGEFGIDTANMFEFWDWVGGRYSLWSAIGLSIVLSIGFDNFVELLSGAHAMDKHFSTTPAE
KNLPVLLALIGIWYNNFFGAETAILPYDQYMHRFAAYFQQGNMESNGKYVDRNGNVVDYQT
GPIIWGEPGTNGQHAFYQLIHQGTKMVPCDFIAPAITHNPLSDHHQKLLSNFFAQTEALAFG
KSREVVEQEYRDQGKDPATLDYVVPFKVFEGNRPTNSILLREITPFSLGALIALYEHKIFTQ
GVILNIFTFDQWGVELGKQLANRILPELKDDKEISSLEVLFQGPHDSSSTNGLINRYKAWRG

SEQ ID NO:32

MKNINPTQTAAWQALQKHFDKMDVTIADLFAKDGDRFSKFSATFDDQMLVDYSKNRITEET
LAKLQDLAKECDLAGAIKSMFSGEKINRTENRAVLHVALRNRSNTPILVDGKDVMPEVNAVL
EKMKTFSIAIISGEWKGYTGKAITDVVNIGIGGSDLGPYMVTEALRPYKNHLNMHFVSNVDG
THIAEVLKKVNPETTFLVASKTFTTQETMTNAHSARDWFLKAAGDEKHHVAKHFAALSTNAK
AVGEFGIDTANMFEFWDWVGGRYSLWSAIGLSIVLSIGFDNFVELLSGAHAMDKHFSTTPAE
KNLPVLLALIGIWYNNFFGAETAILPYDQYMHRFAAYFQQGNMESNGKYVDRNGNVVDYQT
GPIIWGEPGTNGQHAFYQLIHQGTKMVPCDFIAPAITHNPLSDHHQKLLSNFFAQTEALAFG
KSREVVEQEYRDQGKDPATLDYVVPFKVFEGNRPTNSILLREITPFSLGALIALYEHKIFTQ
GVILNIFTFDQWGVELGKQLANRILPELKDDKEISSHDSSSTNGLINRYKLEVLFQGPAWRG

SEQ ID NO:33

ATGGGCCCCAGAAGAAGAATTTCGGCATGAGCCTGATCAAGCATAACTCTTGCGTCATTACCAC
GGAGAATGGTAAGTTCACGGGCTTGGGCGTTTATGACCGTTTCGTCGTGGTTCCGACCCACG
CTGACCCGGGTAAAGAAATCCAGGTTGACGGTATCACGACCAAAGTGATTGATAGCTATGAT
CTCTATAATAAGAACGGCATCAAGCTGGAAATCACGGTGCTGAAACTGGACCGTAATGAAAA
GTTTCGTGATATCCGTCGCTATATTCCGAATAACGAGGATGACTACCCAAATTGCAATCTGG
CGCTGCTGGCAAATCAGCCGGAACCGACGATCATCAACGTGGGTGACGTGGTGAGCTATGGC
AATATCCTGCTGAGCGGTAACCGACCGCGCGTATGCTGAAGTATTCCTATCCGACGAAAAG
CGGCTATTGCGGCGGCGTGCTCTATAAGATTGGTCAAGTCCTGGGCATCCACGTCGGCGGTA
ATGGCCGCGATGGTTTCAGCGCGATGCTGCTGCGTAGCTATTTACCGACGTCCAGTGATAA

SEQ ID NO:34

MGPEEEFGMSLIKHNSCVITTENGKFTGLGVYDRFVVVPTHADPGKEIQVDGITTVIDSYD
LYNKNGIKLEITVLKLDRENEKFRDIRRYIPNNEDDYPNCNLALLANQPEPTIINVGDVVSYG
NILLSGNQTARMLKYSYPTKSGYCGGVLYKIGQVLGIHVGGNGRDGFSAMLLRSYFTDVQ

SEQ ID NO:35

ATGAAAAAACGGCAATTGCGATAGCGGTTGCGCTAGCTGGTTTTGCCACGGTGGCGCAGGC
TGGCCCAGAAGAAGAATTTCGGCATGAGCCTGATCAAGCATAACTCTTGCGTCATTACCACGG
AGAATGGTAAGTTCACGGGCTTGGGCGTTTATGACCGTTTCGTCGTGGTTCCGACCCACGCT
GACCCGGGTAAAGAAATCCAGGTTGACGGTATCACGACCAAAGTGATTGATAGCTATGATCT
CTATAATAAGAACGGCATCAAGCTGGAAATCACGGTGCTGAAACTGGACCGTAATGAAAAGT
TTCGTGATATCCGTCGCTATATTCCGAATAACGAGGATGACTACCCAAATTGCAATCTGGCG

CTGCTGGCAAATCAGCCGGAACCGACGATCATCAACGTGGGTGACGTGGTGAGCTATGGCAA
TATCCTGCTGAGCGGTAACCAGACCGCGCGTATGCTGAAGTATTCCTATCCGACGAAAAGCG
GCTATTGCGGCGGCGTGCTCTATAAGATTGGTCAAGTCCTGGGCATCCACGTCGGCGGTAAT
GGCCGCGATGGTTTCAGCGCGATGCTGCTGCGTAGCTATTTACCGACGTCCAGTGATAA

SEQ ID NO:36

MKKTAAIAAVALAGFATVAQAGPEEEFGMSLIKHNSCVITTENGKFTGLGVYDRFVVVPTHA
DPGKEIQVDGITTVIDSYDLYNKNGIKLEITVLKLDNRNEKFRDIRRYIPNNEDDYPNCNLA
LLANQPEPTIINVGDVVSIGNILLSGNQATARMLKYSYPTKSGYCGGVLYKIGQVLGIHVGGN
GRDGFSAMLLRSYFTDVQ

SEQ ID NO:37

CTGGAAGTGCTGTTTCAGGGTCCG

SEQ ID NO:38

LEVLFQGP

SEQ ID NO:39

EVLFQGP

SEQ ID NO:40

VLFQGP

SEQ ID NO:41

LFQGP

SEQ ID NO:42

LEVLFQG

SEQ ID NO:43

LEVLFQ

SEQ ID NO:44

LEVLF

SEQ ID NO:45

EVLFQG

SEQ ID NO:46

MKKTAIAIAVALAGFATVAQA

SEQ ID NO:47

GTGTCCCGTATTATTATGCTGATCCCTACCGGAACCAGCGTCGGTCTGACCAGCGTCAGCCT
TGGCGTGATCCGTGCAATGGAACGCAAAGGCGTTTCGTCTGAGCGTTTTCAAACCTATCGCTC
AGCCGCGTACCGGTGGCGATGCGCCCGATCAGACTACGACTATCGTGCGTGCGAACTCTTCC
ACCACGACGGCCGCTGAACCGCTGAAAATGAGCTACGTTGAAGGTCTGCTTTCCAGCAATCA
GAAAGATGTGCTGATGGAAGAGATCGTCGAAACTACCACGCTAACACCAAAGACGCTGAAG
TCGTTCTGGTTGAAGGTCTGGTCCCGACACGTAAGCACCAGTTTGCCCAGTCTCTGAACTAC
GAAATCGCTAAAACGCTGAATGCGGAAATCGTCTTCGTTATGTCTCAGGGCACTGACACCCC
GGAACAGCTGAAAGAGCGTATCGAACTGACCCGCAACAGCTTCGGCGGTGCCAAAAACACCA
ACATCACCGGCGTTATCGTTAACAACTGAACGCACCGGTTGATGAACAGGGTCGTACTCGC
CCGGATCTGTCCGAGATTTTCGACGACTCTTCCAAAGCTAAAGTAAACAATGTTGATCCGGC
GAAGCTGCAAGAATCCAGCCCGCTGCCGGTTCTCGGCGCTGTGCCGTGGAGCTTTGACCTGA
TCGCGACTCGTGCGATCGATATGGCTCGCCACCTGAATGCGACCATCATCAACGAAGGCGAC
ATCAATACTCGCCGCGTTAAATCCGTCACCTTTCTGCGCACGCAGCATTCCGCACATGCTGGA
GCACTTCCGTGCCGGTTCTCTGCTGGTGACTTCCGCAGACCGTCTTGACGTGCTGGTGGCCG
CTTGCCCTGGCAGCCATGAACGGCGTAGAAATCGGTGCCCTGCTGCTGACTGGCGGTTACGAA
ATGGACGCGCGCATTTCTAAACTGTGCGAACGTGCTTTCGCTACCGGCCTGCCGGTATTTAT
GGTGAACACCAACACCTGGCAGACCTCTCTGAGCCTGCAGAGCTTCAACCTGGAAGTTCCGG
TTGACGATCACGAACGTATCGAGAAAGTTTCAGGAATACGTTGCTAACTACATCAACGCTGAC
TGGATCGAATCTCTGACTGCCACTTCTGAGCGCAGCCGTCGTCTGTCTCCGCCTGCGTTCGG
TTATCAGCTGACTGAACTTGCGCGCAAAGCGGGCAAACGTATCGTACTGCCGGAAGGTGACG
AACC GCGTACCGTTAAAGCAGCCGCTATCTGTGCTGAACGTGGTATCGCAACTTGCGTACTG
CTGGGTAAATCCGGCAGAGATCAACCGTGTTGCAGCGTCTCAGGGTGTAGA ACTGGGTGCAGG
GATTGAAATCGTTGATCCAGAAGTGGTTCGCGAAAGCTATGTTGGTCGTCTGGTCGAACTGC
GTAAGAACAAAGGCATGACCGAAACCGTTGCCCGCGAACAGCTGGAAGACAACGTGGTGCTC
GGTACGCTGATGCTGGAACAGGATGAAGTTGATGGTCTGGTTTCCGGTGCTGTTACACTAC
CGCAAACACCATCCGTCCGCCGCTGCAGCTGATCAAACTGCACCGGGCAGCTCCCTGGTAT
CTTCCGTGTTCTTCATGCTGCTGCCGGAACAGGTTTACGTTTACGGTGACTGTGCGATCAAC
CCGGATCCGACCGCTGAACAGCTGGCAGAAATCGCGATTTCAGTCCGCTGATTCCGCTGCGGC
CTTCGGTATCGAACCGCGCGTTGCTATGCTCTCCTACTCCACCGGTACTTCTGGTGCAGGTA
GCGACGTAGAAAAAGTTTCGCGAAGCAACTCGTCTGGCGCAGGAAAAACGTCCTGACCTGATG
ATCGACGGTCCGCTGCAGTACGACGCTGCGGTAATGGCTGACGTTGCGAAATCCAAAGCGCC
GAACTCTCCGGTTGCAGGTGCGGCTACCGTGTTTCATCTTCCCGGATCTGAACACCGGTAACA
CCACCTACAAAGCGGTACAGCGTTCTGCCGACCTGATCTCCATCGGGCCGATGCTGCAGGGT
ATGCGCAAGCCGGTTAACGACCTGTCCCGTGCGCACTGGTTGACGATATCGTCTACACCAT
CGCGCTGACTGCGATTTCAGTCTGCACAGCAGCAGTAA

SEQ ID NO:48

VSRIIMLIPTGTSVGLTSVSLGVIRAMERKGVRLSVFKPIAQPRTG DAPDQTTTIVRANSS
TTTAAEPLKMSYVEGLLSSNQKDVLMEEIVANYHANTKDAEVVLVEGLVPTRKHQFAQSLNY
EIAKTLNAEIVFVMSQGTDTPEQLKERIELTRNSFGGAKNTNITGVIVNKL NAPVDEQGRTR
PDLSEIFDDSSKAKVNNVDPKQLQESSPLPVLGAVPWSFDLIATRAIDMARHLNATIINEGD
INTRRVKSVTFCARSI PHMLEHFRAGSLLVTSADRPDVLVAACLAAMNGVEIGALLLTGGYE
MDARISKLCERAFATGLPVFMVNTNTWQTSLSLQSFNLEVPVDDHERIEKVQEYVANYINAD

WIESLTATSESRRLSPPAFRYQLTELARKAGKRIVLPEGDEPRTVKAAAICAERGIATCVL
LGNPAEINRVAASQGVELGAGIEIVDPEVVRESYVGRVLVELRKNKGMTETVAREQLEDNVVL
GTLMLEQDEVDGLVSGAVHTTANTIRPLQLIKTAPGSSLVSSVFFMLLPEQVYVYGDCAIN
PDPTAEQLAEIAIQSADSAAAFGIEPRVAMLSYSTGTSGAGSDVEKVVREATRLAQEKRPDLM
IDGPLQYDAAVMADVAKSKAPNSPVAGRATVFIIPDLNTGNTTYKAVQRSADLISIGPMLQG
MRKPVNDLSRGALVDDIVYTTIALTAIQSAQQQ

SEQ ID NO:49

GTGTCCCGTATTATTATGCTGATCCCTACCGGAACCAGCGTCGGTCTGACCAGCGTCAGCCT
TGGCGTGATCCGTGCAATGGAACGCAAAGGCGTTTCGTCTGAGCGTTTTCAAACCTATCGCTC
AGCCGCGTACCGGTGGCGATGCGCCCGATCAGACTACGACTATCGTGCGTGCGAACTCTTCC
ACCACGACGGCCGCTGAACCGCTGAAAATGAGCTACGTTGAAGGTCTGCTTTCCAGCAATCA
GAAAGATGTGCTGATGGAAGAGATCGTCGAAACTACCACGCTAACACCAAAGACGCTGAAG
TCGTTCTGGTTGAAGGTCTGGTCCCGACACGTAAGCACCAGTTTGCCCAGTCTCTGAACTAC
GAAATCGCTAAAACGCTGAATGCGGAAATCGTCTTCGTTATGTCTCAGGGCACTGACACCCC
GGAACAGCTGAAAGAGCGTATCGAACTGACCCGCAACAGCTTCGGCGGTGCCAAAAACACCA
ACATCACCGGCGTTATCGTTAACAACTGAACGCACCGGTTGATGAACAGGGTCGTACTCGC
CCGGATCTGTCCGAGATTTTCGACGACTCTTCCAAAGCTAAAGTAAACAATGTTGATCCGGC
GAAGCTGCAAGAATCCAGCCCGCTGCCGGTTCTCGGCGCTGTGCCGTGGAGCTTTGACCTGA
TCGCGACTCGTGCGATCGATATGGCTCGCCACCTGAATGCGACCATCATCAACGAAGGCGAC
ATCAATACTCGCCGCGTTAAATCCGTCACTTTCTGCGCACGCAGCATTCCGCACATGCTGGA
GCACTTCCGTGCCGGTTCTCTGCTGGTGACTTCCGCAGACCGTCCTGACGTGCTGGTGGCCG
CTTGCCGTGGCAGCCATGAACGGCGTAGAAATCGGTGCCCTGCTGCTGACTGGCGGTTACGAA
ATGGACGCGCGCATTTCTAAACTGTGCGAACGTGCTTTTCGCTACCGGCCTGCCGGTATTTAT
GGTGAACACCAACACCTGGCAGACCTCTCTGAGCCTGCAGAGCTTCAACCTGGAAGTTCCGG
TTGACGATCACGAACGTATCGAGAAAGTTTCAGGAATACGTTGCTAACTACATCAACGCTGAC
TGGATCGAATCTCTGACTGCCACTTCTCTGGAAGTGCTGTTTCAGGGTCCGGAGCGCAGCCG
TCGTCTGTCTCCGCCTGCGTTCCGTTATCAGCTGACTGAACTTGCGCGCAAAGCGGGCAAAC
GTATCGTACTGCCGGAAGGTGACGAACCGCGTACCGTTAAAGCAGCCGCTATCTGTGCTGAA
CGTGGTATCGCAACTTGCGTACTGCTGGGTAATCCGGCAGAGATCAACCGTGTTGCAGCGTC
TCAGGGTGTAGAACTGGGTGCAGGGATTGAAATCGTTGATCCAGAAGTGTTTCGCGAAAGCT
ATGTTGGTCGTCTGGTCAACTGCGTAAGAACAAAGGCATGACCGAAACCGTTGCCCGCGAA
CAGCTGGAAGACAACGTGGTGCTCGGTACGCTGATGCTGGAACAGGATGAAGTTGATGGTCT
GGTTTCCGGTGCTGTTTACACTACCGCAAACACCATCCGTCCGCCGCTGCAGCTGATCAAAA
CTGCACCGGGCAGCTCCCTGGTATCTTCCGTGTTCTTCATGCTGCTGCCGGAACAGGTTTAC
GTTTACGGTGACTGTGCGATCAACCCGGATCCGACCGCTGAACAGCTGGCAGAAATCGCGAT
TCAGTCCGCTGATTCCGCTGCGGCCTTCGGTATCGAACCGCGCGTTGCTATGCTCTCCTACT
CCACCGGTACTTCTGGTGCAGGTAGCGACGTAGAAAAAGTTCGCGAAGCAACTCGTCTGGCG
CAGGAAAAACGTCCTGACCTGATGATCGACGGTCCGCTGCAGTACGACGCTGCGGTAATGGC
TGACGTTGCGAAATCCAAAGCGCCGAACTCTCCGGTTGCAGGTCGCGCTACCGTGTTTCATCT
TCCCGGATCTGAACACCGGTAACACCACCTACAAAGCGGTACAGCGTTCTGCCGACCTGATC
TCCATCGGGCCGATGCTGCAGGGTATGCGCAAGCCGGTTAACGACCTGTCCCGTGGCGCACT
GGTTGACGATATCGTCTACACCATCGCGCTGACTGCGATTGAGTCTGCACAGCAGCAGTAA

SEQ ID NO:50

VSRIIMLIPTGTSVGLTSVSLGVIRAMERKGVRLSVFKPIAQPRTGGDAPDQTTTIVRANSS
TTTAAEPLKMSYVEGLLSSNQKDVLMEEIVANYHANTKDAEVVLVEGLVPTRKHQFAQSLNY
EIAKTLNAEIVFVMSQGTDTPEQLKERIELTRNSFGGAKNTNITGVIIVNKLNAFVDEQGRTR

PDLSEIFDDSSKAKVNNVDPKQLQESSPLPVLGAVPWSFDLIATRAIDMARHLNATIINEGD
INTRRVKSVTFCARSIPHMLEHFRAAGSLLVTSADRPDVLVAACLAAMNGVEIGALLLTGGYE
MDARISKLCERAFATGLPVFMVNTNTWQTSLSLQSFNLEVPVDDHERIEKVQEYVANYINAD
WIESLTATSLEVLFFQGPERSRRLSPPAFRYQLTELARKAGKRIVLPEGDEPRTVKAAAICAE
RGIATCVLLGNPAEINRVAASQGVELGAGIEIVDPEVVRESYVGRLEVELRKNKGMTETVARE
QLEDNVVLGTLMLEQDEVDGLVSGAVHTTANTIRPPLQLIKTAPGSSSLVSSVFFMLLPEQVY
VYGDCAINPDPTAEQLAEIAIQSADSAAAFGIEPRVAMLSYSTGTSGAGSDVEKVREATRLA
QEKRPDLMDIGPLQYDAVMADVAKSKAPNSPVAGRATVFIKPDNLNTGNTTYKAVQRSADLI
SIGPMLQGMKRPVNDLSRGALVDDIVYTIALTAIQSAQQQ

SEQ ID NO:51

GTGTCCCGTATTATTATGCTGATCCCTACCGGAACCAGCGTCGGTCTGACCAGCGTCAGCCT
TGGCGTGATCCGTGCAATGGAACGCAAAGGCGTTTCGTCTGAGCGTTTTCAAACCTATCGCTC
AGCCGCGTACCGGTGGCGATGCGCCCGATCAGACTACGACTATCGTGCGTGCGAACTCTTCC
ACCACGACGGCCGCTGAACCGCTGAAAATGAGCTACGTTGAAGGTCTGCTTTCCAGCAATCA
GAAAGATGTGCTGATGGAAGAGATCGTCGCAAACCTACCACGCTAACACCAAAGACGCTGAAG
TCGTTCTGGTTGAAGGTCTGGTCCCGACACGTAAGCACCAGTTTGCCAGTCTCTGAACTAC
GAAATCGCTAAAACGCTGAATGCGGAAATCGTCTTCGTTATGTCTCAGGGCACTGACACCCC
GGAACAGCTGAAAGAGCGTATCGAACTGACCCGCAACAGCTTCGGCGGTGCCAAAAACACCA
ACATCACCGGCGTTATCGTTAACAACCTGAACGCACCGGTTGATGAACAGGGTCTGACTCGC
CCGGATCTGTCCGAGATTTTCGACGACTCTTCCAAAGCTAAAGTAAACAATGTTGATCCGGC
GAAGCTGCAAGAATCCAGCCCGCTGCCGGTTCTCGGCGCTGTGCCGTGGAGCTTTGACCTGA
TCGCGACTCGTGCGATCGATATGGCTCGCCACCTGAATGCGACCATCATCAACGAAGGCGAC
ATCAATACTCGCCGCGTTAAATCCGTCACCTTCTGCGCACGCAGCATTCCGCACATGCTGGA
GCACTTCCGTGCCGGTTCTCTGCTGGTGAATTCGCGAGACCGTCTGACGTGCTGGTGGCCG
CTTGCTTGGCAGCCATGAACGGCGTAGAAATCGGTGCCCTGCTGCTGACTGGCGGTTACGAA
ATGGACGCGCGCATTTCTAAACTGTGCGAACGTGCTTTCTGCTACCGGCCTGCCGGTATTTAT
GGTGAACACCAACACCTGGCAGACCTCTCTGAGCCTGCAGAGCTTCAACCTGGAAGTTCCGG
TTGACGATCACGAACGTATCGAGAAAGTTTACGGAATACGTTGCTAACTACATCAACGCTGAC
TGGATCGAATCTCTGACTGCCACTTCTGAGCTGGAAGTGCTGTTTCAGGGTCCGCGCAGCCG
TCGTCTGTCTCCGCTGCGTTCCGTTATCAGCTGACTGAACTTGCGCGCAAAGCGGGCAAAC
GTATCGTACTGCCGGAAGGTGACGAACCGCGTACCGTTAAAGCAGCCGCTATCTGTGCTGAA
CGTGGTATCGCAACTTGCCTACTGCTGGGTAATCCGGCAGAGATCAACCGTGTTGCAGCGTC
TCAGGGTGTAGAATGGGTGCAGGGATTGAAATCGTTGATCCAGAAGTGTTTCGCGAAAGCT
ATGTTGGTCTGCTGGTTCGAACTGCGTAAGAACAAGGCATGACCGAAACCGTTGCCCGCGAA
CAGCTGGAAGACAACGTGGTGCTCGGTACGCTGATGCTGGAACAGGATGAAGTTGATGGTCT
GGTTTCCGGTGCTGTTTACACTACCGCAAACACCATCCGTCCGCCGCTGCAGCTGATCAAAA
CTGCACCGGGCAGCTCCCTGGTATCTTCCGTGTTCTTCATGCTGCTGCCGGAACAGGTTTAC
GTTTACGGTGACTGTGCGATCAACCCGGATCCGACCGCTGAACAGCTGGCAGAAATCGCGAT
TCAGTCCGCTGATTCCGCTGCGGCCTTCGGTATCGAACCGCGCGTTGCTATGCTCTCTACT
CCACCGGTACTTCTGGTGCAGGTAGCGACGTAGAAAAAGTTTCGCGAAGCAACTCGTCTGGCG
CAGGAAAAACGTCCTGACCTGATGATCGACGGTCCGCTGCAGTACGACGCTGCGGTAATGGC
TGACGTTGCGAAATCCAAAGCGCCGAACTCTCCGGTTGCAGGTGCGCTACCGTGTTTCTATCT
TCCCGGATCTGAACACCGGTAACACCACCTACAAAGCGGTACAGCGTTCTGCCGACCTGATC
TCCATCGGGCCGATGCTGCAGGGTATGCGCAAGCCGGTTAACGACCTGTCCCGTGGCGCACT
GGTTGACGATATCGTCTACACCATCGCGCTGACTGCGATTTCAGTCTGCACAGCAGCAGTAA

SEQ ID NO:52

VSRIIMLIPTGTSVGLTSVSLGVIRAMERKGVRLSVFKPIAQPRTGGDAPDQTTTIVRANSS
TTTAAEPLKMSYVEGLLSSNQKDVLMEEIVANYHANTKDAEVVLVEGLVPTRKHQFAQSLNY
EIAKTLNAEIVFVMSQGTDTPEQLKERIELTRNSFGGAKNTNITGVIVNKLNPVDEQGRTR
PDLSEIFDDSSKAKVNNVDPAKLQESSPLPVLGAVPWSFDLIATRAIDMARHLNATIINEGD
INTRRVKSVTFCARSHIPMLEHFRAAGSLLVTSADRPDVLVAACLAAMNGVEIGALLLTGGYE
MDARISKLCERAFATGLPVFMVNTNTWQTSLSLQSFNLEVPVDDHERIEKVQEYVANYINAD
WIESLTATSELEVLFQGPSRRLSPPAFRYQLTELARKAGKRIVLPEGDEPRTVKAAAICAE
RGIATCVLLGNPAEINRVAASQGVELGAGIEIVDPEVVRESYVGRLEVELRKNKGMTETVARE
QLEDNVVLGTLMLEQDEVDGLVSGAVHTTANTIRPPLQLIKTAPGSSLVSSVFFMLLPEQVY
VYGDCAINPDPTAEQLAEIAIQSADSAAAFGIEPRVAMLSYSTGTSGAGSDVEKVVREATR
LAQEKRPDL MIDGPLQYDAAVMADVAKSKAPNSPVAGRATVFI FPD LNTGNTTYKAVQRSADLI
SIGPMLQGMRKPVNDLSRGALVDDIVYTIALTAIQSAQQQ

SEQ ID NO:53

GTGTCCCGTATTATTATGCTGATCCCTACCGGAACCAGCGTCGGTCTGACCAGCGTCAGCCT
TGGCGTGATCCGTGCAATGGAACGCAAAGGCGTTTCGTCTGAGCGTTTTCAAACCTATCGCTC
AGCCGCGTACCGGTGGCGATGCGCCCGATCAGACTACGACTATCGTGCGTGCGAACTCTTCC
ACCACGACGGCCGCTGAACCGCTGAAAATGAGCTACGTTGAAGGTCTGCTTTCCAGCAATCA
GAAAGATGTGCTGATGGAAGAGATCGTCGCAAACCTACCACGCTAACACCAAAGACGCTGAAG
TCGTTCTGGTTGAAGGTCTGGTCCCGACACGTAAGCACCAGTTTGCCCGAGTCTCTGAACTAC
GAAATCGCTAAAACGCTGAATGCGGAAATCGTCTTTCGTTATGTCTCAGGGCACTGACACCCC
GGAACAGCTGAAAGAGCGTATCGAACTGACCCGCAACAGCTTCGGCGGTGCCAAAAACACCA
ACATCACCGGCGTTATCGTTAACAACTGAACGCACCGGTTGATGAACAGGGTCGTACTCGC
CCGGATCTGTCCGAGATTTTCGACGACTCTTCCAAAGCTAAAGTAAACAATGTTGATCCGGC
GAAGCTGCAAGAATCCAGCCCGCTGCCGGTTCTCGGCGCTGTGCCGTGGAGCTTTGACCTGA
TCGCGACTCGTGCGATCGATATGGCTCGCCACCTGAATGCGACCATCATCAACGAAGGCGAC
ATCAATACTCGCCGCGTTAAATCCGTCACCTTCTGCGCACGCAGCATTCCGCACATGCTGGA
GCACTTCCGTGCCGGTTCTCTGCTGGTGACTTCCGCAGACCGTCTTGACGTGCTGGTGGCCG
CTTGCCCTGGCAGCCATGAACGGCGTAGAAATCGGTGCCCTGCTGCTGACTGGCGGTTACGAA
ATGGACGCGCGCATTTCTAAACTGTGCGAACGTGCTTTTCGCTACCGGCCTGCCGGTATTTAT
GGTGAACACCAACACCTGGCAGACCTCTCTGAGCCTGCAGAGCTTCAACCTGGAAGTTCCGG
TTGACGATCACGAACGTATCGAGAAAGTTTCAGGAATACGTTGCTAACTACATCAACGCTGAC
TGATCGAATCTCTGACTGCCACTTCTGAGCGCAGCCGTCGTCTGGAAGTGCTGTTTCAGGG
TCCGTCTCCGCTGCGTTCCGTTATCAGCTGACTGAACTTGCGCGCAAAGCGGGCAAACGTA
TCGTACTGCCGGAAGGTGACGAACCGCGTACCGTTAAAGCAGCCGCTATCTGTGCTGAACGT
GGTATCGCAACTTGCGTACTGCTGGGTAATCCGGCAGAGATCAACCGTGTTGCAGCGTCTCA
GGGTGTAGAACTGGGTGCAGGGATTGAAATCGTTGATCCAGAAGTGTTTCGCGAAAGCTATG
TTGGTCGTCTGGTCGAACTGCGTAAGAACAAGGCATGACCGAAACCGTTGCCCGCGAACAG
CTGGAAGACAACGTGGTGCTCGGTACGCTGATGCTGGAACAGGATGAAGTTGATGGTCTGGT
TTCCGGTGCTGTTACACTACCGCAAACACCATCCGTCCGCCGCTGCAGCTGATCAAACTG
CACCGGGCAGCTCCCTGGTATCTTCCGTGTTCTTCATGCTGCTGCCGGAACAGGTTTACGTT
TACGGTGACTGTGCGATCAACCCGGATCCGACCGCTGAACAGCTGGCAGAAATCGCGATTCA
GTCCGCTGATTCCGCTGCGGCCCTTCGGTATCGAACCGCGCGTTGCTATGCTCTCCTACTCCA
CCGGTACTTCTGGTGCAGGTAGCGACGTAGAAAAAGTTTCGCGAAGCAACTCGTCTGGCGCAG
GAAAAACGTCCTGACCTGATGATCGACGGTCCGCTGCAGTACGACGCTGCGGTAATGGCTGA
CGTTGCGAAATCCAAAGCGCCGAACTCTCCGTTGTCAGGTGCGGCTACCGTGTTTCATCTTCC
CGGATCTGAACACCGGTAACACCACCTACAAAGCGGTACAGCGTTCTGCCGACCTGATCTCC

ATCGGGCCGATGCTGCAGGGTATGCGCAAGCCGGTTAACGACCTGTCCCGTGGCGCACTGGT
TGACGATATCGTCTACACCATCGCGCTGACTGCGATTCACTCTGCACAGCAGCAGTAA

SEQ ID NO:54

VSRIIMLIPTGTSVGLTSVSLGVIRAMERKGVRLSVFKPIAQPRRTGGDAPDQTTTIVRANSS
TTTAAEPLKMSYVEGLLSSNQKDVLMEEIVANYHANTKDAEVVLVEGLVPTRKHQFAQSLNY
EIAKTLNAEIVFVMSQGTDTPEQLKERIELTRNSFGGAKNTNITGVIVNKLNAVPDEQGRTR
PDLSEIFDDSSKAKVNNVDPKQLQESSPLPVLGAVPWSFDLIATRAIDMARHLNATIINEGD
INTRRVKSVTFCARSI PHMLEHFRA G SLLVTSADRPDVLVAACLAAMNGVEIGALLLTGGYE
MDARISKLCERAFATGLPVFMVNTNTWQTSLSLQSFNLEVPVDDHERIEKVQEYVANYINAD
WIESLTATSESRRLLEVLFQGPSPPAFRYQLTELARKAGKRIVLPEGDEPRTVKAAAI CAER
GIATCVLLGNPAEINRVAASQGVELGAGIEIVDPEVVRESYVGRVLVELRKNKGMTETVAREQ
LEDNVVLGTLMLEQDEVDGLVSGAVHTTANTIRPPLQLIKTAPGSSLVSSVFFMLLPEQVYV
YGDCAINPDPTAEQLAEIAIQSADSAAAFGIEPRVAMLSYSTGTSGAGSDVEKVVREATRLAQ
EKRPDLMIDGPLQYDAAVMADVAKSKAPNSPVAGRATVFI FPD LNTGNTTYKAVQRSADLIS
IGPMLQGMKRPVNDLSRGALVDDIVYTIALTAIQSAQQQ

SEQ ID NO:55

GTGTCCCGTATTATTATGCTGATCCCTACCGGAACCAGCGTCGGTCTGACCAGCGTCAGCCT
TGGCGTGATCCGTGCAATGGAACGCAAAGGCGTTTCGTCTGAGCGTTTTCAAACCTATCGCTC
AGCCGCGTACCGGTGGCGATGCGCCCGATCAGACTACGACTATCGTGCGTGCGAACTCTTCC
ACCACGACGGCCGCTGAACCGCTGAAAATGAGCTACGTTGAAGGTCTGCTTTCCAGCAATCA
GAAAGATGTGCTGATGGAAGAGATCGTCGCAAACCTACCACGCTAACACCAAAGACGCTGAAG
TCGTTCTGGTTGAAGGTCTGGTCCCGACACGTAAGCACCAGTTTGCCAGTCTCTGAACTAC
GAAATCGCTAAAACGCTGAATGCGGAAATCGTCTTCGTTATGTCTCAGGGCACTGACACCCC
GGAACAGCTGAAAGAGCGTATCGAACTGACCCGCAACAGCTTCGGCGGTGCCAAAAACACCA
ACATCACCGGCGTTATCGTTAACAACCTGAACGCACCGGTTGATGAACAGGGTCGTACTCGC
CCGGATCTGTCCGAGATTTTCGACGACTCTTCCAAAGCTAAAGTAAACAATGTTGATCCGGC
GAAGCTGCAAGAATCCAGCCCGCTGCCGGTTCTCGGCGCTGTGCCGTGGAGCTTTGACCTGA
TCGCGACTCGTGCGATCGATATGGCTCGCCACCTGAATGCGACCATCATCAACGAAGGCGAC
ATCAATACTCGCCGCGTTAAATCCGTCACCTTCTGCGCACGCAGCATTCCGCACATGCTGGA
GCACTTCCGTGCCGGTTCTCTGCTGGTGACTTCCGCAGACCGTCTTGACGTGCTGGTGGCCG
CTTGCCCTGGCAGCCATGAACGGCGTAGAAAATCGGTGCCCTGCTGCTGACTGGCGGTTACGAA
ATGGACGCGCGCATTTCTAAACTGTGCGAACGTGCTTTTCGCTACCGGCCTGCCGGTATTTAT
GGTGAACACCAACACCTGGCAGACCTCTCTGAGCCTGCAGAGCTTCAACCTGGAAGTTCCGG
TTGACGATCACGAACGTATCGAGAAAGTT CAGGAATACGTTGCTAACTACATCAACGCTGAC
TGGATCGAATCTCTGACTGCCACTTCTGAGCGCAGCCGTCGTCTGTCTCCGCTGCGTTCCG
TTATCAGCTGACTGAACTTGCGCGCAAAGCGGGCAAACGTATCGTACTGGAAGTGCTGTTTC
AGGGTCCGGAAGGTGACGAACCGCGTACCGTTAAAGCAGCCGCTATCTGTGCTGAACGTGGT
ATCGCAACTTGCGTACTGCTGGGTAATCCGGCAGAGATCAACCGTGTTGCAGCGTCTCAGGG
TGTAGAACTGGGTGCAGGGATTGAAATCGTTGATCCAGAAGTGGTTCGCGAAAGCTATGTTG
GTCGTCTGGTCGAACTGCGTAAGAACAAGGCATGACCGAAACCGTTGCCCGCGAACAGCTG
GAAGACAACGTGGTGCTCGGTACGCTGATGCTGGAACAGGATGAAGTTGATGGTCTGGTTTC
CGGTGCTGTTTCACTACCGCAAACACCATCCGTCCGCCGCTGCAGCTGATCAAACTGCAC
CGGGCAGCTCCCTGGTATCTTCCGTGTTCTTCATGCTGCTGCCGGAACAGGTTTACGTTTAC
GGTGACTGTGCGATCAACCCGATCCGACCGCTGAACAGCTGGCAGAAATCGCGATT CAGTC
CGCTGATTCCGCTGCGGCCTTCGGTATCGAACCGCGCGTTGCTATGCTCTCCTACTCCACCG
GTACTTCTGGTGCAGGTAGCGACGTAGAAAAAGTTTCGCGAAGCAACTCGTCTGGCGCAGGAA

AAACGTCCTGACCTGATGATCGACGGTCCGCTGCAGTACGACGCTGCGGTAATGGCTGACGT
TGCGAAATCCAAAGCGCCGAACCTCTCCGGTTGCAGGTGCGGCTACCGTGTTTCATCTTCCCGG
ATCTGAACACCGGTAACACCACCTACAAAGCGGTACAGCGTTCTGCCGACCTGATCTCCATC
GGGCCGATGCTGCAGGGTATGCGCAAGCCGGTTAACGACCTGTCCCGTGCGCGACTGGTTGA
CGATATCGTCTACACCATCGCGCTGACTGCGATTGAGTCTGCACAGCAGCAGTAA

SEQ ID NO:56

VSRIIMLIPTGTSVGLTSVSLGVIRAMERKGVRLSVFKPIAQPRTGGDAPDQTTTIVRANSS
TTTAAEPLKMSYVEGLLSSNQKDVLMEEIVANYHANTKDAEVVLVEGLVPTRKHQFAQSLNY
EIAKTLNAEIVFVMSQGTDTPEQLKERIELTRNSFGGAKNTNITGVIVNKLNPVDEQGRTR
PDLSEIFDDSSKAKVNNVDPKQLQESSPLPVLGAVPWSFDLIATRAIDMARHLNATIINEGD
INTRRVKSVTFCARSIPHMLEHFRAAGSLVTSADRPDVLVAACLAAMNGVEIGALLLTGGYE
MDARISKLCERAFATGLPVFMVNTNTWQTSLSLQSFNLEVPVDDHERIEKVQEYVANYINAD
WIESLTATSESRRLSPPAFRYQLTELARKAGKRIVLEVLFQGPPEGDEPRTVKAAAI CAERG
IATCVLLGNPAEINRVAASQGVELGAGIEIVDPEVVRESYVGRLVELRKNKGMTETVAREQL
EDNVVLGTLMLEQDEVDGLVSGAVHTTANTIRPPLQLIKTAPGSSLVSSVFFMLLPEQVYVY
GDCAINPDPTAEQLAEIAIQSADSAAAFGIEPRVAMLSYSTGTSGAGSDVEKVVREATRLAQE
KRPDLMDGPLQYDAVMADVAKSKAPNSPVAGRATVFI FPDNLNTGNTTYKAVQRSADLISI
GPMLQGMKRPVNDLSRGALVDDIVYTIALTAIQSAQQQ

SEQ ID NO:57

ATGTCCTCACGTAAAGAGCTTGCCAATGCTATTCGTGCGCTGAGCATGGACGCAGTACAGAA
AGCCAAATCCGGTCACCCGGGTGCCCTATGGGTATGGCTGACATTGCCGAAGTCTGTGGC
GTGATTTCTCTGAAACACAACCCGCAGAATCCGTCTGGGCTGACCGTGACCGCTTCGTGCTG
TCCAACGGCCACGGCTCCATGCTGATCTACAGCCTGCTGCACCTCACCGGTTACGATCTGCC
GATGGAAGAAGTGAAGAACTTCCGTCAGCTGCACTCTAAAGTCCGGGTCACCCGGAAGTGG
GTTACACCGCTGGTGTGGAAGCCACCCGGTCCGCTGGGTCAGGGTATTGCCAACGCAGTC
GGTATGGCGATTGCAGAAAAACGCTGGCGGCGCAGTTTAACCGTCCGGGCCACGACATTGT
CGACCACTACACCTACGCCTTCATGGGCGACGGCTGCATGATGGAAGGCATCTCCACGAAG
TTTGCTCTCTGGCGGGTACGCTGAAGCTGGGTAAAGTGAATGCTTCTACGATGACAACGGT
ATTTCTATCGATGGTCACGTTGAAGGCTGGTTCACCGACGACACCGCAATGCGTTTCGAAGC
TTACGGCTGGCACGTTATTTCGCGACATCGACGGTCATGACGCGGCATCTATCAAACGCGCAG
TAGAAGAAGCGCGCGCAGTGACTGACAAACCTTCCCTGCTGATGTGCAAAACCATCATCGGT
TTCGGTTCCCCGAACAAAGCCGGTACCCACGACTCCACGGTGCGCCGCTGGGCGACGCTGA
AATTGCCCTGACCCGCGAACAAGTGGGCTGGAATATGCGCCGTTGAAATCCCGTCTGAAA
TCTATGCTCAGTGGGATGCGAAAGAAGCAGGCCAGGCCGAAAGAATCCGCATGGAACGAGAAA
TTCGCTGCTTACGCGAAAGCTTATCCGCAGGAAGCCGCTGAATTTACCCGCCGTATGAAAGG
CGAAATGCCGTCTGACTTCGACGCTAAAGCGAAAGAGTTTCATCGCTAAAGTGCAGGCTAATC
CGGCGAAAATCGCCAGCCGTAAAGCGTCTCAGAATGCTATCGAAGCGTTCCGGTCCGCTGTTG
CCGGAATTCCTCGGCGGTTCTGCTGACCTGGCGCCGCTAACCTGACCCTGTGGTCTGGTTC
TAAAGCAATCAACGAAGATGCTGCGGGTAACTACATCCACTACGGTGTTCGCGAGTTTCGGTA
TGACCGCGATTGCTAACGGTATCTCCCTGCACGGTGGCTTCCTGCCGTACACCTCCACCTTC
CTGATGTTTCGTGGAATACGCACGTAACGCCGTACGTATGGCTGCGCTGATGAAACAGCGTCA
GGTGATGGTTTACACCCACGACTCCATCGGTCTGGGCGAAGACGGCCCGACTCACCAGCCGG
TTGAGCAGGTGCTTCTCTGCGCGTAACCCCGAACATGTCTACATGGCGTCCGTGTGACCAG
GTTGAATCCGCGGTGCGGTGGAATACGGTGTGAGCGTCAGGACGGCCCGACCGCACTGAT
CCTCTCCCGTCAGAACCTGGCGCAGCAGGAACGAACTGAAGAGCAACTGGCAAACATCGCGC
GCGGTGGTTATGTGCTGAAAGACTGCGCCGGTACGCCGGAAGTGAATTTTCATCGCTACCGGT

TCAGAAAGTTGAACTGGCTGTTGCTGCCTACGAAAACTGACTGCCGAAGGCGTGAAAGCGCG
CGTGGTGTCCATGCCGTCTACCGACGCATTTGACAAGCAGGATGCTGCTTACCGTGAATCCG
TACTGCCGAAAGCGGTTACTGCACGCGTTGCTGTAGAAGCGGGTATTGCTGACTACTGGTAC
AAGTATGTTGGCCTGAACGGTGCTATCGTCGGTATGACCACCTTCGGTGAATCTGCTCCGGC
AGAGCTGCTGTTTGAAGAGTTCGGCTTCACTGTTGATAACGTTGTTGCGAAAGCAAAAGAAC
TGCTGTAA

SEQ ID NO:58

ATGTCCTCACGTAAAGAGCTTGCCAATGCTATTCGTGCGCTGAGCATGGACGCAGTACAGAA
AGCCAAATCCGGTCACCCGGGTGCCCTATGGGTATGGCTGACATTGCCGAAGTCCTGTGGC
GTGATTTCTGAAACACAACCCGCAGAATCCGTCTGGGCTGACCGTGACCGCTTCGTGCTG
TCCAACGGCCACGGCTCCATGCTGATCTACAGCCTGCTGCACCTCACCGGTTACGATCTGCC
GATGGAAGAACTGAAAACTTCCGTCAGCTGCACTCTAAAACTCCGGGTACCCGGAAGTGG
GTTACACCGCTGGTGTGGAACCACCACCGGTCCGCTGGGTGAGGGTATTGCCAACGCAGTC
GGTATGGCGATTGCAGAAAAACGCTGGCGGCGCAGTTTAACCGTCCGGGCCACGACATTGT
CGACCACTACACCTACGCCTTCATGGGCGACGGCTGCATGATGGAAGGCATCTCCCACGAAG
TTTGCTCTCTGGCGGGTACGCTGAAGCTGGGTAACTGATTGCATTCTACGATGACAACGGT
ATTTCTATCGATGGTCACGTTGAAGGCTGGTTCACCGACGACACCGCAATGCGTTTCGAAGC
TTACGGCTGGCACGTTATTTCGCGACATCGACGGTCATGACGCGGCATCTATCAAACGCGCAG
TAGAAGAAGCGCGCGCAGTGACTGACAAACCTTCCCTGCTGATGTGCAAAACCATCATCGGT
TTCGGTTCCCCGAACAAAGCCGGTACCCACGACTCCCACGGTGCGCCGCTGGGCGACGCTGA
AATTGCCCTGACCCGCGAACAACTGGGCTGGAAATATGCGCCGTTGAAATCCCGTCTGAAA
TCTATGCTCAGTGGGATGCGAAAGAAGCAGGCCAGGCGAAAGAATCCGCATGGAACGAGAAA
TTCGCTGCTTACGCGAAAGCTTATCCGCAGGAAGCCGCTGAATTTACCCGCCGTATGAAAGG
CGAAATGCCGTCTGACTTCGACGCTAAAGCGAAAGAGTTCATCGCTAAACTGCAGGCTAATC
CGGCGAAAATCGCCAGCCGTAAAGCGTCTCAGAATGCTATCGAAGCGTTCCGGTCCGCTGTTG
CCGGAATTCTCGGCGGTTCTGCTGACCTGGCGCCGTCTAACCTGACCCTGTGGTCTGGTTC
TAAAGCAATCAACGAAGATGCTGCGGGTAACTACATCCACTACGGTGTTGCGGAGTTCGGTA
TGACCGCGATTGCTAACGGTATCTCCCTGCACGGTGGCTTCCTGCCGTACACCTCCACCTTC
CTGATGTTTCGTGGAATACGCACGTAACGCCGTACGTATGGCTGCGCTGATGAAACAGCGTCA
GGTGATGGTTTACACCCACGACTCCATCGGTCTGGGCGAAGACGGCCCGACTCACCAGCCGG
TTGAGCAGGTGCTTCTCTGCGCGTAACCCCGAACATGTCTACATGGCGTCCGTGTGACCAG
GTTGAATCCGCGGTGCGGTGGAATACGGTGTGAGCGTCAGGACGGCCCGACCGCACTGAT
CCTCTCCCGTCAGAACCTGGCGCAGCAGGAACGAACTGAAGAGCAACTGGCAAACATCGCGC
GCGGTGGTTATGTGCTGAAAGACTGCGCCGGTACGCCGGAAGTGAATTTTCATCGCTACCGGT
TCAGAAAGTTGAACTGGCTGTTGCTGCCTACGAAAACTGACTGCCGAAGGCGTGAAAGCGCG
CGTGGTGTCCATGCCGTCTACCGACGCATTTGACAAGCAGGATGCTGCTTACCGTGAATCCG
TACTGCCGAAAGCGGTTACTGCACGCGTTGCTGTAGAAGCGGGTATTGCTGACTACTGGTAC
AAGTATGTTGGCCTGAACGGTGCTATCGTCGGTATGACCACCTTCCTGGAAGTGCTGTTTCA
GGGTCCGGGTGAATCTGCTCCGGCAGAGCTGCTGTTTGAAGAGTTCGGCTTCACTGTTGATA
ACGTTGTTGCGAAAGCAAAAGAACTGCTGTAA

SEQ ID NO:59

ATGTCCTCACGTAAAGAGCTTGCCAATGCTATTCGTGCGCTGAGCATGGACGCAGTACAGAA
AGCCAAATCCGGTCACCCGGGTGCCCTATGGGTATGGCTGACATTGCCGAAGTCCTGTGGC
GTGATTTCTGAAACACAACCCGCAGAATCCGTCTGGGCTGACCGTGACCGCTTCGTGCTG
TCCAACGGCCACGGCTCCATGCTGATCTACAGCCTGCTGCACCTCACCGGTTACGATCTGCC
GATGGAAGAACTGAAAACTTCCGTCAGCTGCACTCTAAAACTCCGGGTACCCGGAAGTGG

GTTACACCGCTGGTGTGGAACACCACCACCGGTCCGCTGGGTCAGGGTATTGCCAACGCAGTC
GGTATGGCGATTGCAGAAAAACGCTGGCGGCGCAGTTTAACCGTCCGGGGCCACGACATTGT
CGACCACTACACCTACGCCTTCATGGGCGACGGCTGCATGATGGAAGGCATCTCCACGAAG
TTTGCTCTCTGGCGGGTACGCTGAAGCTGGGTAAACTGATTGCATTCTACGATGACAACGGT
ATTTCTATCGATGGTCACGTTGAAGGCTGGTTCACCGACGACACCGCAATGCGTTTCGAAGC
TTACGGCTGGCACGTTATTTCGCGACATCGACGGTCATGACGCGGCATCTATCAAACGCGCAG
TAGAAGAAGCGCGCGCAGTGACTGACAAACCTTCCCTGCTGATGTGCAAAACCATCATCGGT
TTCGGTTCCCCGAACAAAGCCGGTACCCACGACTCCCACGGTGCGCCGCTGGGCGACGCTGA
AATTGCCCTGACCCGCGAACAACTGGGCTGGAATATGCGCCGTTCGAAATCCCGTCTGAAA
TCTATGCTCAGTGGGATGCGAAAGAAGCAGGCCAGGCCGAAAGAATCCGCATGGAACGAGAAA
TTCGCTGCTTACGCGAAAGCTTATCCGCAGGAAGCCGCTGAATTTACCCGCCGTATGAAAGG
CGAAATGCCGTCTGACTTCGACGCTAAAGCGAAAGAGTTTCATCGCTAAACTGCAGGCTAATC
CGGCGAAAATCGCCAGCCGTAAAGCGTCTCAGAATGCTATCGAAGCGTTTCGGTCCGCTGTTG
CCGGAATTCCTCGGCGGTTCTGCTGACCTGGCGCCGTCTAACCTGACCCTGTGGTCTGGTTC
TAAAGCAATCAACGAAGATGCTGCGGGTAACTACATCCACTACGGTGTTTCGCGAGTTTCGGTA
TGACCGCGATTGCTAACGGTATCTCCCTGCACGGTGGCTTCCTGCCGTACACCTCCACCTTC
CTGATGTTTCGTGGAATACGCACGTAACGCCGTACGTATGGCTGCGCTGATGAAACAGCGTCA
GGTGATGGTTTACACCCACGACTCCATCGGTCTGGGCGAAGACGGCCCCGACTCACCAGCCGG
TTGAGCAGGTCGCTTCTCTGCGCGTAACCCCGAACATGTCTACATGGCGTCCGTGTGACCAG
GTTGAATCCGCGGTGCGCTGGAATACGGTGTGAGCGTCAGGACGGCCCCGACCGCACTGAT
CCTCTCCCGTCAAGAACCTGGCGCAGCAGGAACGAACTGAAGAGCAACTGGCAAACATCGCGC
GCGGTGGTTATGTGCTGAAAGACTGCGCCGGTACGCCGGAAGTGAATTTTCATCGCTACCGGT
TCAGAAAGTTGAACTGGCTGTTGCTGCCTACGAAAACTGACTGCCGAAGGCGTGAAAGCGCG
CGTGGTGTCCATGCCGTCTACCGACGCATTTGACAAGCAGGATGCTGCTTACCGTGAATCCG
TACTGCCGAAAGCGGTTACTGCACGCGTTGCTGTAGAAGCGGGTATTGCTGACTACTGGTAC
AAGTATGTTGGCCTGAACGGTGCTATCGTCGGTATGACCACCTTCGGTCTGGAAGTGCTGTT
TCAGGGTCCGGAATCTGCTCCGGCAGAGCTGCTGTTTGAAGAGTTCGGCTTCACTGTTGATA
ACGTTGTTGCGAAAGCAAAAGAACTGCTGTAA

SEQ ID NO:60

ATGTCCTCACGTAAAGAGCTTGCCAATGCTATTTCGTGCGCTGAGCATGGACGCAGTACAGAA
AGCCAAATCCGGTACCCCGGGTGCCCTATGGGTATGGCTGACATTGCCGAAGTCTGTGGC
GTGATTTCTGAAACACAACCCGCAGAATCCGTCTGGGCTGACCGTGACCGCTTCGTGCTG
TCCAACGGCCACGGCTCCATGCTGATCTACAGCCTGCTGCACCTCACC GGTTACGATCTGCC
GATGGAAGAACTGAAAACTTCCGTCAGCTGCACTCTAAAACCTCCGGGTCACCCGGAAGTGG
GTTACACCGCTGGTGTGGAACACCACCACCGGTCCGCTGGGTCAGGGTATTGCCAACGCAGTC
GGTATGGCGATTGCAGAAAAACGCTGGCGGCGCAGTTTAACCGTCCGGGGCCACGACATTGT
CGACCACTACACCTACGCCTTCATGGGCGACGGCTGCATGATGGAAGGCATCTCCACGAAG
TTTGCTCTCTGGCGGGTACGCTGAAGCTGGGTAAACTGATTGCATTCTACGATGACAACGGT
ATTTCTATCGATGGTCACGTTGAAGGCTGGTTCACCGACGACACCGCAATGCGTTTCGAAGC
TTACGGCTGGCACGTTATTTCGCGACATCGACGGTCATGACGCGGCATCTATCAAACGCGCAG
TAGAAGAAGCGCGCGCAGTGACTGACAAACCTTCCCTGCTGATGTGCAAAACCATCATCGGT
TTCGGTTCCCCGAACAAAGCCGGTACCCACGACTCCCACGGTGCGCCGCTGGGCGACGCTGA
AATTGCCCTGACCCGCGAACAACTGGGCTGGAATATGCGCCGTTCGAAATCCCGTCTGAAA
TCTATGCTCAGTGGGATGCGAAAGAAGCAGGCCAGGCCGAAAGAATCCGCATGGAACGAGAAA
TTCGCTGCTTACGCGAAAGCTTATCCGCAGGAAGCCGCTGAATTTACCCGCCGTATGAAAGG
CGAAATGCCGTCTGACTTCGACGCTAAAGCGAAAGAGTTTCATCGCTAAACTGCAGGCTAATC
CGGCGAAAATCGCCAGCCGTAAAGCGTCTCAGAATGCTATCGAAGCGTTTCGGTCCGCTGTTG
CCGGAATTCCTCGGCGGTTCTGCTGACCTGGCGCCGTCTAACCTGACCCTGTGGTCTGGTTC

TAAAGCAATCAACGAAGATGCTGCGGGTAACTACATCCACTACGGTGTTTCGCGAGTTTCGGTA
TGACCGCGATTGCTAACGGTATCTCCCTGCACGGTGGCTTCCTGCCGTACACCTCCACCTTC
CTGATGTTTCGTGGAATACGCACGTAACGCCGTACGTATGGCTGCGCTGATGAAACAGCGTCA
GGTGATGGTTTACACCCACGACTCCATCGGTCTGGGCGAAGACGGCCCGACTCACCAGCCGG
TTGAGCAGGTCGCTTCTCTGCGCGTAACCCCGAACATGTCTACATGGCGTCCGTGTGACCAG
GTTGAATCCGCGGTTCGCGTGGAATACGGTGTTGAGCGTCAGGACGGCCCGACCGCACTGAT
CCTCTCCCGTCAGAACCTGGCGCAGCAGGAACGAACTGAAGAGCAACTGGCAAACATCGCGC
GCGGTGGTTATGTGCTGAAAGACTGCGCCGGTCAGCCGGAAGTGAATTTTCATCGCTACCGGT
TCAGAAAGTTGAACTGGCTGTTGCTGCCTACGAAAACTGACTGCCGAAGGCGTGAAAGCGCG
CGTGGTGTCATGCCGTCTACCGACGCATTTGACAAGCAGGATGCTGCTTACCGTGAATCCG
TACTGCCGAAAGCGGTTACTGCACGCGTTGCTGTAGAAGCGGGTATTGCTGACTACTGGTAC
AAGTATGTTGGCCTGAACGGTGCTATCGTCGGTATGACCACCTTCGGTGAAGTGAAGTGCT
GTTTCAGGGTCCGTCTGCTCCGGCAGAGCTGCTGTTTGAAGAGTTCGGCTTCACTGTTGATA
ACGTTGTTGCGAAAGCAAAAGAACTGCTGTAA

SEQ ID NO:61

ATGTCCTCACGTAAAGAGCTTGCCAATGCTATTCGTGCGCTGAGCATGGACGCAGTACAGAA
AGCCAAATCCGGTCACCCGGGTGCCCTATGGGTATGGCTGACATTGCCGAAGTCTGTGGC
GTGATTTCTGAAACACAACCCGCAGAATCCGTCTGGGCTGACCGTGACCGCTTCGTGCTG
TCCAACGGCCACGGCTCCATGCTGATCTACAGCCTGCTGCACCTCACCGGTTACGATCTGCC
GATGGAAGAACTGAAAACTTCCGTGAGCTGCACTCTAAACTCCGGGTACCCGGAAGTGG
GTTACACCGCTGGTGTGGAACCACCACCGGTCCGCTGGGTGAGGATATTGCCAACGCAGTC
GGTATGGCGATTGCAGAAAAACGCTGGCGGCGCAGTTTAACCGTCCGGGCCACGACATTGT
CGACCACTACACCTACGCCTTCATGGGCGACGGCTGCATGATGGAAGGCATCTCCACGAAG
TTTGCTCTCTGGCGGGTACGCTGAAGCTGGGTAAACTGATTGCATTCTACGATGACAACGGT
ATTTCTATCGATGGTCACGTTGAAGGCTGGTTACCGACGACACCGCAATGCGTTTCGAAGC
TTACGGCTGGCACGTTATTTCGCGACATCGACGGTCATGACGCGGCATCTATCAAACGCGCAG
TAGAAGAAGCGCGCGCAGTGACTGACAAACCTTCCCTGCTGATGTGCAAAACCATCATCGGT
TTCGGTTCCCGAACAAGCCGGTACCCACGACTCCCACGGTGCGCCGCTGGGCGACGCTGA
AATTGCCCTGACCCGCGAACAACCTGGGCTGGAAATATGCGCCGTTGAAATCCCGTCTGAAA
TCTATGCTCAGTGGGATGCGAAAGAAGCAGGCCAGGCGAAAGAATCCGCATGGAACGAGAAA
TTCGCTGCTTACGCGAAAGCTTATCCGCAGGAAGCCGCTGAATTTACCCGCCGTATGAAAGG
CGAAATGCCGTCTGACTTCGACGCTAAAGCGAAAGAGTTCATCGCTAAACTGCAGGCTAATC
CGGCGAAAATCGCCAGCCGTAAAGCGTCTCAGAATGCTATCGAAGCGTTTCGGTCCGCTGTTG
CCGGAATTCCTCGGCGGTTCTGCTGACCTGGCGCCGTCTAACCTGACCCTGTGGTCTGGTTC
TAAAGCAATCAACGAAGATGCTGCGGGTAACTACATCCACTACGGTGTTTCGCGAGTTTCGGTA
TGACCGCGATTGCTAACGGTATCTCCCTGCACGGTGGCTTCCTGCCGTACACCTCCACCTTC
CTGATGTTTCGTGGAATACGCACGTAACGCCGTACGTATGGCTGCGCTGATGAAACAGCGTCA
GGTGATGGTTTACACCCACGACTCCATCGGTCTGGGCGAAGACGGCCCGACTCACCAGCCGG
TTGAGCAGGTCGCTTCTCTGCGCGTAACCCCGAACATGTCTACATGGCGTCCGTGTGACCAG
GTTGAATCCGCGGTTCGCGTGGAATACGGTGTTGAGCGTCAGGACGGCCCGACCGCACTGAT
CCTCTCCCGTCAGAACCTGGCGCAGCAGGAACGAACTGAAGAGCAACTGGCAAACATCGCGC
GCGGTGGTTATGTGCTGAAAGACTGCGCCGGTCAGCCGGAAGTGAATTTTCATCGCTACCGGT
TCAGAAAGTTGAACTGGCTGTTGCTGCCTACGAAAACTGACTGCCGAAGGCGTGAAAGCGCG
CGTGGTGTCATGCCGTCTACCGACGCATTTGACAAGCAGGATGCTGCTTACCGTGAATCCG
TACTGCCGAAAGCGGTTACTGCACGCGTTGCTGTAGAAGCGGGTATTGCTGACTACTGGTAC
AAGTATGTTGGCCTGAACGGTGCTATCGTCGGTATGACCACCTTCGGTGAATCTCTGGAAGT
GCTGTTTCAGGGTCCGGCTCCGGCAGAGCTGCTGTTTGAAGAGTTCGGCTTCACTGTTGATA
ACGTTGTTGCGAAAGCAAAAGAACTGCTGTAA

SEQ ID NO:62

ATGTCCTCACGTAAAGAGCTTGCCAATGCTATTTCGTGCGCTGAGCATGGACGCAGTACAGAA
AGCCAAATCCGGTCACCCGGGTGCCCTATGGGTATGGCTGACATTGCCGAAGTCCTGTGGC
GTGATTTCTTGAAACACAACCCGCAGAATCCGTCTGGGCTGACCGTGACCGCTTCGTGCTG
TCCAACGGCCACGGCTCCATGCTGATCTACAGCCTGCTGCACCTCACCGGTTACGATCTGCC
GATGGAAGAACTGAAAACTTCCGTCAGCTGCACTCTAAAACTCCGGGTCACCCGGAAGTGG
GTTACACCGCTGGTGTGGAAACCACCACCGGTCCGCTGGGTGAGGGTATTGCCAACGCAGTC
GGTATGGCGATTGCAGAAAAACGCTGGCGGCGCAGTTTAACCGTCCGGGCCACGACATTGT
CGACCACTACACCTACGCCTTCATGGGCGACGGCTGCATGATGGAAGGCATCTCCACGAAG
TTTGCTCTCTGGCGGGTACGCTGAAGCTGGGTAAACTGATTGCATTCTACGATGACAACGGT
ATTTCTATCGATGGTCACGTTGAAGGCTGGTTCACCGACGACACCGCAATGCGTTTCGAAGC
TTACGGCTGGCACGTTATTTCGCGACATCGACGGTCATGACGCGGCATCTATCAAACGCGCAG
TAGAAGAAGCGCGCGCAGTGACTGACAAACCTTCCCTGCTGATGTGCAAAACCATCATCGGT
TTCGGTTCCCCGAACAAAGCCGGTACCCACGACTCCCACGGTGCGCCGCTGGGCGACGCTGA
AATTGCCCTGACCCGCGAACAACCTGGGCTGGAAATATGCGCCGTTCGAAATCCCGTCTGAAA
TCTATGCTCAGTGGGATGCGAAAGAAGCAGGCCAGGCGAAAGAATCCGCATGGAACGAGAAA
TTCGCTGCTTACGCGAAAGCTTATCCGCAGGAAGCCGCTGAATTTACCCGCCGTATGAAAGG
CGAAATGCCGTCTGACTTCGACGCTAAAGCGAAAGAGTTCATCGCTAAACTGCAGGCTAATC
CGGCGAAAATCGCCAGCCGTAAAGCGTCTCAGAATGCTATCGAAGCGTTCGGTCCGCTGTTG
CCGGAATTCCTCGGCGGTTCTGCTGACCTGGCGCCGTCTAACCTGACCCTGTGGTCTGGTTC
TAAAGCAATCAACGAAGATGCTGCGGGTAACTACATCCACTACGGTGTTTCGCGAGTTCGGTA
TGACCGCGATTGCTAACGGTATCTCCCTGCACGGTGGCTTCCTGCCGTACACCTCCACCTTC
CTGATGTTTCGTGGAATACGCACGTAACGCCGTACGTATGGCTGCGCTGATGAAACAGCGTCA
GGTGATGGTTTACACCCACGACTCCATCGGTCTGGGCGAAGACGGCCCGACTCACCAGCCGG
TTGAGCAGGTCGCTTCTCTGCGCGTAACCCCGAACATGTCTACATGGCGTCCGTGTGACCAG
GTTGAATCCGCGGTTCGCGTGGAAATACGGTGTGAGCGTCAGGACGGCCCGACCGCACTGAT
CCTCTCCCGTCAGAACCTGGCGCAGCAGGAACGAACTGAAGAGCAACTGGCAAACATCGCGC
GCGGTGGTTATGTGCTGAAAGACTGCGCCGGTCAGCCGGAACCTGATTTTCATCGCTACCGGT
TCAGAAAGTTGAACTGGCTGTTGCTGCCTACGAAAACTGACTGCCGAAGGCGTGAAAGCGCG
CGTGGTGTCCATGCCGTCTACCGACGCATTTGACAAGCAGGATGCTGCTTACCGTGAATCCG
TACTGCCGAAAGCGGTTACTGCACGCGTTGCTGTAGAAGCGGGTATTGCTGACTACTGGTAC
AAGTATGTTGGCCTGAACGGTGCTATCGTCGGTATGACCACCTTCGGTGAATCTGCTCCGCT
GGAAGTGCTGTTTCAGGGTCCGGCAGAGCTGCTGTTTGAAGAGTTCGGCTTCACTGTTGATA
ACGTTGTTGCGAAAGCAAAAGAACTGCTGTAA

SEQ ID NO:63

MSSRKELANAI RAL SMDAVQKAKSGHPGAPMGMA DIAEVLWRDFLKHNPQNPSWADRDRFVL
SNHGGSMLIYSLHLTG YDLPMEELKNFRQLH SKTPGHPEVGYTAGVETTTGPLGQGIANAV
GMAIAEKT LAAQFNRP GHDIVDHYTYAFMGD GCMMEGISH EVCSLAGTLKLGLKLI AFYDDNG
ISIDGHVEGWFTDDTAMRFEAYGWHVIR DIGHDAASIKRAVEEARAVTDKPSLLMCKTIIG
FGSPNKAGTHDSHGAPLGDAEIAL TREQLGWKYAPFEIPSEIYAQWDAKEAGQAKESAWNEK
FAAYAKAYPQEAAEFTRRMKGEMPSDFDAKAKEFI AKLQANPAKIASRKASQNAIEAFGPLL
PEFLGGSADLAPSNLTLWSGSKAINEDAAGNYIHYGVREFGMTAIANGISLHGGFLPYTSTF
LMFVEYARNAVRMAALMKQRQVMVYTHDSIGLGEDGP THQPVEQIASLRVTPNMSTWRPCDQ
VESAVAWKYGVERQDGPTALILSRQNLAQQERTEEQ LANIARGGYVLKDCAGQPELIFIATG
SEVELAVAAYEKLTAEGVKARVVSMPSTDAFDKQDAAYRESVLPKAVTARVAVEAGIADYWY
KYVGLNGAIVGMTTFGESAPAEQLFEFEGFTVDNVVAKAKALL

SEQ ID NO:64

MSSRKELANAI RAL SMDAVQKAKSGHPGAPMG MADI AEVLWRDFLKHNPQNPSWADRDRFVL
SNGHGSMLIY SLLHLTG YDLPMEELKNFRQLH SKTPGHPEVGYTAGVETTTG PLGQGIANAV
GMAIAEKT LAAQFNRP GHDIVDHYTYAFMG DGCMMEG I SHEVCSLAGTLKLGKLI AFYDDNG
I SIDGHVEGWFTDDTAMRFEAYGWHVIR D IDGHDAAS I KRAVEEARAVTDKPSLLMCKT I IG
FGSPNKAGTHDSHGAPLGDAEIAL TREQLGWKYAPFEIPSEIYAQWDAKEAGQAKESAWNEK
FAAYAKAYPQEAAEFTRRMKGEMPSDFDAKAKEF IAKLQANPAKIASRKASQNAIEAFGPLL
PEFLGGSADLAPSNLTLWSGSKAINEDAAGNYIHYGVREFGMTAIANGISLHGGFLPYTSTF
LMFVEYARNAVRMAALMKQRQVMVYTHDSIGLGEDGPTHQPVEQVASLRVTPNMSTWRPCDQ
VESAVAWKYGVERQDGPTALILSRQNL AQQERTEEQ LANIARGGYVLKDCAGQPELIFIATG
SEVELAVAAYEKLTAEGVKARVVSMPSTDAFDKQDAAYRESVLPKAVTARVAVEAGIADYWY
KYVGLNGAIVGMTTFLEVL FQGPGESAPAE LLFEEFGFTVDNVVAKAKELL

SEQ ID NO:65

MSSRKELANAI RAL SMDAVQKAKSGHPGAPMG MADI AEVLWRDFLKHNPQNPSWADRDRFVL
SNGHGSMLIY SLLHLTG YDLPMEELKNFRQLH SKTPGHPEVGYTAGVETTTG PLGQGIANAV
GMAIAEKT LAAQFNRP GHDIVDHYTYAFMG DGCMMEG I SHEVCSLAGTLKLGKLI AFYDDNG
I SIDGHVEGWFTDDTAMRFEAYGWHVIR D IDGHDAAS I KRAVEEARAVTDKPSLLMCKT I IG
FGSPNKAGTHDSHGAPLGDAEIAL TREQLGWKYAPFEIPSEIYAQWDAKEAGQAKESAWNEK
FAAYAKAYPQEAAEFTRRMKGEMPSDFDAKAKEF IAKLQANPAKIASRKASQNAIEAFGPLL
PEFLGGSADLAPSNLTLWSGSKAINEDAAGNYIHYGVREFGMTAIANGISLHGGFLPYTSTF
LMFVEYARNAVRMAALMKQRQVMVYTHDSIGLGEDGPTHQPVEQVASLRVTPNMSTWRPCDQ
VESAVAWKYGVERQDGPTALILSRQNL AQQERTEEQ LANIARGGYVLKDCAGQPELIFIATG
SEVELAVAAYEKLTAEGVKARVVSMPSTDAFDKQDAAYRESVLPKAVTARVAVEAGIADYWY
KYVGLNGAIVGMTTFGLEVL FQGPESAPAE LLFEEFGFTVDNVVAKAKELL

SEQ ID NO:66

MSSRKELANAI RAL SMDAVQKAKSGHPGAPMG MADI AEVLWRDFLKHNPQNPSWADRDRFVL
SNGHGSMLIY SLLHLTG YDLPMEELKNFRQLH SKTPGHPEVGYTAGVETTTG PLGQGIANAV
GMAIAEKT LAAQFNRP GHDIVDHYTYAFMG DGCMMEG I SHEVCSLAGTLKLGKLI AFYDDNG
I SIDGHVEGWFTDDTAMRFEAYGWHVIR D IDGHDAAS I KRAVEEARAVTDKPSLLMCKT I IG
FGSPNKAGTHDSHGAPLGDAEIAL TREQLGWKYAPFEIPSEIYAQWDAKEAGQAKESAWNEK
FAAYAKAYPQEAAEFTRRMKGEMPSDFDAKAKEF IAKLQANPAKIASRKASQNAIEAFGPLL
PEFLGGSADLAPSNLTLWSGSKAINEDAAGNYIHYGVREFGMTAIANGISLHGGFLPYTSTF
LMFVEYARNAVRMAALMKQRQVMVYTHDSIGLGEDGPTHQPVEQVASLRVTPNMSTWRPCDQ
VESAVAWKYGVERQDGPTALILSRQNL AQQERTEEQ LANIARGGYVLKDCAGQPELIFIATG
SEVELAVAAYEKLTAEGVKARVVSMPSTDAFDKQDAAYRESVLPKAVTARVAVEAGIADYWY
KYVGLNGAIVGMTTFGELEVL FQGPSAPAE LLFEEFGFTVDNVVAKAKELL

SEQ ID NO:67

MSSRKELANAI RAL SMDAVQKAKSGHPGAPMG MADI AEVLWRDFLKHNPQNPSWADRDRFVL
SNGHGSMLIY SLLHLTG YDLPMEELKNFRQLH SKTPGHPEVGYTAGVETTTG PLGQGIANAV
GMAIAEKT LAAQFNRP GHDIVDHYTYAFMG DGCMMEG I SHEVCSLAGTLKLGKLI AFYDDNG
I SIDGHVEGWFTDDTAMRFEAYGWHVIR D IDGHDAAS I KRAVEEARAVTDKPSLLMCKT I IG
FGSPNKAGTHDSHGAPLGDAEIAL TREQLGWKYAPFEIPSEIYAQWDAKEAGQAKESAWNEK

FAAYAKAYPQEAAEFTRRMKGEMPSTDFDAKAKEFIAKLQANPAKIASRKASQNAIEAFGPLL
PEFLGGSADLAPSNLTLWSGSKAINEDAAGNYIHYGVREFGMTAIANGISLHGGFLPYTSTF
LMFVEYARNAVRMAALMKQRQVMVYTHDSIGLGEDGPTHQPVEQVASLRVTPNMSTWRPCDQ
VESAVAWKYGVERQDGPTALILSRQNLAQQERTEEQLANIARGGYVLKDCAGQPELIFIATG
SEVELAVAAYEKLTAEGVKARVVSMPSTDAFDKQDAAYRESVLPKAVTARVAVEAGIADYWY
KYVGLNGAIVGMTTFGESLEVLFQGPAPAEELLFEEFGFTVDNVVAKAKELL

SEQ ID NO:68

MSSRKELANAI RALSMDAVQKAKSGHPGAPMGMA DIAEVLWRDFLKHNPQNPSWADRDRFVL
SNGHGSMLIYSLHLTG YDLPMEELKNFRQLH SKTPGHPEVGYTAGVETTTGPLGQGIANAV
GMAIAEKTLLAAQFNRP GHDIVDHYTYAFMGD GCMMEGISHVCSLAGTLKLGLKLI AFYDDNG
ISIDGHVEGWFTDDTAMRFEAYGWHVIR DIDGHDAASIKRAVEEARAVTDKPSLLMCKTIIG
FGSPNKAGTHDSHGAPLGDAEIALTREQLGWKYAPFEIPSEIYAQWDAKEAGQAKESAWNEK
FAAYAKAYPQEAAEFTRRMKGEMPSTDFDAKAKEFIAKLQANPAKIASRKASQNAIEAFGPLL
PEFLGGSADLAPSNLTLWSGSKAINEDAAGNYIHYGVREFGMTAIANGISLHGGFLPYTSTF
LMFVEYARNAVRMAALMKQRQVMVYTHDSIGLGEDGPTHQPVEQVASLRVTPNMSTWRPCDQ
VESAVAWKYGVERQDGPTALILSRQNLAQQERTEEQLANIARGGYVLKDCAGQPELIFIATG
SEVELAVAAYEKLTAEGVKARVVSMPSTDAFDKQDAAYRESVLPKAVTARVAVEAGIADYWY
KYVGLNGAIVGMTTFGESAPLEVLFQGPAPAEELLFEEFGFTVDNVVAKAKELL

SEQ ID NO:69

CTGGAAGTGCTGTTTCAGGGTCCG

SEQ ID NO:70

MKNINPTQTAA

SEQ ID NO:71

ATGAAAAACATCAATCCAACGCAGACCGCTGCC

SEQ ID NO:72

MLEVLFQGPKNINPTQTAA

SEQ ID NO:73

ATGCTGGAAGTGCTGTTTCAGGGTCCGAAAAACATCAATCCAACGCAGACCGCTGCC

SEQ ID NO:74

MKLEVLFQGPNINPTQTAA

SEQ ID NO:75

ATGAAACTGGAAGTGCTGTTTCAGGGTCCGAACATCAATCCAACGCAGACCGCTGCC

SEQ ID NO:76

MKNLEVLFFQGPINPTQTAA

SEQ ID NO:77

ATGAAAAACCTGGAAGTGCTGTTTCAGGGTCCGATCAATCCAACGCAGACCGCTGCC

SEQ ID NO:78

MLEVLFFQGPAA

SEQ ID NO:79

ATGCTGGAAGTGCTGTTTCAGGGTCCGGCTGCC

SEQ ID NO:80

MKLEVLFFQGPA

SEQ ID NO:81

ATGAAACTGGAAGTGCTGTTTCAGGGTCCGGCC

SEQ ID NO:82

MKNLEVLFFQGP

SEQ ID NO:83

ATGAAAAACCTGGAAGTGCTGTTTCAGGGTCCG

SEQ ID NO:84

MKNINLEVLFFQGPTQTAA

SEQ ID NO:85

ATGAAAAACATCAATCTGGAAGTGCTGTTTCAGGGTCCAACGCAGACCGCTGCC

SEQ ID NO:86

TAAWQALEVLFFQGPQKH

SEQ ID NO:87

ACCGCTGCCTGGCAGGCACTAGAAGTGCTGTTTCAGGGTCCGCAGAAACAC

What is claimed is:

CLAIMS

1. A recombinant phosphoglucose isomerase protein with a protease recognition sequence located between two amino acids of the wild-type protein.
2. The recombinant protein of claim 1, wherein the protease recognition sequence is the protease recognition sequence recognized by a protease selected from the group consisting of alanine carboxypeptidase, *Armillaria mellea* astacin, bacterial leucyl aminopeptidase, cancer procoagulant, cathepsin B, clostripain, cytosol alanyl aminopeptidase, elastase, endoproteinase Arg-C, enterokinase, gastricsin, gelatinase, Gly-X carboxypeptidase, glycyl endopeptidase, human rhinovirus 3C protease, hypodermin C, Iga-specific serine endopeptidase, leucyl aminopeptidase, leucyl endopeptidase, lysC, lysosomal pro-X carboxypeptidase, lysyl aminopeptidase, methionyl aminopeptidase, myxobacter, nardilysin, pancreatic endopeptidase E, picornain 2A, picornain 3C, proendopeptidase, prolyl aminopeptidase, proprotein convertase I, proprotein convertase II, russellysin, saccharopepsin, semenogelase, T-plasminogen activator, thrombin, tissue kallikrein, tobacco etch virus (TEV), togavirin, tryptophanyl aminopeptidase, U-plasminogen activator, V8, venombin A, venombin AB, and Xaa-pro aminopeptidase.
3. The recombinant protein of claim 2, wherein the protease recognition sequence is human rhinovirus 3C protease.
4. The recombinant protein of claim 3, wherein the human rhinovirus 3C protease recognition sequence comprises the sequence of SEQ ID NO:38.
5. The recombinant protein of any one of claims 1-4, wherein the protease recognition sequence is located in a C terminal region of the protein.
6. The recombinant protein of any one of claims 1-4, wherein the protease recognition sequence is located in a solvent-exposed loop region of the protein.
7. The recombinant protein of any one of claims 1-6, wherein the protease recognition sequence is located after amino acid 410, 524, 525, 526, 527, 528, 529, 530, 531, 532, or 545 of the sequence of SEQ ID NO:17.
8. The recombinant protein of claim 7, wherein the protease recognition sequence is located after amino acid 526 of the sequence of SEQ ID NO:17.
9. The recombinant protein of any one of claims 1-8, wherein the recombinant protein comprises the sequence of SEQ ID NO:25.

10. A nucleic acid encoding the recombinant protein of any one of claims 1-9.
11. A vector comprising the nucleic acid of claim 10.
12. The vector of claim 11, wherein the vector is a cloning vector or an expression vector.
13. The vector of claim 12, wherein the vector is a plasmid, a fosmid, a phagemid, a virus genome or an artificial chromosome.
14. The vector of claim 13, wherein the vector is a plasmid.
15. A cell comprising the vector of any one of claims 11-14.
16. The cell of claim 15, wherein the cell is a bacterial cell, a fungal cell, a mammalian cell or a plant cell.
17. The cell of claim 16, wherein the cell is a bacterial cell.
18. The cell of claim 17, wherein the bacterial cell is an *Escherichia coli* cell.
19. A cell that expresses the recombinant protein of any one of claims 1-9.
20. A lysate of the cell of any one of claims 15-19.
21. A nucleic acid encoding a recombinant phosphoglucose isomerase protein with a protease recognition sequence located between two codons of the wild-type nucleic acid.
22. The nucleic acid of claim 21, wherein the protease recognition sequence is the protease recognition sequence recognized by a protease selected from the group consisting of alanine carboxypeptidase, *Armillaria mellea* astacin, bacterial leucyl aminopeptidase, cancer procoagulant, cathepsin B, clostripain, cytosol alanyl aminopeptidase, elastase, endoproteinase Arg-C, enterokinase, gastricsin, gelatinase, Gly-X carboxypeptidase, glycyl endopeptidase, human rhinovirus 3C protease, hypodermin C, Iga-specific serine endopeptidase, leucyl aminopeptidase, leucyl endopeptidase, lysC, lysosomal pro-X carboxypeptidase, lysyl aminopeptidase, methionyl aminopeptidase, myxobacter, nardilysin, pancreatic endopeptidase E, picornain 2A, picornain 3C, proendopeptidase, prolyl aminopeptidase, proprotein convertase I, proprotein convertase II, russellysin, saccharopepsin, semenogelase, T-plasminogen activator, thrombin, tissue kallikrein, tobacco etch virus (TEV), togavirin, tryptophanyl aminopeptidase, U-plasminogen activator, V8, venombin A, venombin AB, and Xaa-pro aminopeptidase.

23. The nucleic acid of claim 22, wherein the protease recognition sequence is human rhinovirus 3C protease.
24. The nucleic acid of claim 23, wherein the human rhinovirus 3C protease recognition sequence comprises the sequence of SEQ ID NO:37.
25. The nucleic acid of any one of claims 21-24, wherein the protease recognition sequence is located in a C terminal region of the protein.
26. The nucleic acid of any one of claims 21-24, wherein the protease recognition sequence is located in a region corresponding to a solvent-exposed loop region of the encoded recombinant phosphoglucose isomerase protein.
27. The nucleic acid of any one of claims 21-26, wherein the protease recognition sequence is located after codons 410, 524, 525, 526, 527, 528, 529, 530, 531, 532, or 545 of the sequence of SEQ ID NO:1.
28. The nucleic acid of claim 27, wherein the protease recognition sequence is located after codon 526 of the sequence of SEQ ID NO:1.
29. The nucleic acid of any one of claims 21-28, wherein the nucleic acid comprises the sequence of SEQ ID NO:9.
30. A recombinant phosphoglucose isomerase protein encoded by the nucleic acid of any one of claims 21-29.
31. A vector comprising the nucleic acid of any one of claims 21-29.
32. The vector of claim 31, wherein the vector is a cloning vector or an expression vector.
33. The vector of claim 32, wherein the vector is a plasmid, a fosmid, a phagemid, a virus genome, or an artificial chromosome.
34. The vector of claim 33, wherein the vector is a plasmid.
35. A cell comprising the vector of any one of claims 31-34.
36. The cell of claim 35, wherein the cell is a bacterial cell, a fungal cell, a mammalian cell, or a plant cell.
37. The cell of claim 36, wherein the cell is a bacterial cell.
38. The cell of claim 37, wherein the bacterial cell is an *Escherichia coli* cell.
39. A cell that expresses recombinant phosphoglucose isomerase protein of claim 30.

40. A lysate of the cell of any one of claims 35-39.
41. A recombinant protein with at least one protease recognition sequence located between two amino acids of the wild-type protein in a solvent-exposed loop region of the protein, wherein the protease recognition sequence is cleaved by a cognate protease with single recognition sequence specificity, and wherein the recombinant protein activity in the presence of the cognate protease is reduced by at least 30% of the recombinant protein activity in the absence of the cognate protease.
42. The recombinant protein of claim 41, wherein the recombinant protein is a recombinant phosphoglucose isomerase protein.
43. The recombinant protein of claim 41 or 42, wherein the protease recognition sequence is the protease recognition sequence recognized by a protease selected from the group consisting of alanine carboxypeptidase, *Armillaria mellea*, astacin, bacterial leucyl aminopeptidase, cancer procoagulant, cathepsin B, clostripain, cytosol alanyl aminopeptidase, elastase, endoproteinase Arg-C, enterokinase, gastricsin, gelatinase, Gly-X carboxypeptidase, glycyl endopeptidase, human rhinovirus 3C protease, hypodermin C, Iga-specific serine endopeptidase, leucyl aminopeptidase, leucyl endopeptidase, lysC, lysosomal pro-X carboxypeptidase, lysyl aminopeptidase, methionyl aminopeptidase, myxobacter, nardilysin, pancreatic endopeptidase E, picornain 2A, picornain 3C, proendopeptidase, prolyl aminopeptidase, proprotein convertase I, proprotein convertase II, russellysin, saccharopepsin, semenogelase, T-plasminogen activator, thrombin, tissue kallikrein, tobacco etch virus (TEV), togavirin, tryptophanyl aminopeptidase, U-plasminogen activator, V8, venombin A, venombin AB, Factor X protease, HIV protease, and Xaa-pro aminopeptidase.
44. The recombinant protein of claim 43, wherein the protease recognition sequence is a human rhinovirus 3C protease recognition sequence.
45. The recombinant protein of claim 44, wherein the human rhinovirus 3C protease recognition sequence comprises the sequence of SEQ ID NO:38.
46. A method of engineering a recombinant protein with a protease recognition sequence, the method comprising the steps of:
- (a) transforming cells with

a plurality of nucleic acid variants, wherein each nucleic acid variant contains a nucleotide sequence that encodes a recombinant protein with at least one protease recognition sequence located between two codons of the wild-type nucleic acid, and a nucleic acid with an inducible promoter sequence operably linked to a nucleotide sequence encoding a cognate protease;

- (b) culturing the cells under non-inducing conditions on media that prohibits cell growth of inactive recombinant proteins and collecting cells with a growth rate comparable to wild-type control cells; and
- (c) culturing the cells under conditions that induce expression of the cognate protease and collecting cells that do not grow or that have a reduced growth rate in comparison to wild-type control cells.

47. The method of claim 46, wherein the cells are transformed with the nucleic acid with an inducible promoter sequence after step (b) and before step (c).

48. The method of claim 46 or 47, further comprising sequencing the nucleic acid variants of the cells collected in step (c) to identify the location of the protease recognition sequence.

49. The method of any one of claims 46-48, wherein the wild-type protein is required for cell growth and the genome of the cell lacks a chromosomal copy of the wild-type gene encoding the wild-type protein.

50. The method of any one of claims 46-49, wherein the recombinant protein is a recombinant phosphoglucose isomerase protein, a recombinant phosphotransacetylase protein, or a recombinant transketolase A protein.

51. The method of any one of claims 46-50, wherein the protease recognition sequence is the protease recognition sequence recognized by a protease selected from the group consisting of alanine carboxypeptidase, *Armillaria mellea*, astacin, bacterial leucyl aminopeptidase, cancer procoagulant, cathepsin B, clostripain, cytosol alanyl aminopeptidase, elastase, endoproteinase Arg-C, enterokinase, gastricsin, gelatinase, Gly-X carboxypeptidase, glycyl endopeptidase, human rhinovirus 3C protease, hypodermin C, Iga-specific serine endopeptidase, leucyl aminopeptidase, leucyl endopeptidase, lysC, lysosomal pro-X carboxypeptidase, lysyl aminopeptidase, methionyl aminopeptidase, myxobacter, nardilysin, pancreatic endopeptidase E, picornain 2A, picornain 3C, proendopeptidase, prolyl aminopeptidase, proprotein convertase I, proprotein convertase II, russellysin,

saccharopepsin, semenogelase, T-plasminogen activator, thrombin, tissue kallikrein, tobacco etch virus (TEV), togavirin, tryptophanyl aminopeptidase, U-plasminogen activator, venombin A, venombin AB, Factor X protease, HIV protease, and Xaa-pro aminopeptidase.

52. The method of claim 51, wherein the protease recognition sequence is a human rhinovirus 3C protease recognition sequence.

53. The method of claim 52, wherein the human rhinovirus 3C protease recognition sequence comprises the sequence of SEQ ID NO:37.

54. A method of engineering a plurality of nucleic acid variants encoding recombinant proteins, the method comprising inserting a sequence that encodes a protease recognition sequence after each codon of a nucleic acid encoding a target protein to produce a plurality of nucleic acid variants encoding recombinant proteins, wherein each recombinant protein has a protease recognition sequence at a unique location in its primary amino acid sequence.

55. A method of engineering a recombinant protein with a protease recognition sequence, the method comprising the steps of:

- (a) transforming cells with a plurality of nucleic acid variants produced by the method of claim 54, and a nucleic acid with an inducible promoter sequence operably linked to a nucleotide sequence encoding a cognate protease;
- (b) culturing the cells under non-inducing conditions on media that prohibits cell growth of inactive recombinant proteins and collecting cells with a normal growth rate; and
- (c) culturing the cells under conditions that induce expression of the cognate protease and collecting cells that do not grow or that have a reduced growth rate.

56. The method of claim 55, wherein the cells are transformed with the nucleic acid with an inducible promoter sequence after step (b) and before step (c).

57. The method of claim 55 or 56, further comprising sequencing the nucleic acid variants of the cells collected in step (c) to identify the location of the protease recognition sequence.

58. The method of any one of claims 55-57, wherein the wild-type protein corresponding to the recombinant proteins is required for cell growth and the genome of the cell lacks a chromosomal copy of the wild-type gene encoding the wild-type protein.

59. The method of any one of claims 55-8, wherein the recombinant protein is a recombinant phosphoglucose isomerase protein, a recombinant phosphotransacetylase protein, or a recombinant transketolase A protein.
60. The method of any one of claims 55-59, wherein the protease recognition sequence is the protease recognition sequence recognized by a protease selected from the group consisting of alanine carboxypeptidase, *Armillaria mellea*, astacin, bacterial leucyl aminopeptidase, cancer procoagulant, cathepsin B, clostripain, cytosol alanyl aminopeptidase, elastase, endoproteinase Arg-C, enterokinase, gastricsin, gelatinase, Gly-X carboxypeptidase, glycyl endopeptidase, human rhinovirus 3C protease, hypodermin C, Iga-specific serine endopeptidase, leucyl aminopeptidase, leucyl endopeptidase, lysC, lysosomal pro-X carboxypeptidase, lysyl aminopeptidase, methionyl aminopeptidase, myxobacter, nardilysin, pancreatic endopeptidase E, picornain 2A, picornain 3C, proendopeptidase, prolyl aminopeptidase, proprotein convertase I, proprotein convertase II, russellysin, saccharopepsin, semenogelase, T-plasminogen activator, thrombin, tissue kallikrein, tobacco etch virus (TEV), togavirin, tryptophanyl aminopeptidase, U-plasminogen activator, V8, venombin A, venombin AB, Factor X protease, HIV protease, and Xaa-pro aminopeptidase.
61. The method of claim 60, wherein the protease recognition sequence is a human rhinovirus 3C protease recognition sequence.
62. The method of claim 61, wherein the human rhinovirus 3C protease recognition sequence comprises the sequence of SEQ ID NO:37.
63. A heterogeneous plurality of nucleic acid variants, wherein each nucleic acid variant encodes a recombinant protein that is modified to include at least one protease recognition sequence located between two amino acids of the wild-type protein.
64. A heterogeneous cell population, wherein each cell of the population comprises:
a nucleic acid variant, wherein each nucleic acid variant encodes a recombinant protein that is modified to include at least one protease recognition sequence located between two amino acids of the wild-type protein; and
a nucleic acid encoding a cognate protease operably linked to an inducible promoter.
65. A recombinant phosphotransacetylase protein with a protease recognition sequence located between two amino acids of the wild-type protein.

66. The recombinant protein of claim 65, wherein the protease recognition sequence is the protease recognition sequence recognized by a protease selected from the group consisting of alanine carboxypeptidase, *Armillaria mellea* astacin, bacterial leucyl aminopeptidase, cancer procoagulant, cathepsin B, clostripain, cytosol alanyl aminopeptidase, elastase, endoproteinase Arg-C, enterokinase, gastricsin, gelatinase, Gly-X carboxypeptidase, glycyl endopeptidase, human rhinovirus 3C protease, hypodermin C, Iga-specific serine endopeptidase, leucyl aminopeptidase, leucyl endopeptidase, lysC, lysosomal pro-X carboxypeptidase, lysyl aminopeptidase, methionyl aminopeptidase, myxobacter, nardilysin, pancreatic endopeptidase E, picornain 2A, picornain 3C, proendopeptidase, prolyl aminopeptidase, proprotein convertase I, proprotein convertase II, russellysin, saccharopepsin, semenogelase, T-plasminogen activator, thrombin, tissue kallikrein, tobacco etch virus (TEV), togavirin, tryptophanyl aminopeptidase, U-plasminogen activator, V8, venombin A, venombin AB, and Xaa-pro aminopeptidase.
67. The recombinant protein of claim 66, wherein the protease recognition sequence is human rhinovirus 3C protease.
68. The recombinant protein of claim 67, wherein the human rhinovirus 3C protease recognition sequence comprises the sequence of SEQ ID NO:38.
69. The recombinant protein of any one of claims 65-68, wherein the protease recognition sequence is located after amino acid 381, 382, 387, or 409 of the sequence of SEQ ID NO:48.
70. The recombinant protein of claim 69, wherein the protease recognition sequence is located after amino acid 381 of the sequence of SEQ ID NO:48.
71. The recombinant protein of any one of claims 65-70, wherein the recombinant protein comprises the sequence of SEQ ID NO:50.
72. A nucleic acid encoding the recombinant protein of any one of claims 65-71.
73. A vector comprising the nucleic acid of claim 72.
74. The vector of claim 73, wherein the vector is a cloning vector or an expression vector.
75. The vector of claim 74, wherein the vector is a plasmid, a fosmid, a phagemid, a virus genome or an artificial chromosome.
76. The vector of claim 75, wherein the vector is a plasmid.
77. A cell comprising the vector of any one of claims 73-76.

78. The cell of claim 77, wherein the cell is a bacterial cell, a fungal cell, a mammalian cell or a plant cell.
79. The cell of claim 78, wherein the cell is a bacterial cell.
80. The cell of claim 79, wherein the bacterial cell is an *Escherichia coli* cell.
81. A cell that expresses the recombinant protein of any one of claims 65-71.
82. A lysate of the cell of any one of claims 77-82.
83. A nucleic acid encoding a recombinant phosphotransacetylase protein with a protease recognition sequence located between two codons of the wild-type nucleic acid.
84. The nucleic acid of claim 83, wherein the protease recognition sequence is the protease recognition sequence recognized by a protease selected from the group consisting of alanine carboxypeptidase, *Armillaria mellea* astacin, bacterial leucyl aminopeptidase, cancer procoagulant, cathepsin B, clostripain, cytosol alanyl aminopeptidase, elastase, endoproteinase Arg-C, enterokinase, gastricsin, gelatinase, Gly-X carboxypeptidase, glycyl endopeptidase, human rhinovirus 3C protease, hypodermin C, Iga-specific serine endopeptidase, leucyl aminopeptidase, leucyl endopeptidase, lysC, lysosomal pro-X carboxypeptidase, lysyl aminopeptidase, methionyl aminopeptidase, myxobacter, nardilysin, pancreatic endopeptidase E, picornain 2A, picornain 3C, proendopeptidase, prolyl aminopeptidase, proprotein convertase I, proprotein convertase II, russellysin, saccharopepsin, semenogelase, T-plasminogen activator, thrombin, tissue kallikrein, tobacco etch virus (TEV), togavirin, tryptophanyl aminopeptidase, U-plasminogen activator, V8, venombin A, venombin AB, and Xaa-pro aminopeptidase.
85. The nucleic acid of claim 84, wherein the protease recognition sequence is human rhinovirus 3C protease.
86. The nucleic acid of claim 85, wherein the human rhinovirus 3C protease recognition sequence comprises the sequence of SEQ ID NO:37.
87. The nucleic acid of any one of claims 83-86, wherein the protease recognition sequence is located after codons 381, 382, 387, or 409 of the sequence of SEQ ID NO:47.
88. The nucleic acid of claim 87, wherein the protease recognition sequence is located after codon 381 of the sequence of SEQ ID NO:47.

89. The nucleic acid of any one of claims 83-88, wherein the nucleic acid comprises the sequence of SEQ ID NO:49.
90. A recombinant phosphotransacetylase protein encoded by the nucleic acid of any one of claims 83-89.
91. A vector comprising the nucleic acid of any one of claims 83-89.
92. The vector of claim 91, wherein the vector is a cloning vector or an expression vector.
93. The vector of claim 92, wherein the vector is a plasmid, a fosmid, a phagemid, a virus genome, or an artificial chromosome.
94. The vector of claim 93, wherein the vector is a plasmid.
95. A cell comprising the vector of any one of claims 91-94.
96. The cell of claim 95, wherein the cell is a bacterial cell, a fungal cell, a mammalian cell, or a plant cell.
97. The cell of claim 96, wherein the cell is a bacterial cell.
98. The cell of claim 97, wherein the bacterial cell is an *Escherichia coli* cell.
99. A cell that expresses recombinant phosphotransacetylase protein of claim 90.
100. A lysate of the cell of any one of claims 95-99.
101. A recombinant transketolase A protein with a protease recognition sequence located between two amino acids of the wild-type protein.
102. The recombinant protein of claim 101, wherein the protease recognition sequence is the protease recognition sequence recognized by a protease selected from the group consisting of alanine carboxypeptidase, *Armillaria mellea* astacin, bacterial leucyl aminopeptidase, cancer procoagulant, cathepsin B, clostripain, cytosol alanyl aminopeptidase, elastase, endoproteinase Arg-C, enterokinase, gastricsin, gelatinase, Gly-X carboxypeptidase, glycyl endopeptidase, human rhinovirus 3C protease, hypodermin C, Iga-specific serine endopeptidase, leucyl aminopeptidase, leucyl endopeptidase, lysC, lysosomal pro-X carboxypeptidase, lysyl aminopeptidase, methionyl aminopeptidase, myxobacter, nardilysin, pancreatic endopeptidase E, picornain 2A, picornain 3C, proendopeptidase, prolyl aminopeptidase, proprotein convertase I, proprotein convertase II, russellysin, saccharopepsin, semenogelase, T-plasminogen activator, thrombin, tissue kallikrein, tobacco

etch virus (TEV), togavirin, tryptophanyl aminopeptidase, U-plasminogen activator, V8, venombin A, venombin AB, and Xaa-pro aminopeptidase.

103. The recombinant protein of claim 102, wherein the protease recognition sequence is human rhinovirus 3C protease.

104. The recombinant protein of claim 103, wherein the human rhinovirus 3C protease recognition sequence comprises the sequence of SEQ ID NO:38.

105. The recombinant protein of any one of claims 101-104, wherein the protease recognition sequence is located after amino acid 635, 636, 637, 638, or 640 of the sequence of SEQ ID NO:63.

106. The recombinant protein of claim 105, wherein the protease recognition sequence is located after amino acid 637 of the sequence of SEQ ID NO:63.

107. The recombinant protein of any one of claims 101-106, wherein the recombinant protein comprises the sequence of SEQ ID NO:66.

108. A nucleic acid encoding the recombinant protein of any one of claims 101-107.

109. A vector comprising the nucleic acid of claim 108.

110. The vector of claim 109, wherein the vector is a cloning vector or an expression vector.

111. The vector of claim 110, wherein the vector is a plasmid, a fosmid, a phagemid, a virus genome or an artificial chromosome.

112. The vector of claim 111, wherein the vector is a plasmid.

113. A cell comprising the vector of any one of claims 109-112.

114. The cell of claim 113, wherein the cell is a bacterial cell, a fungal cell, a mammalian cell or a plant cell.

115. The cell of claim 114, wherein the cell is a bacterial cell.

116. The cell of claim 115, wherein the bacterial cell is an *Escherichia coli* cell.

117. A cell that expresses the recombinant protein of any one of claims 101-107.

118. A lysate of the cell of any one of claims 113-117.

119. A nucleic acid encoding a recombinant transketolase A protein with a protease recognition sequence located between two codons of the wild-type nucleic acid.
120. The nucleic acid of claim 119, wherein the protease recognition sequence is the protease recognition sequence recognized by a protease selected from the group consisting of alanine carboxypeptidase, *Armillaria mellea* astacin, bacterial leucyl aminopeptidase, cancer procoagulant, cathepsin B, clostripain, cytosol alanyl aminopeptidase, elastase, endoproteinase Arg-C, enterokinase, gastricsin, gelatinase, Gly-X carboxypeptidase, glycyl endopeptidase, human rhinovirus 3C protease, hypodermin C, Iga-specific serine endopeptidase, leucyl aminopeptidase, leucyl endopeptidase, lysC, lysosomal pro-X carboxypeptidase, lysyl aminopeptidase, methionyl aminopeptidase, myxobacter, nardilysin, pancreatic endopeptidase E, picornain 2A, picornain 3C, proendopeptidase, prolyl aminopeptidase, proprotein convertase I, proprotein convertase II, russellysin, saccharopepsin, semenogelase, T-plasminogen activator, thrombin, tissue kallikrein, tobacco etch virus (TEV), togavirin, tryptophanyl aminopeptidase, U-plasminogen activator, V8, venombin A, venombin AB, and Xaa-pro aminopeptidase.
121. The nucleic acid of claim 120, wherein the protease recognition sequence is human rhinovirus 3C protease.
122. The nucleic acid of claim 121, wherein the human rhinovirus 3C protease recognition sequence comprises the sequence of SEQ ID NO:37.
123. The nucleic acid of any one of claims 119-122, wherein the protease recognition sequence is located after codons 635, 636, 637, 638, or 640 of the sequence of SEQ ID NO:57.
124. The nucleic acid of claim 123, wherein the protease recognition sequence is located after codon 637 of the sequence of SEQ ID NO:57.
125. The nucleic acid of any one of claims 119-124, wherein the nucleic acid comprises the sequence of SEQ ID NO:60.
126. A recombinant transketolase A protein encoded by the nucleic acid of any one of claims 119-125.
127. A vector comprising the nucleic acid of any one of claims 119-125.
128. The vector of claim 127, wherein the vector is a cloning vector or an expression vector.

129. The vector of claim 128, wherein the vector is a plasmid, a fosmid, a phagemid, a virus genome, or an artificial chromosome.
130. The vector of claim 129, wherein the vector is a plasmid.
131. A cell comprising the vector of any one of claims 127-130.
132. The cell of claim 131, wherein the cell is a bacterial cell, a fungal cell, a mammalian cell, or a plant cell.
133. The cell of claim 132, wherein the cell is a bacterial cell.
134. The cell of claim 133, wherein the bacterial cell is an *Escherichia coli* cell.
135. A cell that expresses recombinant transketolase A protein of claim 126.
136. A lysate of the cell of any one of claims 131-135.