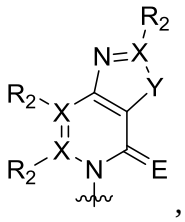
CLAIMS

WHAT IS CLAIMED IS:

1. A nucleobase of the structure:



,

wherein:

each X is independently carbon or nitrogen;

R_2 is present when X is carbon and is independently hydrogen, alkyl, alkenyl, alkynyl, methoxy, methanethiol, methaneseleno, halogen, cyano, or azide group;

Y is sulfur, oxygen, selenium, or secondary amine; and

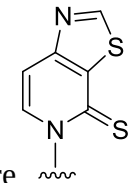
E is oxygen, sulfur, or selenium;

wherein the wavy line indicates a point of bonding to a ribosyl, deoxyribosyl, or dideoxyribosyl moiety or an analog thereof, wherein the ribosyl, deoxyribosyl, or dideoxyribosyl moiety or analog thereof is in free form, is connected to a mono-phosphate, diphosphate, triphosphate, α-thiotriphosphate, β-thiotriphosphate, or γ-thiotriphosphate group, or is included in an RNA or a DNA or in an RNA analog or a DNA analog.

2. The nucleobase of claim 1, wherein X is carbon.

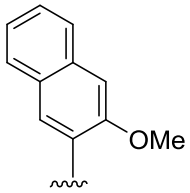
3. The nucleobase of claim 1 or 2, wherein E is sulfur.

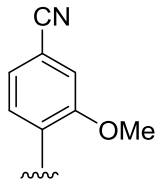
4. The nucleobase of any one of claims 1 to 3, wherein Y is sulfur.

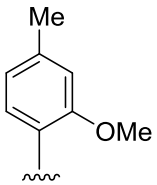
5. The nucleobase of claim 1, which has the structure .

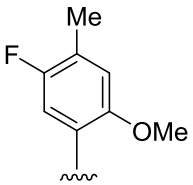
6. The nucleobase of any one of claims 1 to 5, which is bound to a complementary base-pairing nucleobase to form an unnatural base pair (UBP).

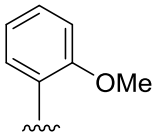
7. The nucleobase of claim 6, wherein the complementary base-pairing nucleobase is selected from:

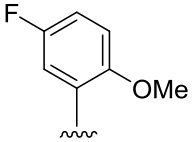


, 

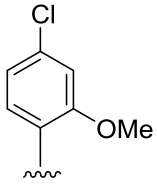
, 

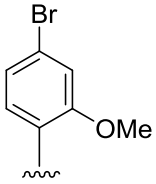
, 

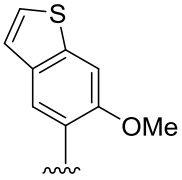
, 

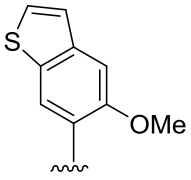
, 

,



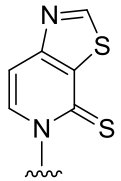
, 

, 

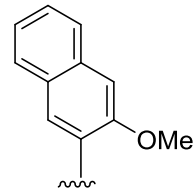
, and 

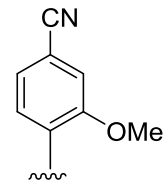
.

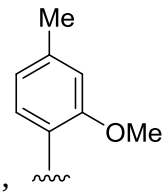
8. A double stranded oligonucleotide duplex wherein a first oligonucleotide strand comprises the nucleobase of any one of claims 1 to 5, and a second complementary oligonucleotide strand comprises a complementary base-pairing nucleobase in a complementary base-pairing site thereof.

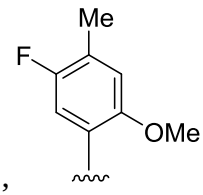
9. The double stranded oligonucleotide duplex of claim 8, wherein the first oligonucleotide strand comprises 

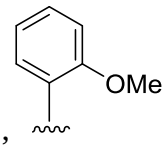
; and the second strand comprises a complementary base pairing nucleobase selected from:

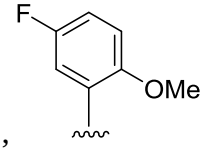


, 

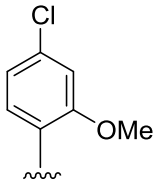
, 

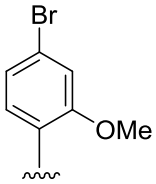
, 

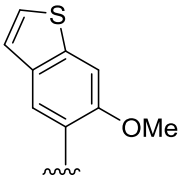
, 

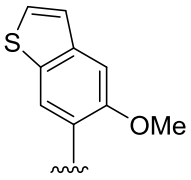
, 

,

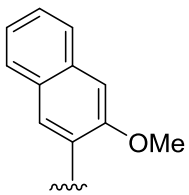


, 

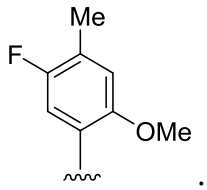
, 

, and 

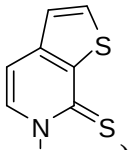
in a complementary base-pairing site thereof.

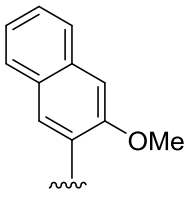
10. The double stranded oligonucleotide duplex of claim 9, wherein the second strand comprises the complementary base pairing nucleobase 

.

11. The double stranded oligonucleotide duplex of claim 9, wherein the second strand comprises the complementary base pairing nucleobase 

.

12. A plasmid comprising a gene encoding a transfer RNA (tRNA) and/or a gene encoding a protein of interest, wherein the gene comprises at least one nucleobase of any one of claims 1 to 5 or TPT3 (

) and at least one complementary base-pairing nucleobase of claim 7 or NaM (

), wherein the complementary base-pairing nucleobase is in a complementary base-pairing site.

13. An mRNA encoded by the plasmid of claim 10 encoding the tRNA.

14. An mRNA encoded by the plasmid of claim 10 encoding the protein.

15. A transfer RNA (tRNA) comprising the nucleobase of any one of claims 1 to 5, comprising:

an anticodon, wherein the anticodon comprises the nucleobase, optionally wherein the nucleobase is at the first, second, or third position of the anticodon; and

a recognition element, wherein the recognition element promotes selective charging of the tRNA with an unnatural amino acid by an aminoacyl tRNA synthetase.

16. The tRNA of claim 15, wherein the aminoacyl tRNA synthetase is derived from *Methanosarcina*, or a variant thereof, or *Methanococcus* (*Methanocaldococcus*) or a variant thereof.

17. The tRNA of claim 15, wherein the unnatural amino acid comprises an aromatic moiety.

18. The tRNA of claim 15, wherein the unnatural amino acid is a lysine or phenylalanine derivative.

19. A structure comprising the formula:

N1 - Zx - N2

wherein:

each Z is independently a nucleobase of any one of claims 1 to 7, which is bonded to a ribosyl or deoxyribosyl or analog thereof;

N1 is one or more nucleotides or analogs thereof, or a terminal phosphate group attached at the 5’ end of the ribosyl or deoxyribosyl or analog thereof of Z;

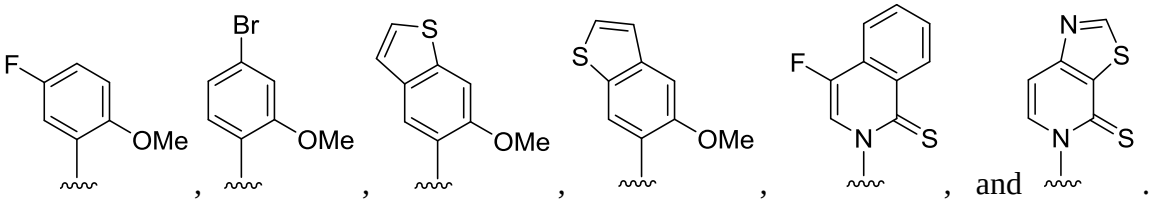
N2 is one or more nucleotides or analogs thereof, or a terminal hydroxyl group attached to the 3’ end of the ribosyl or deoxyribosyl or analog thereof of Z; and

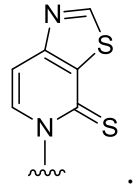
x is an integer from 1 to 20.

20. The structure of claim 19, wherein the structure encodes a gene, optionally wherein Zx is located in a translated region of the gene or wherein Zx is located in an untranslated region of the gene.

21. A polynucleotide library, wherein the library comprises at least 5000 unique polynucleotides, and wherein each polynucleotide comprises at least one nucleobase of any one of claims 1 to 5.

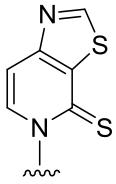
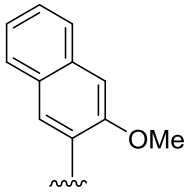
22. A nucleoside triphosphate comprising a nucleobase, wherein the nucleobase is selected from:

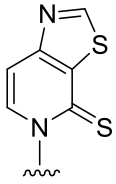
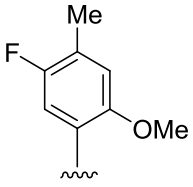




23. The nucleoside triphosphate of claim 22, wherein the nucleobase is .

24. The nucleoside triphosphate of claim 22 or 23, wherein the nucleoside comprises ribose or deoxyribose.

25. A DNA comprising a nucleobase having the structure , and a complementary base-pairing nucleobase having the structure .

26. A DNA comprising a nucleobase having the structure , and a complementary base-pairing nucleobase having the structure .

~~27. A method of transcribing a DNA to a tRNA or an mRNA encoding a protein comprising:~~

~~contacting a DNA comprising a gene encoding the tRNA or protein with ribonucleoside triphosphates and an RNA polymerase, wherein the gene encoding the tRNA or protein comprises a first unnatural base paired to and forming a first unnatural base pair with a second unnatural base, and wherein the ribonucleoside triphosphates comprise a third unnatural base capable of forming a second unnatural base pair with the first unnatural base, wherein the first unnatural base pair and the second unnatural base pair are not the same.~~

~~28. The method of claim 27, wherein the ribonucleoside triphosphates further comprise a fourth unnatural base, wherein the fourth unnatural base is capable of forming a second unnatural base pair with the third unnatural base.~~

~~29. The method of claim 28, wherein the first unnatural base pair and the second unnatural base pair are not the same.~~

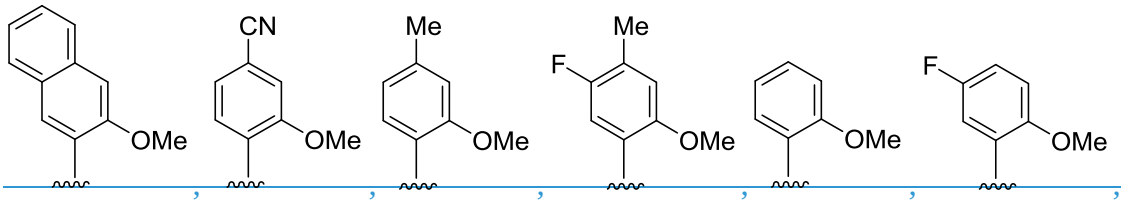
~~30. The method of any one of claims 27-29, further comprising, before contacting the DNA with the ribonucleoside triphosphates and RNA polymerase, replicating the DNA by contacting the DNA with deoxyribonucleoside triphosphates and a DNA polymerase,~~

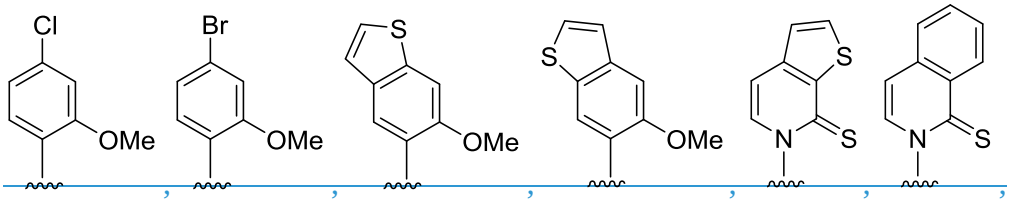
wherein the ribonucleoside triphosphates comprise a fifth unnatural base capable of forming a fifth unnatural base pair with the first unnatural base, wherein the first unnatural base pair and the fifth unnatural base pair are not the same.

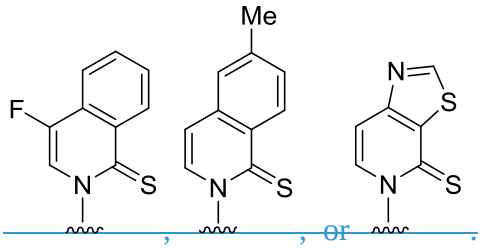
31. The method of any one of claims 27-30, wherein the first unnatural base comprises TPT3, the second unnatural base comprises CNMO or NaM, the third unnatural base comprises TAT1, and the fourth unnatural bases comprises NaM or 5FM.

32. The method of any one of claims 27-31, wherein the method comprises use of a semi-synthetic organism, optionally wherein the organism is a bacterium, optionally wherein the bacterium is *Escherichia coli*.

33. The method of any one of claims 27-32, wherein at least one of the first unnatural base, the second unnatural base, the third unnatural base, or the fourth unnatural base comprises:

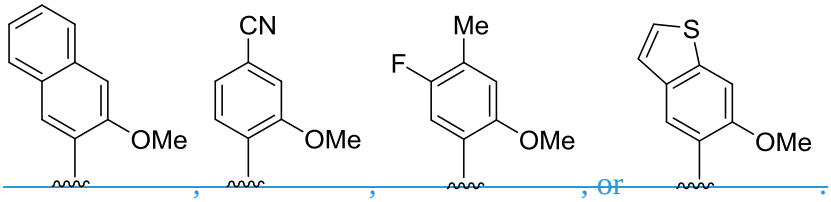






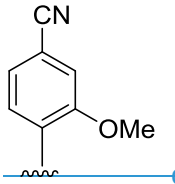
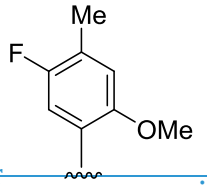
, , , or .

34. The method of claim 33, wherein at least one of the first unnatural base, the second unnatural base, the third unnatural base, or the fourth unnatural base comprises:



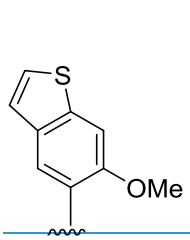
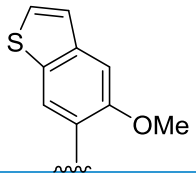
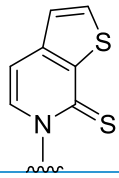
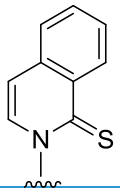
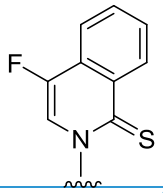
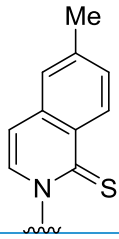
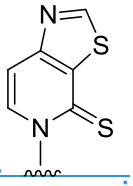
, , , , or .

35. The method of claim 34, wherein at least one of the first unnatural base, the second unnatural base, the third unnatural base, or the fourth unnatural base comprises:

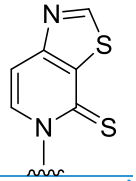



or .

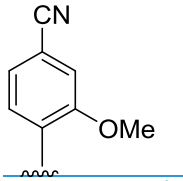
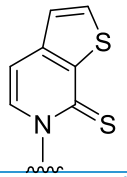
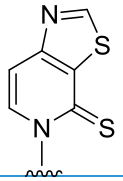
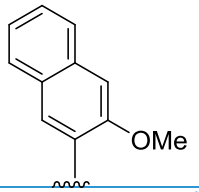
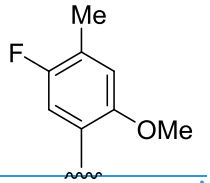
36. The method of claim 34, wherein at least one of the first unnatural base, the second unnatural base, the third unnatural base, or the fourth unnatural base comprises:

, , , , , , or .

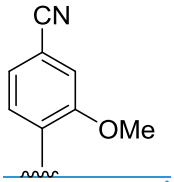
37. The method of claim 34, wherein at least one of the first unnatural base, the second unnatural base, the third unnatural base, or the fourth unnatural base is .

38. The method of claim 34, wherein at least one of the first unnatural base, the second unnatural base, the third unnatural base, or the fourth unnatural base comprises:

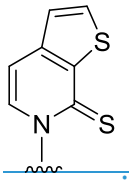
, , , , or .

39. The method of any one of claims 34-38, wherein the first or second unnatural base is

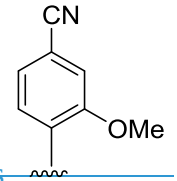
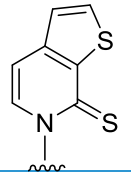
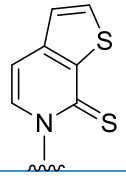
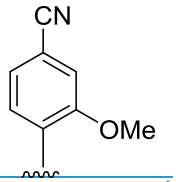


.

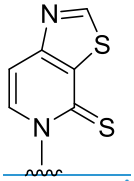
40. The method of any one of claims 34-39, wherein the first or second unnatural base is



.

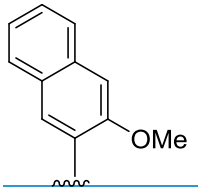
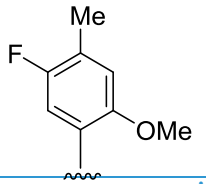
41. The method of any one of claims 34-38, wherein the first unnatural base is  and the second unnatural base is  or the first unnatural base is  and the second unnatural base is .

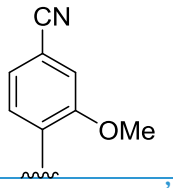
42. The method of any one of claims 34-41, wherein the third or fourth unnatural base is

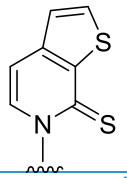


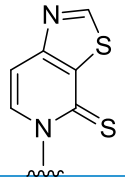
.

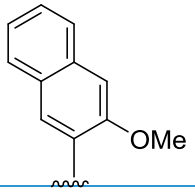
43. The method of any one of claims 34-42, wherein the third or fourth unnatural base is

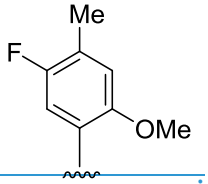
 or .

44. The method of any one of claims 34-38, wherein the first unnatural base is 

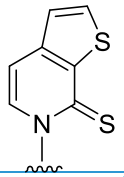
, the second unnatural base is 

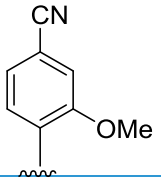
, the third unnatural base is 

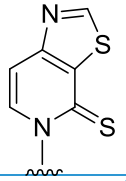
, and the fourth unnatural base is 

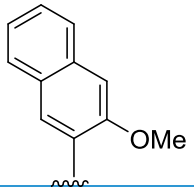
or 

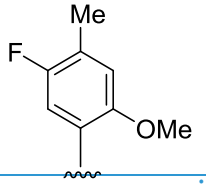
.

45. The method of any one of claims 34-38, wherein the first unnatural base is 

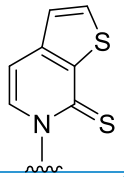
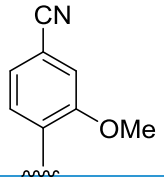
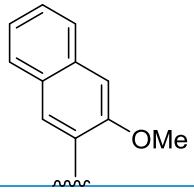
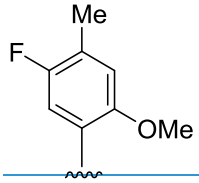
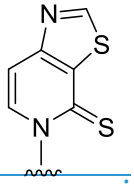
, the second unnatural base is 

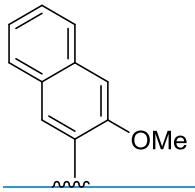
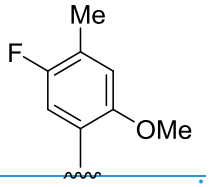
, the third unnatural base is 

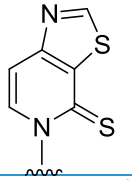
, and the fourth unnatural base is 

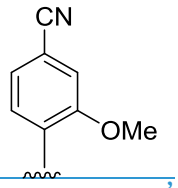
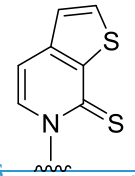
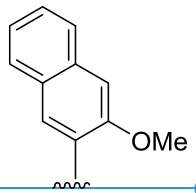
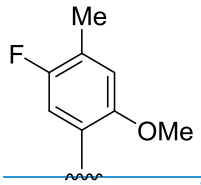
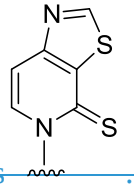
or 

.

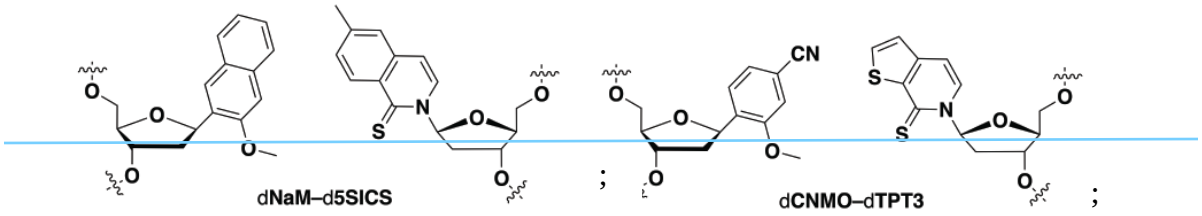
46. The method of any one of claims 34-38, wherein the first unnatural base is , the second unnatural base is , the third unnatural base is  or  and the fourth unnatural base is .

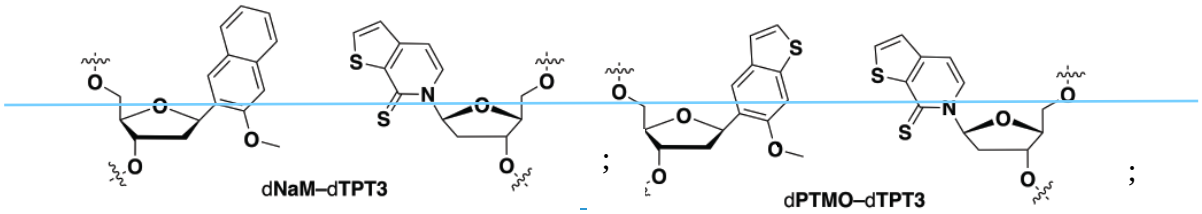
47. The method of any one of claims 34-43, wherein the third unnatural base is  or .

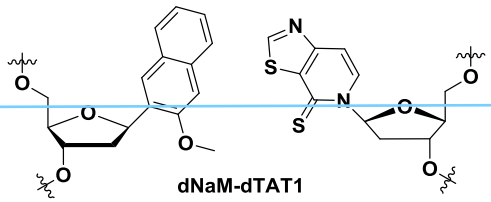
48. The method of any one of claims 34-47, wherein the fourth unnatural base is .

49. The method of any one of claims 34-38, wherein the first unnatural base is , the second unnatural base is , the third unnatural base is  or , and the fourth unnatural base is .

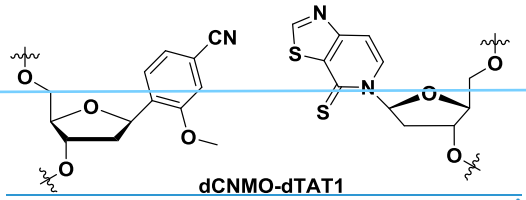
50. The method of any one of claims 34-49, wherein the DNA comprises at least one unnatural base pair (UBP) selected from:



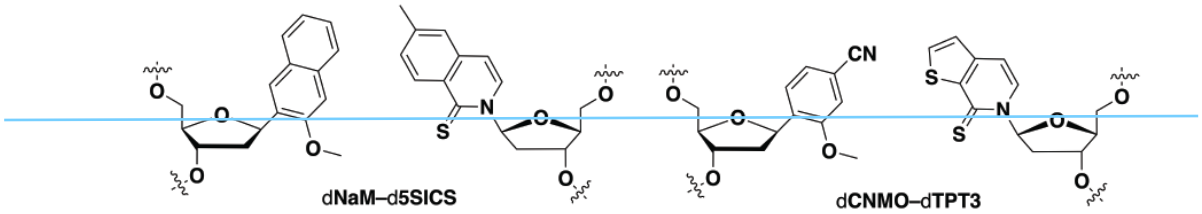


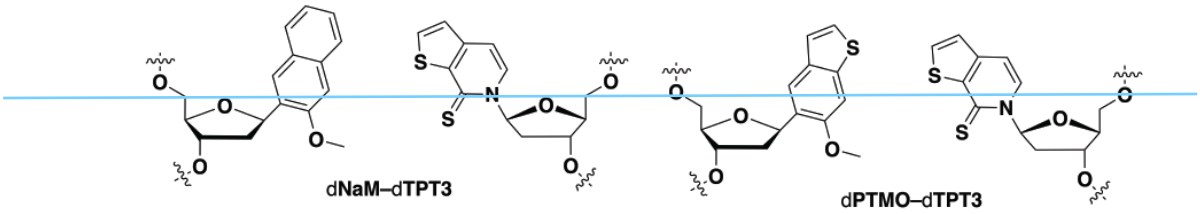


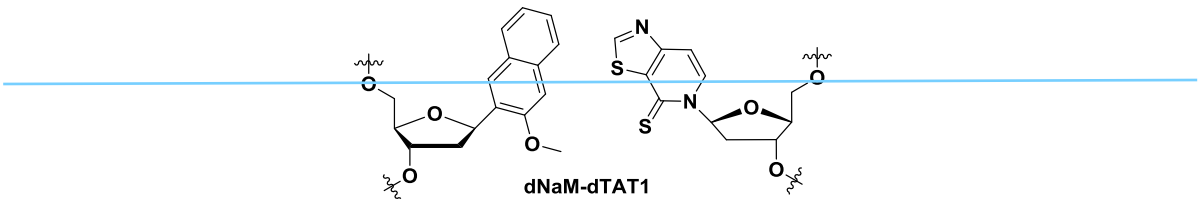
; and

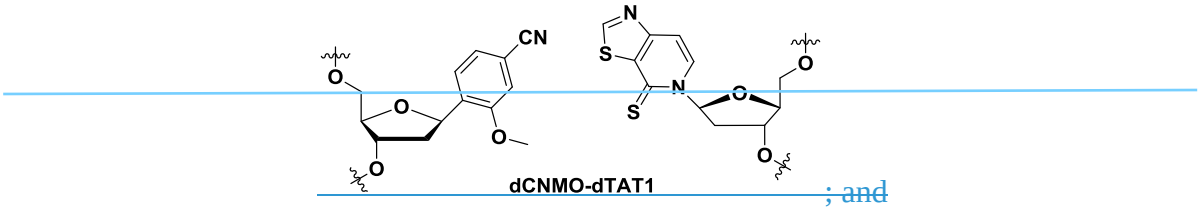


51. The method of any one of claims 34-50, wherein the DNA comprises at least one unnatural base pair (UBP) selected from



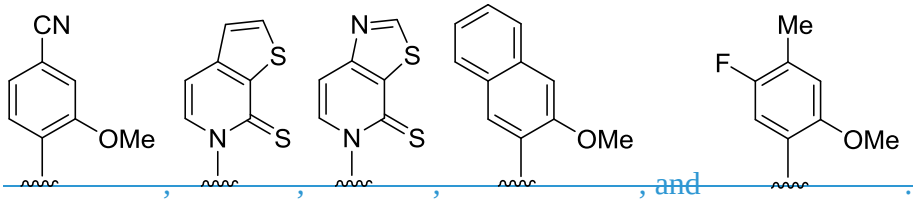






; and

wherein the mRNA and/or the tRNA comprise at least one unnatural base selected from:



, , , , and .

52. The method of any one of the claims 27-51, wherein the first unnatural base comprises dCNMO, and the second unnatural base comprises dTPT3.

53. The method of any one of the claims 27-52, wherein the third unnatural base comprises NaM, and the second unnatural base comprises TAT1.

54. The method of any one of claims 27-53, wherein the mRNA is transcribed, and the method further comprises translating the mRNA into a protein, wherein the protein comprises an unnatural amino acid at a position corresponding to a codon of the mRNA that comprises the third unnatural base.

55. The method of any one of the claims 27-54, wherein the protein comprises at least two unnatural amino acids.

56. The method of any one of the claims 27-54, wherein the protein comprises at least three unnatural amino acids.

57. The method of any one of the claims 27-54, wherein the protein comprises at least two different unnatural amino acids.

58. The method of any one of the claims 27-54, wherein the protein comprises at least three different unnatural amino acids.

59. The method of any one of the claims 27-58, wherein the at least one unnatural amino acid:

is a lysine analogue;

comprises an aromatic side chain;

comprises an azido group;

comprises an alkyne group; or

comprises an aldehyde or ketone group.

60. The method of any one of the claims 27-59, wherein the at least one unnatural amino acid does not comprise an aromatic side chain.

61. The method of claim 59 or 60, wherein the at least one unnatural amino acid comprises N6-azidoethoxy-carbonyl-L-lysine (AzK) or N6-propargylethoxy-carbonyl-L-lysine (PraK).

62. The method of claim 61, wherein the at least one unnatural amino acid comprises N6-azidoethoxy-carbonyl-L-lysine (AzK).

63. The method of claim 61, wherein the at least one unnatural amino acid comprises N6-propargylethoxy-carbonyl-L-lysine (PraK).

64. An mRNA produced from the method of any one of claims 27-63.

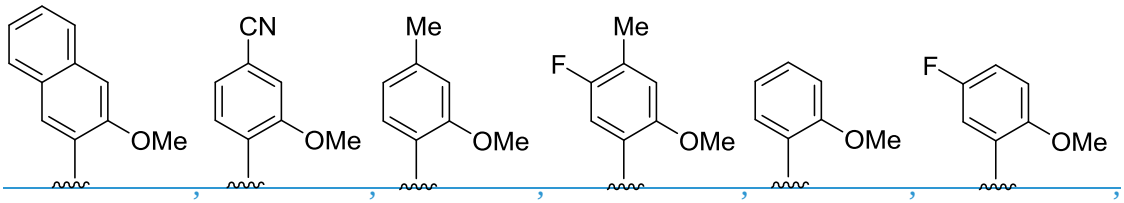
65. A tRNA produced from the method of any one of claims 27-63.

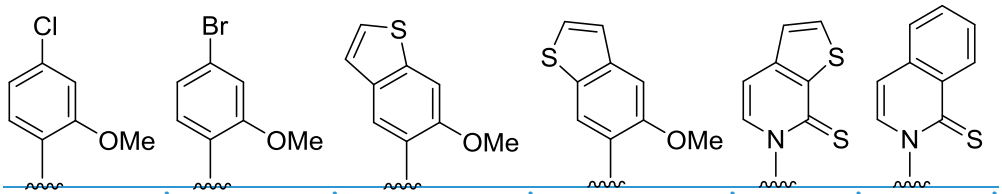
66. A protein encoded by the mRNA of claim 64 comprising an unnatural amino acid at a position corresponding to a codon of the mRNA that comprises the third unnatural base.

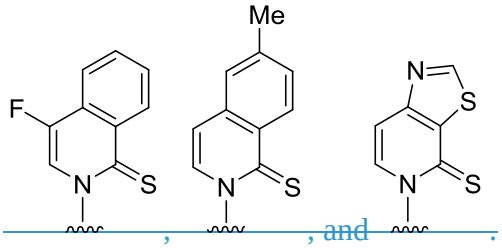
67. A semi-synthetic organism comprising an expanded genetic alphabet, wherein the genetic alphabet comprises at least three distinct unnatural bases.

68. The semi-synthetic organism of claim 67, wherein the organism comprises a microorganism, optionally wherein the microorganism is an *Escherichia coli*.

69. The semi-synthetic organism of claim 67 or 68, wherein the organism comprises DNA comprising at least one unnatural nucleobase selected from:



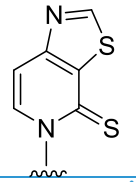




, and .

70. The semi-synthetic organism of any one of the claims 67-69, wherein the DNA comprises at least one unnatural base pair (UBP),

wherein the unnatural base pair (UBP) is dCNMO-dTPT3, dNaM-dTPT3, dCNMO-dTAT1, d5FM-dTAT1, or dNaM-dTAT1.

71. The semi-synthetic organism of claim 67, wherein the DNA comprises at least one unnatural nucleobase which is 

.

72. The semi-synthetic organism of any one of claims 67-71, wherein the organism expresses a heterologous nucleoside triphosphate transporter.

73. The semi-synthetic organism of claim 72, wherein the heterologous nucleoside triphosphate transporter is *Pt*NTT2.

74. The semi-synthetic organism of any one of claims 67-73, wherein the organism further expresses a heterologous tRNA synthetase.

75. The semi-synthetic organism of claim 74, wherein the heterologous tRNA synthetase is *M. barkeri* pyrrolysyl-tRNA synthetase (*Mb* PylRS).

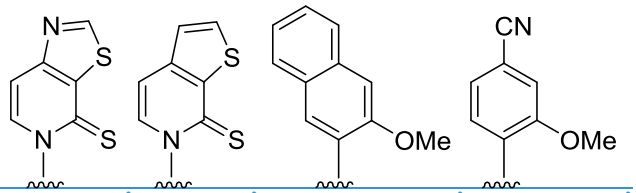
76. The semi-synthetic organism of any one of claims 67-75, wherein the organism further expresses a heterologous RNA polymerase.

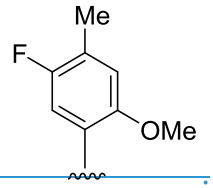
77. The semi-synthetic organism of claim 76, wherein the heterologous RNA polymerase is T7 RNAP.

78. The semi-synthetic organism of any one of claims 67-77, wherein the organism does not express a protein having the function of DNA recombinational repair.

79. The semi-synthetic organism of claim 78, wherein the organism does not express RecA.

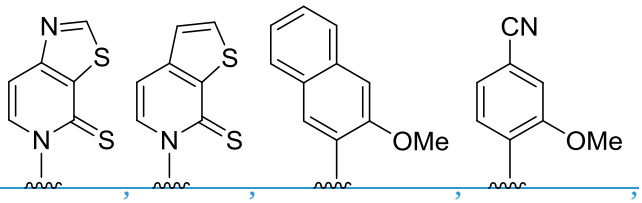
80. The semi-synthetic organism of any one of claims 67-79, further comprising a heterologous mRNA.

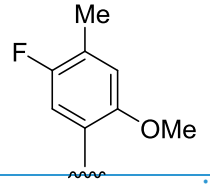
81. The semi-synthetic organism of claim 80, wherein the heterologous mRNA comprises at least one unnatural base selected from: 

, , , , 

and .

82. The semi-synthetic organism of any one of claims 67-81, further comprising a heterologous tRNA.

83. The semi-synthetic organism of claim 82, wherein the heterologous tRNA comprises at least one unnatural base selected from: 

, , , , 

and .

84. A method of transcribing DNA, the method comprising:

providing one or more DNAs comprising (1) a gene encoding a protein, wherein a template strand of the gene encoding the protein comprises a first unnatural base and (2) a gene encoding a tRNA, wherein a template strand of the gene encoding the tRNA comprises a second unnatural base capable of forming a base pair with the first unnatural base;

transcribing the gene encoding the protein to incorporate a third unnatural base into an mRNA, the third unnatural base capable of forming a first unnatural base pair with the first unnatural base;

transcribing the gene encoding a tRNA to incorporate a fourth unnatural base into a tRNA, wherein the fourth unnatural base is capable of forming a second unnatural base pair with the second unnatural base, wherein the first unnatural base pair and the second unnatural base pair are not the same.

85. The method of claim 84, further comprising translating a protein from the mRNA utilizing the tRNA, wherein said protein comprises an unnatural amino acid at a position corresponding to a codon comprising the third unnatural base in the mRNA.

86. A method of replicating a DNA, the method comprising:

providing a DNA comprising (1) a gene encoding a protein, wherein a template strand of the gene encoding the protein comprises a first unnatural base and (2) a gene encoding a tRNA, wherein a template strand of the gene encoding the tRNA comprises a second unnatural base capable of forming a base pair with the first unnatural base; and

~~replicating the DNA to incorporate a first substitute unnatural base in place of the first unnatural base, and/or to incorporate a second substitute unnatural base in place of the second unnatural base;~~

~~wherein the method optionally further comprises:~~

~~transcribing the gene encoding a protein to incorporate a third unnatural base into an mRNA, the third unnatural base being capable of forming a first unnatural base pair with the first unnatural base and/or the first substitute unnatural base; and/or~~

~~transcribing the gene encoding a tRNA to incorporate a fourth unnatural base into a tRNA, wherein the fourth unnatural base is capable of forming a second unnatural base pair with the second unnatural base and/or the second substitute unnatural base, wherein the first unnatural base pair and the second unnatural base pair are not the same.~~

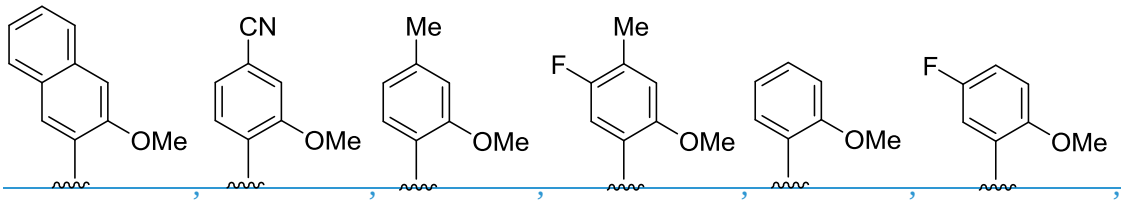
~~87. The method of claim 86, further comprising transcribing the gene encoding a protein to incorporate a third unnatural base into an mRNA, the third unnatural base being capable of forming a first unnatural base pair with the first unnatural base and/or the first substitute unnatural base.~~

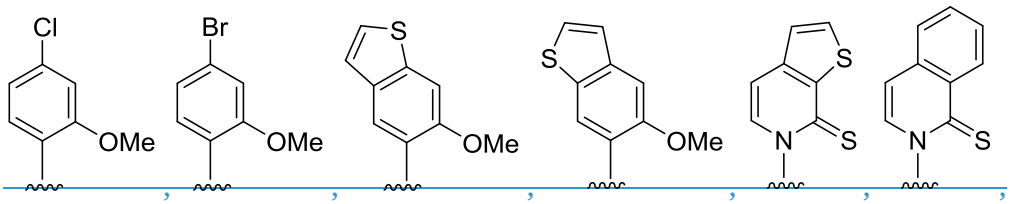
~~88. The method of claim 86 or 87, further comprising transcribing the gene encoding a tRNA to incorporate a fourth unnatural base into a tRNA, wherein the fourth unnatural base is capable of forming a second unnatural base pair with the second unnatural base and/or the second substitute unnatural base, wherein the first unnatural base pair and the second unnatural base pair are not the same.~~

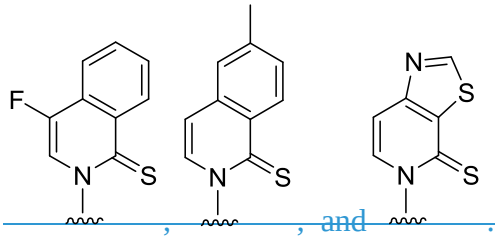
~~89. The method of any one of claims 84-88, wherein the wherein the method is an *in vivo* method comprising use of a semi-synthetic organism which is a bacterium.~~

~~90. The method of claim 89, wherein the organism comprises an *Escherichia coli*.~~

~~91. The method of any one of claims 84-90, wherein at least one of the first unnatural base, the second unnatural base, the third unnatural base, or the fourth unnatural base comprises:~~

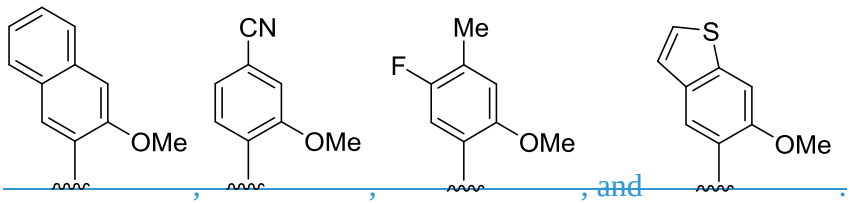






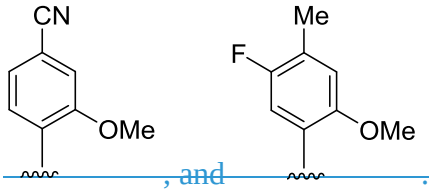
, and .

92. The method of any one of claims 84-91, wherein at least one of the first unnatural base, the second unnatural base, the third unnatural base, or the fourth unnatural base comprises:



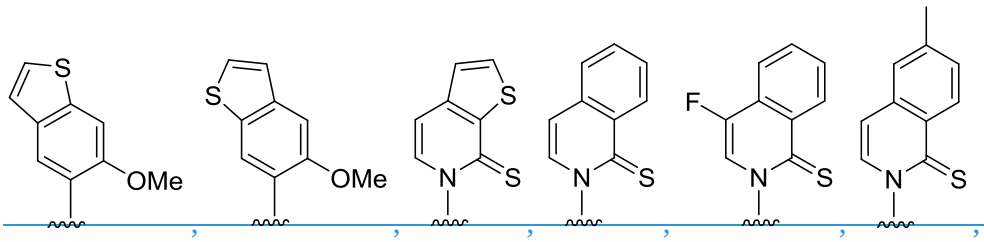
, and .

93. The method of any one of claims 84-92, wherein at least one of the first unnatural base, the second unnatural base, the third unnatural base, or the fourth unnatural base comprises:

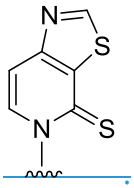


, and .

94. The method of any one of claims 84-93, wherein at least one of the first unnatural base, the second unnatural base, the third unnatural base, or the fourth unnatural base comprises

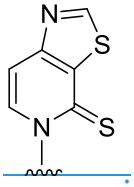


, and



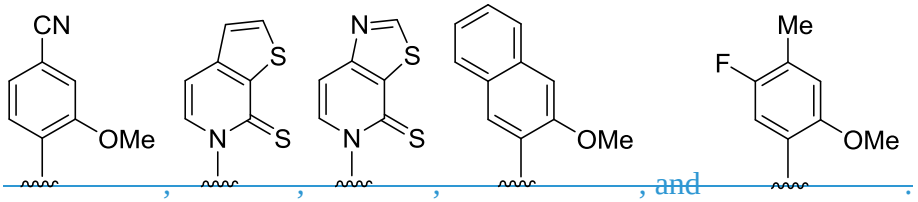
.

95. The method of any one of claims 84-94, wherein at least one of the first unnatural base, the second unnatural base, the third unnatural base, or the fourth unnatural base is

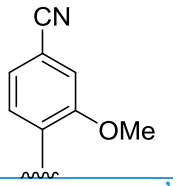
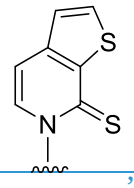
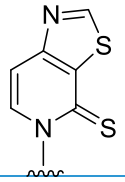
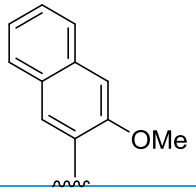
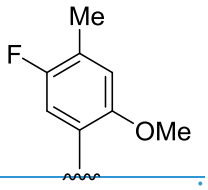


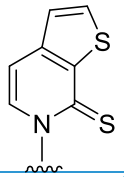
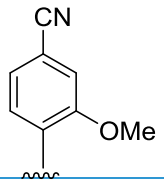
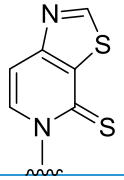
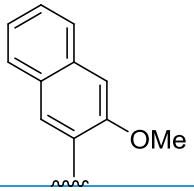
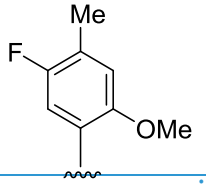
.

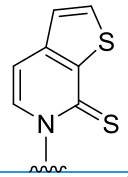
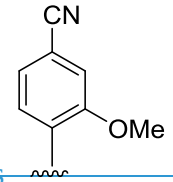
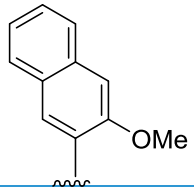
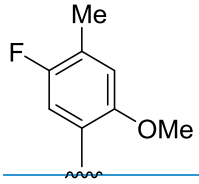
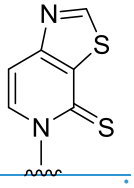
96. The method of any one of claims 84-95, wherein at least one of the first unnatural base, the second unnatural base, the third unnatural base, or the fourth unnatural base comprises:



, and .

97. The method of any one of claims 84-96, wherein the first unnatural base is , the second unnatural base is , the third unnatural base is , and the fourth unnatural base is  or .

98. The method of any one of claims 84-97, wherein the first unnatural base is , the second unnatural base is , the third unnatural base is , and the fourth unnatural base is  or .

99. The method of any one of claims 84-98, wherein the first unnatural base is , the second unnatural base is , the third unnatural base is  or  and the fourth unnatural base is .

100. The method of any one of claims 84-99, wherein the first unnatural base comprises dCNMO, and the second unnatural base comprises dTPT3.

101. The method of any one of claims 84-100, wherein the third unnatural base comprises NaM, and the second unnatural base comprises TAT1.

102. The method of any one of claims 84-101, wherein the protein comprises at least two unnatural amino acids.