SYNTHETIC TRNA SYNTHETASES AND CELLS COMPRISING SYNTHETIC MOLECULES FOR PRODUCTION OF POLYPEPTIDES

CROSS-REFERENCE TO RELATED APPLICATIONS

[1] This patent application claims priority to US Provisional Patent Application No. 63/569,650, filed March 25, 2024, the contents of which are incorporated herein by reference in its entirety for all purposes.

SEQUENCE LISTING

[2] The instant application contains a Sequence Listing which has been submitted electronically in XML format and is hereby incorporated by reference in its entirety. Said XML copy, created on March 20, 2025, is named "01183-0192-00PCT.xml" and is 18,282 bytes in size.

INTRODUCTION AND SUMMARY

[3] The natural genetic code consists of 64 codons made possible by four letters of the genetic alphabet. Three codons are used as stop codons, leaving 61 sense codons that are recognized by a transfer RNA (tRNA) charged by a cognate amino acyl tRNA synthetase (also referred to herein simply as a tRNA synthetase) with one of the 20 proteogenic amino acids. While the canonical amino acids have enabled the remarkable diversity of living organisms, there are many chemical functionalities and associated reactivities that they do not provide. The ability to expand the genetic code to include unnatural or non-canonical amino acids (ncAAs) likely bestows the protein with a desired function or activity and dramatically facilitates many known and emerging applications of proteins such as therapeutic development. Current methods of synthesizing unnatural proteins or unnatural polypeptides containing unnatural amino acids, such as interleukins with an unnatural amino acid, have limitations.

[4] For example, expression of a heterologous tRNA synthetase in cells producing polypeptides comprising an unnatural amino acid can complicate purification of the polypeptides comprising an unnatural amino acid and limit yields. For example, low solubility and/or high expression levels may cause a significant portion of the heterologous tRNA synthetase to partition with insoluble target polypeptides. Providing a heterologous tRNA synthetase with improved catalytic activity can also be desirable, e.g., in that less expression of the tRNA synthetase would be needed to support robust target protein production. Accordingly, there is a need for tRNA synthetases with improved solubility and/or activity.

[5] There also remains a need for improved cells that can carry out *in vivo* synthesis of unnatural polypeptides containing unnatural amino acids and produce the unnatural

polypeptides containing unnatural amino acids with sufficient titer and/or fidelity. The present disclosure aims to meet one or more of these needs or provide other benefits. Accordingly, the following embodiments are provided.

[6] Described herein are tRNA synthetases and cells that may be used for producing polypeptides comprising one or more unnatural amino acids.

Embodiment 1 is a prokaryotic cell comprising:

a first chromosomally integrated DNA sequence encoding a transfer RNA synthetase (tRNA synthetase) operably linked to an inducible promoter and a 5' UTR;

a second DNA sequence encoding a promoter for a repressor of an inducible promoter,

a third DNA sequence encoding at least one unnatural transfer RNA (tRNA) molecule;

a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising a codon recognized by the unnatural tRNA molecule;

wherein the 5' UTR enables increased tRNA synthetase expression compared to SEQ ID NO: 1.

Embodiment 2 is a prokaryotic cell comprising:

a DNA sequence encoding a *Pt*NTT2 operably linked to a P_{lac} promoter, wherein (i) the DNA sequence does not comprise an oppositely oriented promoter within about 1.3-1.5 kb downstream of the 3' end of the *Pt*NTT2 coding sequence or (ii) the DNA sequence comprises a transcriptional terminator that terminates anterograde *Pt*NTT2 transcription, and further comprises a transcription terminator element downstream of the stop codon of the DNA sequence encoding the *Pt*NTT2 and the transcription terminator element is oriented to terminate retrograde transcription into the DNA sequence encoding the *Pt*NTT2.

Embodiment 3 is a prokaryotic cell comprising:

a DNA sequence encoding a tRNA synthetase;

a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed;

a DNA sequence encoding a non-self-cleaving LexA, optionally wherein the LexA comprises a S119 mutation, optionally wherein the S119 mutation is S119A; and

a *recA* hypomorph, optionally wherein the *recA* hypomorph results from RecX overexpression.

Embodiment 4 is the cell of embodiment 1, wherein the unnatural tRNA molecule is an unnatural pyrrolysyl tRNA molecule.

Embodiment 5 is the cell of embodiment 1 or embodiment 4, wherein the 5' UTR comprises the DNA sequence of SEQ ID NO: 2.

Embodiment 5.1 is the cell of embodiment 1 or embodiment 4, wherein the 5' UTR comprises a lac operator and/or the DNA sequence of SEQ ID NO: 13.

Embodiment 6 is the cell of any one of embodiments 1, 4, or 5, wherein the second DNA sequence is chromosomally integrated.

Embodiment 7 is the cell of any one of embodiments 1 or 4-6, wherein the third DNA sequence is on a plasmid.

Embodiment 8 is the cell of any one of embodiments 1 or 4-7, wherein the fourth DNA sequence is on a plasmid.

Embodiment 9 is the cell of any one of embodiments 1 or 3-8, wherein the tRNA synthetase comprises a chimeric pyrrolysyl-tRNA synthetase comprising a pyrrolysyl-tRNA synthetase N-terminal domain from a first species and a pyrrolysyl-tRNA synthetase C-terminal domain from a second species.

Embodiment 9.1 is the cell of any one of embodiments 1 or 3-9, wherein the first species is *Methanosarcina barkeri*, and the second species is *Methanosarcina mazei*.

Embodiment 9.2 is the cell of any one of embodiments 1 or 3-9.1, wherein the tRNA synthetase comprises a chimeric pyrrolysyl-tRNA synthetase comprising a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD).

Embodiment 10 is the cell of any one of embodiments 1 or 3-8, wherein the tRNA synthetase comprises a split chimeric pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the first polypeptide chain comprises a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD), and wherein the second polypeptide chain comprises a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD).

Embodiment 11 is the cell of any one of embodiments 1, 3-8, or 10, wherein the tRNA synthetase comprises a split chimeric pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) comprises an amino acid sequence having at least 80% identity to positions 1-137 of SEQ ID NO: 4, and wherein the *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD) comprises an amino acid sequence having at least 80% identity to positions 150-419 of SEQ ID NO: 4, optionally wherein the C-terminal amino acid of the *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase NTD corresponds to P137 or S149 of SEQ ID NO: 4.

Embodiment 12 is the cell of any one of embodiments 1 or 3-8, wherein the tRNA synthetase comprises a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase.

Embodiment 13 is the cell of any one of embodiments 1 or 3-8, wherein the tRNA synthetase comprises a split *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising two

separate polypeptide chains, wherein the first polypeptide chain comprises a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD), and wherein the second polypeptide chain comprises a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase C-terminal domain (CTD).

Embodiment 14 is the cell of embodiment any one of embodiments 1, 3-8, or 13, wherein the tRNA synthetase comprises a split *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) comprises an amino acid sequence having at least 80% identity to positions 1-110, 1-137, or 1-149 of SEQ ID NO: 3, and wherein the *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase C-terminal domain (CTD) comprises an amino acid sequence having at least 80% identity to positions 150-419 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase NTD corresponds to S110, P137, or S149 of SEQ ID NO: 3.

Embodiment 15 is a cell comprising:

a DNA sequence encoding a split *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the first polypeptide chain comprises part of a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 1-110, 1-137, or 1-149 of SEQ ID NO: 3, and wherein the second polypeptide chain comprises part of a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 150-419 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to S110, P137, or S149 of SEQ ID NO: 3.

Embodiment 16 is the cell of embodiment 15, wherein the split *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprises an amino acid sequence having at least 80% identity to positions 1-110 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to S110.

Embodiment 17 is the cell of embodiment 15, wherein the split *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprises an amino acid sequence having at least 80% identity to positions 1-137 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to P137.

Embodiment 18 is the cell of embodiment 15, wherein the split *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprises an amino acid sequence having at least 80% identity to positions 1-149 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to S149.

Embodiment 19 is a cell comprising:

a DNA sequence encoding a split chimeric pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the first polypeptide chain comprises part of a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 1-137 of SEQ ID NO: 4, and wherein the second polypeptide chain comprises part of a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 150-419 of SEQ ID NO: 4, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to P137 of SEQ ID NO: 4.

Embodiment 20 is a cell comprising:

a DNA sequence encoding a split pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the first polypeptide subunit comprises part of a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 1-138 or 1-172 of SEQ ID NO: 5, and wherein the second polypeptide subunit comprises part of a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 185-454 of SEQ ID NO: 5, optionally wherein the C-terminal amino acid of the first polypeptide subunit corresponds to I138 or N172 and of SEQ ID NO: 5.

Embodiment 21 is the cell of any one of embodiments 9-11 or 19, wherein the tRNA synthetase comprises at least the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, and A100E.

Embodiment 22 is the cell of any one of embodiments 9-11, 19, or 21, wherein the tRNA synthetase comprises at least the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F.

Embodiment 23 is the cell of any one of embodiments 9-11, 19, 21, or 22, wherein the tRNA synthetase comprises an amino acid sequence having at least 80% identity to SEQ ID NO: 6.

Embodiment 24 is the cell of any one of embodiments 9-11, 19, or 21-23, wherein the tRNA synthetase comprises at least one, two, three, four, five, six, seven, or eight of the following mutations with reference to SEQ ID NO: 4: P5T or P5L, V31I, T56P, H62Y, A100E, E302A, C313V, and Y349F.

Embodiment 25 is the cell of any one of embodiments 9-11, 19, or 21-24, wherein the tRNA synthetase comprises at least the following mutations with reference to SEQ ID NO: 4: P5T or P5L, V31I, T56P, H62Y, A100E, E302A, C313V, and Y349F.

Embodiment 26 is the cell of any one of embodiments of any one of embodiments 9-11, 19, or 21-25, wherein the tRNA synthetase comprises an amino acid sequence having at least 80% identity to SEQ ID NO: 7.

Embodiment 27 is the cell of any one of embodiments 1 or 3-26, wherein the pyrrolysyl-tRNA synthetase is fused to *Escherichia coli* thioredoxin.

Embodiment 28 is the cell of any one of embodiments 1, 4-9.2, or 21-27, wherein the first sequence comprises a DNA sequence that is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 8.

Embodiment 29 is the cell of any one of embodiments 1, 4-9.2, or 21-24, wherein the first sequence comprises a DNA sequence comprising SEQ ID NO: 8.

Embodiment 30 is the cell of any one of embodiments 1, 4-14, or 21-29, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ1} promoter.

Embodiment 31 is the cell of any one of embodiments 1, 4-14, or 21-29, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ} promoter.

Embodiment 32 is the cell of any one of embodiments 1, 2, or 4-31, wherein the cell does not comprise a functional *recA* gene.

Embodiment 33 is a prokaryotic cell comprising:

a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to an inducible promoter, and a 5' UTR comprising SEQ ID NO: 2;

a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ1} promoter;

a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid; and

a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising a codon recognized by the unnatural pyrrolysyl tRNA, wherein the fourth DNA sequence is on a plasmid; and a fifth DNA sequence encoding a nucleoside triphosphate transporter operably linked to a P_{lacUV5} promoter, optionally wherein the prokaryotic cell further comprises an antibiotic resistance gene.

Embodiment 33.1 is a prokaryotic cell comprising:

a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to an inducible promoter, and a 5' UTR comprising a lac operator and/or the DNA sequence of SEQ ID NO: 13;

a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ1} promoter;

a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein

the third DNA sequence is on a plasmid; and

a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising a codon recognized by the unnatural pyrrolysyl tRNA, wherein the fourth DNA sequence is on a plasmid; and a fifth DNA sequence encoding a nucleoside triphosphate transporter operably linked to a P_{lacUV5} promoter, optionally wherein the prokaryotic cell further comprises an antibiotic resistance gene.

Embodiment 34 is a prokaryotic cell comprising:

a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to an inducible promoter, and a 5' UTR comprising SEQ ID NO: 2;

a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ1} promoter;

a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid; and

a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising an unnatural codon recognized by the unnatural pyrrolysyl tRNA, wherein the fourth DNA sequence is on a plasmid, and a DNA sequence encoding a nucleoside triphosphate transporter, wherein the DNA sequence encoding the nucleoside triphosphate transporter is operably linked to a P_{lac} promoter and does not comprise an oppositely oriented promoter within 1.3-1.5 kb downstream of the 3' end of the nucleoside triphosphate transporter coding sequence.

Embodiment 34.1 is a prokaryotic cell comprising:

a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to an inducible promoter, and a 5' UTR comprising a lac operator and/or SEQ ID NO: 13;

a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ1} promoter;

a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid; and

a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising an unnatural codon recognized by the unnatural pyrrolysyl tRNA, wherein the fourth DNA sequence is on a plasmid, and a DNA sequence encoding a nucleoside triphosphate transporter, wherein the DNA sequence encoding the nucleoside triphosphate transporter is operably linked to a P_{lac} promoter and does not comprise an oppositely oriented promoter within 1.3-1.5 kb downstream of the 3' end of the nucleoside triphosphate transporter coding sequence.

Embodiment 35 is a prokaryotic cell comprising:

a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to an inducible promoter, and a 5' UTR comprising SEQ ID NO: 2;

a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ1} promoter;

a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid;

a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising an unnatural codon recognized by the unnatural pyrrolysyl tRNA, wherein the fourth DNA sequence is on a plasmid;

a fifth DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed;

a sixth DNA sequence encoding a non-self-cleaving LexA, optionally wherein the LexA comprises a S119 mutation, optionally wherein the S119 mutation is S119A;

a DNA sequence encoding a nucleoside triphosphate transporter, wherein the DNA sequence encoding the nucleoside triphosphate transporter is operably linked to a P_{lac} promoter and does not comprise an oppositely oriented promoter within 1.3-1.5 kb downstream of the 3' end of the nucleoside triphosphate transporter coding sequence; and

a *recA* hypomorph, optionally wherein the *recA* hypomorph results from RecX overexpression.

Embodiment 35.1 is a prokaryotic cell comprising:

a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to an inducible promoter, and a 5' UTR comprising a lac operator and/or SEQ ID NO: 13;

a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ1} promoter;

a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid;

a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising an unnatural codon recognized by the unnatural pyrrolysyl tRNA, wherein the fourth DNA sequence is on a plasmid;

a fifth DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed;

a sixth DNA sequence encoding a non-self-cleaving LexA, optionally wherein the LexA comprises a S119 mutation, optionally wherein the S119 mutation is S119A;

a DNA sequence encoding a nucleoside triphosphate transporter, wherein the DNA sequence encoding the nucleoside triphosphate transporter is operably linked to a P_{lac} promoter and does not comprise an oppositely oriented promoter within 1.3-1.5 kb downstream of the 3' end of the nucleoside triphosphate transporter coding sequence; and

a *recA* hypomorph, optionally wherein the *recA* hypomorph results from RecX overexpression.

Embodiment 36 is a prokaryotic cell comprising:

a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to an inducible promoter, and a 5' UTR comprising SEQ ID NO: 2;

a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ} promoter;

a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid;

a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising an unnatural codon recognized by the unnatural pyrrolysyl tRNA molecule, wherein the fourth DNA sequence is on a plasmid.

Embodiment 36.1 is a prokaryotic cell comprising:

a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to an inducible promoter, and a 5' UTR comprising a lac operator and/or SEQ ID NO: 13;

a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ} promoter;

a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid;

a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising an unnatural codon recognized by the unnatural pyrrolysyl tRNA molecule, wherein the fourth DNA sequence is on a plasmid.

Embodiment 37 is a prokaryotic cell comprising:

a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to an inducible promoter, and a 5' UTR comprising SEQ ID NO: 2;

a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ1} promoter;

a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid;

a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising a codon recognized by the unnatural pyrrolysyl tRNA molecule, wherein the fourth DNA

sequence is on a plasmid;

wherein the cell does not comprise a functional *recA* gene.

Embodiment 37.1 is a prokaryotic cell comprising:

a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to an inducible promoter, and a 5' UTR comprising a lac operator and/or SEQ ID NO: 13;

a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ1} promoter;

a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid;

a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising a codon recognized by the unnatural pyrrolysyl tRNA molecule, wherein the fourth DNA sequence is on a plasmid;

wherein the cell does not comprise a functional *recA* gene.

Embodiment 38 is a prokaryotic cell comprising:

a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to a inducible promoter, and a 5' UTR comprising SEQ ID NO: 2;

a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ} promoter;

a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid;

a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising a codon recognized by the unnatural pyrrolysyl tRNA molecule, wherein the fourth DNA sequence is on a plasmid;

wherein the cell does not comprise a functional *recA* gene.

Embodiment 38.1 is a prokaryotic cell comprising:

a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to a inducible promoter, and a 5' UTR comprising a lac operator and/or SEQ ID NO: 13;

a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ} promoter;

a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid;

a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising a codon recognized by the unnatural pyrrolysyl tRNA molecule, wherein the fourth DNA

sequence is on a plasmid;

wherein the cell does not comprise a functional *recA* gene.

Embodiment 39 is the cell of any one of embodiments 1-38, wherein the cell further comprises a DNA sequence encoding a T7 bacteriophage RNA polymerase, optionally wherein the DNA sequence encoding the T7 bacteriophage RNA polymerase is operably linked to an inducible P_{lacUV5} promoter.

Embodiment 40 is the cell of any one of embodiments 1, 3-32, or 36-39, wherein the cell comprises a DNA sequence encoding a nucleoside triphosphate transporter.

Embodiment 41 is the cell of any one of embodiments 2, 33-35, or 40, wherein the nucleoside triphosphate transporter comprises a truncated *Pt*NTT2.

Embodiment 42 is the cell of embodiment 41, wherein the amino acid sequence encoding the truncated *Pt*NTT2 is at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 9.

Embodiment 43 is the cell of embodiment 41, wherein the amino acid sequence encoding the truncated *Pt*NTT2 is SEQ ID NO: 9.

Embodiment 44 is the cell of any one of embodiments 1, 2, 4-34, or 36-43, further comprising a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed.

Embodiment 45 is the cell of any one of embodiments 1, 2, 4-34, or 36-43, further comprising a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is constitutively overexpressed.

Embodiment 46 is the cell of any one of embodiments 1, 2, 4-34, or 36-45, further comprising a DNA sequence encoding a LexA that does not undergo RecA-stimulated cleavage.

Embodiment 47 is the cell of any one of embodiments 1, 2, 4-31, 33, 34, 36, or 39-46, further comprising a *recA* hypomorph.

Embodiment 48 is the cell of embodiment 1, 2, 4-31, 33, 34, 36, or 39-47, wherein the *recA* hypomorph results from RecX overexpression.

Embodiment 49 is the cell of any one of embodiments 1-48, further comprising an antibiotic resistance gene.

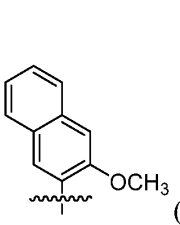
Embodiment 50 is the cell of embodiment 49, wherein the antibiotic resistance gene is a chloramphenicol resistance gene.

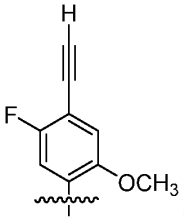
Embodiment 51 is the cell of any one of embodiments 1, 4-14, or 21-50,

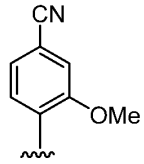
wherein the third DNA sequence encoding at least one unnatural tRNA molecule comprises a first unnatural nucleotide comprising a first unnatural base; and

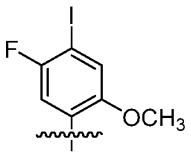
wherein the fourth DNA sequence encoding at least one unnatural mRNA molecule comprises a second unnatural nucleotide comprising a second unnatural base.

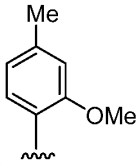
Embodiment 52 is the cell of embodiment 51, wherein the first unnatural nucleotide and the second unnatural nucleotide each comprise an unnatural base independently selected from the group consisting of:

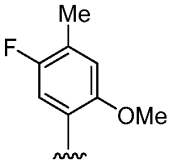


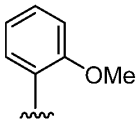
(NaM),


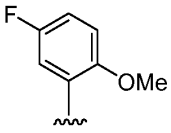
(FEMO),


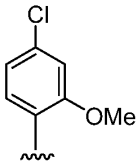
(CNMO),


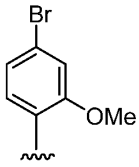


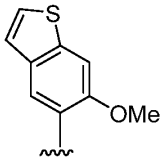
(FIMO),


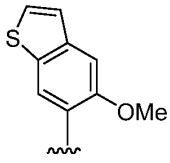
(MMO2),


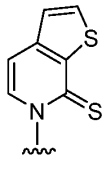
(5FM),


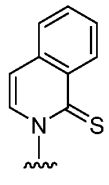
(2OMe),


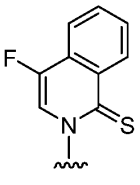
(5F2OMe),


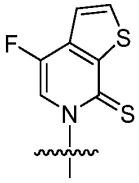
(ClMO),


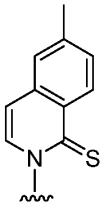
(BrMO),


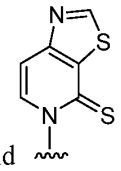
(PTMO),


(MTMO),


(TPT3),


(SICS),
(FSICS),


(FTPT3),


(5SICS), and


(TAT1).

Embodiment 53 is the cell of embodiment 51 or 52, wherein the first unnatural nucleotide and the second unnatural nucleotide each comprise an unnatural base selected from the group consisting of CNMO, TPT3, NAM, or TAT1.

Embodiment 54 is the cell of any one of embodiments 51-53, wherein the first unnatural base comprises CNMO.

Embodiment 55 is the cell of any one of embodiments 51-53, wherein the first unnatural base comprises TPT3.

Embodiment 56 is the cell of any one of embodiments 51-53, wherein the first unnatural base comprises NAM.

Embodiment 57 is the cell of any one of embodiments 51-53, wherein the first unnatural base comprises TAT1.

Embodiment 58 is the cell of any one of embodiments 1, 4-14, or 21-50,

wherein the third DNA sequence encoding at least one unnatural tRNA molecule comprises a first unnatural base pair;

wherein the fourth DNA sequence encoding at least one unnatural mRNA molecule comprises a second unnatural base pair.

Embodiment 59 is the cell of embodiment 58, wherein the first unnatural base pair is CNMO/TPT3.

Embodiment 60 is the cell of embodiment 58, wherein the first unnatural base pair is NaM/TPT3.

Embodiment 61 is the cell of embodiment 58, wherein the first unnatural base pair is CNMO/TAT1.

Embodiment 62 is the cell of embodiment 58, wherein the first unnatural base pair is NaM/TAT1.

Embodiment 63 is the cell of embodiment 58, wherein the first unnatural base pair is NaM/NaM.

Embodiment 64 is the cell of any one of embodiments 1, 4-14, or 21-50, wherein the codon recognized by the unnatural tRNA, when read from a 5’ to 3’ direction, comprises NNX or NXN, wherein N is any natural nucleotide, and X is an unnatural nucleotide.

Embodiment 65 is the cell of embodiment 64, wherein the codon recognized by the unnatural tRNA read from a 5’ to 3’ direction comprises UUX, UGX, CGX, AGX, GAX, CAX, AUX, CUX, GUX, UAX, GGX, GXU, CXU, GXG, AXG, GXC, AXC, GXA, CXC, or UXC.

Embodiment 66 is the cell of embodiment 64, wherein the codon read from a 5’ to 3’ direction comprises CGX, AGX, GAX, GXU, CXU, GXC, AXC, or GXA.

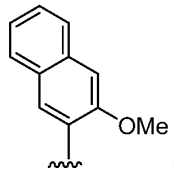
Embodiment 67 is the cell of embodiment 64, wherein the codon read from a 5’ to 3’ direction comprises CGX, AGX, GAX, GXU, GXC, or AXC.

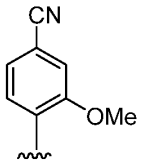
Embodiment 68 is the cell of embodiment 64, wherein the codon read from a 5’ to 3’ direction is GXU.

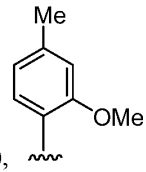
Embodiment 69 is the cell of embodiment 64, wherein the codon read from a 5’ to 3’ direction is GXC.

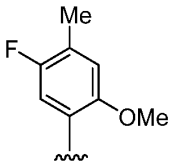
Embodiment 70 is the cell of embodiment 64, wherein the codon read from a 5' to 3' direction is AXC.

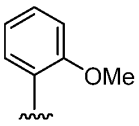
Embodiment 71 is the cell of any one of embodiments 64-70, wherein X comprises any one of the following bases:

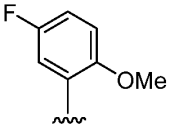


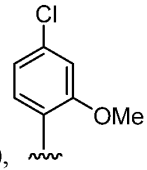
(NaM),


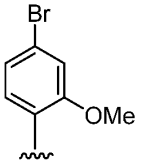
(CNMO),


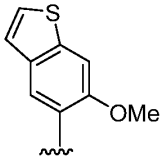
(MMO2),


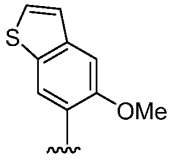


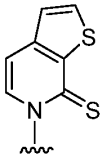
(5FM),
(2OMe),


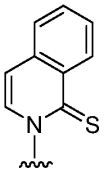
(5F2OMe),


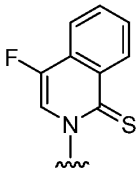
(ClMO),


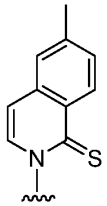
(BrMO),


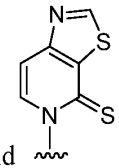
(PTMO),


(MTMO),




(TPT3),
(SICS),


(FSICS),


(5SICS), and


(TAT1).

Embodiment 72 is the cell of any one of embodiments 64-70, wherein X is NaM.

Embodiment 73 is the cell of any one of embodiments 64-70, wherein X is CNMO.

Embodiment 74 is the cell of any one of embodiments 64-73, wherein Y is TPT3.

Embodiment 75 is the cell of any one of embodiments 64-73, wherein Y is TAT1.

Embodiment 76 is the cell of any one of embodiments 2, 3, 9-32, or 39-50, wherein the cell comprises unnatural dNTPs.

Embodiment 77 is the cell of any one of embodiments 1-76, wherein the cell comprises at least one unnatural amino acid.

Embodiment 78 is the cell of embodiment 77, wherein the unnatural amino acid comprises

a lysine analogue;

an aromatic side chain;

an azido group;

an alkyne group; or

an aldehyde or ketone group.

Embodiment 79 is the cell of embodiment 77, wherein the unnatural amino acid does not comprise an aromatic side chain.

Embodiment 80 is the cell of embodiment 77, wherein the unnatural amino acid is selected from N6-azidoethoxy-carbonyl-L-lysine, N6-propargylethoxy-carbonyl-L-lysine (PraK), N6-(propargyloxy)-carbonyl-L-lysine (PrK), p-azido-phenylalanine(*p*AzF), BCN-L-lysine, norbornene lysine, TCO-lysine, methyltetrazine lysine, allyloxycarbonyllysine, 2-amino-8-oxononanoic acid, 2-amino-8-oxooctanoic acid, p-acetyl-L-phenylalanine, p-azidomethyl-L-phenylalanine (pAMF), p-iodo-L-phenylalanine, m-acetylphenylalanine, 2-amino-8-oxononanoic acid, p-propargyloxyphenylalanine, p-propargyl-phenylalanine, 3-methyl-phenylalanine, L-Dopa, fluorinated phenylalanine, isopropyl-L-phenylalanine, p-azido-L-phenylalanine, p-acyl-L-phenylalanine, p-benzoyl-L-phenylalanine, p-bromophenylalanine, p-amino-L- phenylalanine, isopropyl-L-phenylalanine, O-allyltyrosine, O-methyl-L-tyrosine, O-4-allyl-L-tyrosine, 4-propyl-L-tyrosine, phosphonotyrosine, tri-O-acetyl-GlcNAcp-serine, L-phosphoserine, phosphonoserine, L-3-(2-naphthyl)alanine, 2-amino-3-((2-((3-(benzyloxy)-3-oxopropyl)amino)ethyl)selanyl)propanoic acid, 2-amino-3-(phenylselanyl)propanoic, selenocysteine, N6-(((2-azidobenzyl)oxy)carbonyl)-L-lysine, N6-(((3-azidobenzyl)oxy)carbonyl)-L-lysine, and N6-(((4-azidobenzyl)oxy)carbonyl)-L-lysine.

Embodiment 81 is the cell of embodiment 77, wherein the unnatural amino acid is an azidolysine (AzK).

Embodiment 82 is the cell of any one of embodiments 1-81, wherein the cell is a gamma proteobacterium.

Embodiment 83 is the cell of any one of embodiments 1-82, wherein the cell is a member of order Enterobacterales.

Embodiment 84 is the cell of any one of embodiments 1-83, wherein the cell is a member of family *Enterobacteriaceae*.

Embodiment 85 is the cell of any one of embodiments 1-84, wherein the cell is a member of genus *Escherichia*.

Embodiment 86 is the cell of any one of embodiments 1-85, wherein the cell is an *Escherichia coli* cell.

Embodiment 87 is the cell of any one of embodiments 77-86, wherein the cell translates at least one unnatural polypeptide comprising the unnatural amino acid.

Embodiment 87.1 is the cell of any one of embodiments 77-86, wherein the cell can synthesize at least one unnatural polypeptide comprising the unnatural amino acid.

Embodiment 88 is the cell of any one of embodiments 77-87.1, wherein the cell exhibits improved titer, as measured by production of unnatural polypeptides comprising the unnatural amino acid, optionally wherein titer is measured in mg/L of unnatural polypeptides comprising the unnatural amino acid, further optionally wherein titer is measured using HPLC, or titer is measured in optical density (OD) of the cells, as compared to a reference prokaryotic cell comprising a functional plasmid comprising a DNA sequence encoding an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising SEQ ID NO: 3 and wherein the reference prokaryotic cell comprises a wild-type *recA* gene, a DNA sequence encoding a *Pt*NTT2 operably linked to a P_{lac} promoter, and a 5' UTR of SEQ ID NO: 1 and is otherwise isogenic with the cell of any one of embodiments 1, 33, or 36-38.

Embodiment 89 is the cell of any one of embodiments 77-87, wherein the cell exhibits improved fidelity in synthesizing an unnatural polypeptide comprising the unnatural amino acid, wherein fidelity is percent unnatural amino acid incorporation at an intended site, optionally wherein fidelity is measured using LC-MS, as compared to a reference prokaryotic cell comprising a functional plasmid comprising a DNA sequence encoding an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising SEQ ID NO: 3 and wherein the reference prokaryotic cell comprises a wild-type *recA* gene, a DNA sequence encoding a *Pt*NTT2 operably linked to a P_{lac} promoter, and a 5' UTR of SEQ ID NO: 1 and is otherwise isogenic with the cell of any one of embodiments 1, 33, or 36-38.

Embodiment 90 is a cell culture comprising the cells of any of embodiments 77-87, wherein the culture is characterized by reduced colony size heterogeneity as compared to a reference prokaryotic cell comprising a functional plasmid comprising a DNA sequence encoding an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising SEQ ID NO: 3 and wherein the reference prokaryotic cell comprises a wild-type *recA* gene, a DNA sequence encoding a *Pt*NTT2 operably linked to a P_{lac} promoter, and a 5' UTR of SEQ ID NO: 1 and is otherwise isogenic with the cell of any one of embodiments 1, 33, or 36-38.

Embodiment 91 is a chimeric pyrrolysyl-tRNA synthetase comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F.

Embodiment 92 is a chimeric pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to SEQ ID NO: 6, wherein the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 31 of SEQ ID NO: 6 is isoleucine; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 56 of SEQ ID NO: 6 is proline; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 62 of SEQ ID NO: 6 is tyrosine; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 100 of SEQ ID NO: 6 is glutamate; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 313 of SEQ ID NO: 6 is valine; and the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 349 of SEQ ID NO: 6 is phenylalanine.

Embodiment 93 is a chimeric pyrrolysyl-tRNA synthetase comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: P5T or P5L, E302A, V31I, T56P, H62Y, A100E, C313V, and Y349F.

Embodiment 94 is a chimeric pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to SEQ ID NO: 7, wherein the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 5 of SEQ ID NO: 6 is threonine or leucine; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 31 of SEQ ID NO: 6 is isoleucine; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 56 of SEQ ID NO: 6 is proline; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 62 of SEQ ID NO: 6 is tyrosine; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 100 of SEQ ID NO: 6 is glutamate; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 302 of SEQ ID NO: 6 is alanine; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 313 of SEQ ID NO: 6 is valine; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 349 of SEQ ID NO: 6 is phenylalanine.

Embodiment 95 is a chimeric pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the first polypeptide chain comprises part of a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 1-137 of SEQ ID NO: 4, and wherein the second polypeptide chain comprises part of a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 150-419 of SEQ ID NO:

4, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to P137 of SEQ ID NO: 4.

Embodiment 96 is a pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the first polypeptide subunit comprises part of a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 1-110, 1-137, or 1-149 of SEQ ID NO: 3, and wherein the second polypeptide chain comprises part of a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 150-419 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to S110, P137, or S149 of SEQ ID NO: 3.

Embodiment 97 is the tRNA synthetase of embodiment 96, wherein the pyrrolysyl-tRNA synthetase comprises an amino acid sequence having at least 80% identity to positions 1-110 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the first polypeptide subunit corresponds to S110 of SEQ ID NO: 3.

Embodiment 98 is the tRNA synthetase of embodiment 96, wherein the pyrrolysyl-tRNA synthetase comprises an amino acid sequence having at least 80% identity to positions 1-137 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the first polypeptide subunit corresponds to P137 of SEQ ID NO: 3.

Embodiment 99 is the tRNA synthetase of embodiment 96, wherein the pyrrolysyl-tRNA synthetase comprises an amino acid sequence having at least 80% identity to positions 1-149 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the first polypeptide subunit corresponds to S149 of SEQ ID NO: 3.

Embodiment 100 is the tRNA synthetase of embodiment 95, wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F.

Embodiment 101 is the tRNA synthetase of embodiment 95, wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: P5T or P5L, V31I, T56P, H62Y, A100E, E302A, C313V, and Y349F.

Embodiment 102 is a pyrrolysyl tRNA synthetase comprising two separate polypeptide chains, wherein the first polypeptide subunit comprises part of a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 1-138 or 1-172 of SEQ ID NO: 5, and wherein the second polypeptide subunit comprises part of a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 185-454 of SEQ ID NO: 5,

optionally wherein the C-terminal amino acid of the first polypeptide subunit corresponds to I138 or N172 of SEQ ID NO: 5.

Embodiment 103 is the tRNA synthetase of any one of embodiments 91-102, wherein the pyrrolysyl-tRNA synthetase is fused to *Escherichia coli* thioredoxin.

Embodiment 104 is an isolated nucleic acid encoding the tRNA synthetase of any one of embodiments 91-103.

Embodiment 105 is a vector comprising the nucleic acid of embodiment 104.

Embodiment 106 is a host cell comprising the vector of embodiment 105.

Embodiment 107 is a host cell that produces the tRNA synthetase of any one of embodiments 91-102.

Embodiment 108 is a method for making a tRNA synthetase, comprising culturing the host cell of embodiment 106 or 107 under conditions suitable for expression of the tRNA synthetase.

Embodiment 109 is the method of embodiment 108, further comprising recovering the tRNA synthetase produced by the host cell.

Embodiment 110 is a method of preparing a cell that can express a polypeptide comprising an unnatural amino acid with improved titer and/or fidelity, the cell comprising one or more of a DNA sequence encoding LexA, a DNA sequence encoding RecA and RecX, a DNA sequence encoding Pol II, a DNA sequence encoding tRNA synthetase that can charge a tRNA with an unnatural amino acid, and/or a DNA sequence encoding *Pt*NTT2; and the method comprising one or more of:

modifying the DNA sequence encoding LexA so that LexA does not undergo RecA-stimulated cleavage;

modifying the cell to decrease expression of RecA;

modifying the cell to increase expression of Pol II;

modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression; and/or

modifying the cell to increase expression of the tRNA synthetase.

Embodiment 111 is the method of embodiment 110, the method comprising two or more of:

modifying the DNA sequence encoding LexA so that LexA does not undergo RecA-stimulated cleavage;

modifying the cell to decrease expression of RecA;

modifying the cell to increase expression of Pol II;

modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression; and/or

modifying the cell to increase expression of the tRNA synthetase.

Embodiment 112 is the method of embodiment 110, the method comprising three or more of:

modifying the DNA sequence encoding LexA so that LexA does not undergo RecA-stimulated cleavage;

modifying the cell to decrease expression of RecA;

modifying the cell to increase expression of Pol II;

modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression; and/or

modifying the cell to increase expression of the tRNA synthetase.

Embodiment 113 is the method of embodiment 110, the method comprising four or more of:

modifying the DNA sequence encoding LexA so that LexA does not undergo RecA-stimulated cleavage;

modifying the cell to decrease expression of RecA;

modifying the cell to increase expression of Pol II;

modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression; and/or

modifying the cell to increase expression of the tRNA synthetase.

Embodiment 114 is the method of embodiment 110, the method comprising each of:

modifying the DNA sequence encoding LexA so that LexA does not undergo RecA-stimulated cleavage;

modifying the cell to decrease expression of RecA;

modifying the cell to increase expression of Pol II;

modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression; and

modifying the cell to increase expression of the tRNA synthetase.

Embodiment 115 is the method of any one of embodiments 110-114, wherein modifying the DNA sequence encoding LexA so that LexA does not undergo Rec-A stimulated cleavage comprises a S119 mutation.

Embodiment 116 is the method of embodiment 115, wherein the S119 mutation is S119A.

Embodiment 117 is the method of any one of embodiments 118-124, wherein modifying the cell to decrease expression of RecA results in a *recA* hypomorph.

Embodiment 118 is the method of any one of embodiments 110-117, wherein modifying the cell to decrease expression of RecA comprises modifying the DNA sequence encoding RecA and/or RecX.

Embodiment 119 is the method of embodiment 117 or 118, wherein modifying the cell to decrease expression of RecA comprises modifying the DNA sequence encoding RecA and RecX so that RecX is overexpressed.

Embodiment 120 is the method of any one of embodiments 117-119 wherein the DNA sequences encoding RecA and RecX are on a single transcript, and wherein modifying the cell to decrease expression of RecA comprises deleting a terminator between the DNA sequence encoding RecA and the DNA sequence encoding RecX.

Embodiment 121 is the method of any one of embodiments 110-120, wherein modifying the cell to increase expression of Pol II comprises modifying the promoter for Pol II so that the promoter for Pol II is derepressed.

Embodiment 122 is the method of any one of embodiments 110-121, wherein modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression comprises modifying the DNA sequence encoding *Pt*NTT2 to be operably linked to a P_{lac} promoter and so that the DNA sequence does not comprise an oppositely oriented promoter within about 1.3-1.5 kb downstream of the 3' end of the *Pt*NTT2 coding sequence.

Embodiment 123 is the method of any one of embodiments 110-121, wherein modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression comprises modifying the DNA sequence encoding *Pt*NTT2 to be operably linked to a P_{lac} promoter and so that the DNA sequence comprises a transcriptional terminator that terminates anterograde PtNTT2 transcription, and further comprises a transcription terminator element downstream of the stop codon of the DNA sequence encoding the *Pt*NTT2 and the transcription terminator element is oriented to terminate retrograde transcription into the DNA sequence encoding the *Pt*NTT2.

Embodiment 124 is the method of any one of embodiments 110-123, wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to a 5' UTR, wherein the 5' UTR enables increased tRNA synthetase expression compared to SEQ ID NO: 1.

Embodiment 125 is the method of embodiment 124, wherein the 5' UTR comprises SEQ ID NO: 2.

Embodiment 125.1 is the method of embodiment 124, wherein the 5' UTR comprises a lac operator and/or SEQ ID NO: 13.

Embodiment 126 is the method any one of embodiments 110-123, wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to an inducible promoter and providing a DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the

promoter for the repressor of the inducible promoter is a P_{lacIQ} promoter or a P_{lacIQ1} promoter.

Embodiment 127 is the method of embodiment 126, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ} promoter.

Embodiment 128 is the method of embodiment 126, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ1} promoter.

Embodiment 129 is the method of any one of embodiments 110-123, the method comprising modifying the cell to increase expression of the tRNA synthetase:

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to a 5' UTR, wherein the 5' UTR comprises SEQ ID NO: 2; and

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to an inducible promoter and providing a DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ1} promoter.

Embodiment 129.1 is the method of any one of embodiments 110-123, the method comprising modifying the cell to increase expression of the tRNA synthetase:

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to a 5' UTR, wherein the 5' UTR comprises a lac operator and/or SEQ ID NO: 13; and

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to an inducible promoter and providing a DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ1} promoter.

Embodiment 130 is the method of embodiment 110, the method comprising modifying the cell to increase expression of the tRNA synthetase and modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression:

wherein modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression comprises modifying the DNA sequence encoding *Pt*NTT2 to be operably linked to a P_{lac} promoter and so that the DNA sequence does not comprise an oppositely oriented promoter within about 1.3-1.5 kb downstream of the 3' end of the *Pt*NTT2 coding sequence;

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to a 5' UTR, wherein the 5' UTR comprises SEQ ID NO: 2; and

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to an inducible promoter and providing a DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ1} promoter.

Embodiment 130.1 is the method of embodiment 110, the method comprising modifying the cell to increase expression of the tRNA synthetase and modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression:

wherein modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression comprises modifying the DNA sequence encoding *Pt*NTT2 to be operably linked to a P_{lac} promoter and so that the DNA sequence does not comprise an oppositely oriented promoter within about 1.3-1.5 kb downstream of the 3' end of the *Pt*NTT2 coding sequence;

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to a 5' UTR, wherein the 5' UTR comprises a lac operator and/or SEQ ID NO: 13; and

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to an inducible promoter and providing a DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ1} promoter.

Embodiment 131 is the method of embodiment 110, the method comprising modifying the cell to increase expression of the tRNA synthetase, modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression, modifying the DNA sequence encoding LexA so that LexA does not undergo RecA-stimulated cleavage, modifying the cell to increase expression of Pol II, and modifying the cell to decrease expression of RecA:

wherein modifying the DNA sequence encoding LexA so that LexA does not undergo Rec-A stimulated cleavage comprises a S119A mutation;

wherein modifying the cell to decrease expression of RecA results in a *recA* hypomorph by modifying the DNA sequence encoding RecA and RecX so that RecX is overexpressed;

wherein modifying the cell to increase expression of Pol II comprises modifying the promoter for Pol II so that the promoter for Pol II is derepressed;

wherein modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression comprises modifying the DNA sequence encoding *Pt*NTT2 to be operably linked to a P_{lac} promoter and so that the DNA sequence does not comprise an oppositely oriented promoter within about 1.3-1.5 kb downstream of the 3' end of the *Pt*NTT2 coding sequence;

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to a 5' UTR, wherein the 5' UTR comprises SEQ ID NO: 2; and

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to an inducible promoter and providing a DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ1} promoter.

Embodiment 131.1 is the method of embodiment 110, the method comprising modifying the cell to increase expression of the tRNA synthetase, modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression, modifying the DNA sequence encoding LexA so that LexA does not undergo RecA-stimulated cleavage, modifying the cell to increase expression of Pol II, and modifying the cell to decrease expression of RecA:

wherein modifying the DNA sequence encoding LexA so that LexA does not undergo Rec-A stimulated cleavage comprises a S119A mutation;

wherein modifying the cell to decrease expression of RecA results in a *recA* hypomorph by modifying the DNA sequence encoding RecA and RecX so that RecX is overexpressed;

wherein modifying the cell to increase expression of Pol II comprises modifying the promoter for Pol II so that the promoter for Pol II is derepressed;

wherein modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression comprises modifying the DNA sequence encoding *Pt*NTT2 to be operably linked to a P_{lac} promoter and so that the DNA sequence does not comprise an oppositely oriented promoter within about 1.3-1.5 kb downstream of the 3' end of the *Pt*NTT2 coding sequence;

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to a 5' UTR, wherein the 5' UTR comprises a lac operator and/or SEQ ID NO: 13; and

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to an inducible promoter and providing a DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ1} promoter.

Embodiment 132 is the method of embodiment 110, the method comprising modifying the cell to increase expression of the tRNA synthetase:

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to a 5' UTR, wherein the 5' UTR comprises SEQ ID NO: 2; and

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to an inducible promoter and providing a DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ} promoter.

Embodiment 132.1 is the method of embodiment 110, the method comprising modifying the cell to increase expression of the tRNA synthetase:

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to a 5' UTR, wherein the 5' UTR comprises a lac operator and/or SEQ ID NO: 13; and

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to an inducible promoter and providing a DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ} promoter.

Embodiment 133 is the method of embodiment 110, the method comprising modifying the cell to increase expression of the tRNA synthetase and modifying the cell to decrease expression of RecA:

wherein modifying the cell to decrease expression of RecA comprises deleting DNA sequence encoding RecA;

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to a 5' UTR, wherein the 5' UTR comprises SEQ ID NO: 2; and

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to an inducible promoter and providing a DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ1} promoter.

Embodiment 133.1 is the method of embodiment 110, the method comprising modifying the cell to increase expression of the tRNA synthetase and modifying the cell to decrease expression of RecA:

wherein modifying the cell to decrease expression of RecA comprises deleting DNA sequence encoding RecA;

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to a 5' UTR, wherein the 5' UTR comprises a lac operator and/ or SEQ ID NO: 13; and

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to an inducible promoter and providing a DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ1} promoter.

Embodiment 134 is the method of embodiment 110, the method comprising modifying the cell to increase expression of the tRNA synthetase and modifying the cell to decrease expression of RecA:

wherein modifying the cell to decrease expression of RecA comprises deleting DNA sequence encoding RecA;

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to a 5' UTR, wherein the 5' UTR comprises SEQ ID NO: 2; and

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to an inducible promoter and providing a DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ} promoter.

Embodiment 134.1 is the method of embodiment 110, the method comprising modifying the cell to increase expression of the tRNA synthetase and modifying the cell to decrease expression of RecA:

wherein modifying the cell to decrease expression of RecA comprises deleting DNA sequence encoding RecA;

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to a 5' UTR, wherein the 5' UTR comprises a lac operator and/or SEQ ID NO: 13; and

wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to an inducible promoter and providing a DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ} promoter.

Embodiment 135 is the cell of any one of embodiments 1-89, wherein the cell is a bacterial cell, e.g., *E. coli*, optionally wherein the cell is a bacterial cell comprising a BL21(DE3) background.

Embodiment 136 is the cell of any one of embodiments 1-89 and 135, wherein the cell comprises a DNA sequence in which the sequence encoding IS1 transposase is deleted, optionally wherein the sequence encoding IS1 transposase is replaced with an inactive recombination sequence, optionally wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence.

Embodiment 137 is the cell of any one of embodiments 1-89, 135, and 136, wherein the cell comprises a DNA sequence in which *gpE* encoding a lambda DE3 phage capsid protein is deleted, optionally wherein *gpE* is replaced with an inactive recombination sequence, further optionally wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence.

Embodiment 138 is the cell of any one of embodiments 1-89 and 135-137, wherein the cell comprises a chimeric pyrrolysyl-tRNA synthetase, e.g., a chimeric pyrrolysyl-tRNA synthetase comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F, wherein the tRNA synthetase is integrated into a chromosomal *arsB* locus for the cell (e.g., bacterial cell such as *E. coli*, optionally wherein the *E. coli* comprises a BL21(DE3) background)

under the control of an inducible promoter (e.g., an IPTG-inducible promoter), optionally wherein the inducible promoter (e.g., an IPTG-inducible promoter) is Tac or TacO.

Embodiment 139 is the cell of any one of embodiments 1-89 and 135-138, wherein a DNA sequence encoding a repressor of an inducible promoter (e.g., an IPTG-inducible promoter) is integrated into a chromosomal *arsB* locus, wherein the repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is LacI repressor protein, and wherein the promoter for the sequence encoding the repressor of the inducible promoter is a P_{lacIQ1} promoter, optionally wherein the chromosomal *arsB* locus is the chromosomal *arsB* locus for the cell (e.g., bacterial cell such as *E. coli*, optionally wherein the *E. coli* comprises a BL21(DE3) background).

Embodiment 140 is the cell of any one of embodiments 1-89 and 135-139, wherein the cell comprises a 5' UTR comprising SEQ ID NO: 2.

Embodiment 140.1 is the cell of any one of embodiments 1-89 and 135-139, wherein the cell comprises a 5' UTR comprising a lac operator and/or SEQ ID NO: 13.

Embodiment 141 is the cell of any one of embodiments 1-89 and 135-140, wherein a DNA sequence encoding a truncated *Pt*NTT2 is integrated into a chromosomal *lacZYA* locus of the cell (e.g., bacterial cell such as *E. coli*, optionally wherein the *E. coli* comprises a BL21(DE3) background), optionally replacing *lacZY*, optionally wherein the DNA sequence encoding the truncated *Pt*NTT2 is operably linked to a promoter, optionally wherein the promoter is P_{lacUV5} or P_{lac}.

Embodiment 142 is the cell of any one of embodiments 1-89 and 135-141, wherein the DNA sequence encoding the truncated *Pt*NTT2 does not comprise an oppositely oriented promoter within 1.3-1.5 kb downstream of the 3' end of the truncated *Pt*NTT2 coding sequence.

Embodiment 143 is the cell of any one of embodiments 1-89 and 135-142, wherein the cell comprises a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed, optionally wherein the derepressed promoter is a *polB* promoter with a LexA binding box deletion or an inactivated LexA binding box.

Embodiment 144 is the cell of any one of embodiments 1-89 and 135-143, wherein RecX is overexpressed, optionally wherein the DNA sequences encoding RecA and RecX are on a single transcript, and wherein a terminator between the DNA sequence encoding RecA and the DNA sequence encoding RecX is deleted.

Embodiment 145 is the cell of any one of embodiments 1-89 and 135-144, wherein the cell comprises a DNA sequence encoding LexA comprising a S119A mutation and/or wherein the cell comprises a DNA sequence comprising a *relA* deletion.

Embodiment 146 is a prokaryotic cell comprising:

a DNA sequence in which the sequence encoding IS1 transposase is deleted and replaced with an inactive recombination sequence, wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence;

a DNA sequence in which *gpE* encoding a lambda DE3 phage capsid protein is deleted and replaced with an inactive recombination sequence, wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence;

a chimeric pyrrolysyl-tRNA synthetase comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F, wherein the tRNA synthetase is integrated into a chromosomal *arsB* locus under the control of an inducible promoter;

a DNA sequence encoding a repressor of the inducible promoter integrated into the chromosomal *arsB* locus; and

a DNA sequence encoding a truncated *Pt*NTT2 integrated into a chromosomal *lacZYA* locus of the cell replacing *lacZY*, wherein the DNA sequence encoding the truncated *Pt*NTT2 is operably linked to a promoter, optionally wherein the promoter is P_{lacUV5}.

Embodiment 147 is a prokaryotic cell comprising:

a DNA sequence in which the sequence encoding IS1 transposase is deleted and replaced with an inactive recombination sequence, wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence;

a DNA sequence in which *gpE* encoding a lambda DE3 phage capsid protein is deleted and replaced with an inactive recombination sequence, wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence;

a chimeric pyrrolysyl-tRNA synthetase comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F, wherein the tRNA synthetase is integrated into a chromosomal *arsB* locus under the control of an inducible promoter;

a DNA sequence encoding a repressor of the inducible promoter integrated into the chromosomal *arsB* locus; and

a DNA sequence encoding a truncated *Pt*NTT2 integrated into a chromosomal *lacZYA* locus of the cell replacing *lacZY*, wherein the DNA sequence encoding the truncated *Pt*NTT2 is operably linked to a promoter, optionally wherein the promoter is P_{lac}, and wherein the DNA sequence encoding the truncated *Pt*NTT2 does not comprise an oppositely oriented promoter within 1.3-1.5 kb downstream of the 3' end of the truncated PtNTT2 coding sequence.

Embodiment 148 is a prokaryotic cell comprising:

a DNA sequence in which the sequence encoding IS1 transposase is deleted and replaced with an inactive recombination sequence, wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence;

a DNA sequence in which *gpE* encoding a lambda DE3 phage capsid protein is deleted and replaced with an inactive recombination sequence, wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence;

a chimeric pyrrolysyl-tRNA synthetase comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F, wherein the tRNA synthetase is integrated into a chromosomal *arsB* locus under the control of an inducible promoter;

a DNA sequence encoding a repressor of the inducible promoter integrated into the chromosomal *arsB* locus;

a DNA sequence encoding a truncated *Pt*NTT2 integrated into a chromosomal *lacZYA* locus of the cell replacing *lacZY*, wherein the DNA sequence encoding the truncated *Pt*NTT2 is operably linked to a promoter, optionally wherein the promoter is P_{lac}, and wherein the DNA sequence encoding the truncated *Pt*NTT2 does not comprise an oppositely oriented promoter within 1.3-1.5 kb downstream of the 3' end of the truncated PtNTT2 coding sequence; and

a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed, further wherein the derepressed promoter is a *polB* promoter with a LexA binding box deletion or an inactivated LexA binding box.

Embodiment 149 is a prokaryotic cell comprising:

a DNA sequence in which the sequence encoding IS1 transposase is deleted and replaced with an inactive recombination sequence, wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence;

a DNA sequence in which *gpE* encoding a lambda DE3 phage capsid protein is deleted and replaced with an inactive recombination sequence, wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence;

a chimeric pyrrolysyl-tRNA synthetase comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F, wherein the tRNA synthetase is integrated into a chromosomal *arsB* locus under the control of an inducible promoter;

a DNA sequence encoding a repressor of the inducible promoter integrated into the chromosomal *arsB* locus;

a DNA sequence encoding a truncated *Pt*NTT2 integrated into a chromosomal *lacZYA* locus of the cell replacing *lacZY*, wherein the DNA sequence encoding the truncated *Pt*NTT2 is operably linked to a promoter, optionally wherein the promoter is P_{lac}, and wherein the DNA sequence encoding the truncated *Pt*NTT2 does not comprise an oppositely oriented promoter within 1.3-1.5 kb downstream of the 3' end of the truncated PtNTT2 coding sequence;

a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed, further wherein the derepressed promoter is a *polB* promoter with a LexA binding box deletion or an inactivated LexA binding box; and

DNA sequences encoding RecA and RecX on a single transcript, wherein a terminator between the DNA sequence encoding RecA and the DNA sequence encoding RecX is deleted.

Embodiment 150 is a prokaryotic cell comprising:

a DNA sequence in which the sequence encoding IS1 transposase is deleted and replaced with an inactive recombination sequence, wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence;

a DNA sequence in which *gpE* encoding a lambda DE3 phage capsid protein is deleted and replaced with an inactive recombination sequence, wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence;

a chimeric pyrrolysyl-tRNA synthetase comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I,

T56P, H62Y, A100E, C313V, and Y349F, wherein the tRNA synthetase is integrated into a chromosomal *arsB* locus under the control of an inducible promoter;

a DNA sequence encoding a repressor of the inducible promoter integrated into the chromosomal *arsB* locus;

a DNA sequence encoding a truncated *Pt*NTT2 integrated into a chromosomal *lacZYA* locus of the cell replacing *lacZY*, wherein the DNA sequence encoding the truncated *Pt*NTT2 is operably linked to a promoter, optionally wherein the promoter is P_{lac}, and wherein the DNA sequence encoding the truncated *Pt*NTT2 does not comprise an oppositely oriented promoter within 1.3-1.5 kb downstream of the 3' end of the truncated PtNTT2 coding sequence;

a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed, further wherein the derepressed promoter is a *polB* promoter with a LexA binding box deletion or an inactivated LexA binding box;

DNA sequences encoding RecA and RecX on a single transcript, wherein a terminator between the DNA sequence encoding RecA and the DNA sequence encoding RecX is deleted; and

a DNA sequence encoding LexA comprising a S119A mutation.

Embodiment 151 is a prokaryotic cell comprising:

a DNA sequence in which the sequence encoding IS1 transposase is deleted and replaced with an inactive recombination sequence, wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence;

a DNA sequence in which *gpE* encoding a lambda DE3 phage capsid protein is deleted and replaced with an inactive recombination sequence, wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence;

a chimeric pyrrolysyl-tRNA synthetase comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F, wherein the tRNA synthetase is integrated into a chromosomal *arsB* locus under the control of an inducible promoter;

a DNA sequence encoding a repressor of the inducible promoter integrated into the chromosomal *arsB* locus;

a DNA sequence encoding a truncated *Pt*NTT2 integrated into a chromosomal *lacZYA* locus of the cell replacing *lacZY*, wherein the DNA sequence encoding the truncated *Pt*NTT2 is operably linked to a promoter, optionally wherein the promoter is

P_{lac}, and wherein the DNA sequence encoding the truncated *Pt*NTT2 does not comprise an oppositely oriented promoter within 1.3-1.5 kb downstream of the 3' end of the truncated PtNTT2 coding sequence;

a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed, further wherein the derepressed promoter is a *polB* promoter with a LexA binding box deletion or an inactivated LexA binding box;

DNA sequences encoding RecA and RecX on a single transcript, wherein a terminator between the DNA sequence encoding RecA and the DNA sequence encoding RecX is deleted;

a DNA sequence encoding LexA comprising a S119A mutation; and

a DNA sequence comprising a *relA* deletion or inactivated *relA*.

Embodiment 152 is the cell of any one of embodiments 146-151, wherein the cell is a bacterial cell.

Embodiment 153 is the cell of embodiment 152, wherein the bacterial cell is *E. coli*.

Embodiment 154 is the cell of embodiment 152 or 153, wherein the cell comprises a BL21(DE3) background.

Embodiment 155 is the cell of any one of embodiments 146-154, wherein the inducible promoter is an IPTG-inducible promoter.

Embodiment 156 is the cell of embodiment 155, wherein the IPTG-inducible promoter is Tac or TacO.

Embodiment 157 is the cell of embodiment 155 or embodiment 156, wherein the repressor of the IPTG-inducible promoter is LacI repressor protein.

Embodiment 158 is the cell of embodiment 157, wherein the promoter for the sequence encoding the repressor of the inducible promoter is a P_{lacIQ1} promoter.

Embodiment 159 is the cell of any one of embodiments 146-158, wherein the chimeric pyrrolysyl-tRNA synthetase is operably linked to a 5' UTR comprising SEQ ID NO: 2.

Embodiment 159.1 is the cell of any one of embodiments 146-158, wherein the chimeric pyrrolysyl-tRNA synthetase is operably linked to a 5' UTR comprising a lac operator and/or SEQ ID NO: 13.

Embodiment 160 is the cell of any one of embodiments 1-89 and 135-159, wherein the cell comprises a DNA sequence encoding an mRNA molecule comprising at least one unnatural codon encoding an unnatural amino acid.

Embodiment 161 is the cell of any one of embodiments 1-89 and 135-159, wherein the cell comprises a DNA sequence encoding an mRNA molecule comprising at least first and second unnatural codons each encoding an unnatural amino acid.

Embodiment 162 is the cell of any one of embodiments 1-89 and 135-159, wherein the cell comprises a DNA sequence encoding an mRNA molecule comprising first and second unnatural codons, wherein the first and second unnatural codons encode a first and second unnatural amino acid.

Embodiment 163 is the cell of embodiment 161 or 162, wherein the first unnatural amino acid is different from the second unnatural amino acid.

Embodiment 164 is the cell of embodiment 163, wherein the cell comprises a second tRNA and a second tRNA synthetase, wherein the second tRNA synthetase can aminoacylate the second tRNA and forms a (second) orthogonal tRNA synthetase/tRNA pair.

Embodiment 165 is the cell of embodiment 164, wherein the second tRNA synthetase is a *Methanocaldococcus jannaschii* tRNA synthetase.

Embodiment 166 is the cell of embodiment 165, wherein the second tRNA is a tRNA from *Methanocaldococcus jannaschii*.

Embodiment 167 is the cell of any one of embodiments 163-166, wherein the second unnatural amino acid is para-Acetyl-L-phenylalanine (*p*AcPhe).

Embodiment 168 is the cell of any one of embodiments 163-167, wherein the first unnatural amino acid is AzK and the second unnatural amino acid is *p*AcPhe.

Embodiment 169 is the cell of embodiment 167 or 168, wherein the second unnatural codon read from a 5' to 3' direction comprises AXC, wherein X is an unnatural nucleotide.

Embodiment 170 is the cell of embodiment 169, wherein the first unnatural codon read from a 5' to 3' direction comprises GXC.

Embodiment 171 is the cell of any one of embodiments 160-170, wherein the mRNA molecule encodes a polypeptide comprising a wild-type OmpA signal peptide.

Embodiment 172 is the cell of any one of embodiments 160-171, wherein the wild-type OmpA signal peptide is from *E. coli*.

BRIEF DESCRIPTION OF THE DRAWINGS

[7] Figure 1 shows a western blot of the pyrrolysyl tRNA synthetase (labeled as "Synthetase") from *Methosarcina barkeri* (SEQ ID NO: 3). The "Target" is a polypeptide having an unnatural amino acid.

[8] Figure 2 is an illustration of a wild-type *M. barkeri* pyrrolysyl tRNA synthetase ("Control Synthetase") (e.g., SEQ ID NO: 3).

[9] Figure 3 is an illustration of a chimeric pyrrolysyl tRNA synthetase ("Chimeric Synthetase") (e.g., SEQ ID NO: 4) having an N-terminal domain ("NTD") from the *M. barkeri*

pyrrolysyl tRNA synthetase fused to the C-terminal domain ("CTD") from the *M. mazei* pyrrolysyl tRNA synthetase.

[10] Figure 4 is an illustration of a split pyrrolysyl tRNA synthetase ("Split Synthetase") having an N-terminal domain ("NTD") from the *M. barkeri* pyrrolysyl tRNA synthetase and separately the C-terminal domain ("CTD") from the *M. barkeri* pyrrolysl tRNA synthetase.

[11] Figure 5 is an illustration of a split chimeric pyrrolysyl tRNA synthetase ("Split Chimeric Synthetase") having an N-terminal domain ("NTD") from the *M. barkeri* pyrrolysyl tRNA synthetase and separately the C-terminal domain ("CTD") from the *M. mazei* pyrrolysyl tRNA synthetase.

[12] Figure 6 shows results from a superfolder green fluorescent protein (GFP) fluorescence experiment testing expression of a polypeptide with a non-canonical amino acid at position 151 of the superfolder GFP gene. Four candidate pyrrolysyl tRNA synthetases were tested: Control (wild-type *M. barkeri* synthetase); Split #1 (*M. barkeri* pyrrolysyl-tRNA synthetase having two separate polypeptide chains, wherein the first polypeptide chain has part of a *M. barkeri* pyrrolysyl-tRNA synthetase with an amino acid sequence of positions 1-110 of SEQ ID NO: 3, and wherein the second polypeptide chain has part of a *M. barkeri* pyrrolysyl-tRNA synthetase with an amino acid sequence of positions 150-419 of SEQ ID NO: 3); Split #2 (*M. barkeri* pyrrolysyl-tRNA synthetase having two separate polypeptide chains, wherein the first polypeptide chain has part of a *M. barkeri* pyrrolysyl-tRNA synthetase with an amino acid sequence of positions 1-137 of SEQ ID NO: 3, and wherein the second polypeptide chain has part of a *M. barkeri* pyrrolysyl-tRNA synthetase with an amino acid sequence of positions 150-419 of SEQ ID NO: 3); and Split #3 (*M. barkeri* pyrrolysyl-tRNA synthetase having two separate polypeptide chains, wherein the first polypeptide chain has part of a *M. barkeri* pyrrolysyl-tRNA synthetase with an amino acid sequence of positions 1-149 of SEQ ID NO: 3, and wherein the second polypeptide chain has part of a *M. barkeri* pyrrolysyl-tRNA synthetase with an amino acid sequence of positions 150-419 of SEQ ID NO: 3).

[13] Figure 7 is an illustration of *E. coli* thioredoxin (TrxA) C-terminally fused to the N-terminal domain of a split *M. barkeri* pyrrolysyl synthetase.

[14] Figure 8 shows results from a GFP fluorescence experiment testing expression of polypeptide with a non-canonical amino acid at position 151 of the superfolder GFP gene. Six candidate pyrrolysyl tRNA synthetases were tested, three of which had TrxA C-terminally fused to the N-terminal domains of split pyrrolysyl synthetases: Split #1 (*M. barkeri* pyrrolysyl-tRNA synthetase having two separate polypeptide chains, wherein the first polypeptide chain has part of a *M. barkeri* pyrrolysyl-tRNA synthetase with an amino acid

sequence of positions 1-110 of SEQ ID NO: 3, and wherein the second polypeptide chain has part of a *M. barkeri* pyrrolysyl-tRNA synthetase with an amino acid sequence of positions 150-419 of SEQ ID NO: 3); Split #1 fused to TrxA(N); Split #2 (*M. barkeri* pyrrolysyl-tRNA synthetase having two separate polypeptide chains, wherein the first polypeptide chain has part of a *M. barkeri* pyrrolysyl-tRNA synthetase with an amino acid sequence of positions 1-137 of SEQ ID NO: 3, and wherein the second polypeptide chain has part of a *M. barkeri* pyrrolysyl-tRNA synthetase with an amino acid sequence of positions 150-419 of SEQ ID NO: 3), Split #2 fused to TrxA(N), Split #3 (*M. barkeri* pyrrolysyl-tRNA synthetase having two separate polypeptide chains, wherein the first polypeptide chain has part of a *M. barkeri* pyrrolysyl-tRNA synthetase with an amino acid sequence of positions 1-149 of SEQ ID NO: 3, and wherein the second polypeptide chain has part of a *M. barkeri* pyrrolysyl-tRNA synthetase with an amino acid sequence of positions 150-419 of SEQ ID NO: 3), and Split #3 fused to TrxA(N).

[15] Figure 9 is an illustration of higher-stringency constructs used to detect enzymatic activity of the pyrrolysyl tRNA synthetases using GFP fluorescence. Either one, two, or three TAG stop codons (1xTAG, 2xTAG, or 3xTAG) were inserted in the sequence for production of GFPs. The 1xTAG construct included a TAG codon at position 151, the 2xTAG construct included TAG codons at positions 151 and 190, and the 3xTAG construct included TAG codons at positions 151, 190, and 200. The GFP constructs all included a T7 promoter.

[16] Figure 10 shows results from an experiment in which higher-stringency constructs with a T7 promoter were used to detect enzymatic activity of four pyrrolysyl tRNA synthetases using GFP fluorescence. Results are shown for constructs with one, two, and three TAG stop codons (1xTAG, 2xTAG, or 3xTAG) for each of the following four pyrrolysyl tRNA synthetases: Control (*M. barkeri* pyrrolysyl tRNA synthetase); Split (split pyrrolysyl tRNA synthetase having an NTD from the *M. barkeri* pyrrolysyl tRNA synthetase and separately the CTD from the *M. barkeri* pyrrolysl tRNA synthetase), Chimeric (chimeric pyrrolysyl tRNA synthetase having an NTD from the *M. barkeri* pyrrolysyl tRNA synthetase fused to the CTD from the *M. mazei* pyrrolysl tRNA synthetase), and Split Chimeric (split chimeric pyrrolysyl tRNA synthetase having an NTD from the *M. barkeri* pyrrolysyl tRNA synthetase and separately the CTD from the *M. mazei* pyrrolysyl tRNA synthetase).

[17] Figure 11 shows a western blot from an experiment in which four pyrrolysyl tRNA synthetases were tested for insoluble contamination when producing Molecule B with a T7 promoter. “Hi” meaning high expression refers to expression level of synthetase variants under the standard ptac promoter: “Control” (wild-type *M. barkeri* synthetase); Split #2; “Chimeric”—*M. barkeri* PylRS NTD fused to *M. mazei* PylRS CTD; and “Split Chimeric”—

M. barkeri PylRS NTD and, separately, *M. mazei* PylRS CTD. Titer, amount measured, and fidelity results are shown for expression of Molecule B.

[18] Figure 12 shows results from an experiment in which higher-stringency constructs with a T7 promoter were used to detect enzymatic activity of four pyrrolysyl tRNA synthetases using GFP fluorescence. Results are shown for constructs with one, two, and three TAG stop codons (1xTAG, 2xTAG, or 3xTAG) for each of the following four pyrrolysyl tRNA synthetases: Control (*M. barkeri* pyrrolysyl tRNA synthetase); Split (split pyrrolysyl tRNA synthetase having an NTD from the *M. barkeri* pyrrolysyl tRNA synthetase and separately the CTD from the *M. barkeri* pyrrolysl tRNA synthetase), Chimeric (chimeric pyrrolysyl tRNA synthetase having an NTD from the *M. barkeri* pyrrolysyl tRNA synthetase fused to the CTD from the *M. mazei* pyrrolysl tRNA synthetase), and Split Chimeric (split chimeric pyrrolysyl tRNA synthetase having an NTD from the *M. barkeri* pyrrolysyl tRNA synthetase and separately the CTD from the *M. mazei* pyrrolysl tRNA synthetase).

[19] Figure 13 shows results from an experiment in which higher-stringency GFP constructs and lower expression synthetase constructs were used to detect enzymatic activity of four pyrrolysyl tRNA synthetases using GFP fluorescence. Results are shown for constructs with one, two, and three TAG stop codons (1xTAG, 2xTAG, or 3xTAG) for each of the following four pyrrolysyl tRNA synthetases: Control (*M. barkeri* pyrrolysyl tRNA synthetase); Split #2 (split pyrrolysyl tRNA synthetase having an NTD from the *M. barkeri* pyrrolysyl tRNA synthetase and separately the CTD from the *M. barkeri* pyrrolysl tRNA synthetase), Chimeric (chimeric pyrrolysyl tRNA synthetase having an NTD from the *M. barkeri* pyrrolysyl tRNA synthetase fused to the CTD from the *M. mazei* pyrrolysl tRNA synthetase), and Split Chimeric (split chimeric pyrrolysyl tRNA synthetase having an NTD from the *M. barkeri* pyrrolysyl tRNA synthetase and separately the CTD from the *M. mazei* pyrrolysl tRNA synthetase).

[20] Figure 14 shows a western blot from an experiment in which four pyrrolysyl tRNA synthetases were tested for insoluble contamination when producing Molecule B with a a T7 promoter. Synthetase variants were expressed with standard ptac promoter for “Hi” (meaning high expression) or a modified promoter with a SalI digestion site deleted for “Low” expression: “Control” (wild-type *M. barkeri* synthetase); Split #2; “Chimeric”—*M. barkeri* PylRS NTD fused to *M. mazei* PylRS CTD; and “Split Chimeric”—*M. barkeri* PylRS NTD and, separately, *M. mazei* PylRS CTD. Titer, amount measured, and fidelity results are shown for expression of Molecule B.

[21] Figure 15 shows results from an experiment in which higher-stringency constructs under a T7 promoter were used to detect enzymatic activity of four pyrrolysyl tRNA

synthetases with "low" expression using GFP fluorescence. Results are shown for constructs with one, two, and three TAG stop codons (1xTAG, 2xTAG, or 3xTAG) for each of the following three pyrrolysyl tRNA synthetases: "Chimeric Synthetase"—*M. barkeri* PylRS NTD fused to *M. mazei* PylRS CTD; "Chimeric (IPYE) Synthetase"—the "Chimeric Synthetase" with the IPYE mutations; and "Chimeric (IPYEVF) Synthetase"—the "Chimeric Synthetase" with the IPYEVF mutations.

[22] Figure 16 shows titer results for production of Molecule A. Two pyrrolysyl tRNA synthetases were tested: the Control Synthetase and the "Chimeric (IPYEVF) Synthetase." The Control Synthetase (wild-type *M. barkeri* synthetase) was under a high expression promoter, the T7 promoter. The "Chimeric (IPYEVF) Synthetase"—*M. barkeri* PylRS NTD fused to *M. mazei* PylRS CTD with the IPYEVF mutations—was under a low expression promoter.

[23] Figure 17 shows fidelity results for production of Molecule A. Two pyrrolysyl tRNA synthetases were tested: the Control Synthetase and the "Chimeric (IPYEVF) Synthetase." The Control Synthetase (wild-type *M. barkeri* synthetase) was under a high expression promoter, the T7 promoter. The "Chimeric (IPYEVF) Synthetase"—*M. barkeri* PylRS NTD fused to *M. mazei* PylRS CTD with the IPYEVF mutations—was under a low expression promoter.

[24] Figure 18 shows titer results for production of Molecule B. Two pyrrolysyl tRNA synthetases were tested: the Control Synthetase and the "Chimeric (IPYEVF) Synthetase." The Control Synthetase (wild-type *M. barkeri* synthetase) was under a high expression promoter, the T7 promoter. The "Chimeric (IPYEVF) Synthetase"—*M. barkeri* PylRS NTD fused to *M. mazei* PylRS CTD with the IPYEVF mutations—was under a low expression promoter.

[25] Figure 19 shows fidelity results for production of Molecule B. Two pyrrolysyl tRNA synthetases were tested: the Control Synthetase and the "Chimeric (IPYEVF) Synthetase." The Control Synthetase (wild-type *M. barkeri* synthetase) was under a high expression promoter, the T7 promoter. The "Chimeric (IPYEVF) Synthetase"—*M. barkeri* PylRS NTD fused to *M. mazei* PylRS CTD with the IPYEVF mutations—was under a low expression promoter.

[26] Figure 20 shows HPLC results for production of Molecule B. One pyrrolysyl tRNA synthetase was tested: the Control Synthetase (wild-type *M. barkeri* synthetase) under a high expression promoter, the T7 promoter.

[27] Figure 21 shows HPLC results for production of Molecule B. One pyrrolysyl tRNA synthetase was tested: "Chimeric (IPYEVF) Synthetase"—*M. barkeri* PylRS NTD fused to *M. mazei* PylRS CTD with the IPYEVF mutations—under a low expression promoter.

[28] Figure 22 shows titer/OD (mg/L/OD) and Final OD results for production of Molecule B with various concentrations of AzK using the "Chimeric (IPYEVF) Synthetase"—*M. barkeri* PylRS NTD fused to *M. mazei* PylRS CTD with the IPYEVF mutations.

[29] Figure 23 shows fidelity (%) and titer (mg/L) results for production of Molecule B with various concentration of AzK using the "Chimeric (IPYEVF) Synthetase"—*M. barkeri* PylRS NTD fused to *M. mazei* PylRS CTD with the IPYEVF mutations.

[30] Figure 24 shows titer (mg/L) and titer/OD results for Strain J (with a Chimeric (IPYEVF) Synthetase) and Strain J with a Chimeric (IPYEVF) Synthetase that includes two additional mutations, P5T and E302A.

[31] Figure 25 shows titer (mg/L) and titer/OD results for Strain J (with a Chimeric (IPYEVF) Synthetase) and Strain J with a Chimeric (IPYEVF) Synthetase that includes two additional mutations, P5T and E302A.

[32] Figure 26 shows OD results for production of Molecule B with Strain J (with a Chimeric (IPYEVF) Synthetase) and Strain J with a Chimeric (IPYEVF) Synthetase that includes two additional mutations, P5T and E302A, at various AzK concentrations.

[33] Figure 27 shows titer (mg/L) results for production of Molecule B with Strain J (with a Chimeric (IPYEVF) Synthetase) and Strain J with a Chimeric (IPYEVF) Synthetase that includes two additional mutations, P5T and E302A, at various AzK concentrations.

[34] Figure 28 shows fidelity (%) results for production of Molecule B with Strain J (with a Chimeric (IPYEVF) Synthetase) and Strain J with a Chimeric (IPYEVF) Synthetase that includes two additional mutations, P5T and E302A, at various AzK concentrations.

[35] Figure 29 shows titer/OD (mg/L/OD) results for production of Molecule B with Strain J (with a Chimeric (IPYEVF) Synthetase) and Strain J with a Chimeric (IPYEVF) Synthetase that includes two additional mutations, P5T and E302A, at various AzK concentrations.

[36] Figure 30 shows a western blot with results for production of Molecule B with Strain J (with a Chimeric (IPYEVF) Synthetase) and Strain J with a Chimeric (IPYEVF) Synthetase that includes two additional mutations, P5T and E302A, at various AzK concentrations.

[37] Figure 31 is an illustration that shows the basic components of a strain used for producing certain protein products having an unnatural amino acid, where X and Y each represent an unnatural nucleotide.

[38] Figure 32 is an illustration that shows a control strain with two plasmids: one encoding Molecule A (or Molecule B), a tRNA, and a zeocin antibiotic resistance marker, and the other encoding a chimeric pyrrolysyl tRNA synthetase and a tetracycline antibiotic resistance marker.

[39] Figure 33 is an illustration that shows a strain with a chromosomally integrated pyrrolysyl tRNA synthetase and one plasmid encoding Molecule A (or Molecule B), a tRNA, and a zeocin antibiotic resistance marker.

[40] Figure 34 shows an illustration of a construct with a plasmid encoding the chimeric pyrrolysyl tRNA synthetase (IPYEVF) with a standard 5' UTR (SEQ ID NO: 1), which includes a lac operator, a tac promoter, and a tetracycline resistance marker.

[41] Figure 35 shows an illustration of a construct with a chromosomally integrated chimeric pyrrolysyl tRNA synthetase (IPYEVF) with a standard 5' UTR (SEQ ID NO: 1), which includes a lac operator, and a tac promoter.

[42] Figure 36 shows an illustration of a construct with a chromosomally integrated chimeric pyrrolysyl tRNA synthetase (IPYEVF) with a 5' UTR that was modified to include SEQ ID NO: 2, which includes a lac operator, and a tac promoter.

[43] Figure 37 shows an illustration of a construct used to detect enzymatic activity of the pyrrolysyl tRNA synthetases using GFP fluorescence with one TAG stop codon inserted at position 151 in the sequence for production of GFP.

[44] Figure 38 shows results from an experiment in which GFP fluorescence was used to detect enzymatic activity of three pyrrolysyl tRNA synthetases. The first construct ("chPylRS(IPYEVF) plasmid RBS orig") included a plasmid encoding the chimeric pyrrolysyl tRNA synthetase (IPYEVF) with a standard 5' UTR (SEQ ID NO: 1), which includes a lac operator, a tac promoter, and a tetracycline resistance marker. The second construct ("chPylRS(IPYEVF) chromosomal RBS orig") included a chromosomally integrated chimeric pyrrolysyl tRNA synthetase (IPYEVF) with a standard 5' UTR (SEQ ID NO: 1), which includes a lac operator, and a tac promoter. The third construct ("chPylRS(IPYEVF) chromosomal RBS opt") included a chromosomally integrated chimeric pyrrolysyl tRNA synthetase (IPYEVF) with a 5' UTR that was modified to include SEQ ID NO: 2, which includes a lac operator, and a tac promoter.

[45] Figure 39 shows an illustration of a construct with a chromosomally integrated chimeric pyrrolysyl tRNA synthetase (IPYEVF) with a 5' UTR that was modified to include SEQ ID NO: 2, which includes a lac operator, a tac promoter, and also lacIQ promoter for the lacI repressor and a lacI repressor.

[46] Figure 40 shows an illustration of a construct with a chromosomally integrated chimeric pyrrolysyl tRNA synthetase (IPYEVF) with a 5' UTR that was modified to include SEQ ID NO: 2, which includes a lac operator, a tac promoter, and also lacIQ1 promoter for the lacI repressor and a lacI repressor.

[47] Figure 41 shows photographs of plates of cells of Strain A, Strain E, and Strain F at different stages in the cell expansion process.

[48] Figure 42 shows a western blot with results from expression of Molecule A precursor by the following strains: Strain E, Strain F, Strain G, Strain H, Strain B, Strain C, and Strain A. Characteristics of each strain are shown below the western blot.

[49] Figures 43A-43C show results from expression of Molecule A by strains Strain E, Strain F, Strain G, Strain H, Strain B, Strain C, and Strain A. Titer results (mg/L) are shown in Figure 43A. Fidelity (% AzK at P65) are shown in Figure 43B. Final OD_{600} results are shown in Figure 43C.

[50] Figure 44 shows a western blot with results from expression of Molecule A precursor by Strain F, Strain D, and Strain A.

[51] Figure 45 shows fidelity (%) results from expression of Molecule A at different stages in the cell expansion process by Strain A and Strain F.

[52] Figure 46 shows photographs of plates of cells of Strain F and Strain A at different stages of the cell expansion process.

[53] Figure 47 shows photographs of plates of cells of Strain A (+ expression plasmid for a polypeptide with an unnatural amino acid) and Strain F at different stages of the cell expansion process.

[54] Figure 48 shows illustrations of three constructs encoding *Pt*NTT2: a construct for Strain F, a construct for Strain I, and a construct for Strain J.

[55] Figure 49 shows results from an experiment testing ATP export (RLU/s/OD) for a strain with no nucleotide triphosphate transporter, Strain I, Strain J, and Strain F.

[56] Figure 50 shows photographs of plates of cells of a strain with no nucleotide triphosphate transporter, Strain I, Strain J, and Strain F. In addition, Molecule B % fidelity results for expression of Molecule B for each of Strain I, Strain J, and Strain F are shown.

[57] Figures 51A-51C show fidelity results for expression of Molecule A (Figure 51A), Molecule B (Figure 51B), and Molecule C (Figure 51C) for Strain A, Strain F, and/or Strain J at different stages of the cell expansion process. Figure 51A shows fidelity results for expression of Molecule A by Strain A and Strain F. Figure 51B shows fidelity results for expression of Molecule B by Strain F and Strain J. Figure 51C shows fidelity results for expression of Molecule C by Strain F and Strain J.

[58] Figure 52A-52B show results from an experiment testing ATP export (RLU/s/OD) for Strain F (Figure 52A) and Strain J (Figure 52B) at different stages of the cell expansion process.

[59] Figure 53 shows photographs of plates of cells for Strain F and Strain J at different stages of the cell expansion process.

[60] Figure 54 shows photographs of plates of cells for Strain J and Strain J with an expression plasmid (for expressing a polypeptide with an unnatural amino acid) at different stages of the cell expansion process.

[61] Figure 55 shows photographs of plates of cells for Strain J and Strain F at different stages of the cell expansion process.

[62] Figure 56 shows photographs of plates of cells for Strain F and Strain J at different stages of the cell expansion process. Figure 56 also shows a diagram of the media used for growing the cells at different stages of the cell expansion process.

[63] Figures 57A-57D shows results for experiments in which Molecule B was expressed using Strain J or Strain K with different options for a deoxyribonucleotide (dX/dYTP) bolus reduction (none, pre-ind, post-ind, or both). Each experiment was conducted with either dNaMTP as one of the deoxyribonucleotides or dCNMOTP. Figure 57A shows fidelity (% AzK at H16); Figure 57B shows titer (mg/L), Figure 57C shows Titer / OD (mg/L/OD), and Figure 57D shows final OD.

[64] Figure 58A-58D shows results for experiments in which Molecule B was expressed using Strain J or Strain K with different options for a deoxyribonucleotide (dX/dYTP) bolus reduction (none, pre-ind, post-ind, or both). Each experiment was conducted with either dNaMTP as one of the deoxyribonucleotides or dCNMOTP. Figure 58A shows fidelity (% AzK at H16); Figure 58B shows titer (mg/L), Figure 58C shows Titer / OD (mg/L/OD), and Figure 58D shows final OD.

[65] Figures 59A-59B show results from expression of Molecule B by Strain J in an experiment performed to investigate the effect of substituting dCNMOTP for dNaMTP. Figure 59A shows titer results (mg/L) and Figure 59B shows fidelity (%) results.

[66] Figures 60A-60D show results from experiments conducted to determine whether increasing the expression of pyrrolysyl tRNA synthetase would improve Strain J's tolerance to when and how much AzK was added during the culture process. Different promoters were used for the synthetase—TacO, 23119, 0.26x T7, and 3.65x T7. Figure 60A shows results in terms of Final OD, OD Final / OD Induction, Titer (mg/L), % Fidelity, and Titer/OD. FIGS 60B-60D show four different shades of bars to represent different concentrations of AzK relative to standard concentration (standard, half, or quarter) for the first

bolus and the second bolus. Figure 60B shows titer (mg/L) results. Figure 60C shows (mg/L per OD results). Figure 60D shows fidelity (%) results.

TABLE 1 – DESCRIPTION OF THE SEQUENCES

Description	Sequence	SEQ ID NO
Original 5′ UTR (untranslated region) in Strain A	AATTGTGAGCGGATAACAATTCACACAGGAAACAGTATTC	1
Improved 5′ UTR for chromosomally integrated pyrrolysyl tRNA synthetase	AATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACAT	2
Methanosarcina barkeri pyrrolysyl-tRNA synthetase	MDKKPLDVLISATGLWMSRTGTLHKIKHHEVSRSKIYIEMACGDHLVVNNSRSCRTARAFRHHKYRKTCKRCRVSDEDINNFLTRSTE SKNSVKVRVVSAPKVKKAMPKSVSRAPKPLENSVSAKASTNTSRSVPSPAKSTPNSSVPASAPAPSLTRSQLDRVEALLSPEDKISIN MAKPFRELEPELVTRRKNDFQRLYTNDREDYLGKLERDITKFFVDRGFLEIKSPILIPAEYVERMGINNDTELSKQIFRVDKNLCLRP MLAPTLYNYLRKLDRILPGPIKIFEVGPCYRKESDGKEHLEEFTMVNFCQMGSGCTRENLEALIKEFLDYLEIDFEIVGDSCMVYGDT LDIMHGDLELSSAVVGPVSLDREWGIDKPWIGAGFGLERLLKVMHGFKNIKRASRSESYYNGISTNL*	3
Chimeric pyrrolysyl-tRNA synthetase (*Methanosarcina barkeri* is bold, *Methanosarcina mazei* is underlined)	**MDKKPLDVLISATGLWMSRTGTLHKIKHHEVSRSKIYIEMACGDHLVVNNSRSCRTARAFRHHKYRKTCKRCRVSDEDINNFLTRSTE** **SKNSVKVRVVSAPKVKKAMPKSVSRAPKPLENSVSAKASTNTSRSVPSPAKSTPNSSVPAS**ASAPALTKSQTDRLEVLLNPKDEISLN SGKPFRELESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDKNFCLRP MLAPNLYNYLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFTMLNFCQMGSGCTRENLESIITDFLNHLGIDFKIVGDSCMVYGDT LDVMHGDLELSSAVVGPIPLDREWGIDKPWIGAGFGLERLLKVKHDFKNIKRAARSESYYNGISTNL*	4
Methanosarcina mazei pyrrolysyl-tRNA synthetase	MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTARALRHHKYRKTCKRCRVSDEDLNKFLTKANE DQTSVKVKVVSAPTRTKKAMPKSVARAPKPLENTEAAQAQPSGSKFSPAIPVSTQESVSVPASVSTSISSISTGATASALVKGNTNPI TSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPFRELESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKS PILIPLEYIERMGIDNDTELSKQIFRVDKNFCLRPMLAPNLYNYLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFTMLNFCQMGS GCTRENLESIITDFLNHLGIDFKIVGDSCMVYGDTLDVMHGDLELSSAVVGPIPLDREWGIDKPWIGAGFGLERLLKVKHDFKNIKRA ARSESYYNGISTNL*	5
Chimeric pyrrolysyl-tRNA synthetase *with IPYEVF mutations*	MDKKPLDVLISATGLWMSRTGTLHKIKHHEISRSKIYIEMACGDHLVVNNSRSCRPARAFRYHKYRKTCKRCRVSDEDINNFLTRSTE SKNSVKVRVVSEPKVKKAMPKSVSRAPKPLENSVSAKASTNTSRSVPSPAKSTPNSSVPASASAPALTKSQTDRLEVLLNPKDEISLN SGKPFRELESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDKNFCLRP	6

(V31I, T56P, H62Y, A100E, C313V, and Y349F)	MLAPNLYNYLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFTMLNFVQMGSGCTRENLESIITDFLNHLGIDFKIVGDSCMVFGDT LDVMHGDLELSSAVVGPIPLDREWGIDKPWIGAGFGLERLLKVKHDFKNIKRAARSESYYNGISTNL*	
Chimeric pyrrolysyl-tRNA synthetase with IPYEVF + TA mutations (V31I, T56P, H62Y, A100E, C313V, and Y349F + P5T and E302A)	MDKKTLDVLISATGLWMSRTGTLHKIKHHEISRSKIYIEMACGDHLVVNNSRSCRPARAFRYHKYRKTCKRCRVSDEDINNFLTRSTE SKNSVKVRVVSEPKVKKAMPKSVSRAPKPLENSVSAKASTNTSRSVPSPAKSTPNSSVPASASAPALTKSQTDRLEVLLNPKDEISLN SGKPFRELESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRVDKNFCLRP MLAPNLYNYLRKLDRALPDPIKIFEIGPCYRKESDGKAHLEEFTMLNFVQMGSGCTRENLESIITDFLNHLGIDFKIVGDSCMVFGDT LDVMHGDLELSSAVVGPIPLDREWGIDKPWIGAGFGLERLLKVKHDFKNIKRAARSESYYNGISTNL*	
DNA sequence encoding chimeric pyrrolysyl-tRNA synthetase (IPYEVF) operably linked to an inducible promoter and an improved 5′ UTR (underlined)	TTGACAATTAATCATCGGCTCGTATAATGTGTGG**AATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGA** **AGGAGATATACAT**ATGGATAAAAAACCGCTGGACGTTCTGATCTCCGCTACGGGTCTGTGGATGAGCCGCACGGGTACGCTGCATAAA ATTAAACACCACGAAATTTCACGTTCGAAAATCTATATCGAAATGGCGTGCGGTGATCATCTGGTGGTTAACAATAGCCGTTCTTGTC GCCCCGCGCGTGCCTTTCGCTATCACAAATACCGCCAAAACGTGCAAACGTTGTCGCGTCAGATGAAGACATTAACAATTTCCTGAC CCGTAGTACGGAATCCAAAACTCAGTGAAAGTTCGCGTCGTGAGTGAGCCGAAAGTTAAAAAAGCGATGCCGAAAAGTGTCTCCCGT GCCCCGAAACCGCTGGAAAACTCAGTGTCGGCAAAAGCTTCCACCAATACGAGCCGCTCTGTTCCGTCGCCGGCAAAAAGCACCCCGA ACAGCTCTGTCCCGGCAAGCGCATCTGCACCGGCACTTACAAAGTCCCAAACCGACCGTTTAGAAGTCTTGCTTAATCCGAAGGATGA AATTTCGTTAAACTCTGGCAAGCCGTTTCGTGAGCTGGAGTCAGAACTTCTTTCACGGCGTAAAAAAGATCTTCAGCAGATATATGCT GAGGAGCGCGAGAATTACTTAGGAAAGCTGGAACGTGAAATAACTCGTTTTTTGTTGACAGAGGGTTCTTGGAGATAAAGTCCCCAA TACTGATACCGTTAGAATACATCGAAAGAATGGGTATAGATAATGATACCGAGTTAAGTAACAAATATTTAGAGTTGATAAGAATTT TTGTTTGCGGCCGATGTTGGCGCCCAATCTTTACAACTATCTTCGGAAGCTGGACCGCGCCTTCCTGACCCGATCAAAATCTTTGAA ATTGGTCCCTGTTATCGGAAGGAATCGGATGGCAAGGAACATTTGGAGGAGTTCACCATGCTGAACTTTGTGCAAATGGGGTCTGGCT GCACGCGGGAAAATTTGGAAAGCATAATTACGGATTTTCTGAACCACCTTGGCATAGACTTTAAGATCGTTGGAGATTCTTGCATGGT TTTTGGGGATACGCTGGATGTAATGCACGGAGATTTAGAGCTTTCAAGTGCAGTAGTAGGACCTATTCCGTTAGACAGAGAATGGGGA ATTGACAAGCCATGGATCGGTGCGGGTTTCGGCTTGGAGCGCCTGCTTAAGGTAAAGCACGACTTTAAAAACATAAAGCGGGCGGCAA GATCTGAGAGTTATTACAACGGAATATCAACAAATTTGTAA	
Truncated *Pt*NTT2 (66-575) nucleoside triphosphate transporter (protein sequence)	MGGSTVAPTTPLATGGALRKVRQAVFPIYGNQEVTKFLLIGSIKFFIILALTLTRDTKDTLIVTQCGAEAIAFLKIYGVLPAATAFIA LYSKMSNAMGKKMLFYSTCIPFFTFFGLFDVFIYPNAERLHPSLEAVQAILPGGAASGGMAVLAKIATHWTSALFYVMAEIYSSVSVG LLFWQFANDVVNVDQAKRFYPLFAQMSGLAPVLAGQYVVVRFASKAVNFEASMHRLTAAVTFAGVEQAIAFLKIYGVLPAATAFIA NEQSIKPKKKKPKMSMVESGKFLASSQYLRLIAMLVLGYGLSINFTEIMWKSLVKKQYPDPLDYQRFMGNFSSAVGLSTCIVIFFGVH VIRLLGWKVGALATPGIMAILALPFFACILLGLDSPARLEIAVIFGTIQSLLSKTSKYALFDPTTQMAYIPLDDESKVKGKAAIDVLG SRIGKSGGSLIQQGLVFVFGNIINAAPVVGVVYYSVLVAWMSAAGRLSGLFQAQTEMDKADKMEAKTNKEK*	
Wild-type, full length *Pt*NTT2 nucleoside triphosphate	MRPYPTIALISVFLSAATRISATSSHQASALPVKKGTHVPDSPKLSKLYIMAKTKSVSSSFDPPRGGSTVAPTTPLATGGALRKVRQA VFPIYGNQEVTKFLLIGSIKFFIILALTLTRDTKDTLIVTQCGAEAIAFLKIYGVLPAATAFIALYSKMSNAMGKKMLFYSTCIPFFT FFGLFDVFIYPNAERLHPSLEAVQAILPGGAASGGMAVLAKIATHWTSALFYVMAEIYSSVSVGLLFWQFANDVVNVDQAKRFYPLFA QMSGLAPVLAGQYVVVRFASKAVNFEASMHRLTAAVTFAGIMICIFYQLSSSYVERTESAKPAADNEQSIKPKKKKPKMSMVESGKFLA	

transporter (protein sequence)	SSQYLRLIAMLVLGYGLSINFTEIMWKSLVKKQYPDPLDYQRFMGNFSSAVGLSTCIVIFFGVHVIRLLGWKVGALATPGIMAILALP FFACILLGLDSPARLEIAVIFGTIQSLLSKTSKYALFDPTTQMAYIPLDDESKVKGKAAIDVLGSRIGKSGGSLIQQGLVFVFGNIIN AAPVVGVVYYSVLVAWMSAAGRLSGLFQAQTEMDKADKMEAAKTNKEK	
DNA sequence encoding chimeric pyrrolysyl-tRNA synthetase (*Methanosarcina barkeri is bold, Methanosarcina mazei is underlined) with IPYEVF mutations with IPYEVF + TA mutations (V31I, T56P, H62Y, A100E, C313V, and Y349F + P5T*	**ATGGATAAAAAAAACCCTGGACGTTCTGATCTCCGCTACGGGTCTGTGGATGAGCCGCACGGGTACGCTGCATAAAATTAAACACCACG** **AAATTTCACGTTCGAAAATCTATATCGAAATGGCGTGCGGTGATCATCTGGTGGTTAACAATAGCCGTTCTTGTCGCCCCGCGCGTGC** **CTTTCGCTATCACAAATACCGCAAAACGTTGTCGCGTGTCAGATGAAGACATTAACAATTTCCTGACCCGTAGTACGGAA** **TCCAAAAACTCAGTGAAAGTTCGCGTCGTGAGTGAGCCGAAAGTTAAAAAAGCGATGCCGAAAAGTGTCTCCCGTGCCCCGAAACCGC** **TGGAAAACTCAGTGTCGGCAAAAGCTTCCACCAATACGAGCCGCTCTGTTCCGTCGCCGGCAAAAAGCACCCCGAACAGCTCTGTCCC** **GGCAAGC**GCATCTGCACCGGCACTTACAAAGTCCCAAACCGACCGTTTAGAAGTCTTGCTTAATCCGAAGGATGAAATTTCGTTAAAC TCTGGCAAGCCGTTTCGTGAGCTGGAGTCAGAACTTCTTTCACGGCGTAAAAAAAGATCTTCAGCAGATATATGCTGAGGAGCGCGAGA ATTACTTAGGAAAGCTGGAACGTGAAATAACTCGTTTTTTTGTTGACAGAGGGTTCTTGGAGATAAAGTCCCCAATACTGATACCGTT AGAATACATCGAAAGAATGGGTATAGATAATGATACCGAGTTAAGTAAACAAATATTTAGAGTTGATAAGAATTTTGTTTGCGGCCG ATGTTGGCGCCCAATCTTTACAACTATCTTCGGAAGCTGGACCGCGCCCTTCCTGACCCGATCAAAATCTTTGAAATTGGTCCCTGTT ATCGGAAGGAATCGGATGGCAAGGCCCATTTGGAGGAGTTCACCATGCTGAACTTTGTGCAAATGGGGTCTGGCTGCACGCGGGAAAA TTTGGAAAGCATAATTACGGATTTCTGAACCACCTTGGCATAGACTTTAAGATCGTTGGAGATTCTTGCATGGTTTTTGGGGATACG CTGGATGTAATGCACGGAGATTTAGAGCTTTCAAGTGCAGTAGTAGGACCTATTCCGTTAGACAGAGAATGGGGAATTGACAAGCCAT GGATCGGTGCGGGTTTCGGCTTGGAGCGCCTGCTTAAGGTAAAGCACGACTTTAAAAACATAAAGCGGGCGGCAAGATCTGAGAGTTA TTACAACGGAATATCAACAAATTTGTAA	
LacO operator sequence in Improved 5′ UTR for chromosomally integrated pyrrolysyl tRNA synthetase	AATTGTGAGCGGATAACAATT	
Improved 5′ UTR for chromosomally integrated pyrrolysyl tRNA synthetase without the LacO operator sequence	CCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACAT	

DETAILED DESCRIPTION

I. Definitions

[67] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which the claimed subject matter belongs. It is to be understood that the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of any subject matter claimed. To the extent any material incorporated herein by reference is inconsistent with the express content of this disclosure, the express content controls. In this application, the use of the singular includes the plural unless specifically stated otherwise. It must be noted that, as used in the specification and the appended claims, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise. In this application, the use of "or" means "and/or" unless stated otherwise. Furthermore, use of the term "including" as well as other forms, such as "include", "includes," and "included," is not limiting.

[68] Although various features of the invention may be described in the context of a single embodiment, the features may also be provided separately or in any suitable combination. Conversely, although the invention may be described herein in the context of separate embodiments for clarity, the invention may also be implemented in a single embodiment.

[69] Reference in the specification to "some embodiments", "an embodiment", "one embodiment" or "other embodiments" means that a particular feature, structure, or characteristic described in connection with the embodiments is included in at least some embodiments, but not necessarily all embodiments, of the inventions.

[70] As used herein, ranges and amounts can be expressed as "about" a particular value or range. About also includes the exact amount. Hence "about 5 μL" means "about 5 μL" and also "5 μL." Generally, the term "about" includes an amount that would be expected to be within experimental error, such as for example, within 15%, 10%, or 5%.

[71] The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described.

[72] The term "antibody" herein is used in the broadest sense and encompasses various antibody structures, including but not limited to monoclonal antibodies, polyclonal antibodies, multispecific antibodies (e.g., bispecific antibodies), and antibody fragments so

long as they exhibit the desired antigen-binding activity. An "antibody fragment" refers to a molecule other than an intact antibody that comprises a portion of an intact antibody that binds the antigen to which the intact antibody binds. Examples of antibody fragments include but are not limited to Fv, Fab, Fab', Fab'-SH, F(ab')2; diabodies; linear antibodies; single-chain antibody molecules (e.g., scFv); camelid heavy-chain variable domains (e.g., VHH), shark variable new antigen receptor (VNAR); and multispecific antibodies formed from antibody fragments.

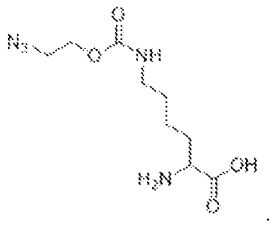
[73] As used herein, the terms "first DNA sequence," "second DNA sequence," "third DNA sequence," and so forth with respect to DNA sequences does not necessarily mean that the DNA sequences are on separate molecules. The use of first, second, third, etc., also does not imply that the DNA sequences are arranged in any particular way. In some embodiments, a first DNA sequence may be on the same chromosome or the same plasmid as a second and/or third DNA sequence; a second DNA sequence may be on the same chromosome or the same plasmid as a first and/or third DNA sequence; or a third DNA sequence may be on the same chromosome or the same plasmid as a first and/or second DNA sequence.

[74] The disclosure provides nucleic acid sequences and amino acid sequences having a certain degree of identity to a given nucleic acid sequence or amino acid sequence, respectively (a reference sequence). "Sequence identity" between first and second nucleic acid sequences indicates the percentage of nucleotides that are identical between the sequences; for example, if a first nucleic acid sequence is 95% identical to a second nucleic acid sequence, then the first nucleic acid sequence contains matches to 95% of the nucleotides in the second nucleic acid sequence. "Sequence identity" between two amino acid sequences indicates the percentage of amino acids that are identical between the sequences; for example, if a first amino acid sequence is 95% identical to a second amino acid sequence, then the first amino acid sequence contains matches to 95% of the nucleotides in the second amino acid sequence. The terms "% identical", "% identity" or similar terms are intended to refer, in particular, to the percentage of nucleotides or amino acids which are determined to be identical using an optimal alignment between the sequences to be compared. Said percentage is purely statistical, and the differences between the two sequences may be but are not necessarily randomly distributed over the entire length of the sequences to be compared. Comparisons of two sequences are usually carried out by comparing said sequences after optimal alignment. The optimal alignment for a comparison may be carried out manually or with the aid of an appropriate algorithm such as the alignment algorithm by Needleman and Wunsch, 1970, J. Mol. Biol. 48, 443, or with the aid of computer programs using said algorithms (e.g., GAP,

BESTFIT, and FASTA in Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Drive, Madison, Wis.).

[75] As used herein, the term "unnatural amino acid," "non-natural amino acid," "non-canonical amino acid," or "ncAA" refers to an amino acid other than one of the 20 naturally occurring amino acids. Exemplary unnatural amino acids are described in Young et al., "Beyond the canonical 20 amino acids: expanding the genetic lexicon," *J. of Biological Chemistry* **285**(15): 11039-11044 (2010), the disclosure of which is herein incorporated by reference. The term "unnatural protein" or "unnatural polypeptide" refers to a polypeptide comprising at least one unnatural amino acid.

[76] As used herein, the term "azidolysine" or "AzK" refers to any amino acid in which the side chain of lysine has been derivatized with a moiety comprising an azido, such as an azidoethoxylysine, e.g., N^6-((2-azidoethoxy)carbonyl)lysine. In some embodiments, AzK refers to N^6-((2-azidoethoxy)carbonyl)lysine. AzK has the following chemical structure:



[77] As used herein, the term "fidelity" refers to a measure of non-canonical amino acid incorporation at an intended site (versus natural substitutions) in a polypeptide comprising an unnatural amino acid. The intended site may be described as a certain numbered position within a sequence. It is often expressed as a percentage, with 100% fidelity meaning that the non-canonical amino acid was always incorporated at the intended site. In some embodiments, fidelity is measured using LC-MS.

[78] As used herein, the term "titer" refers to amount of polypeptide product, for example, an unnatural polypeptide product. In some embodiments, it refers to an amount of polypeptide product produced in a certain volume of liquid, i.e., a concentration. In some embodiments, titer is measured in mg/L. In some embodiments, titer is measured using optical density (OD) of the cells. In some embodiments, titer is measured using HPLC.

[79] As used herein, a "promoter" refers to a region of DNA that can facilitate the transcription of a particular gene, by providing a start site for the synthesis of RNA corresponding to a gene.

[80] As used herein, the term "operably linked" as used herein with respect to promoters, refers to a relationship between a coding sequence and a promoter element. The

promoter is operably linked with the coding sequence when expression from the coding sequence via transcription is regulated, or controlled by, the promoter element.

[81] As used herein, "5' untranslated region" or "5' UTR" refers to a region of an mRNA that is between the transcription initiation point and the start codon. A 5' UTR can originate from any suitable nucleic acid, such as genomic DNA, plasmid DNA, RNA or mRNA, for example, from any suitable organism (e.g., virus, bacterium, yeast, fungi, plant, insect or mammal). The artisan may select appropriate elements for the 5' UTR based upon the chosen expression system (e.g., expression in a chosen organism, or expression in a cell-free system, for example). A 5' UTR sometimes comprises one or more of the following elements known to the artisan: enhancer sequences (e.g., transcriptional or translational), transcription initiation site, transcription factor binding site, translation regulation site, translation initiation site, translation factor binding site, accessory protein binding site, feedback regulation agent binding sites, Pribnow box, TATA box, -35 element, E-box (helix-loop-helix binding element), ribosome binding site, replicon, internal ribosome entry site (IRES), silencer element and the like. A 5' UTR can comprise a translational enhancer nucleotide sequence. A translational enhancer sequence often binds to a ribosome, sometimes is an 18S rRNA-binding ribonucleotide sequence (i.e., a 40S ribosome binding sequence) and sometimes is an internal ribosome entry sequence (IRES). An IRES generally forms an RNA scaffold with RNA tertiary structures that contact a 40S ribosomal subunit via a number of specific intermolecular interactions. Examples of ribosomal enhancer sequences are known and can be identified by the artisan (e.g., Mignone et al., Nucleic Acids Research 33: D141-D146 (2005); Paulous et al., Nucleic Acids Research 31: 722-733 (2003); Akbergenov et al., Nucleic Acids Research 32: 239-247 (2004); Mignone et al., Genome Biology 3(3): reviews0004.1-0001.10 (2002); Gallie, Nucleic Acids Research 30: 3401-3411 (2002); Shaloiko et al., DOI: 10.1002/bit.20267; and Gallie et al., Nucleic Acids Research 15: 3257-3273 (1987)). A translational enhancer sequence sometimes is a prokaryotic sequence, such as a Shine-Dalgarno consensus sequence. In certain embodiments, the translational enhancer sequence is a viral nucleotide sequence.

[82] As used herein, "mRNA" or "messenger RNA" refers to a type of single-stranded RNA involved in polypeptide synthesis that carries protein information from DNA to a ribosome that reads the mRNA sequence and translates each codon into its corresponding amino acid in a growing polypeptide chain.

[83] As used herein, an "unnatural mRNA molecule" or "unnatural messenger RNA molecule" is an mRNA molecule that contains a least one unnatural nucleotide.

[84] As used herein, an "unnatural nucleotide" is a nucleotide which contains a modification to either the sugar or phosphate moieties or an unnatural base.

[85] As used herein, an "unnatural base" is a base other than adenine, guanine, cytosine, thymine, or uracil that does not ordinarily occur in a given organism (e.g., 5-methylcytosine is considered natural). In some embodiments, an unnatural base contains an unnatural ring or unnatural ring system, e.g., one or more rings containing an arrangement of 5 or 6 ring atoms and bonds that does not occur in A, T, C, G, or U.

[86] As used herein, a "tRNA" or "transfer RNA" refers to a molecule with an anticodon that can match with a codon in an mRNA molecule during translation. The tRNA can also be charged with an aminoacyl moiety, which is encoded by the codon to which the tRNA binds.

[87] As used herein, an "unnatural tRNA" or "unnatural transfer RNA" or "unnatural transfer RNA (tRNA) molecule" or "unnatural tRNA molecule" refers to a tRNA with an anticodon that contains an unnatural nucleotide and can match with a codon in an mRNA molecule during translation. The tRNA can also be charged with an unnatural amino acid, which is encoded by the codon to which the unnatural tRNA binds, by a tRNA synthetase.

[88] As used herein, "tRNA synthetase" or "aminoacyl tRNA synthetase" refers to an enzyme that pairs (or, "charges") a tRNA with an amino acid.

[89] As used herein, a "pyrrolysyl tRNA synthetase" or "PylRS" is an aminoacyl-tRNA synthetase that displays high substrate side chain promiscuity, low selectivity toward its substrate α-amine, and low selectivity toward the anticodon of $tRNA^{Pyl}$. A PylRS can be used to charge a $tRNA^{Pyl}$, such as an unnatural $tRNA^{Pyl}$, with an unnatural amino acid. In some embodiments, the PylRS is from *Methanosarcina barkeri* or *Methanococcus jannaschii*.

[90] As used herein, *Pt*NTT2 is a transporter from *Phaeodactylum tricornutum* capable of transporting various natural and unnatural nucleosides and nucleotides into cells across the cell membrane. In some embodiments, *Pt*NTT2 is a truncated polypeptide. In some embodiments, it has at least 80% identity to any one or more of SEQ ID NOs 1, 4, 6, and 8 of WO 2017/223528 A1. Examples of *Pt*NTT2s can be found in WO 2015/157555 A2 and WO 2017/223528 A1.

[91] As used herein, "polymerase" refers to an enzyme that catalyzes the polymerization of a nucleic acid strand using an existing nucleic acid as a template. Examples of useful polymerases include DNA polymerases and RNA polymerases.

[92] As used herein, "T7 bacteriophage RNA polymerase" or "T7 polymerase" refers to the DNA-dependent RNA polymerase from T7 bacteriophages that consists of one

subunit that is capable of catalyzing RNA synthesis/transcription. *See* Kochetkov, S.N. et al., FEBS Letters vol. 440(3), pages 264-267 (1998). See also the T7 RNA sequence in the UniProt entry, P00573 RPOL_BPT7.

[93] As used herein, "Polymerase II" or "Pol II" refers to DNA polymerase II, an enzyme encoded by the *polB* gene, also known as the *dinA* gene, as well as modified versions thereof. *See* Campbell, J.L., DNA Polymerase II, Bacterial, Encyclopedia of Biological Chemistry, pages 726-728 (2004).

[94] As used herein, "antibiotic resistance gene" refers to a nucleic acid sequence encoding a gene that allows the cell in which it is present to survive in the presence of, and/or under selective pressure of, an antibiotic. In some embodiments, the antibiotic resistance gene is a chloramphenicol resistance gene that allows the cell to survive in the presence of chloramphenicol. In some embodiments, the antibiotic resistance gene is a tetracycline resistance gene that allows the cell to survive in the presence of tetracycline. In some embodiments, the antibiotic resistance gene is a zeocin resistance gene that allows the cell to survive in the presence of zeocin.

[95] As used herein, "inducible promoter" refers to a promoter that shows increased activity in response to specific transcription factors, small molecules, stimuli, or other elements.

[96] As used herein, a "repressor of an inducible promoter" is a molecule that blocks or otherwise inhibits an inducible promoter from promoting transcription. In some embodiments, the repressor does so by blocking RNA polymerase from binding to the nucleic acid.

[97] As used herein, "LacI" refers to a protein that, in the absence of lactose, acts as a repressor of transcription from the *lac* promoter. It does so by binding tightly to the lac operator and interfering with transcription by RNA polymerase. Transcription can be induced upon binding of lactose (or an analog thereof, such as isopropyl β-D-1-thiogalactopyranoside (IPTG)), causing LacI to dissociate from the operator and allowing RNA polymerase to proceed with transcription.

[98] As used herein, "P_{lacI}" refers to a wild-type promoter for LacI. An exemplary P_{lacI} sequence may be found in Figure 4 of Glascock, C.B. & Weickert, M.J., Gene vol. 223, pages 221-231 (1998).

[99] As used herein, "P_{lacIQ}" refers to a mutant promoter for LacI that is about 10 times stronger than the wild-type promoter for LacI, resulting in higher levels of transcription

of LacI. An exemplary P_{lacIQ} sequence may be found in Figure 4 of Glascock, C.B. & Weickert, M.J., Gene vol. 223, pages 221-231 (1998).

[100] As used herein, "P_{lacIQ1}" refers to a mutant promoter for LacI that is about 100 times stronger than the wild-type promoter for LacI, resulting in higher levels of transcription of LacI. An exemplary P_{lacIQ1} sequence may be found in Figure 4 of Glascock, C.B. & Weickert, M.J., Gene vol. 223, pages 221-231 (1998).

[101] As used herein, a "P_{lacUV5}" is a promoter that is a stronger variant of the wild-type *lac* promoter for the *lac* operon. It is controlled by the lac repressor, LacI. An exemplary P_{lacUV5} promoter may be found in Figure 1(A)ii. of Rincon, A. & Farny, N.G., Microbial Biotechnology vol. 16, pages 961-976 (2023).

[102] In some embodiments, "IPTG" or "isopropyl β-D-1-thiogalactopyranoside" refers to a molecular analog of allolactose used to induce protein expression in a gene under the control of the lac operator.

[103] As used herein, "induction" in the context of culturing cells refers to the step of adding a molecule that induces the cells to express a polypeptide product, for example, the step of adding IPTG to media for cells that have a gene encoding a polypeptide product that is controlled by the lac operon.

[104] As used herein, a "transcriptional terminator" or "transcription terminator" or "terminator" refers to a DNA sequence that signals the polymerase transcribing the sequence to stop.

[105] As used herein, "oppositely oriented promoter" refers to a promoter that, in relation to a given gene, promotes transcription in the antisense direction to the given gene.

[106] As used herein, "retrograde transcription" refers to antisense transcription relative to the sense strand of a given gene.

[107] As used herein, "anterograde transcription" refers to sense transcription relative to the sense strand of a given gene.

[108] As used herein, "RecA" refers to the recombinase A protein. RecA is an enzyme involved in the DNA repair process. It is present in *Escherichia coli* and it is conserved among bacteria. RecA promotes recombinational DNA repair by catalyzing a DNA strand exchange reaction.

[109] As used herein, "RecX" refers to a protein that interacts directly with the RecA protein and negatively modulates the recombinase, ATPase, and coprotease activities of RecA in *E. coli*. In some bacteria these inhibitory activities are performed by blocking RecA filament extension. The *recX* gene that encodes the RecX protein is located downstream from *recA* in

the wild-type *E. coli* genome, forming a single operon. In *E. coli*, the *recA* and *recX* genes are separated by a terminator, meaning that the transcript often, but does not always, end at *recA*. When the transcript reads through, both RecA and RecX are produced.

[110] As used herein "RecA hypomorph" refers to an allele that results in reduced expression and/or activity of the RecA protein as compared to wild-type RecA. In some embodiments, the RecA hypomorph is created by deletion of the terminator between the *recA* and *recX* genes, resulting in a readthrough mRNA transcript, and further resulting in increased production of the RecX protein that negatively modulates RecA.

[111] As used herein, "RecX overexpression" refers to expression of RecX at levels that are greater than or equal to 250-fold lower, such as 100-fold lower, than RecA expression. In wild-type *E. coli*, expression of RecX is typically downregulated at both transcriptional and translational levels resulting in about a 500-fold lower protein level compared to RecA. *See* Alekseev, A. et al., Elife vol. 11: e78409 (2022).

[112] As used herein, "LexA" refers to a polypeptide that functions as a repressor of the SOS response during normal growth of *E. coli*. In response to DNA damage, RecA is activated and induces the cleavage of LexA, resulting in de-repression of genes in the SOS regulon. LexA can also cleave itself in the absence of RecA under certain conditions.

[113] As used herein, "LexA that does not undergo RecA-stimulated cleavage" refers to LexA that is not able to be induced by RecA to cleave the bond to be able to de-repress genes in the SOS regulon, so it is unable to induce SOS response. In some embodiments, a LexA that does not undergo RecA-stimulated cleavage is created by mutating an amino acid residue involved in cleavage, such as Ser-119 or Lys-156. In some embodiments, the LexA that does not undergo RecA-stimulated cleavage is LexA with a S119A mutation. LexA contains two types of sites involved in cleavage. The substrate cleavage site is composed of the bond to be cleaved (between Ala-84 and Gly-85 in wild-type LexA) and likely the neighboring groups that confer specificity on that site. The active site (contained within the C-terminal domain) is composed of a catalytic center that carries out the chemistry of cleavage and a binding pocket that binds the substrate and positions it optimally with respect to the catalytic center. The chemical mechanism is thought to involve a serine nucleophile (Ser-119), activated by a neutral general base lysine (Lys-156). Ser-119 and Lys-156 are completely conserved in wild-type members of the LexA superfamily.

[114] As used herein, "derepressed" refers to a gene that is transcribed because repressor control has been lifted. The state of the gene under standard environmental

conditions is that it is repressed by a repressor for at least some amount of time and is not transcribed.

[115] As used herein, "constitutively overexpressed" refers to expression of a gene that always takes place (i.e., is always "on") regardless of environmental conditions.

[116] As used herein, "isogenic" means that two cells, strains, or the like have the same or substantially the same genotype other than any specifically noted mutations or other differences.

[117] As used herein, "colony size heterogeneity" refers to, among multiple colonies of bacteria, the property of different colonies having varying sizes (as measured by, for example, diameter or radius of the colony). Colony size heterogeneity can indicate genetic instability in a strain.

[118] As used herein, "OD600" or "optical density 600" refers to a measurement of light scattering by a liquid culture measured at a wavelength of 600 nm. In some embodiments, a spectrophotometer is used to measure OD600, and/or OD600 is measured using a 1 cm path length.

[119] As used herein, "nucleotide" refers to a compound comprising a nucleoside moiety and a phosphate moiety. Exemplary natural nucleotides include, without limitation, adenosine triphosphate (ATP), uridine triphosphate (UTP), cytidine triphosphate (CTP), guanosine triphosphate (GTP), adenosine diphosphate (ADP), uridine diphosphate (UDP), cytidine diphosphate (CDP), guanosine diphosphate (GDP), adenosine monophosphate (AMP), uridine monophosphate (UMP), cytidine monophosphate (CMP), and guanosine monophosphate (GMP), deoxyadenosine triphosphate (dATP), deoxythymidine triphosphate (dTTP), deoxycytidine triphosphate (dCTP), deoxyguanosine triphosphate (dGTP), deoxyadenosine diphosphate (dADP), thymidine diphosphate (dTDP), deoxycytidine diphosphate (dCDP), deoxyguanosine diphosphate (dGDP), deoxyadenosine monophosphate (dAMP), deoxythymidine monophosphate (dTMP), deoxycytidine monophosphate (dCMP), and deoxyguanosine monophosphate (dGMP). Exemplary natural deoxyribonucleotides, which comprise a deoxyribose as the sugar moiety, include dATP, dTTP, dCTP, dGTP, dADP, dTDP, dCDP, dGDP, dAMP, dTMP, dCMP, and dGMP. Exemplary natural ribonucleotides, which comprise a ribose as the sugar moiety, include ATP, UTP, CTP, GTP, ADP, UDP, CDP, GDP, AMP, UMP, CMP, and GMP.

II. Overview

[120] Provided herein are cells, tRNA synthetases, and methods useful for expressing polypeptides comprising an unnatural amino acid. In some embodiments, cells are provided

that are capable of expressing a polypeptide comprising an unnatural amino acid with high yield and/or fidelity. In some embodiments, tRNA synthetases are provided that have high solubility and/or catalytic activity for charging a tRNA, such as an unnatural tRNA, with an unnatural amino acid. In some embodiments, methods are provided that are capable of producing a polypeptide comprising an unnatural amino acid with high yield and/or fidelity. In some embodiments, cells are provided with improved genetic stability (which can be observed, e.g., though improved colony size homogeneity).

III. Unnatural tRNA/Unnatural tRNA Synthetase Pairs

[121] In some embodiments, the tRNA synthetase comprises a chimeric pyrrolysyl-tRNA synthetase comprising a *Methanosarcina barkeri* pyrroylysyl-tRNA synthetase N-terminal domain (NTD) fused to a *Methanosarcina mazei* pyrroylysl-tRNA synthetase C-terminal domain (CTD). In some embodiments, the tRNA synthetase comprises a split chimeric pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the first polypeptide chain comprises a *Methanosarcina barkeri* pyrroylysl-tRNA synthetase N-terminal domain (NTD), and wherein the second polypeptide chain comprises a *Methanosarcina mazei* pyrroylysl-tRNA synthetase C-terminal domain (CTD). In some embodiments, the tRNA synthetase comprises two separate polypeptide chains, wherein the *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) comprises comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 1-137 of SEQ ID NO: 4, and wherein the *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD) comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 150-419 of SEQ ID NO: 4, optionally wherein the C-terminal amino acid of *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase NTD corresponds to P137 or S149 of SEQ ID NO: 4.

[122] In some embodiments, the tRNA synthetase comprises a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase. In some embodiments, the tRNA synthetase comprises a split *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprises an N-terminal domain (NTD), and comprises a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase C-terminal domain (CTD). In some embodiments, the tRNA synthetase comprises two separate polypeptide chains, wherein the first polypeptide chain comprising a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase NTD comprises an amino acid sequence

having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 1-110, 1-137, or 1-149 of SEQ ID NO: 3, and wherein the second polypeptide chain comprising a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase CTD comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 150-419 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase NTD corresponds to S110, P137, or S149 of SEQ ID NO: 3.

[123] In some embodiments, the tRNA synthetase comprises a first polypeptide chain comprising the N-terminal domain (NTD) and a second polypeptide chain comprising the C-terminal domain (CTD), e.g., as separate chains.

[124] In some embodiments, a cell comprises a DNA sequence encoding a split *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the first polypeptide chain comprises part of a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 1-110, 1-137, or 1-149 of SEQ ID NO: 3, and wherein the second polypeptide chain comprises part of a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 150-419 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to S110, P137, or S149 of SEQ ID NO: 3. In some embodiments, the split *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 1-110 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to S110. In some embodiments, the split *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 1-137 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to P137. In some embodiments, the split *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 1-149 of SEQ ID

NO: 3, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to S149.

[125] In some embodiments, a cell comprises a DNA sequence encoding a split chimeric pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the first polypeptide chain comprises part of a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 1-137 of SEQ ID NO: 4, and wherein the second polypeptide chain comprises part of a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80%,81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 150-419 of SEQ ID NO: 4, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to P137 of SEQ ID NO: 4.

[126] In some embodiments, a cell comprises a DNA sequence encoding a split pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the first polypeptide chain comprises part of a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 1-138 or 1-172 of SEQ ID NO: 5, and wherein the second polypeptide chain comprises part of a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 185-454 of SEQ ID NO: 5, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to I138 or N172 and of SEQ ID NO: 5.

[127] In some embodiments, the tRNA synthetase comprises at least the following mutations with reference to SEQ ID NO: 3, 4, or 5: V31I, T56P, H62Y, and A100E. In some embodiments, the tRNA synthetase comprises at least the following mutations with reference to SEQ ID NO: 3 or 4: V31I, T56P, H62Y, A100E, C313V, and Y349F. In some embodiments, the tRNA synthetase comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 6. In some embodiments, the tRNA synthetase comprises at least one, two, three, four, five, six, seven, or eight of the following mutations with reference to SEQ ID NO: 3, 4, or 5: P5T or P5L, V31I, T56P, H62Y, A100E, E302A, C313V, and Y349F. In some embodiments, the tRNA synthetase comprises at least the

following mutations with reference to SEQ ID NO: 3 or 4: P5T or P5L, V31I, T56P, H62Y, A100E, E302A, C313V, and Y349F. In some embodiments, the tRNA synthetase comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 7.

[128] In some embodiments, the pyrrolysyl-tRNA synthetase is fused to *Escherichia coli* thioredoxin.

[129] In some embodiments, provided herein is a chimeric pyrrolysyl-tRNA synthetase comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F.

[130] In some embodiments, provided herein is a chimeric pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 6, wherein the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 31 of SEQ ID NO: 6 is isoleucine; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 56 of SEQ ID NO: 6 is proline; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 62 of SEQ ID NO: 6 is tyrosine; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 100 of SEQ ID NO: 6 is glutamate; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 313 of SEQ ID NO: 6 is valine; and the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 349 of SEQ ID NO: 6 is phenylalanine.

[131] In some embodiments, provided herein is a chimeric pyrrolysyl-tRNA synthetase comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: P5T or P5L, E302A, V31I, T56P, H62Y, A100E, C313V, and Y349F.

[132] In some embodiments, provided herein is a chimeric pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 7, wherein the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 5 of SEQ ID NO: 6 is threonine or leucine; the position of

the chimeric pyrrolysyl-tRNA synthetase corresponding to position 31 of SEQ ID NO: 6 is isoleucine; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 56 of SEQ ID NO: 6 is proline; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 62 of SEQ ID NO: 6 is tyrosine; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 100 of SEQ ID NO: 6 is glutamate; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 302 of SEQ ID NO: 6 is alanine; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 313 of SEQ ID NO: 6 is valine; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 349 of SEQ ID NO: 6 is phenylalanine.

[133] In some embodiments, provided herein is a split chimeric pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the first polypeptide chain comprises part of a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 1-137 of SEQ ID NO: 4, and wherein the second polypeptide chain comprises part of a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 150-419 of SEQ ID NO: 4, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to P137 of SEQ ID NO: 4.

[134] In some embodiments, provided herein is a pyrrolysyl-tRNA synthetase comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising the following mutations with reference to SEQ ID NO: 3: V31I, T56P, H62Y, A100E, C313V, and Y349F.

[135] In some embodiments, provided herein is a pyrrolysyl-tRNA synthetase comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 3: P5T or P5L, V31I, T56P, H62Y, E302A, A100E, C313V, and Y349F.

[136] In some embodiments, provided herein is a split pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the first polypeptide chain comprises part of a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 1-110, 1-137, or 1-149 of SEQ ID NO: 3, and wherein the second polypeptide chain comprises part of a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising an amino acid sequence

having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 150-419 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to S110, P137, or S149 of SEQ ID NO: 3. In some embodiments, the split pyrrolysyl-tRNA synthetase comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 1-110 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to S110 of SEQ ID NO: 3. In some embodiments, the split pyrrolysyl-tRNA synthetase comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 1-137 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to P137 of SEQ ID NO: 3. In some embodiments, the split pyrrolysyl-tRNA synthetase comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 1-149 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to S149 of SEQ ID NO: 3.

[137] In some embodiments, the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 3 or 4: V31I, T56P, H62Y, A100E, C313V, and Y349F.

[138] In some embodiments, the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 3 or 4: P5T or P5L, V31I, T56P, H62Y, A100E, E302A, C313V, and Y349F.

[139] In some embodiments, provided herein is a split pyrrolysyl tRNA synthetase comprising two separate polypeptide chains, wherein the first polypeptide chain comprises part of a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 1-138 or 1-172 of SEQ ID NO: 5, and wherein the second polypeptide chain comprises part of a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to positions 185-454 of SEQ ID NO: 5, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to I138 or N172 of SEQ ID NO: 5.

[140] In some embodiments, the pyrrolysyl-tRNA synthetase is fused to *Escherichia coli* thioredoxin.

[141] In some instances, incorporation of the unnatural amino acid into a protein is mediated by an orthogonal, modified synthetase/tRNA pair. Such orthogonal pairs comprise a natural or mutated synthetase that is capable of charging the unnatural tRNA with a specific unnatural amino acid, often while minimizing charging of a) other endogenous amino acids or alternate unnnatural amino acids onto the unnatural tRNA and b) any other (including endogenous) tRNAs. Such orthogonal pairs comprise tRNAs that are capable of being charged by the synthetase, while avoiding being charged with other endogenous amino acids by endogenous synthetases. In some embodiments, such pairs are identified from various organisms, such as bacteria, yeast, Archaea, or human sources. In some embodiments, an orthogonal synthetase/tRNA pair comprises components from a single organism. In general, nucleotide or amino acid sequence is considered to be "from" an organism if the sequence matches a sequence found in that organism or is at least 97%, 98%, or 99% identical thereto. In some embodiments, a sequence from an organism is identical to a sequence found in that organism. In some embodiments, an orthogonal synthetase/tRNA pair comprises components from two different organisms. In some embodiments, an orthogonal synthetase/tRNA pair comprising components that prior to modification, promote translation of different amino acids. In some embodiments, an orthogonal synthetase is a modified alanine synthetase. In some embodiments, an orthogonal synthetase is a modified arginine synthetase. In some embodiments, an orthogonal synthetase is a modified asparagine synthetase. In some embodiments, an orthogonal synthetase is a modified aspartic acid synthetase. In some embodiments, an orthogonal synthetase is a modified cysteine synthetase. In some embodiments, an orthogonal synthetase is a modified glutamine synthetase. In some embodiments, an orthogonal synthetase is a modified glutamic acid synthetase. In some embodiments, an orthogonal synthetase is a modified alanine glycine. In some embodiments, an orthogonal synthetase is a modified histidine synthetase. In some embodiments, an orthogonal synthetase is a modified leucine synthetase. In some embodiments, an orthogonal synthetase is a modified isoleucine synthetase. In some embodiments, an orthogonal synthetase is a modified lysine synthetase. In some embodiments, an orthogonal synthetase is a modified methionine synthetase. In some embodiments, an orthogonal synthetase is a modified phenylalanine synthetase. In some embodiments, an orthogonal synthetase is a modified proline synthetase. In some embodiments, an orthogonal synthetase is a modified serine synthetase. In some embodiments, an orthogonal synthetase is a modified threonine synthetase.

In some embodiments, an orthogonal synthetase is a modified tryptophan synthetase. In some embodiments, an orthogonal synthetase is a modified tyrosine synthetase. In some embodiments, an orthogonal synthetase is a modified valine synthetase. In some embodiments, an orthogonal synthetase is a modified phosphoserine synthetase. In some embodiments, an orthogonal tRNA is a modified alanine tRNA. In some embodiments, an orthogonal tRNA is a modified arginine tRNA. In some embodiments, an orthogonal tRNA is a modified asparagine tRNA. In some embodiments, an orthogonal tRNA is a modified aspartic acid tRNA. In some embodiments, an orthogonal tRNA is a modified cysteine tRNA. In some embodiments, an orthogonal tRNA is a modified glutamine tRNA. In some embodiments, an orthogonal tRNA is a modified glutamic acid tRNA. In some embodiments, an orthogonal tRNA is a modified alanine glycine. In some embodiments, an orthogonal tRNA is a modified histidine tRNA. In some embodiments, an orthogonal tRNA is a modified leucine tRNA. In some embodiments, an orthogonal tRNA is a modified isoleucine tRNA. In some embodiments, an orthogonal tRNA is a modified lysine tRNA. In some embodiments, an orthogonal tRNA is a modified methionine tRNA. In some embodiments, an orthogonal tRNA is a modified phenylalanine tRNA. In some embodiments, an orthogonal tRNA is a modified proline tRNA. In some embodiments, an orthogonal tRNA is a modified serine tRNA. In some embodiments, an orthogonal tRNA is a modified threonine tRNA. In some embodiments, an orthogonal tRNA is a modified tryptophan tRNA. In some embodiments, an orthogonal tRNA is a modified tyrosine tRNA. In some embodiments, an orthogonal tRNA is a modified valine tRNA. In some embodiments, an orthogonal tRNA is a modified phosphoserine tRNA. In some embodiments, the cell comprises a second tRNA and a second tRNA synthetase, wherein the second tRNA synthetase can aminoacylate the second tRNA and forms a (second) orthogonal tRNA synthetase/tRNA pair. In some embodiments, the second tRNA synthetase is a *Methanocaldococcus jannaschii* tRNA synthetase. In some embodiments, the second tRNA from *Methanocaldococcus jannaschii*. In some embodiments, the second unnatural amino acid is para-Acetyl-L-phenylalanine (*p*AcPhe). In some embodiments, the first unnatural amino acid is AzK and the second unnatural amino acid is *p*AcPhe. In some embodiments, the second unnatural codon read from a 5’ to 3’ direction comprises AXC, wherein X is an unnatural nucleotide. In some embodiments, the first unnatural codon read from a 5’ to 3’ direction comprises GXC. *Methanocaldococcus jannaschii* and *Methanococcus jannaschii* refer to the same organism.

[142] In some embodiments, the unnatural amino acid can be incorporated into an unnatural polypeptide or an unnatural protein by an aminoacyl (aaRS or RS)-tRNA synthetase-

tRNA pair. Exemplary aaRS-tRNA pairs include, but are not limited to, *Methanococcus jannaschii* (*Mj-Tyr*) aaRS/tRNA pairs, *Methanococcus jannaschii* (*M. jannaschii*) TyrRS variant *p*AzFRS (*Mjp*AzFRS), *E. coli* TyrRS (*Ec-Tyr*)/*B. stearothermophilus* $tRNA_{CUA}$ pairs, *E. coli* LeuRS (*Ec-Leu*)/*B. stearothermophilus* $tRNA_{CUA}$ pairs, and pyrrolysyl-tRNA pairs, such as *Methanosarcina barkeri* (*M. barkeri*) pyrrolysyl-tRNA synthetase/*Methanosarcina mazei* (*M. mazei*) $tRNA^{Pyl}$ pairs, chimeric pyrrolysyl-tRNA synthetase (in some embodiments comprising a pyrrolysyl-tRNA synthetase N-terminal domain from a first species and a pyrrolysyl-tRNA synthetase C-terminal domain from a second species, in some embodiments the first species is *M. barkeri*, and the second species is *M. mazei*, and in further embodiments an *M. barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) is fused to an *M. mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD)) /*M. mazei* $tRNA^{Pyl}$ pairs, *M. barkeri* pyrrolysyl-tRNA synthetase/ *M. barkeri* $tRNA^{Pyl}$ pairs, chimeric pyrrolysyl-tRNA synthetase/*M. barkeri* $tRNA^{Pyl}$ pairs, *Methanosarcina barkeri* (*M. barkeri*) pyrrolysyl-tRNA synthetase/chimeric $tRNA^{Pyl}$ pairs, chimeric pyrrolysyl-tRNA synthetase (in some embodiments comprising a pyrrolysyl-tRNA synthetase N-terminal domain from a first species and a pyrrolysyl-tRNA synthetase C-terminal domain from a second species, in some embodiments the first species is *M. barkeri*, and the second species is *M. mazei*, and in further embodiments an *M. barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) is fused to an *M. mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD)) / chimeric $tRNA^{Pyl}$ pairs, *M. barkeri* pyrrolysyl-tRNA synthetase/ chimeric $tRNA^{Pyl}$ pairs, and chimeric pyrrolysyl-tRNA synthetase/ chimeric $tRNA^{Pyl}$ pairs. In any of the foregoing embodiments comprising a chimeric $tRNA^{Pyl}$, the sequence of the chimeric $tRNA^{Pyl}$ may comprise or consist of a portion of sequence from *M. barkeri* and a portion of sequence from *M. mazei*. In some embodiments, the chimeric pyrrolysyl-tRNA synthetase in the above-mentioned pairs comprises an amino acid sequence having at least 80% identity to SEQ ID NO: 6 or SEQ ID NO: 7. In some embodiments, the chimeric pyrrolysyl-tRNA synthetase in the above-mentioned pairs comprises a split chimeric pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, in which in some embodiments the *M. barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) comprises an amino acid sequence having at least 80% identity to positions 1-137 of SEQ ID NO: 4, and in some embodiments the *M. mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD) comprises an amino acid sequence having at least 80% identity to positions 150-419 of SEQ ID NO: 4. In some embodiments, the chimeric pyrrolysyl-tRNA synthetase in the above-mentioned pairs comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, and A100E. In some embodiments, the chimeric

pyrrolysyl-tRNA synthetase in the above-mentioned pairs comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F. In some embodiments, the chimeric pyrrolysyl-tRNA synthetase in the above-mentioned pairs comprises the following mutations with reference to SEQ ID NO: 4: P5T or P5L, E302A, V31I, T56P, H62Y, A100E, C313V, and Y349F. In some embodiments the chimeric pyrrolysyl tRNA synthetase is a chimeric pyrrolysyl tRNA synthetase according to any embodiment thereof described elsewhere herein. In some embodiments, in the *M. barkeri pyrrolysyl-tRNA synthetase/ M. mazei* $tRNA^{Pyl}$ pair mentioned above, the tRNA synthetase comprises a split *M. barkeri* pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein in some embodiments the first polypeptide chain comprises a *M. barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD), and in some embodiments the second polypeptide chain comprises a *M. barkeri* pyrrolysyl-tRNA synthetase C-terminal domain (CTD). In some embodiments, the *M. barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) comprises an amino acid sequence having at least 80% identity to positions 1-110, 1-137, or 1-149 of SEQ ID NO: 3, and the *M. barkeri* pyrrolysyl-tRNA synthetase C-terminal domain (CTD) comprises an amino acid sequence having at least 80% identity to positions 150-419 of SEQ ID NO: 3. In some embodiments the split *M. barkeri* pyrrolysyl tRNA synthetase is a split *M. barkeri* pyrrolysyl tRNA synthetase according to any embodiment thereof described elsewhere herein. In some instances, the unnatural amino acid is incorporated into an unnatural polypeptide or an unnatural protein by a *Mj-Tyr*RS/tRNA pair. Exemplary unnatural amino acids (UAAs) that can be incorporated by a *Mj-Tyr*RS/tRNA pair include, but are not limited to, para-substituted phenylalanine derivatives such as *p*-Azido-L-Phenylalanine (*p*AzF), N6-(((2-azidobenzyl)oxy)carbonyl)-L-lysine, N6-(((3-azidobenzyl)oxy)carbonyl)-L-lysine, N6-(((4-azidobenzyl)oxy)carbonyl)-L-lysine, *p*-aminophenylalanine and *p*-methoyphenylalanine; meta-substituted tyrosine derivatives such as 3-aminotyrosine, 3-nitrotyrosine, 3,4-dihydroxyphenylalanine, and 3-iodotyrosine; phenylselenocysteine; *p*-boronophenylalanine; and *o*-nitrobenzyltyrosine.

[143] In some embodiments, a tRNA and a tRNA synthetase (e.g., in a cell described herein) form an orthogonal pair. In some embodiments, an orthogonal pair of a tRNA and a tRNA synthetase is a tRNA synthetase/tRNA pair in which (i) the tRNA synthetase of the pair aminoacylates the tRNA of the pair but does not aminoacylate endogenous tRNAs present in the cell and (ii) the tRNA of the pair does not undergo aminoacylation by endogenous tRNA synthetases present in the cell. In some embodiments, a tRNA synthetase of an orthogonal pair is considered not to aminoacylate any endogenous tRNA if the extent of aminoacylation of

any endogenous tRNA by the tRNA synthetase of the orthogonal pair is less than 10%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.2%, or 0.1%, of the extent of aminoacylation using the orthogonal tRNA synthetase/tRNA pair , e.g., under reaction conditions in which the endogenous tRNA is equimolar with the tRNA of the orthogonal pair and optionally in some embodiments under reaction conditions in which about 30-70% of the tRNA of the orthogonal pair is aminoacylated by the tRNA synthetase of the orthogonal pair. In some embodiments, a tRNA of an orthogonal pair is considered not to undergo aminoacylation by any endogenous tRNA synthetase or is considered not a substrate for any endogenous tRNA synthetase if the extent of aminoacylation of the tRNA of the orthogonal pair by any endogenous tRNA synthetase is less than 10%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.2%, or 0.1%, of the extent of aminoacylation using the orthogonal tRNA synthetase/tRNA pair , e.g., under reaction conditions in which the endogenous tRNA is equimolar with the tRNA of the orthogonal pair and optionally in some embodiments under reaction conditions in which about 30-70% of the tRNA of the orthogonal pair is aminoacylated by the tRNA synthetase of the orthogonal pair. Exemplary orthogonal pairs include, e.g., the pairs listed above and those described in the Examples, e.g., such as *M. barkeri* pyrrolysyl-tRNA synthetase/ *M. mazei* $tRNA^{Pyl}$ pairs, chimeric pyrrolysyl-tRNA synthetase/ *M. mazei* $tRNA^{Pyl}$ pairs, *M. barkeri* pyrrolysyl-tRNA synthetase/ *M. barkeri* $tRNA^{Pyl}$ pairs, and chimeric pyrrolysyl-tRNA synthetase/*M. barkeri* $tRNA^{Pyl}$ pairs, and *Methanocaldococcus jannaschii* tyrosyl tRNA synthetase/ *Methanocaldococcus jannaschii* $tRNA^{Tyr}$ pairs.

[144] In some instances, the unnatural amino acid can be incorporated into an unnatural polypeptide or an unnatural protein by an *Ec-Tyr*/$tRNA_{CUA}$ or an *Ec-Leu*/$tRNA_{CUA}$ pair. Exemplary UAAs that can be incorporated by an *Ec-Tyr*/$tRNA_{CUA}$ or an *Ec-Leu*/$tRNA_{CUA}$ pair include, but are not limited to, phenylalanine derivatives containing benzophenone, ketone, iodide, or azide substituents; *O*-propargyltyrosine; α-aminocaprylic acid, O-methyl tyrosine, O-nitrobenzyl cysteine; and 3-(naphthalene-2-ylamino)-2-amino-propanoic acid.

[145] In some instances, the unnatural amino acid can be incorporated into an unnatural polypeptide or an unnatural protein by a pyrrolysyl-tRNA pair. In some cases, the PylRS can be obtained from an archaebacterial species, e.g., from a methanogenic archaebacterium. In some cases, the PylRS can be obtained from *Methanosarcina barkeri*, *Methanosarcina mazei*, or *Methanosarcina acetivorans*. In some cases, the PylRS can be a chimeric PylRS. Exemplary UAAs that can be incorporated by a pyrrolysyl-tRNA pair include, but are not limited to, amide and carbamate substituted lysines such as N6-(2-azidoethoxy)-

carbonyl-L-lysine (AzK), N6-(((2-azidobenzyl)oxy)carbonyl)-L-lysine, N6-(((3-azidobenzyl)oxy)carbonyl)-L-lysine, N6-(((4-azidobenzyl)oxy)carbonyl)-L-lysine, 2-amino-6-((R)-tetrahydrofuran-2-carboxamido)hexanoic acid, *N*-ε-D-prolyl-L-lysine, and *N*-ε-cyclopentyloxycarbonyl-L-lysine; *N*-ε-Acryloyl-L-lysine; *N*-ε-[(1-(6-nitrobenzo[d][1,3]dioxol-5-yl)ethoxy)carbonyl]-L-lysine; and *N*-ε-(1-methylcyclopro-2-enecarboxamido)lysine.

[146] In some instances, an unnatural amino acid can be incorporated into an unnatural polypeptide or unnatural protein described herein by a synthetase disclosed in US 9,988,619 and US 9,938,516. Exemplary UAAs that can be incorporated by such synthetases include para-methylazido-L-phenylalanine, aralkyl, heterocyclyl, heteroaralkyl unnatural amino acids, and others. In some embodiments, such UAAs comprise pyridyl, pyrazinyl, pyrazolyl, triazolyl, oxazolyl, thiazolyl, thiophenyl, or other heterocycle. Such amino acids in some embodiments comprise azides, tetrazines, or other chemical group capable of conjugation to a coupling partner, such as a water soluble moiety. In some embodiments, such synthetases are expressed and used to incorporate UAAs into proteins *in vivo*. In some embodiments, such synthetases are used to incorporate UAAs into proteins using a cell-free translation system.

[147] In some instances, an unnatural amino acid can be incorporated into an unnatural polypeptide or unnatural protein described herein by a naturally occurring synthetase. In some embodiments, an unnatural amino acid is incorporated into an unnatural polypeptide or unnatural protein by an organism that is auxotrophic for one or more amino acids. In some embodiments, synthetases corresponding to the auxotrophic amino acid are capable of charging the corresponding tRNA with an unnatural amino acid. In some embodiments, the unnatural amino acid is selenocysteine, or a derivative thereof. In some embodiments, the unnatural amino acid is selenomethionine, or a derivative thereof. In some embodiments, the unnatural amino acid is an aromatic amino acid, wherein the aromatic amino acid comprises an aryl halide, such as an iodide. In embodiments, the unnatural amino acid is structurally similar to the auxotrophic amino acid.

IV. Engineered Cells

[148] In some embodiments, provided herein is a prokaryotic cell comprising: a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to an inducible promoter and a 5' UTR; a second DNA sequence encoding a promoter for a repressor of an inducible promoter, a third DNA sequence encoding at least one unnatural transfer RNA (tRNA) molecule; a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising a codon recognized by the unnatural tRNA molecule;

wherein the 5' UTR enables increased tRNA synthetase expression compared to SEQ ID NO: 1.

[149] In some embodiments, the unnatural tRNA molecule is an unnatural pyrrolysyl tRNA molecule. In some embodiments, the unnatural pyrrolysyl tRNA molecule is from *M. barkeri*. In some embodiments, the unnatural pyrrolysyl tRNA molecule is from *M. mazei*. In some embodiments, the unnatural pyrrolysyl tRNA molecule is a chimeric $tRNA^{Pyl}$, e.g., wherein the sequence of the chimeric $tRNA^{Pyl}$ comprises or consists of a portion of sequence from *M. barkeri* and a portion of sequence from *M. mazei*. In some embodiments, the 5' UTR comprises the DNA sequence of SEQ ID NO: 2. In some embodiments, the 5' UTR comprises the DNA sequence of a lac operator and/or SEQ ID NO: 13. In some embodiments, the lac operator comprises the DNA sequence of SEQ ID NO: 12. In some embodiments, the second DNA sequence is chromosomally integrated. In some embodiments, the third DNA sequence is on a plasmid. In some embodiments, the fourth DNA sequence is on a plasmid. In some embodiments, the plasmid has a p15A origin of replication.

[150] In some embodiments, the first sequence comprises a DNA sequence that is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 8. In some embodiments, the first sequence comprises a DNA sequence comprising SEQ ID NO: 8.

[151] In some embodiments, the promoter for the repressor of the inducible promoter is a P_{lacIQ1} promoter. In some embodiments, the promoter for the repressor of the inducible promoter is a P_{lacIQ} promoter.

[152] In some embodiments, the cell does not comprise a functional *recA* gene.

[153] In some embodiments, provided herein is a prokaryotic cell comprising: a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to an inducible promoter, and a 5' UTR comprising SEQ ID NO: 2; a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ1} promoter; a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid; and a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising a codon recognized by the unnatural pyrrolysyl tRNA, wherein the fourth DNA sequence is on a plasmid; and a fifth DNA sequence encoding a nucleoside triphosphate transporter operably linked to a P_{lacUV5} promoter, optionally wherein the prokaryotic cell further comprises an antibiotic resistance gene. In some embodiments, the inducible promoter operably linked to the DNA sequence encoding the tRNA synthetase is a TacO promoter or a Tac promoter, also known as tac promoter, Ptac promoter, ptac promoter,

or P_{tac} promoter. *See* G. L. Rosano & E. A. Ceccarelli, Frontiers in Microbiology 5(172):1-17 (2014), incorporated herein by reference for discussion of tac promoters. In some embodiments, the inducible promoter operably linked to the DNA sequence encoding the tRNA synthetase is a T7 promoter. *See* Hanna Tegel, J. Ottosson, & S. Hober, the FEBS Journal 278: 729-739 (2011), incorporated herein by reference for discussion of T7 promoters. In some embodiments, the inducible promoter operably linked to the DNA sequence encoding the tRNA synthetase is a 23119 promoter, also known as a J23119 promoter. *See* Q. Yang & S. Fong, J. Biological Engineering 11: 33 (2017), incorporated herein by reference for discussion of 23119 promoters.

[154] In some embodiments, provided herein is a prokaryotic cell comprising: a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to an inducible promoter, and a 5' UTR comprising SEQ ID NO: 2; a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ1} promoter; a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid; and a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising an unnatural codon recognized by the unnatural pyrrolysyl tRNA, wherein the fourth DNA sequence is on a plasmid, and a DNA sequence encoding a nucleoside triphosphate transporter, wherein the DNA sequence encoding the nucleoside triphosphate transporter is operably linked to a P_{lac} promoter and does not comprise an oppositely oriented promoter within 1.3-1.5 kb downstream of the 3' end of the nucleoside triphosphate transporter coding sequence. In some embodiments, the inducible promoter operably linked to the DNA sequence encoding the tRNA synthetase is a TacO promoter or a Tac promoter, also known as tac promoter, Ptac promoter, ptac promoter, or P_{tac} promoter. *See* G. L. Rosano & E. A. Ceccarelli, Frontiers in Microbiology 5(172):1-17 (2014), incorporated herein by reference for discussion of tac promoters. In some embodiments, the inducible promoter operably linked to the DNA sequence encoding the tRNA synthetase is a T7 promoter. *See* Hanna Tegel, J. Ottosson, & S. Hober, the FEBS Journal 278: 729-739 (2011), incorporated herein by reference for discussion of T7 promoters. In some embodiments, the inducible promoter operably linked to the DNA sequence encoding the tRNA synthetase is a 23119 promoter, also known as a J23119 promoter. *See* Q. Yang & S. Fong, J. Biological Engineering 11: 33 (2017), incorporated herein by reference for discussion of 23119 promoters.

[155] In some embodiments, provided herein is a prokaryotic cell comprising a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to an

inducible promoter, and a 5' UTR comprising SEQ ID NO: 2; a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ1} promoter; a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid; a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising an unnatural codon recognized by the unnatural pyrrolysyl tRNA, wherein the fourth DNA sequence is on a plasmid; a fifth DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed; a sixth DNA sequence encoding a LexA that does not undergo RecA-stimulated cleavage, optionally wherein the LexA comprises a S119 mutation, optionally wherein the S119 mutation is S119A; a DNA sequence encoding a nucleoside triphosphate transporter, wherein the DNA sequence encoding the nucleoside triphosphate transporter is operably linked to a P_{lac} promoter and does not comprise an oppositely oriented promoter within 1.3-1.5 kb downstream of the 3' end of the nucleoside triphosphate transporter coding sequence; and a *recA* hypomorph, optionally wherein the *recA* hypomorph results from RecX overexpression. In some embodiments, the inducible promoter operably linked to the DNA sequence encoding the tRNA synthetase is a TacO promoter or a Tac promoter, also known as tac promoter, Ptac promoter, ptac promoter, or P_{tac} promoter. *See* G. L. Rosano & E. A. Ceccarelli, Frontiers in Microbiology 5(172):1-17 (2014), incorporated herein by reference for discussion of tac promoters. In some embodiments, the inducible promoter operably linked to the DNA sequence encoding the tRNA synthetase is a T7 promoter. *See* Hanna Tegel, J. Ottosson, & S. Hober, the FEBS Journal 278: 729-739 (2011), incorporated herein by reference for discussion of T7 promoters. In some embodiments, the inducible promoter operably linked to the DNA sequence encoding the tRNA synthetase is a 23119 promoter, also known as a J23119 promoter. *See* Q. Yang & S. Fong, J. Biological Engineering 11: 33 (2017), incorporated herein by reference for discussion of 23119 promoters.

[156] In some embodiments, provided herein is a prokaryotic cell comprising a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to an inducible promoter, and a 5' UTR comprising SEQ ID NO: 2; a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ} promoter; a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid; a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising an unnatural codon recognized by the unnatural pyrrolysyl tRNA molecule, wherein the fourth DNA sequence is on a plasmid. In some embodiments, the inducible promoter operably linked to the DNA sequence encoding the

tRNA synthetase is a TacO promoter or a Tac promoter, also known as tac promoter, Ptac promoter, ptac promoter, or P_{tac} promoter. *See* G. L. Rosano & E. A. Ceccarelli, Frontiers in Microbiology 5(172):1-17 (2014), incorporated herein by reference for discussion of tac promoters. In some embodiments, the inducible promoter operably linked to the DNA sequence encoding the tRNA synthetase is a T7 promoter. *See* Hanna Tegel, J. Ottosson, & S. Hober, the FEBS Journal 278: 729-739 (2011), incorporated herein by reference for discussion of T7 promoters. In some embodiments, the inducible promoter operably linked to the DNA sequence encoding the tRNA synthetase is a 23119 promoter, also known as a J23119 promoter. *See* Q. Yang & S. Fong, J. Biological Engineering 11: 33 (2017), incorporated herein by reference for discussion of 23119 promoters.

[157] In some embodiments, provided herein is a prokaryotic cell comprising a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to an inducible promoter, and a 5' UTR comprising SEQ ID NO: 2; a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ1} promoter; a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid; a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising a codon recognized by the unnatural pyrrolysyl tRNA molecule, wherein the fourth DNA sequence is on a plasmid; wherein the cell does not comprise a functional *recA* gene. In some embodiments, the inducible promoter operably linked to the DNA sequence encoding the tRNA synthetase is a TacO promoter or a Tac promoter, also known as tac promoter, Ptac promoter, ptac promoter, or P_{tac} promoter. *See* G. L. Rosano & E. A. Ceccarelli, Frontiers in Microbiology 5(172):1-17 (2014), incorporated herein by reference for discussion of tac promoters. In some embodiments, the inducible promoter operably linked to the DNA sequence encoding the tRNA synthetase is a T7 promoter. *See* Hanna Tegel, J. Ottosson, & S. Hober, the FEBS Journal 278: 729-739 (2011), incorporated herein by reference for discussion of T7 promoters. In some embodiments, the inducible promoter operably linked to the DNA sequence encoding the tRNA synthetase is a 23119 promoter, also known as a J23119 promoter. *See* Q. Yang & S. Fong, J. Biological Engineering 11: 33 (2017), incorporated herein by reference for discussion of 23119 promoters.

[158] In some embodiments, provided herein is a DNA sequence encoding a *Pt*NTT2 operably linked to a P_{lac} promoter, wherein (i) the DNA sequence does not comprise an oppositely oriented promoter within about 1.3-1.5 kb downstream of the 3' end of the *Pt*NTT2 coding sequence or (ii) the DNA sequence comprises a transcriptional terminator that

terminates anterograde *Pt*NTT2 transcription, and further comprises a transcription terminator element downstream of the stop codon of the DNA sequence encoding the *Pt*NTT2 and the transcription terminator element is oriented to terminate retrograde transcription into the DNA sequence encoding the *Pt*NTT2.

[159] In some embodiments, provided herein is a DNA sequence encoding a tRNA synthetase; a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed; a DNA sequence encoding a LexA that does not undergo RecA-stimulated cleavage, optionally wherein the LexA comprises a S119 mutation, optionally wherein the S119 mutation is S119A; and a *recA* hypomorph, optionally wherein the *recA* hypomorph results from RecX overexpression.

[160] In some embodiments, provided herein is a prokaryotic cell comprising a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to a inducible promoter, and a 5' UTR comprising SEQ ID NO: 2; a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ} promoter; a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid; a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising a codon recognized by the unnatural pyrrolysyl tRNA molecule, wherein the fourth DNA sequence is on a plasmid; wherein the cell does not comprise a functional *recA* gene.

[161] In some embodiments, the cell further comprises a DNA sequence encoding a T7 bacteriophage RNA polymerase, optionally wherein the DNA sequence encoding the T7 bacteriophage RNA polymerase is operably linked to an inducible $Plac_{UV5}$ promoter. In some embodiments, the T7 bacteriophage RNA polymerase is operably linked to an inducible promoter. In some embodiments, the T7 bacteriophage RNA polymerase is operably linked to a lac inducible promoter.

[162] In some embodiments, the cell comprises a DNA sequence encoding a nucleoside triphosphate transporter. In some embodiments, the nucleoside triphosphate transporter comprises a truncated *Pt*NTT2. In some embodiments, the amino acid sequence encoding the truncated *Pt*NTT2 is at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 9. In some embodiments the amino acid sequence encoding the truncated *Pt*NTT2 is SEQ ID NO: 9.

[163] In some embodiments, the amino acid sequence encoding the *Pt*NTT2 is at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 10.

[164] In some embodiments, the cell further comprises a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed. In some embodiments, the cell further comprises a DNA sequence comprising a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is constitutively overexpressed. In some embodiments, the derepressed or constitutively overexpressed promoter is a *polB* promoter with a LexA binding box deletion or an inactivated LexA binding box.

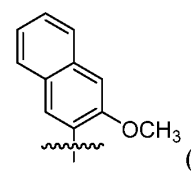
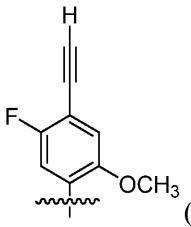
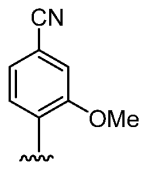
[165] In some embodiments, the cell further comprises a DNA sequence encoding a LexA that does not undergo RecA-stimulated cleavage.

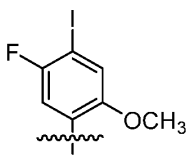
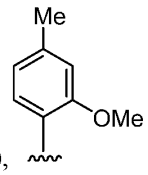
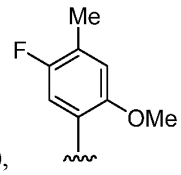
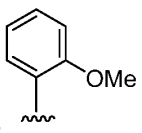
[166] In some embodiments, the cell further comprises a *recA* hypomorph. In some embodiments, the *recA* hypomorph results from RecX overexpression. In some embodiments, the RecA hypomorph is expressed by a promoter comprising a LexA binding site.

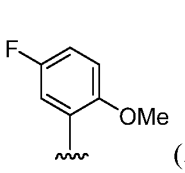
[167] In some embodiments, the prokaryotic cell further comprises an antibiotic resistance gene. In some embodiments, the antibiotic resistance gene is a chloramphenicol resistance gene.

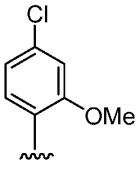
[168] In some embodiments, the third DNA sequence encoding at least one unnatural transfer RNA (tRNA) molecule comprises a first unnatural nucleotide comprising a first unnatural base; and the fourth DNA sequence encoding at least one unnatural mRNA molecule comprises a second unnatural nucleotide comprising a second unnatural base.

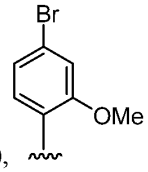
[169] In some embodiments, the first unnatural nucleotide and the second unnatural nucleotide each comprise an unnatural base independently selected from the group consisting of:

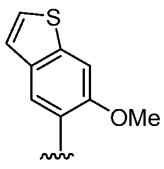
[170]  (NaM),  (FEMO),  (CNMO),

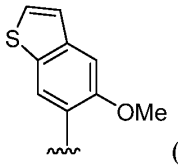
 (FIMO),  (MMO2),  (5FM),  (2OMe),

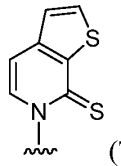


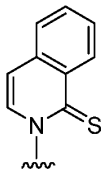
(5F2OMe),


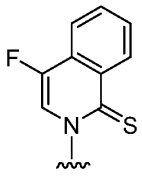
(ClMO),


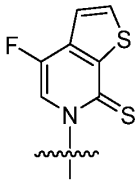
(BrMO),


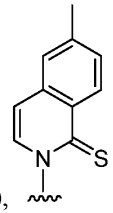
(PTMO),


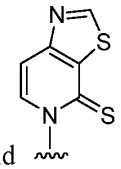
(MTMO),


(TPT3),


(SICS),


(FSICS),


(FTPT3),


(5SICS), and


(TAT1).

[171] In some embodiments, the first unnatural nucleotide and the second unnatural nucleotide each comprise an unnatural base selected from the group consisting of CNMO, TPT3, NAM, or TAT1.

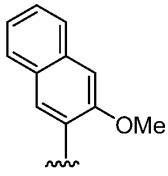
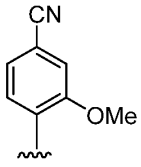
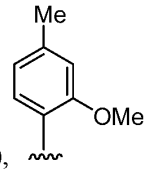
[172] In some embodiments, the first unnatural base comprises CNMO. In some embodiments, the first unnatural base comprises TPT3. In some embodiments, the first unnatural base comprises NAM. In some embodiments, the first unnatural base comprises TAT1.

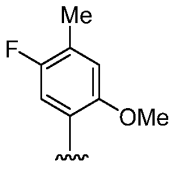
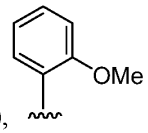
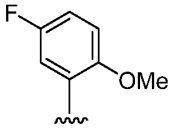
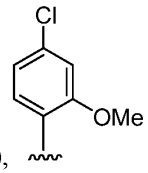
[173] In some embodiments, the third DNA sequence encoding at least one unnatural transfer RNA (tRNA) molecule comprises a first unnatural base pair, and the fourth DNA sequence encoding at least one unnatural mRNA molecule comprises a second unnatural base pair. In some embodiments, the first unnatural base pair is CNMO/TPT3. In some embodiments, the first unnatural base pair is NaM/TPT3. In some embodiments, the first unnatural base pair is CNMO/TAT1. In some embodiments, the first unnatural base pair is NaM/TAT1. In some embodiments, the first unnatural base pair is NaM/NaM.

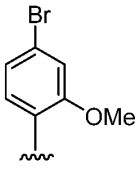
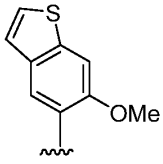
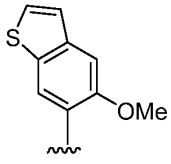
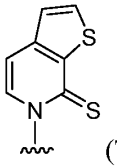
[174] In some embodiments, the codon recognized by the unnatural tRNA comprises a first unnatural nucleotide positioned at a first position, a second position, or a third position of the codon. In some embodiments, the first unnatural nucleotide is positioned at a second position or a third position of the codon.

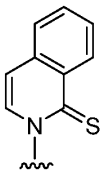
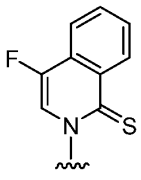
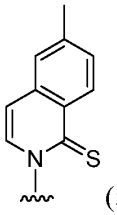
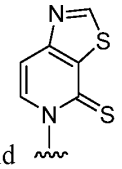
[175] In some embodiments, the codon recognized by the unnatural tRNA, when read from a 5' to 3' direction, comprises NNX or NXN, wherein N is any natural nucleotide, and X is an unnatural nucleotide. In some embodiments, the codon recognized by the unnatural tRNA

read from a 5' to 3' direction comprises UUX, UGX, CGX, AGX, GAX, CAX, AUX, CUX, GUX, UAX, GGX, GXU, CXU, GXG, AXG, GXC, AXC, GXA, CXC, or UXC. In some embodiments, the codon read from a 5' to 3' direction comprises CGX, AGX, GAX, GXU, CXU, GXC, AXC, or GXA. In some embodiments, the codon read from a 5' to 3' direction comprises CGX, AGX, GAX, GXU, GXC, or AXC. In some embodiments, the codon read from a 5' to 3' direction is GXU. In some embodiments, the codon read from a 5' to 3' direction is GXC. In some embodiments, the codon read from a 5' to 3' direction is AXC. In some embodiments, X comprises any one of the following bases:

[176]  (NaM),  (CNMO),  (MMO2),

 (5FM),  (2OMe),  (5F2OMe),  (ClMO),

 (BrMO),  (PTMO),  (MTMO),  (TPT3),

 (SICS),  (FSICS),  (5SICs), and  (TAT1).

[177] In some embodiments, X is NaM. In some embodiments, X is CNMO. In some embodiments, Y is TPT3. In some embodiments, Y is TAT1. In some embodiments, the cell comprises unnatural dNTPs.

[178] In some embodiments, the cell translates at least one unnatural polypeptide comprising the unnatural amino acid. In some embodiments, the cell exhibits improved titer, as measured by production of unnatural polypeptides comprising the unnatural amino acid, optionally wherein titer is measured in mg/L of unnatural polypeptides comprising the unnatural amino acid or optical density (OD) the cells, further optionally wherein titer is measured using HPLC, as compared to a reference prokaryotic cell comprising a functional plasmid comprising a DNA sequence encoding an *Methanosarcina barkeri* pyrrolysyl-tRNA

synthetase comprising SEQ ID NO: 3 and wherein the reference prokaryotic cell comprises a wild-type *recA* gene, a DNA sequence encoding a *Pt*NTT2 operably linked to a P_{lac} promoter, and a 5' UTR of SEQ ID NO: 1 and is otherwise isogenic with a cell described herein. In some embodiments, the functional plasmid has a p15A origin of replication.

[179] In some embodiments, the cell exhibits improved fidelity in synthesizing an unnatural polypeptide comprising the unnatural amino acid, wherein fidelity is percent unnatural amino acid incorporation at an intended site, optionally wherein fidelity is measured using LC-MS, as compared to a reference prokaryotic cell comprising a functional plasmid comprising a DNA sequence encoding an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising SEQ ID NO: 3 and wherein the reference prokaryotic cell comprises a wild-type *recA* gene, a DNA sequence encoding a *Pt*NTT2 operably linked to a P_{lac} promoter, and a 5' UTR of SEQ ID NO: 1 and is otherwise isogenic with a cell described herein. In some embodiments, the functional plasmid has a p15A origin of replication.

[180] In some embodiments, provided herein is a cell culture comprising any of the cells described herein, wherein the culture is characterized by reduced colony size heterogeneity, as compared to a reference prokaryotic cell comprising a functional plasmid comprising a DNA sequence encoding an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising SEQ ID NO: 3 and wherein the reference prokaryotic cell comprises a wild-type *recA* gene, a DNA sequence encoding a *Pt*NTT2 operably linked to a P_{lac} promoter, and a 5' UTR of SEQ ID NO: 1 and is otherwise isogenic with a cell described herein. In some embodiments, the functional plasmid has a p15A origin of replication.

[181] In some embodiments, a cell described herein is a prokaryotic cell. In some embodiments, such a prokaryotic cell is an *E. coli* cell. In some embodiments, such a prokaryotic cell is a competent cell comprising a BL21 background, a BL21(DE3) background, a DH5α background, a DH10B background, a HB101 background, or JM109 background. *See* Son, M.S. & Taylor, R.K., *Curr. Protoc.* Jan; (1)(1): e20 (2021) (e.g., at Table 5A.4.1); Jeong, H., Kim, H.J., & Lee, S.J., *Genome Announc.* Mar-Apr 3(2): e00134-15 (2015); Kim, S.K. et al., *Nucleic Acids Res.* 45(9): 5285-5293 (2017); Chen, J. et al., *Genome Announc.* Mar 6(10): e00097-18 (2018); Durfee, T. et al., *J. Bacteriol.* 190(7): 2597-2606 (2008). In some embodiments, the cell comprises a DNA sequence in which the sequence encoding IS1 transposase is deleted. In some embodiments, the sequence encoding IS1 transposase is replaced with an inactive recombination sequence. In some embodiments, the inactive recombination sequence is an inactive FLP recognition (FRT) sequence. In some embodiments, the cell comprises a DNA sequence in which *gpE* encoding a lambda DE3

phage capsid protein is deleted. In some embodiments, the *gpE* sequence is replaced with an inactive recombination sequence. In some embodiments, the inactive recombination sequence is an inactive FLP recognition (FRT) sequence. In some embodiments, the cell comprises a soluble tRNA synthetase (i.e., substantially soluble, including, e.g., no more than 10%, no more than 9%, no more than 8%, no more than 7%, no more than 6%, no more than 5%, no more than 4%, no more than 3%, no more than 2%, no more than 1%, no more than 0.5%, or less, accumulation of such a tRNA synthetase into inclusion bodies), which in some embodiments comprises a chimeric pyrrolysyl-tRNA synthetase comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), and the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F. In some embodiments, the tRNA synthetase is integrated into a chromosomal *arsB* locus for *E. coli* under the control of an inducible promoter, for example, in some embodiments an IPTG-inducible promoter. In some embodiments, the inducible promoter (e.g., an IPTG-inducible promoter) is Tac or TacO. In some embodiments, the *arsB* locus is a chromosomal *arsB* locus for *E. coli*, e.g., in some embodiments for *E. coli* BL21(DE3). In some embodiments, a DNA sequence encoding a repressor of an IPTG-inducible promoter is integrated into a chromosomal *arsB* locus. In some embodiments, the repressor of the IPTG-inducible promoter is LacI repressor protein. In some embodiments, the promoter for the sequence encoding the repressor of the inducible promoter is a P_{lacIQ1} promoter. In some embodiments, a chromosomal *arsB* locus is a chromosomal *arsB* locus for *E. coli*, e.g., in some embodiments for *E. coli* BL21(DE3). In some embodiments, the tRNA synthetase included the cell is operably linked to a 5' UTR comprising SEQ ID NO: 2. In some embodiments, a DNA sequence encoding a truncated *Pt*NTT2 is integrated into a chromosomal *lacZYA* locus of *E. coli*. In some embodiments, the DNA sequence encoding the truncated *Pt*NTT2 is integrated so that it replaces *lacZY*. In some embodiments, the DNA sequence encoding the truncated *Pt*NTT2 is operably linked to a promoter. In some embodiments, the promoter is P_{lacUV5} or P_{lac}. In some embodiments, a chromosomal *lacZYA* locus is a chromosomal *lacZYA* locus for *E. coli*, e.g., in some embodiments for *E. coli* BL21(DE3). In some embodiments, the DNA sequence encoding the truncated *Pt*NTT2 does not comprise an oppositely oriented promoter within 1.3-1.5 kb downstream of the 3' end of the truncated *Pt*NTT2 coding sequence. In some embodiments, the cell comprises a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed. In some embodiments, the derepressed promoter is a *polB* promoter with a LexA

binding box deletion or an inactivated LexA binding box. In some embodiments, RecX is overexpressed. In some embodiments, the DNA sequences encoding RecA and RecX are on a single transcript, and a terminator between the DNA sequence encoding RecA and the DNA sequence encoding RecX is deleted. In some embodiments, the cell comprises a DNA sequence encoding LexA comprising a S119A mutation.

[182] In some embodiments, a prokaryotic cell is provided wherein the cell is an *E. coli*c cell, e.g., in some embodiments a competent *E. coli* cell comprising a BL21 background, a BL21(DE3) background, a DH5α background, a DH10B background, a HB101 background, or JM109 background; wherein the cell comprises a DNA sequence in which the sequence encoding IS1 transposase is deleted and replaced with an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a DNA sequence in which *gpE* encoding a lambda DE3 phage capsid protein is deleted and replaced with an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a soluble tRNA synthetase, which in some embodiments comprises a chimeric pyrrolysyl-tRNA synthetase, e.g., in some embodiments comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F, wherein the tRNA synthetase is integrated into a chromosomal *arsB* locus under the control of an inducible promoter, for example, in some embodiments, an IPTG-inducible promoter, further optionally wherein the IPTG-inducible promoter is Tac or TacO; wherein a DNA sequence encoding a repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is integrated into a chromosomal *arsB* locus, wherein the repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is LacI repressor protein, and optionally wherein the promoter for the sequence encoding the repressor of the inducible promoter is a P_{lacIQ1} promoter; and wherein a DNA sequence encoding a truncated *Pt*NTT2 is integrated into a chromosomal *lacZYA* locus of *E. coli* replacing *lacZY*, wherein the DNA sequence encoding the truncated *Pt*NTT2 is operably linked to a promoter, which in some embodiments, such a promoter is P_{lacUV5}. In some embodiments, a prokaryotic cell is provided wherein the cell is an *E. coli*c cell, e.g., in some embodiments a competent *E. coli* cell comprising a BL21 background, a BL21(DE3) background, a DH5α background, a DH10B background, a HB101 background, or JM109 background; wherein the cell comprises a DNA sequence in which the sequence encoding IS1 transposase is deleted and replaced with

an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a DNA sequence in which *gpE* encoding a lambda DE3 phage capsid protein is deleted and replaced with an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a soluble tRNA synthetase, which in some embodiments comprises a chimeric pyrrolysyl-tRNA synthetase, e.g., in some embodiments comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F, wherein the tRNA synthetase is integrated into a chromosomal *arsB* locus under the control of an inducible promoter, for example, in some embodiments, an IPTG-inducible promoter, further optionally wherein the IPTG-inducible promoter is Tac or TacO; wherein a DNA sequence encoding a repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is integrated into a chromosomal *arsB* locus, wherein the repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is LacI repressor protein, and optionally wherein the promoter for the sequence encoding the repressor of the inducible promoter is a P_{lacIQ1} promoter; wherein the tRNA synthetase included the cell is operably linked to a 5' UTR comprising SEQ ID NO: 2; and wherein a DNA sequence encoding a truncated *Pt*NTT2 is integrated into a chromosomal *lacZYA* locus of *E. coli* replacing *lacZY*, wherein the DNA sequence encoding the truncated *Pt*NTT2 is operably linked to a promoter, which in some embodiments, such a promoter is P_{lacUV5}.

[183] In some embodiments, a prokaryotic cell is provided wherein the cell is an *E. coli* cell, e.g., in some embodiments a competent *E. coli* cell comprising a BL21 background, a BL21(DE3) background, a DH5α background, a DH10B background, a HB101 background, or JM109 background; wherein the cell comprises a DNA sequence in which the sequence encoding IS1 transposase is deleted and replaced with an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a DNA sequence in which *gpE* encoding a lambda DE3 phage capsid protein is deleted and replaced with an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a soluble tRNA synthetase, which in some embodiments comprises a chimeric pyrrolysyl-tRNA synthetase, e.g., in some embodiments comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an

Methanosarcina mazei pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F, wherein the tRNA synthetase is integrated into a chromosomal *arsB* locus under the control of an inducible promoter, for example, in some embodiments an IPTG-inducible promoter, further optionally wherein the IPTG-inducible promoter is Tac or TacO; wherein a DNA sequence encoding a repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is integrated into a chromosomal *arsB* locus, wherein the repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is LacI repressor protein, and optionally wherein the promoter for the sequence encoding the repressor of the inducible promoter is a P_{lacIQ1} promoter; and wherein a DNA sequence encoding a truncated *Pt*NTT2 is integrated into a chromosomal *lacZYA* locus of *E. coli* replacing *lacZY*, wherein the DNA sequence encoding the truncated *Pt*NTT2 is operably linked to a promoter, which in some embodiments such a promoter is $P_{lac.}$ In some embodiments, a prokaryotic cell is provided wherein the cell is an *E. coli* cell, e.g., in some embodiments a competent *E. coli* cell comprising a BL21 background, a BL21(DE3) background, a DH5α background, a DH10B background, a HB101 background, or JM109 background; wherein the cell comprises a DNA sequence in which the sequence encoding IS1 transposase is deleted and replaced with an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a DNA sequence in which *gpE* encoding a lambda DE3 phage capsid protein is deleted and replaced with an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a soluble tRNA synthetase, which in some embodiments comprises a chimeric pyrrolysyl-tRNA synthetase, e.g., in some embodiments comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F, wherein the tRNA synthetase is integrated into a chromosomal *arsB* locus under the control of an inducible promoter, for example, in some embodiments an IPTG-inducible promoter, further optionally wherein the IPTG-inducible promoter is Tac or TacO; wherein a DNA sequence encoding a repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is integrated into a chromosomal *arsB* locus, wherein the repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is LacI repressor protein, and optionally wherein the promoter for the sequence encoding the repressor of the inducible promoter is a P_{lacIQ1} promoter; wherein the

tRNA synthetase included the cell is operably linked to a 5' UTR comprising SEQ ID NO: 2; wherein a DNA sequence encoding a truncated *Pt*NTT2 is integrated into a chromosomal *lacZYA* locus of *E. coli* replacing *lacZY*, wherein the DNA sequence encoding the truncated *Pt*NTT2 is operably linked to a promoter, which in some embodiments such a promoter is P_{lac}; and wherein the DNA sequence encoding the truncated *Pt*NTT2 does not comprise an oppositely oriented promoter within 1.3-1.5 kb downstream of the 3' end of the truncated *Pt*NTT2 coding sequence.

[184] In some embodiments, a prokaryotic cell is provided wherein the cell is an *E. coli* cell, e.g., in some embodiments a competent *E. coli* cell comprising a BL21 background, a BL21(DE3) background, a DH5α background, a DH10B background, a HB101 background, or JM109 background; wherein the cell comprises a DNA sequence in which the sequence encoding IS1 transposase is deleted and replaced with an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a DNA sequence in which *gpE* encoding a lambda DE3 phage capsid protein is deleted and replaced with an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a soluble tRNA synthetase, which in some embodiments comprises a chimeric pyrrolysyl-tRNA synthetase, e.g., in some embodiments comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F, wherein the tRNA synthetase is integrated into a chromosomal *arsB* locus under the control of an inducible promoter, for example, in some embodiments an IPTG-inducible promoter, further optionally wherein the IPTG-inducible promoter is Tac or TacO; wherein a DNA sequence encoding a repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is integrated into a chromosomal *arsB* locus, wherein the repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is LacI repressor protein, and optionally wherein the promoter for the sequence encoding the repressor of the inducible promoter is a P_{lacIQ1} promoter; wherein a DNA sequence encoding a truncated *Pt*NTT2 is integrated into a chromosomal *lacZYA* locus of *E. coli* replacing *lacZY*, wherein the DNA sequence encoding the truncated *Pt*NTT2 is operably linked to a promoter, which in some embodiments such a promoter is P_{lac}; and wherein the cell comprises a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed, optionally further wherein the derepressed promoter is a *polB* promoter with a

LexA binding box deletion or an inactivated LexA binding box. In some embodiments, a prokaryotic cell is provided wherein the cell is an *E. coli* cell, e.g., in some embodiments a competent *E. coli* cell comprising a BL21 background, a BL21(DE3) background, a DH5α background, a DH10B background, a HB101 background, or JM109 background; wherein the cell comprises a DNA sequence in which the sequence encoding IS1 transposase is deleted and replaced with an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a DNA sequence in which *gpE* encoding a lambda DE3 phage capsid protein is deleted and replaced with an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a soluble tRNA synthetase, which in some embodiments comprises a chimeric pyrrolysyl-tRNA synthetase, e.g., in some embodiments comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F, wherein the tRNA synthetase is integrated into a chromosomal *arsB* locus under the control of an inducible promoter, for example, in some embodiments an IPTG-inducible promoter, further optionally wherein the IPTG-inducible promoter is Tac or TacO; wherein a DNA sequence encoding a repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is integrated into a chromosomal *arsB* locus, wherein the repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is LacI repressor protein, and optionally wherein the promoter for the sequence encoding the repressor of the inducible promoter is a P_{lacIQ1} promoter; wherein the tRNA synthetase included the cell is operably linked to a 5' UTR comprising SEQ ID NO: 2; wherein a DNA sequence encoding a truncated *Pt*NTT2 is integrated into a chromosomal *lacZYA* locus of *E. coli* replacing *lacZY*, wherein the DNA sequence encoding the truncated *Pt*NTT2 is operably linked to a promoter, which in some embodiments such a promoter is P_{lac}; wherein the DNA sequence encoding the truncated *Pt*NTT2 does not comprise an oppositely oriented promoter within 1.3-1.5 kb downstream of the 3' end of the truncated *Pt*NTT2 coding sequence; and wherein the cell comprises a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed, optionally further wherein the derepressed promoter is a *polB* promoter with a LexA binding box deletion or an inactivated LexA binding box.

[185] In some embodiments, a prokaryotic cell is provided wherein the cell is an *E. coli* cell, e.g., in some embodiments a competent *E. coli* cell comprising a BL21 background, a

BL21(DE3) background, a DH5α background, a DH10B background, a HB101 background, or JM109 background; wherein the cell comprises a DNA sequence in which the sequence encoding IS1 transposase is deleted and replaced with an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a DNA sequence in which *gpE* encoding a lambda DE3 phage capsid protein is deleted and replaced with an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a chimeric pyrrolysyl-tRNA synthetase, e.g., in some embodiments comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F, wherein the tRNA synthetase is integrated into a chromosomal *arsB* locus under the control of an inducible promoter, for example, in some embodiments an IPTG-inducible promoter, further optionally wherein the IPTG-inducible promoter is Tac or TacO; wherein a DNA sequence encoding a repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is integrated into a chromosomal *arsB* locus, wherein the repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is LacI repressor protein, and optionally wherein the promoter for the sequence encoding the repressor of the inducible promoter is a P_{lacIQ1} promoter; wherein a DNA sequence encoding a truncated *Pt*NTT2 is integrated into a chromosomal *lacZYA* locus of *E. coli* replacing *lacZY*, wherein the DNA sequence encoding the truncated *Pt*NTT2 is operably linked to a promoter, which in some embodiments such a promoter is P_{lac}; wherein the cell comprises a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed, optionally further wherein the derepressed promoter is a *polB* promoter with a LexA binding box deletion or an inactivated LexA binding box; and wherein RecX is overexpressed, optionally further wherein the DNA sequences encoding RecA and RecX are on a single transcript and/or a terminator between the DNA sequence encoding RecA and the DNA sequence encoding RecX is deleted. In some embodiments, a prokaryotic cell is provided wherein the cell is an *E. coli* cell, e.g., in some embodiments a competent *E. coli* cell comprising a BL21 background, a BL21(DE3) background, a DH5α background, a DH10B background, a HB101 background, or JM109 background; wherein the cell comprises a DNA sequence in which the sequence encoding IS1 transposase is deleted and replaced with an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a DNA sequence in

which *gpE* encoding a lambda DE3 phage capsid protein is deleted and replaced with an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a chimeric pyrrolysyl-tRNA synthetase, e.g., in some embodiments comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F, wherein the tRNA synthetase is integrated into a chromosomal *arsB* locus under the control of an inducible promoter, for example, in some embodiments an IPTG-inducible promoter, further optionally wherein the IPTG-inducible promoter is Tac or TacO; wherein a DNA sequence encoding a repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is integrated into a chromosomal *arsB* locus, wherein the repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is LacI repressor protein, and optionally wherein the promoter for the sequence encoding the repressor of the inducible promoter is a P_{lacIQ1} promoter; wherein the tRNA synthetase included the cell is operably linked to a 5' UTR comprising SEQ ID NO: 2; wherein a DNA sequence encoding a truncated *Pt*NTT2 is integrated into a chromosomal *lacZYA* locus of *E. coli* replacing *lacZY*, wherein the DNA sequence encoding the truncated *Pt*NTT2 is operably linked to a promoter, which in some embodiments such a promoter is P_{lac}; wherein the DNA sequence encoding the truncated *Pt*NTT2 does not comprise an oppositely oriented promoter within 1.3-1.5 kb downstream of the 3' end of the truncated *Pt*NTT2 coding sequence; wherein the cell comprises a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed, optionally further wherein the derepressed promoter is a *polB* promoter with a LexA binding box deletion or an inactivated LexA binding box; and wherein RecX is overexpressed, optionally further wherein the DNA sequences encoding RecA and RecX are on a single transcript and/or a terminator between the DNA sequence encoding RecA and the DNA sequence encoding RecX is deleted.

[186] In some embodiments, a prokaryotic cell is provided wherein the cell is *an E. coli* cell, e.g., in some embodiments a competent *E. coli* cell comprising a BL21 background, a BL21(DE3) background, a DH5α background, a DH10B background, a HB101 background, or JM109 background; wherein the cell comprises a DNA sequence in which the sequence encoding IS1 transposase is deleted and replaced with an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a DNA sequence in which *gpE* encoding a lambda DE3

phage capsid protein is deleted and replaced with an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a soluble tRNA synthetase, which in some embodiments comprises a chimeric pyrrolysyl-tRNA synthetase, e.g., in some embodiments comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F, wherein the tRNA synthetase is integrated into a chromosomal *arsB* locus under the control of an inducible promoter, for example, in some embodiments an IPTG-inducible promoter, optionally further wherein the IPTG-inducible promoter is Tac or TacO; wherein a DNA sequence encoding a repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is integrated into a chromosomal *arsB* locus, wherein the repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is LacI repressor protein, and optionally wherein the promoter for the sequence encoding the repressor of the inducible promoter is a P_{lacIQ1} promoter; wherein a DNA sequence encoding a truncated *Pt*NTT2 is integrated into a chromosomal *lacZYA* locus of *E. coli* replacing *lacZY*, wherein the DNA sequence encoding the truncated *Pt*NTT2 is operably linked to a promoter, which in some embodiments such a promoter is P_{lac}; wherein the cell comprises a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed, optionally further wherein the derepressed promoter is a *polB* promoter with a LexA binding box deletion or an inactivated LexA binding box; and wherein RecX is overexpressed, optionally further wherein the DNA sequences encoding RecA and RecX are on a single transcript and/or a terminator between the DNA sequence encoding RecA and the DNA sequence encoding RecX is deleted; and wherein the cell comprises a DNA sequence encoding LexA comprising a S119A mutation. In some embodiments, a prokaryotic cell is provided wherein the cell is *an E. coli* cell, e.g., in some embodiments a competent *E. coli* cell comprising a BL21 background, a BL21(DE3) background, a DH5α background, a DH10B background, a HB101 background, or JM109 background; wherein the cell comprises a DNA sequence in which the sequence encoding IS1 transposase is deleted and replaced with an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a DNA sequence in which *gpE* encoding a lambda DE3 phage capsid protein is deleted and replaced with an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a soluble tRNA

synthetase, which in some embodiments comprises a chimeric pyrrolysyl-tRNA synthetase, e.g., in some embodiments comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F, wherein the tRNA synthetase is integrated into a chromosomal *arsB* locus under the control of an inducible promoter, for example, in some embodiments an IPTG-inducible promoter, optionally further wherein the IPTG-inducible promoter is Tac or TacO; wherein a DNA sequence encoding a repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is integrated into a chromosomal *arsB* locus, wherein the repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is LacI repressor protein, and optionally wherein the promoter for the sequence encoding the repressor of the inducible promoter is a P_{lacIQ1} promoter; wherein the tRNA synthetase included the cell is operably linked to a 5' UTR comprising SEQ ID NO: 2; wherein a DNA sequence encoding a truncated *Pt*NTT2 is integrated into a chromosomal *lacZYA* locus of *E. coli* replacing *lacZY*, wherein the DNA sequence encoding the truncated *Pt*NTT2 is operably linked to a promoter, which in some embodiments such a promoter is P_{lac}; wherein the DNA sequence encoding the truncated *Pt*NTT2 does not comprise an oppositely oriented promoter within 1.3-1.5 kb downstream of the 3' end of the truncated *Pt*NTT2 coding sequence; wherein the cell comprises a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed, optionally further wherein the derepressed promoter is a *polB* promoter with a LexA binding box deletion or an inactivated LexA binding box; and wherein RecX is overexpressed, optionally further wherein the DNA sequences encoding RecA and RecX are on a single transcript and/or a terminator between the DNA sequence encoding RecA and the DNA sequence encoding RecX is deleted; and wherein the cell comprises a DNA sequence encoding LexA comprising a S119A mutation.

[187] In some embodiments, a prokaryotic cell is provided wherein the cell is an *E. coli* cell, e.g., in some embodiments a competent *E. coli* cell comprising a BL21 background, a BL21(DE3) background, a DH5α background, a DH10B background, a HB101 background, or JM109 background; wherein the cell comprises a DNA sequence in which the sequence encoding IS1 transposase is deleted and replaced with an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a DNA sequence in which *gpE* encoding a lambda DE3 phage capsid protein is deleted and replaced with an inactive recombination sequence, further

wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a soluble tRNA synthetase, which in some embodiments comprises a chimeric pyrrolysyl-tRNA synthetase, e.g., in some embodiments comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F, wherein the tRNA synthetase is integrated into a chromosomal *arsB* locus under the control of an inducible promoter, for example, in some embodiments an IPTG-inducible promoter, optionally further wherein the IPTG-inducible promoter is Tac or TacO; wherein a DNA sequence encoding a repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is integrated into a chromosomal *arsB* locus, wherein the repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is LacI repressor protein, and optionally wherein the promoter for the sequence encoding the repressor of the inducible promoter is a P_{lacIQ1} promoter; wherein a DNA sequence encoding a truncated *Pt*NTT2 is integrated into a chromosomal *lacZYA* locus of *E. coli* replacing *lacZY*, wherein the DNA sequence encoding the truncated *Pt*NTT2 is operably linked to a promoter, which in some embodiments such a promoter is P_{lac}; wherein the cell comprises a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed, optionally further wherein the derepressed promoter is a *polB* promoter with a LexA binding box deletion or an inactivated LexA binding box; and wherein RecX is overexpressed, optionally further wherein the DNA sequences encoding RecA and RecX are on a single transcript and/or a terminator between the DNA sequence encoding RecA and the DNA sequence encoding RecX is deleted; wherein the cell comprises a DNA sequence encoding LexA comprising a S119A mutation; and wherein the cell comprises a *relA* deletion or inactivated *relA*, optionally wherein the *relA* deletion or inactivated *relA* improves titer of a target polypeptide of interest produced by such a cell, and/or reduces stringent response. In some embodiments, a prokaryotic cell is provided wherein the cell is an *E. coli* cell, e.g., in some embodiments a competent *E. coli* cell comprising a BL21 background, a BL21(DE3) background, a DH5α background, a DH10B background, a HB101 background, or JM109 background; wherein the cell comprises a DNA sequence in which the sequence encoding IS1 transposase is deleted and replaced with an inactive recombination sequence, further wherein the inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a DNA sequence in which *gpE* encoding a lambda DE3 phage capsid protein is deleted and replaced with an inactive recombination sequence, further wherein the

inactive recombination sequence is an inactive FLP recognition (FRT) sequence; wherein the cell comprises a soluble tRNA synthetase, which in some embodiments comprises a chimeric pyrrolysyl-tRNA synthetase, e.g., in some embodiments comprising an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F, wherein the tRNA synthetase is integrated into a chromosomal *arsB* locus under the control of an inducible promoter, for example, in some embodiments an IPTG-inducible promoter, optionally further wherein the IPTG-inducible promoter is Tac or TacO; wherein a DNA sequence encoding a repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is integrated into a chromosomal *arsB* locus, wherein the repressor of the inducible promoter (e.g., an IPTG-inducible promoter) is LacI repressor protein, and optionally wherein the promoter for the sequence encoding the repressor of the inducible promoter is a P_{lacIQ1} promoter; wherein the tRNA synthetase included the cell is operably linked to a 5' UTR comprising SEQ ID NO: 2; wherein a DNA sequence encoding a truncated *Pt*NTT2 is integrated into a chromosomal *lacZYA* locus of *E. coli* replacing *lacZY*, wherein the DNA sequence encoding the truncated *Pt*NTT2 is operably linked to a promoter, which in some embodiments such a promoter is P_{lac}; wherein the DNA sequence encoding the truncated *Pt*NTT2 does not comprise an oppositely oriented promoter within 1.3-1.5 kb downstream of the 3' end of the truncated *Pt*NTT2 coding sequence; wherein the cell comprises a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed, optionally further wherein the derepressed promoter is a *polB* promoter with a LexA binding box deletion or an inactivated LexA binding box; and wherein RecX is overexpressed, optionally further wherein the DNA sequences encoding RecA and RecX are on a single transcript and/or a terminator between the DNA sequence encoding RecA and the DNA sequence encoding RecX is deleted; wherein the cell comprises a DNA sequence encoding LexA comprising a S119A mutation; and wherein the cell comprises a *relA* deletion or inactivated *relA*, optionally wherein the *relA* deletion or inactivated *relA* improves titer of a target polypeptide of interest produced by such a cell, and/or reduces stringent response.

[188] In some embodiments, the cell comprises a DNA sequence encoding an mRNA molecule comprising at least one unnatural codon encoding an unnatural amino acid. In some embodiments, the cell comprises a DNA sequence encoding at least first and second unnatural codons each encoding an unnatural amino acid. In some embodiments, the cell comprises a DNA sequence encoding an mRNA molecule comprising first and second unnatural codons,

wherein the first and second unnatural codons encode a first and second unnatural amino acid. In some embodiments, the first unnatural amino acid is different from the second unnatural amino acid. In some embodiments, the mRNA molecule encodes a polypeptide comprising a wild-type OmpA signal peptide. In some embodiments, such a wild-type OmpA signal peptide is from *E. coli*.

V. Isolated Nucleic Acids, Vectors, Host Cells, and Methods

[189] In some embodiments, provided herein is an isolated nucleic acid encoding a tRNA synthetase.

[190] In some embodiments, the tRNA synthetase is encoded by gene on a plasmid. In some embodiments, the tRNA synthetase is encoded by a gene that is chromosomally integrated.

[191] In some embodiments, provided herein is a vector comprising any of the nucleic acids disclosed herein.

[192] In some embodiments, provided herein is a host cell comprising any of the vectors disclosed herein.

[193] In some embodiments, provided herein is a host cell that produces any one of the tRNA synthetases described herein.

[194] In some embodiments, provide herein is a method for making a tRNA synthetase, comprising culturing any of the host cells described herein under conditions suitable for expression of the tRNA synthetase. In some embodiments, the method further comprises recovering the tRNA synthetase produced by the host cell.

VI. Nucleic Acid Molecules

[195] In some embodiments, a nucleic acid (e.g., also referred to herein as nucleic acid molecule of interest) is from any source or composition, such as DNA, cDNA, gDNA (genomic DNA), RNA, siRNA (short inhibitory RNA), RNAi, tRNA, mRNA or rRNA (ribosomal RNA), for example, and is in any form (e.g., linear, circular, supercoiled, single-stranded, double-stranded, and the like). In some embodiments, nucleic acids comprise nucleotides, nucleosides, or polynucleotides. In some cases, nucleic acids comprise natural and unnatural nucleic acids. In some cases, a nucleic acid also comprises unnatural nucleic acids, such as DNA or RNA analogs (e.g., containing base analogs, sugar analogs and/or a non-native backbone and the like). It is understood that the term “nucleic acid” does not refer to or infer a specific length of the polynucleotide chain, thus polynucleotides and oligonucleotides are also included in the definition. Exemplary natural nucleotides include, without limitation, ATP, UTP, CTP, GTP, ADP, UDP, CDP, GDP, AMP, UMP, CMP, GMP, dATP, dTTP, dCTP,

dGTP, dADP, dTDP, dCDP, dGDP, dAMP, dTMP, dCMP, and dGMP. Exemplary natural deoxyribonucleotides include dATP, dTTP, dCTP, dGTP, dADP, dTDP, dCDP, dGDP, dAMP, dTMP, dCMP, and dGMP. Exemplary natural ribonucleotides include ATP, UTP, CTP, GTP, ADP, UDP, CDP, GDP, AMP, UMP, CMP, and GMP. For natural RNA, the uracil base is uridine. A nucleic acid sometimes is a vector, plasmid, phagemid, autonomously replicating sequence (ARS), centromere, artificial chromosome, yeast artificial chromosome (e.g., YAC) or other nucleic acid able to replicate or be replicated in a host cell. In some cases, an unnatural nucleic acid is a nucleic acid analogue. In additional cases, an unnatural nucleic acid is from an extracellular source. In other cases, an unnatural nucleic acid is available to the intracellular space of an organism provided herein, e.g., a genetically modified organism. In some embodiments, an unnatural nucleotide is not a natural nucleotide. In some embodiments, a nucleotide that does not comprise a natural base comprises an unnatural nucleobase.

VII. Unnatural Nucleic Acids

[196] A nucleotide analog, or unnatural nucleotide, comprises a nucleotide which contains some type of modification to either the base, sugar, or phosphate moieties that constitutes a difference from naturally occurring nucleotides. In some embodiments, a modification comprises a chemical modification. In some cases, modifications occur at the 3'OH or 5'OH group, at the backbone, at the sugar component, or at the nucleotide base. Modifications, in some instances, optionally include non-naturally occurring linker molecules and/or of interstrand or intrastrand cross links. In one aspect, the modified nucleic acid comprises modification of one or more of the 3'OH or 5'OH group, the backbone, the sugar component, or the nucleotide base, and /or addition of non-naturally occurring linker molecules. In one aspect, a modified backbone comprises a backbone other than a phosphodiester backbone. In one aspect, a modified sugar comprises a sugar other than deoxyribose (in modified DNA) or other than ribose (modified RNA). In one aspect, a modified base comprises a base other than adenine, guanine, cytosine or thymine (in modified DNA) or a base other than adenine, guanine, cytosine or uracil (in modified RNA).

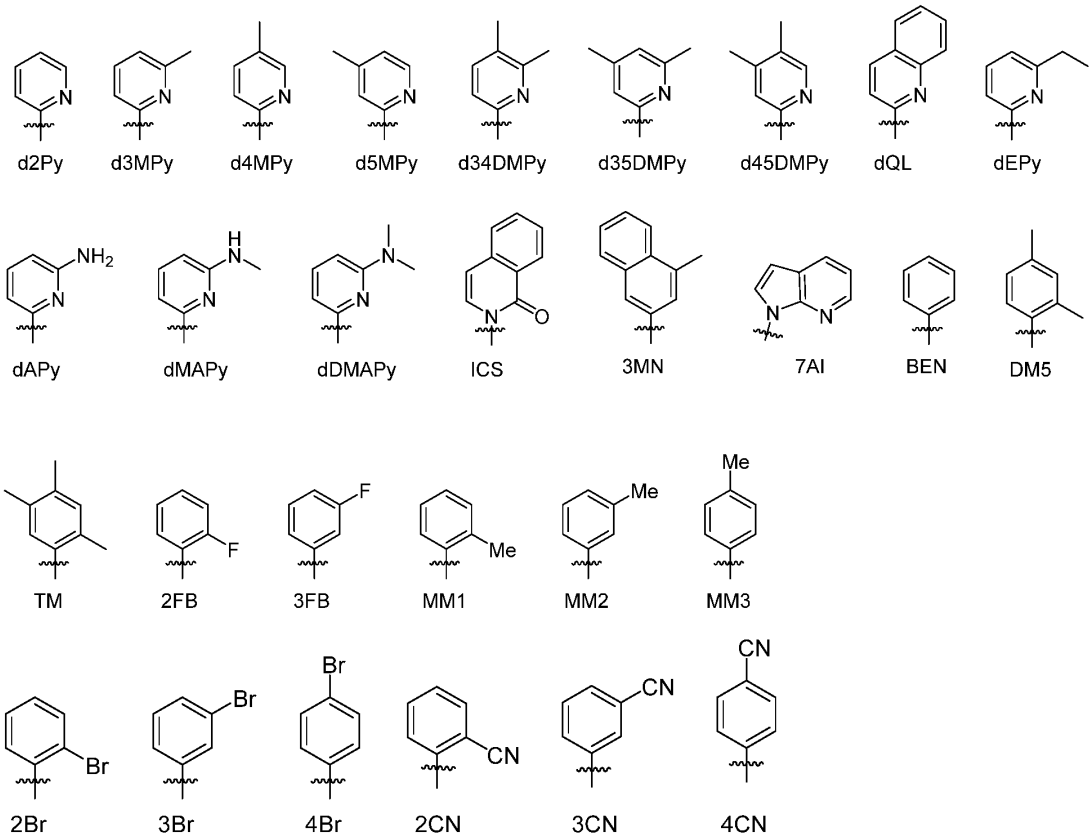
[197] In some embodiments, the nucleic acid comprises at least one modified base. In some instances, the nucleic acid comprises 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or more modified bases. In some cases, modifications to the base moiety include natural and synthetic modifications of A, C, G, and T/U as well as different purine or pyrimidine bases. In some embodiments, a modification is to a modified form of adenine, guanine cytosine or thymine (in modified DNA) or a modified form of adenine, guanine cytosine or uracil (modified RNA).

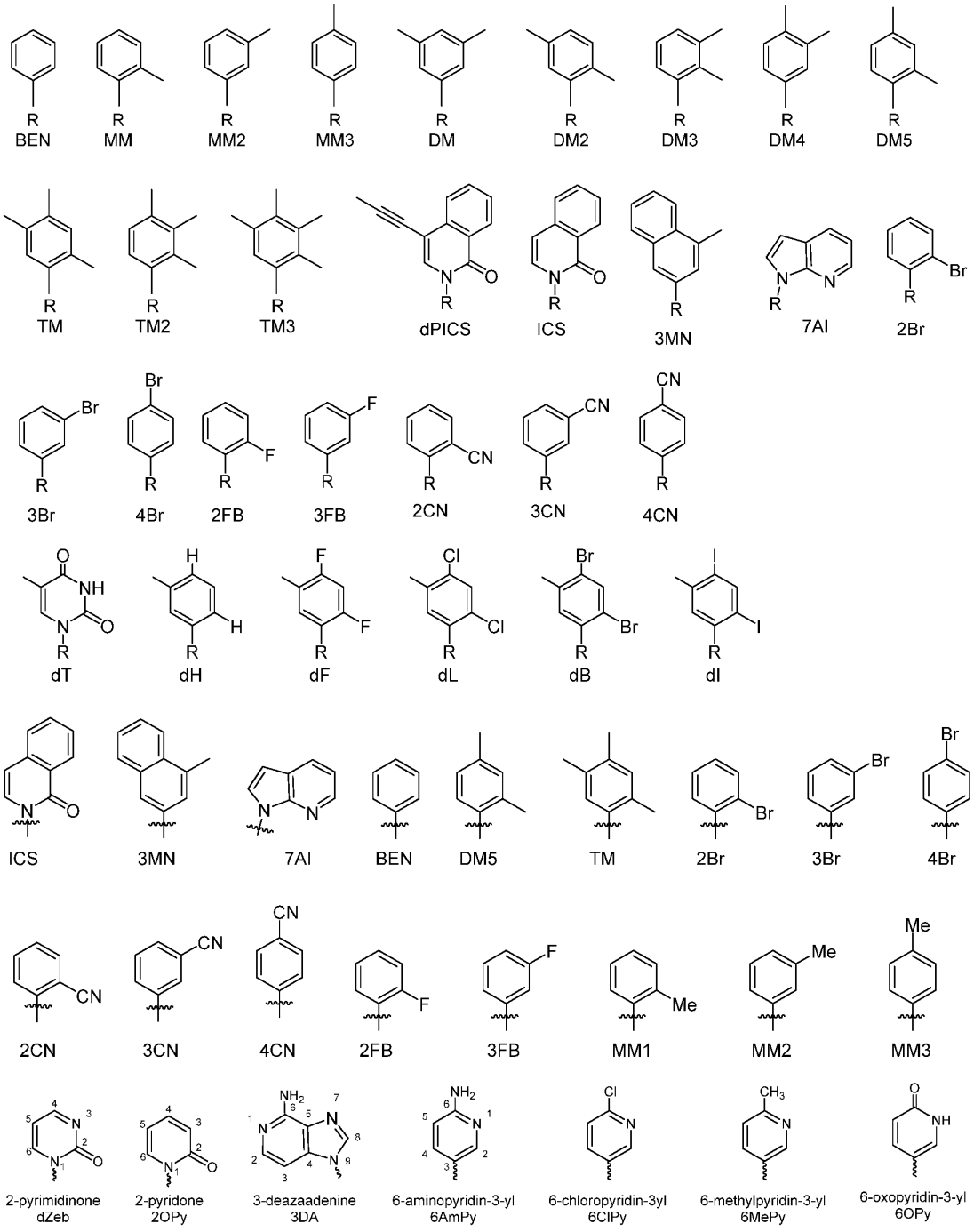
[198] A modified base of a unnatural nucleic acid includes, but is not limited to, uracil-5-yl, hypoxanthin-9-yl (I), 2-aminoadenin-9-yl, 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and 2-thiocytosine, 5-halouracil and cytosine, 5-propynyl uracil and cytosine, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 8-azaguanine and 8-azaadenine, 7-deazaguanine and 7-deazaadenine and 3-deazaguanine and 3-deazaadenine. Certain unnatural nucleic acids, such as 5-substituted pyrimidines, 6-azapyrimidines and N-2 substituted purines, N-6 substituted purines, O-6 substituted purines, 2-aminopropyladenine, 5-propynyluracil, 5-propynylcytosine, 5-methylcytosine, those that increase the stability of duplex formation, universal nucleic acids, hydrophobic nucleic acids, promiscuous nucleic acids, size-expanded nucleic acids, fluorinated nucleic acids, 5-substituted pyrimidines, 6-azapyrimidines and N-2, N-6 and O-6 substituted purines, including 2-aminopropyladenine, 5-propynyluracil and 5-propynylcytosine. 5-methylcytosine (5-me-C), 5- hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl, other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and 2-thiocytosine, 5-halouracil, 5-halocytosine, 5-propynyl (-C≡C-CH_3) uracil, 5-propynyl cytosine, other alkynyl derivatives of pyrimidine nucleic acids, 6-azo uracil, 6-azo cytosine, 6-azo thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5 -halo particularly 5-bromo, 5-trifluoromethyl, other 5-substituted uracils and cytosines, 7-methylguanine, 7- methyladenine, 2-F-adenine, 2-amino-adenine, 8-azaguanine, 8-azaadenine, 7-deazaguanine, 7- deazaadenine, 3-deazaguanine, 3-deazaadenine, tricyclic pyrimidines, phenoxazine cytidine([5,4-b][l,4]benzoxazin-2(3H)-one), phenothiazine cytidine (1H- pyrimido[5,4-b][l,4]benzothiazin-2(3H)-one), G-clamps, phenoxazine cytidine (e.g. 9- (2-aminoethoxy)-H-pyrimido[5,4-b][l,4]benzoxazin-2(3H)-one), carbazole cytidine (2H-pyrimido[4,5- b]indol-2-one), pyridoindole cytidine (H-pyrido[3',2':4,5]pyrrolo[2,3-d]pyrimidin-2-one), those in which the purine or pyrimidine base is replaced with other heterocycles, 7-deaza-adenine, 7-deazaguanosine, 2-aminopyridine, 2-pyridone, azacytosine, 5-bromocytosine, bromouracil, 5-chlorocytosine, chlorinated cytosine, cyclocytosine, cytosine arabinoside, 5- fluorocytosine, fluoropyrimidine, fluorouracil, 5,6-dihydrocytosine, 5-iodocytosine, hydroxyurea, iodouracil, 5-nitrocytosine, 5- bromouracil, 5-

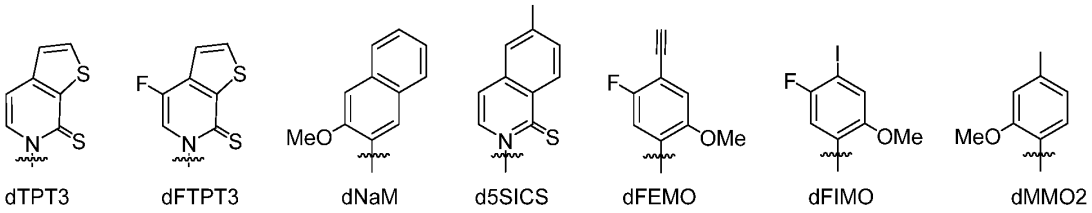
chlorouracil, 5- fluorouracil, and 5-iodouracil, 2-amino-adenine, 6-thio-guanine, 2-thio-thymine, 4-thio-thymine, 5-propynyl-uracil, 4-thio-uracil, N4-ethylcytosine, 7-deazaguanine, 7-deaza-8- azaguanine, 5-hydroxycytosine, 2'-deoxyuridine, 2-amino-2'-deoxyadenosine, and those described in U.S. Patent Nos. 3,687,808; 4,845,205; 4,910,300; 4,948,882; 5,093,232; 5,130,302; 5,134,066; 5,175,273; 5,367,066; 5,432,272; 5,457,187; 5,459,255; 5,484,908; 5,502,177; 5,525,711; 5,552,540; 5,587,469; 5,594,121; 5,596,091; 5,614,617; 5,645,985; 5,681,941; 5,750,692; 5,763,588; 5,830,653 and 6,005,096; WO 99/62923; Kandimalla et al., (2001) Bioorg. Med. Chem. 9:807-813; The Concise Encyclopedia of Polymer Science and Engineering, Kroschwitz, J.I., Ed., John Wiley & Sons, 1990, 858- 859; Englisch et al., Angewandte Chemie, International Edition, 1991, 30, 613; and Sanghvi, Chapter 15, Antisense Research and Applications, Crooke and Lebleu Eds., CRC Press, 1993, 273-288. Additional base modifications can be found, for example, in U.S. Pat. No. 3,687,808; Englisch et al., Angewandte Chemie, International Edition, 1991, 30, 613.

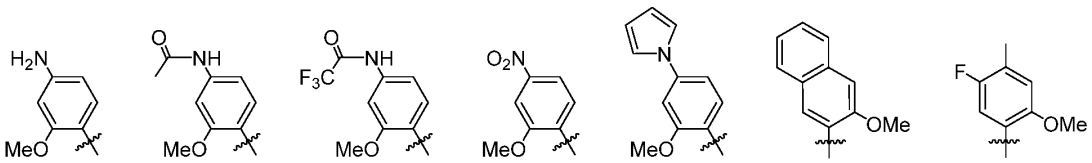
[199] Unnatural nucleic acids comprising various heterocyclic bases and various sugar moieties (and sugar analogs) are available in the art, and the nucleic acid in some cases include one or several heterocyclic bases other than the principal five base components of naturally-occurring nucleic acids. For example, the heterocyclic base includes, in some cases, uracil-5-yl, cytosin-5-yl, adenin-7-yl, adenin-8-yl, guanin-7-yl, guanin-8-yl, 4- aminopyrrolo [2.3-d] pyrimidin-5-yl, 2-amino-4-oxopyrolo [2, 3-d] pyrimidin-5-yl, 2- amino-4-oxopyrrolo [2.3-d] pyrimidin-3-yl groups, where the purines are attached to the sugar moiety of the nucleic acid via the 9-position, the pyrimidines via the 1 -position, the pyrrolopyrimidines via the 7-position and the pyrazolopyrimidines via the 1-position.

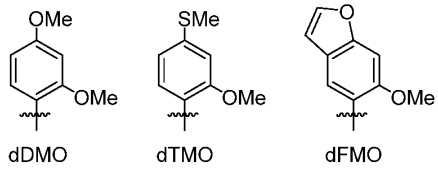
[200] In some embodiments, a modified base of an unnatural nucleic acid is depicted below, wherein the wavy line or R identifies a point of attachment to the deoxyribose or ribose.

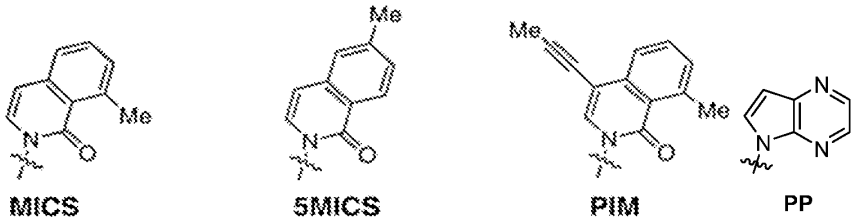

d2Py
d3MPy
d4MPy
d5MPy
d34DMPy
d35DMPy
d45DMPy
dQL
dEPy
dAPy
dMAPy
dDMAPy
ICS
3MN
7AI
BEN
DM5
TM
2FB
3FB
MM1
MM2
MM3
2Br
3Br
4Br
2CN
3CN
4CN

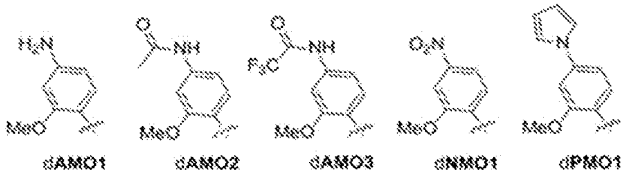

R BEN
R MM
R MM2
R MM3
R DM
R DM2
R DM3
R DM4
R DM5
R TM
R TM2
R TM3
dPICS
ICS
3MN
7AI
2Br
3Br
4Br
2FB
3FB
2CN
3CN
4CN
dT
dH
dF
dL
dB
dI
ICS
3MN
7AI
BEN
DM5
TM
2Br
3Br
4Br
2CN
3CN
4CN
2FB
3FB
MM1
MM2
MM3
2-pyrimidinone dZeb
2-pyridone 2OPy
3-deazaadenine 3DA
6-aminopyridin-3-yl 6AmPy
6-chloropyridin-3yl 6ClPy
6-methylpyridin-3-yl 6MePy
6-oxopyridin-3-yl 6OPy

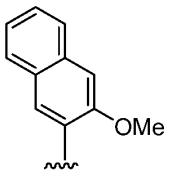
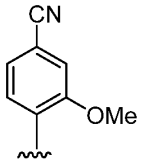
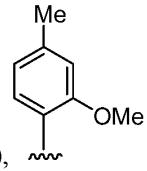


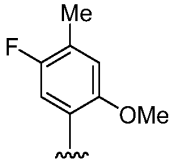
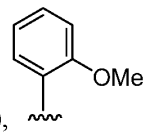
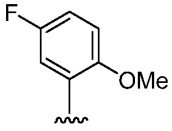
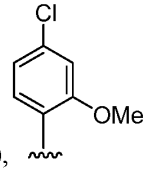


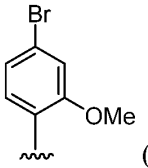


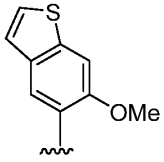


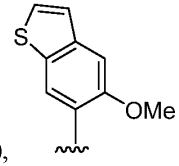


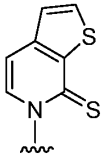
[201]  (NaM),  (CNMO),  (MMO2),

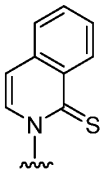
 (5FM),  (2OMe),  (5F2OMe),  (ClMO),

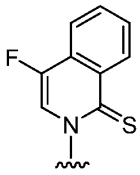


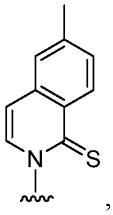
(BrMO),


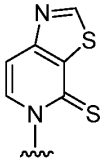
(PTMO),


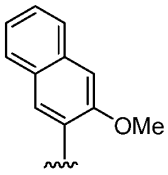
(MTMO),


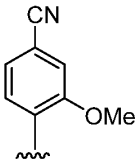
(TPT3),


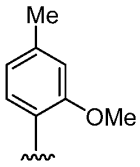
(SICS),


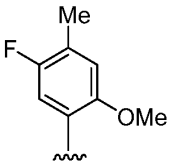
(FSICS),


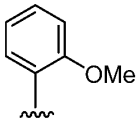
,


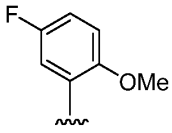
(TAT1),


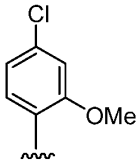
(dNaM),


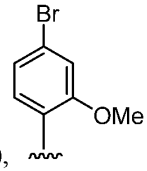
(dCNMO),


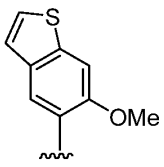
(dMMO2),


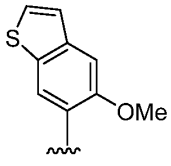
(d5FM),


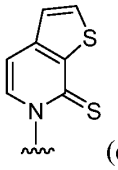
(d2OMe),


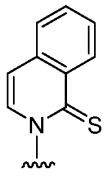
(d5F2OMe),


(dClMO),


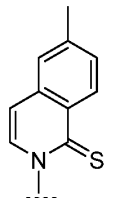
(dBrMO),


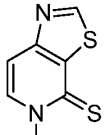
(dPTMO),


(dMTMO),


(dTPT3),


(dSICS),

(dFSICS),
,
and


(dTAT1).

[202] In some embodiments, nucleotide analogs are also modified at the phosphate moiety. Modified phosphate moieties include, but are not limited to, those with modification at the linkage between two nucleotides and contains, for example, a phosphorothioate, chiral phosphorothioate, phosphorodithioate, phosphotriester, aminoalkylphosphotriester, methyl and other alkyl phosphonates including 3'-alkylene phosphonate and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates. It is understood that these phosphate or modified phosphate linkage between two nucleotides are through a 3'-5' linkage or a 2'-5'

linkage, and the linkage contains inverted polarity such as 3'-5' to 5'-3' or 2'-5' to 5'-2'. Various salts, mixed salts and free acid forms are also included. Numerous United States patents teach how to make and use nucleotides containing modified phosphates and include but are not limited to, 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050.

[203] In some embodiments, unnatural nucleic acids include 5'-substituted DNA and RNA derivatives (PCT/US2011/033961; Saha et al., J. Org Chem., 1995, 60, 788-789; Wang et al., Bioorganic & Medicinal Chemistry Letters, 1999, 9, 885-890; and Mikhailov et al., Nucleosides & Nucleotides, 1991, 10(1-3), 339-343; Leonid et al., 1995, 14(3-5), 901-905; and Eppacher et al., Helvetica Chimica Acta, 2004, 87, 3004-3020; PCT/JP2000/004720; PCT/JP2003/002342; PCT/JP2004/013216; PCT/JP2005/020435; PCT/JP2006/315479; PCT/JP2006/324484; PCT/JP2009/056718; PCT/JP2010/067560), or 5'-substituted monomers made as the monophosphate with modified bases (Wang et al., Nucleosides Nucleotides & Nucleic Acids, 2004, 23 (1 & 2), 317-337).

[204] In some embodiments, unnatural nucleic acids include modifications at the 5'-position and the 2'-position of the sugar ring (PCT/US94/02993), such as 5'-CH_2-substituted 2'-O-protected nucleosides (Wu et al., Helvetica Chimica Acta, 2000, 83, 1127-1143 and Wu et al., Bioconjugate Chem. 1999, 10, 921-924). In some cases, unnatural nucleic acids include amide linked nucleoside dimers have been prepared for incorporation into oligonucleotides wherein the 3' linked nucleoside in the dimer (5' to 3') comprises a 2'-OCH_3 and a 5'-(S)-CH_3 (Mesmaeker et al., Synlett, 1997, 1287-1290). Unnatural nucleic acids can include 2'-substituted 5'-CH_2 (or O) modified nucleosides (PCT/US92/01020). Unnatural nucleic acids can include 5'-methylenephosphonate DNA and RNA monomers, and dimers (Bohringer et al., Tet. Lett., 1993, 34, 2723-2726; Collingwood et al., Synlett, 1995, 7, 703-705; and Hutter et al., Helvetica Chimica Acta, 2002, 85, 2777-2806). Unnatural nucleic acids can include 5'-phosphonate monomers having a 2'-substitution (US2006/0074035) and other modified 5'-phosphonate monomers (WO1997/35869). Unnatural nucleic acids can include 5'-modified methylenephosphonate monomers (EP614907 and EP629633). Unnatural nucleic acids can include analogs of 5' or 6'-phosphonate ribonucleosides comprising a hydroxyl group at the 5' and/or 6'-position (Chen et al., Phosphorus, Sulfur and Silicon, 2002, 777, 1783-1786; Jung et al., Bioorg. Med. Chem., 2000, 8, 2501-2509; Gallier et al., Eur. J. Org. Chem., 2007, 925-933; and Hampton et al., J. Med. Chem., 1976, 19(8), 1029-1033). Unnatural nucleic acids can

include 5'-phosphonate deoxyribonucleoside monomers and dimers having a 5'-phosphate group (Nawrot et al., Oligonucleotides, 2006, 16(1), 68-82). Unnatural nucleic acids can include nucleosides having a 6'-phosphonate group wherein the 5' or/and 6'-position is unsubstituted or substituted with a thio-tert-butyl group ($SC(CH_3)_3$) (and analogs thereof); a methyleneamino group (CH_2NH_2) (and analogs thereof) or a cyano group (CN) (and analogs thereof) (Fairhurst et al., Synlett, 2001, 4, 467-472; Kappler et al., J. Med. Chem., 1986, 29, 1030-1038; Kappler et al., J. Med. Chem., 1982, 25, 1179-1184; Vrudhula et al., J. Med. Chem., 1987, 30, 888-894; Hampton et al., J. Med. Chem., 1976, 19, 1371-1377; Geze et al., J. Am. Chem. Soc, 1983, 105(26), 7638-7640; and Hampton et al., J. Am. Chem. Soc, 1973, 95(13), 4404-4414).

[205] In some embodiments, unnatural nucleic acids also include modifications of the sugar moiety. In some embodiments, the unnatural base pairs comprise at least one unnatural nucleotide comprising an unnatural sugar moiety. In some cases, nucleic acids contain one or more nucleosides wherein the sugar group has been modified. Such sugar modified nucleosides may impart enhanced nuclease stability, increased binding affinity, or some other beneficial biological property. In certain embodiments, nucleic acids comprise a chemically modified ribofuranose ring moiety. Examples of chemically modified ribofuranose rings include, without limitation, addition of substituent groups (including 5' and/or 2' substituent groups; bridging of two ring atoms to form bicyclic nucleic acids (BNA); replacement of the ribosyl ring oxygen atom with S, N(R), or $C(R_1)(R_2)$ (R = H, C_1-C_{12} alkyl or a protecting group); and combinations thereof. Examples of chemically modified sugars can be found in WO2008/101157, US2005/0130923, and WO2007/134181.

[206] In some instances, a modified nucleic acid comprises modified sugars or sugar analogs. Thus, in addition to ribose and deoxyribose, the sugar moiety can be pentose, deoxypentose, hexose, deoxyhexose, glucose, arabinose, xylose, lyxose, or a sugar "analog" cyclopentyl group. The sugar can be in a pyranosyl or furanosyl form. The sugar moiety may be the furanoside of ribose, deoxyribose, arabinose or 2'-O-alkylribose, and the sugar can be attached to the respective heterocyclic bases either in [alpha] or [beta] anomeric configuration. Sugar modifications include, but are not limited to, 2'-alkoxy-RNA analogs, 2'-amino-RNA analogs, 2'-fluoro-DNA, and 2'-alkoxy- or amino-RNA/DNA chimeras. For example, a sugar modification may include 2'-O-methyl-uridine or 2'-O-methyl-cytidine. Sugar modifications include 2'-O-alkyl-substituted deoxyribonucleosides and 2'-O-ethyleneglycol like ribonucleosides. The preparation of these sugars or sugar analogs and the respective "nucleosides" wherein such sugars or analogs are attached to a heterocyclic base (nucleic acid

base) is known. Sugar modifications may also be made and combined with other modifications.

[207] Modifications to the sugar moiety include natural modifications of the ribose and deoxy ribose as well as unnatural modifications. Sugar modifications include, but are not limited to, the following modifications at the 2' position: OH; F; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; O-, S- or N-alkynyl; or O-alkyl-O-alkyl, wherein the alkyl, alkenyl and alkynyl may be substituted or unsubstituted C_1 to C_{10}, alkyl or C_2 to C_{10} alkenyl and alkynyl. 2' sugar modifications also include but are not limited to $-O[(CH_2)_nO]_m CH_3$, $-O(CH_2)_nOCH_3$, $-O(CH_2)_nNH_2$, $-O(CH_2)_nCH_3$, $-O(CH_2)_nONH_2$, and $-O(CH_2)_nON[(CH_2)n\ CH_3)]_2$, where n and m are from 1 to about 10.

[208] Other modifications at the 2' position include but are not limited to: C_1 to C_{10} lower alkyl, substituted lower alkyl, alkaryl, aralkyl, O-alkaryl, O-aralkyl, SH, SCH_3, OCN, Cl, Br, CN, CF_3, OCF_3, $SOCH_3$, $SO_2\ CH_3$, ONO_2, NO_2, N_3, NH_2, heterocycloalkyl, heterocycloalkaryl, aminoalkylamino, polyalkylamino, substituted silyl, an RNA cleaving group, a reporter group, an intercalator, a group for improving the pharmacokinetic properties of an oligonucleotide, or a group for improving the pharmacodynamic properties of an oligonucleotide, and other substituents having similar properties. Similar modifications may also be made at other positions on the sugar, particularly the 3' position of the sugar on the 3' terminal nucleotide or in 2'-5' linked oligonucleotides and the 5' position of the 5' terminal nucleotide. Modified sugars also include those that contain modifications at the bridging ring oxygen, such as CH_2 and S. Nucleotide sugar analogs may also have sugar mimetics such as cyclobutyl moieties in place of the pentofuranosyl sugar. There are numerous United States patents that teach the preparation of such modified sugar structures and which detail and describe a range of base modifications, such as U.S. Patent Nos. 4,981,957; 5,118,800; 5,319,080; 5,359,044; 5,393,878; 5,446,137; 5,466,786; 5,514,785; 5,519,134; 5,567,811; 5,576,427; 5,591,722; 5,597,909; 5,610,300; 5,627,053; 5,639,873; 5,646,265; 5,658,873; 5,670,633; 4,845,205; 5,130,302; 5,134,066; 5,175,273; 5,367,066; 5,432,272; 5,457,187; 5,459,255; 5,484,908; 5,502,177; 5,525,711; 5,552,540; 5,587,469; 5,594,121, 5,596,091; 5,614,617; 5,681,941; and 5,700,920, each of which is herein incorporated by reference in its entirety.

[209] Examples of nucleic acids having modified sugar moieties include, without limitation, nucleic acids comprising 5'-vinyl, 5'-methyl (R or S), 4'-S, 2'-F, 2'-OCH_3, and 2'-$O(CH_2)_2OCH_3$ substituent groups. The substituent at the 2' position can also be selected from allyl, amino, azido, thio, O-allyl, O-(C_1-C_{10} alkyl), OCF_3, $O(CH_2)_2SCH_3$, $O(CH_2)_2$-O-

$N(R_m)(R_n)$, and $O\text{-}CH_2\text{-}C(=O)\text{-}N(R_m)(R_n)$, where each R_m and R_n is, independently, H or substituted or unsubstituted C_1-C_{10} alkyl.

[210] In some embodiments, the unnatural base pairs comprise at least one unnatural sugar moiety, wherein the unnatural sugar moiety comprises any one of the following moieties: (i) a modification at the 2' position comprising: OH, substituted lower alkyl, alkaryl, aralkyl, O-alkaryl or O-aralkyl, SH, SCH_3, OCN, Cl, Br, CN, CF_3, OCF_3, $SOCH_3$, SO_2CH_3, ONO_2, NO_2, N_3,or NH_2F; O-alkyl, S-alkyl, or N-alkyl; O-alkenyl, S-alkenyl, or N-alkenyl; O-alkynyl, S-alkynyl, or N-alkynyl; O-alkyl-O-alkyl, 2'-F, 2'-OCH_3, or 2'-$O(CH_2)_2OCH_3$, wherein the alkyl, alkenyl and alkynyl may be substituted or unsubstituted C_1-C_{10}, alkyl, C_2-C_{10} alkenyl, C_2-C_{10} alkynyl, $-O[(CH_2)_nO]_mCH_3$, $-O(CH_2)_nOCH_3$, $-O(CH_2)_nNH_2$, $-O(CH_2)_nCH_3$, $-O(CH_2)_n\text{-}NH_2$, or$-O(CH_2)_nON[(CH_2)_nCH_3)]_2$, wherein n and m are from 1 to about 10; (ii) a modification at the 5' position comprising: 5'-vinyl, or 5'-methyl (R or S); (iii) a modification at the 4' position, 4'-S, heterocycloalkyl, heterocycloalkaryl, aminoalkylamino, polyalkylamino, substituted silyl, an RNA cleaving group, a reporter group, an intercalator, a group for improving the pharmacokinetic properties of an oligonucleotide, or a group for improving the pharmacodynamic properties of an oligonucleotide; or (iv) any combination thereof.

[211] In certain embodiments, nucleic acids described herein include one or more bicyclic nucleic acids. In certain such embodiments, the bicyclic nucleic acid comprises a bridge between the 4' and the 2' ribosyl ring atoms. In certain embodiments, nucleic acids provided herein include one or more bicyclic nucleic acids wherein the bridge comprises a 4' to 2' bicyclic nucleic acid. Examples of such 4' to 2' bicyclic nucleic acids include, but are not limited to, one of the formulae: 4'-(CH_2)-O-2' (LNA); 4'-(CH_2)-S-2'; 4'-$(CH_2)_2$-O-2' (ENA); 4'-$CH(CH_3)$-O-2' and 4'-$CH(CH_2OCH_3)$-O-2', and analogs thereof (see, U.S. Patent No. 7,399,845); 4'-$C(CH_3)(CH_3)$-O-2'and analogs thereof, (see WO2009/006478, WO2008/150729, US2004/0171570, U.S. Patent No. 7,427,672, Chattopadhyaya et al., J. Org. Chem., 209, 74, 118-134, and WO2008/154401). Also see, for example: Singh et al., Chem. Commun., 1998, 4, 455-456; Koshkin et al., Tetrahedron, 1998, 54, 3607-3630; Wahlestedt et al., Proc. Natl. Acad. Sci. U. S. A., 2000, 97, 5633-5638; Kumar et al., Bioorg. Med. Chem. Lett., 1998, 8, 2219-2222; Singh et al., J. Org. Chem., 1998, 63, 10035-10039; Srivastava et al., J. Am. Chem. Soc., 2007, 129(26) 8362-8379; Elayadi et al., Curr. Opinion Invens. Drugs, 2001, 2, 558-561; Braasch et al., Chem. Biol, 2001, 8, 1-7; Oram et al., Curr. Opinion Mol. Ther., 2001, 3, 239-243; U.S. Patent Nos. 4,849,513; 5,015,733; 5,118,800; 5,118,802; 7,053,207; 6,268,490; 6,770,748; 6,794,499; 7,034,133; 6,525,191; 6,670,461; and 7,399,845;

International Publication Nos. WO2004/106356, WO1994/14226, WO2005/021570, WO2007/090071, and WO2007/134181; U.S. Patent Publication Nos. US2004/0171570, US2007/0287831, and US2008/0039618; U.S. Provisional Application Nos. 60/989,574, 61/026,995, 61/026,998, 61/056,564, 61/086,231, 61/097,787, and 61/099,844; and International Applications Nos. PCT/US2008/064591, PCT US2008/066154, PCT US2008/068922, and PCT/DK98/00393.

[212] In certain embodiments, nucleic acids comprise linked nucleic acids. Nucleic acids can be linked together using any inter nucleic acid linkage. The two main classes of inter nucleic acid linking groups are defined by the presence or absence of a phosphorus atom. Representative phosphorus containing inter nucleic acid linkages include, but are not limited to, phosphodiesters, phosphotriesters, methylphosphonates, phosphoramidate, and phosphorothioates (P=S). Representative non-phosphorus containing inter nucleic acid linking groups include, but are not limited to, methylenemethylimino (-CH_2-N(CH_3)-O-CH_2-), thiodiester (-O-C(O)-S-), thionocarbamate (-O-C(O)(NH)-S-); siloxane (-O-Si(H)$_2$-O-); and N,N*-dimethylhydrazine (-CH_2-N(CH_3)-N(CH_3)). In certain embodiments, inter nucleic acids linkages having a chiral atom can be prepared as a racemic mixture, as separate enantiomers, e.g., alkylphosphonates and phosphorothioates. Unnatural nucleic acids can contain a single modification. Unnatural nucleic acids can contain multiple modifications within one of the moieties or between different moieties.

[213] Backbone phosphate modifications to nucleic acid include, but are not limited to, methyl phosphonate, phosphorothioate, phosphoramidate (bridging or non-bridging), phosphotriester, phosphorodithioate, phosphodithioate, and boranophosphate, and may be used in any combination. Other non- phosphate linkages may also be used.

[214] In some embodiments, backbone modifications (e.g., methylphosphonate, phosphorothioate, phosphoroamidate and phosphorodithioate internucleotide linkages) can confer immunomodulatory activity on the modified nucleic acid and/or enhance their stability *in vivo*.

[215] In some instances, a phosphorous derivative (or modified phosphate group) is attached to the sugar or sugar analog moiety in and can be a monophosphate, diphosphate, triphosphate, alkylphosphonate, phosphorothioate, phosphorodithioate, phosphoramidate or the like. Exemplary polynucleotides containing modified phosphate linkages or non-phosphate linkages can be found in Peyrottes et al., 1996, Nucleic Acids Res. 24: 1841-1848; Chaturvedi et al., 1996, Nucleic Acids Res. 24:2318-2323; and Schultz et al., (1996) Nucleic Acids Res. 24:2966-2973; Matteucci, 1997, “Oligonucleotide Analogs: an Overview” in Oligonucleotides

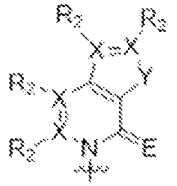
as Therapeutic Agents, (Chadwick and Cardew, ed.) John Wiley and Sons, New York, NY; Zon, 1993, "Oligonucleoside Phosphorothioates" in Protocols for Oligonucleotides and Analogs, Synthesis and Properties, Humana Press, pp. 165-190; Miller et al., 1971, JACS 93:6657-6665; Jager et al., 1988, Biochem. 27:7247-7246; Nelson et al., 1997, JOC 62:7278-7287; U.S. Patent No. 5,453,496; and Micklefield, 2001, Curr. Med. Chem. 8: 1157-1179.

[216] In some cases, backbone modification comprises replacing the phosphodiester linkage with an alternative moiety such as an anionic, neutral or cationic group. Examples of such modifications include: anionic internucleoside linkage; N3' to P5' phosphoramidate modification; boranophosphate DNA; prooligonucleotides; neutral internucleoside linkages such as methylphosphonates; amide linked DNA; methylene(methylimino) linkages; formacetal and thioformacetal linkages; backbones containing sulfonyl groups; morpholino oligos; peptide nucleic acids (PNA); and positively charged deoxyribonucleic guanidine (DNG) oligos (Micklefield, 2001, Current Medicinal Chemistry 8: 1157-1179). A modified nucleic acid may comprise a chimeric or mixed backbone comprising one or more modifications, e.g. a combination of phosphate linkages such as a combination of phosphodiester and phosphorothioate linkages.

[217] Substitutes for the phosphate include, for example, short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH_2 component parts. Numerous United States patents disclose how to make and use these types of phosphate replacements and include but are not limited to U.S. Patent Nos. 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437; and 5,677,439. It is also understood in a nucleotide substitute that both the sugar and the phosphate moieties of the nucleotide can be replaced, by for example an amide type linkage (aminoethylglycine) (PNA). United States Patent Nos. 5,539,082; 5,714,331; and 5,719,262 teach how to make and use PNA molecules, each of which is herein incorporated by reference. See also Nielsen et al., Science, 1991, 254, 1497-1500. It is also

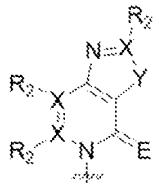
possible to link other types of molecules (conjugates) to nucleotides or nucleotide analogs to enhance for example, cellular uptake. Conjugates can be chemically linked to the nucleotide or nucleotide analogs. Such conjugates include but are not limited to lipid moieties such as a cholesterol moiety (Letsinger et al., Proc. Natl. Acad. Sci. USA, 1989, 86, 6553-6556), cholic acid (Manoharan et al., Bioorg. Med. Chem. Let., 1994, 4, 1053-1060), a thioether, e.g., hexyl-S-tritylthiol (Manoharan et al., Ann. KY. Acad. Sci., 1992, 660, 306-309; Manoharan et al., Bioorg. Med. Chem. Let., 1993, 3, 2765-2770), a thiocholesterol (Oberhauser et al., Nucl. Acids Res., 1992, 20, 533-538), an aliphatic chain, e.g., dodecandiol or undecyl residues (Saison-Behmoaras et al., EM5OJ, 1991, 10, 1111-1118; Kabanov et al., FEBS Lett., 1990, 259, 327-330; Svinarchuk et al., Biochimie, 1993, 75, 49-54), a phospholipid, e.g., di-hexadecyl-rac-glycerol or triethylammonium l-di-O-hexadecyl-rac-glycero-S-H-phosphonate (Manoharan et al., Tetrahedron Lett., 1995, 36, 3651-3654; Shea et al., Nucl. Acids Res., 1990, 18, 3777-3783), a polyamine or a polyethylene glycol chain (Manoharan et al., Nucleosides & Nucleotides, 1995, 14, 969-973), or adamantane acetic acid (Manoharan et al., Tetrahedron Lett., 1995, 36, 3651-3654), a palmityl moiety (Mishra et al., Biochem. Biophys. Acta, 1995, 1264, 229-237), or an octadecylamine or hexylamino-carbonyl-oxycholesterol moiety (Crooke et al., J. Pharmacol. Exp. Ther., 1996, 277, 923-937). Numerous United States patents teach the preparation of such conjugates and include, but are not limited to U.S. Patent Nos. 4,828,979; 4,948,882; 5,218,105; 5,525,465; 5,541,313; 5,545,730; 5,552,538; 5,578,717, 5,580,731; 5,580,731; 5,591,584; 5,109,124; 5,118,802; 5,138,045; 5,414,077; 5,486,603; 5,512,439; 5,578,718; 5,608,046; 4,587,044; 4,605,735; 4,667,025; 4,762,779; 4,789,737; 4,824,941; 4,835,263; 4,876,335; 4,904,582; 4,958,013; 5,082,830; 5,112,963; 5,214,136; 5,082,830; 5,112,963; 5,214,136; 5,245,022; 5,254,469; 5,258,506; 5,262,536; 5,272,250; 5,292,873; 5,317,098; 5,371,241, 5,391,723; 5,416,203, 5,451,463; 5,510,475; 5,512,667; 5,514,785; 5,565,552; 5,567,810; 5,574,142; 5,585,481; 5,587,371; 5,595,726; 5,597,696; 5,599,923; 5,599,928 and 5,688,941.

[218] Described herein are nucleobases used in the compositions and methods for replication, transcription, translation, and incorporation of unnatural amino acids into proteins.



In some embodiments, a nucleobase described herein comprises the structure:

wherein each X is independently carbon or nitrogen; R_2 is optional and when present is independently hydrogen, alkyl, alkenyl, alkynyl; methoxy, methanethiol, methaneseleno,

halogen, cyano, or azide group; wherein each Y is independently sulfur, oxygen, selenium, or secondary amine; wherein each E is independently oxygen, sulfur or selenium; and 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, connected to a mono-phosphate, diphosphate, or triphosphate group, optionally comprising an 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. In some embodiments, R_2 is lower alkyl (e.g., C_1-C_6), hydrogen, or halogen. In some embodiments of a nucleobase described herein, R_2 is fluoro. In some embodiments of a nucleobase described herein, X is carbon. In some embodiments of a nucleobase described herein, E is sulfur. In some embodiments of a nucleobase described herein, Y is sulfur. In some embodiments of a nucleobase described herein, a nucleobase has the structure: . In some embodiments of a nucleobase described herein, E is sulfur and Y is sulfur. In some embodiments of a nucleobase described herein, the wavy line indicates a point of bonding to a ribosyl or deoxyribosyl moiety. In some embodiments of a nucleobase described herein, the wavy line indicates a point of bonding to a ribosyl or deoxyribosyl moiety, connected to a triphosphate group. In some embodiments of a nucleobase described herein is a component of a nucleic acid polymer. In some embodiments of a nucleobase described herein, the nucleobase is a component of a tRNA. In some embodiments of a nucleobase described herein, the nucleobase is a component of an anticodon in a tRNA. In some embodiments of a nucleobase described herein, the nucleobase is a component of an mRNA. In some embodiments of a nucleobase described herein, the nucleobase is a component of a codon of an mRNA. In some embodiments of a nucleobase described herein, the nucleobase is a component of RNA or DNA. In some embodiments of a nucleobase described herein, the nucleobase is a component of a codon in DNA. In some embodiments of a nucleobase described herein, the nucleobase forms a nucleobase pair with another complementary nucleobase.

[219] In some embodiments, X and Y are unnatural bases, and X and Y are independently selected from any one of: (i) 2-thiouracil, 2'-deoxyuridine, 4-thio-uracil, uracil-5-yl, hypoxanthin-9-yl (I), 5-halouracil; 5-propynyl-uracil, 6-azo-uracil, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, pseudouracil, uracil-5-oxacetic acid methylester, uracil-5-oxacetic acid, 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-

carboxypropyl) uracil, 5-methyl-2-thiouracil, 4-thiouracil, 5-methyluracil, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, uracil-5-oxyacetic acid, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, or dihydrouracil; (ii) 5-hydroxymethyl cytosine, 5-trifluoromethyl cytosine, 5-halocytosine, 5-propynyl cytosine, 5-hydroxycytosine, cyclocytosine, cytosine arabinoside, 5,6-dihydrocytosine, 5-nitrocytosine, 6-azo cytosine, azacytosine, N4-ethylcytosine, 3-methylcytosine, 5-methylcytosine, 4-acetylcytosine, 2-thiocytosine, phenoxazine cytidine([5,4-b][l,4]benzoxazin-2(3H)-one), phenothiazine cytidine (1H-pyrimido[5,4-b][l, 4]benzothiazin-2(3H)-one), phenoxazine cytidine (9-(2-aminoethoxy)-H-pyrimido[5,4-b][l,4]benzoxazin-2(3H)-one), carbazole cytidine (2H-pyrimido[4,5- b]indol-2-one), or pyridoindole cytidine (H-pyrido [3',2':4,5]pyrrolo [2,3-d]pyrimidin-2-one); (iii) 2-aminoadenine, 2-propyl adenine, 2-amino-adenine, 2-F-adenine, 2-amino-propyl-adenine, 2-amino-2'-deoxyadenosine, 3-deazaadenine, 7-methyladenine, 7-deaza-adenine, 8-azaadenine, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, and 8-hydroxyl substituted adenines, N6-isopentenyladenine, 2-methyladenine, 2,6-diaminopurine, 2-methythio-N6-isopentenyladenine, or 6-aza-adenine; (iv) 2-methylguanine, 2-propyl and alkyl derivatives of guanine, 3-deazaguanine, 6-thio-guanine, 7-methylguanine, 7-deazaguanine, 7-deazaguanosine, 7-deaza-8-azaguanine, 8-azaguanine, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, and 8-hydroxyl substituted guanines, 1-methylguanine, 2,2-dimethylguanine, 7-methylguanine, or 6-aza-guanine; and (v) hypoxanthine, xanthine, 1-methylinosine, queosine, beta-D-galactosylqueosine, inosine, beta-D-mannosylqueosine, wybutoxosine, hydroxyurea, (acp3)w, 2-aminopyridine, or 2-pyridone.

[220] In some embodiments, at least one unnatural nucleotide base is recognized by an RNA polymerase during transcription.

VIII. Nucleic Acid Base Pairing Properties

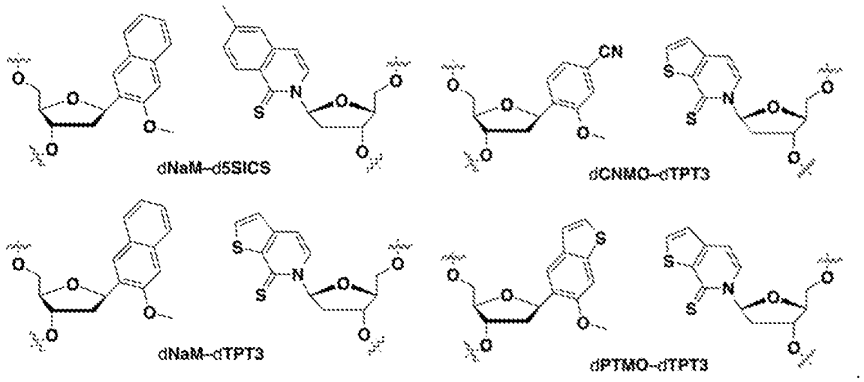
[221] In some embodiments, an unnatural nucleotide forms a base pair (an unnatural base pair; UBP) with another unnatural nucleotide during or after incorporation into DNA or RNA. In some embodiments, a stably integrated unnatural nucleic acid is an unnatural nucleic acid that can form a base pair with another nucleic acid, e.g., a natural or unnatural nucleic acid. In some embodiments, a stably integrated unnatural nucleic acid is an unnatural nucleic acid that can form a base pair with another unnatural nucleic acid (unnatural nucleic acid base pair (UBP)). For example, a first unnatural nucleic acid can form a base pair with a second unnatural nucleic acid. For example, one pair of unnatural nucleoside triphosphates that can base pair during and after incorporation into nucleic acids include a triphosphate of (d)5SICS

((d)5SICSTP) and a triphosphate of (d)NaM ((d)NaMTP). Other examples include but are not limited to: a triphosphate of (d)CNMO ((d)CNMOTP) and a triphosphate of (d)TPT3 ((d)TPT3TP). As another example, unnatural nucleoside triphosphates may self-pair, for example, (d)NaM can pair with (d)NaM and (d)TPT3 can pair with (d)TPT3. Such unnatural nucleotides can have a ribose or deoxyribose sugar moiety (indicated by the "(d)"). For example, one pair of unnatural nucleoside triphosphates that can base pair when incorporated into nucleic acids includes a triphosphate of TAT1 (TAT1TP) and a triphosphate of NaM (NaMTP). In some embodiments, one pair of unnatural nucleoside triphosphates that can base pair when incorporated into nucleic acids includes a triphosphate of dCNMO (dCNMOTP) and a triphosphate of TAT1 (TAT1TP). In some embodiments, one pair of unnatural nucleoside triphosphates that can base pair when incorporated into nucleic acids includes a triphosphate of dTPT3 (dTPT3TP) and a triphosphate of NaM (NaMTP). In some embodiments, an unnatural nucleic acid does not substantially form a base pair with a natural nucleic acid (A, T, G, C). In some embodiments, a stably integrated unnatural nucleic acid can form a base pair with a natural nucleic acid.

[222] In some embodiments, a stably integrated unnatural (deoxy)ribonucleotide is an unnatural (deoxy)ribonucleotide that can form a UBP but does not substantially form a base pair with each any of the natural (deoxy)ribonucleotides. In some embodiments, a stably integrated unnatural (deoxy)ribonucleotide is an unnatural (deoxy)ribonucleotide that can form a UBP but does not substantially form a base pair with one or more natural nucleic acids. For example, a stably integrated unnatural nucleic acid may not substantially form a base pair with A, T, and, C, but can form a base pair with G. For example, a stably integrated unnatural nucleic acid may not substantially form a base pair with A, T, and, G, but can form a base pair with C. For example, a stably integrated unnatural nucleic acid may not substantially form a base pair with C, G, and, A, but can form a base pair with T. For example, a stably integrated unnatural nucleic acid may not substantially form a base pair with C, G, and, T, but can form a base pair with A. For example, a stably integrated unnatural nucleic acid may not substantially form a base pair with A and T, but can form a base pair with C and G. For example, a stably integrated unnatural nucleic acid may not substantially form a base pair with A and C, but can form a base pair with T and G. For example, a stably integrated unnatural nucleic acid may not substantially form a base pair with A and G, but can form a base pair with C and T. For example, a stably integrated unnatural nucleic acid may not substantially form a base pair with C and T, but can form a base pair with A and G. For example, a stably integrated unnatural nucleic acid may not substantially form a base pair with C and G, but can form a base pair with

T and G. For example, a stably integrated unnatural nucleic acid may not substantially form a base pair with T and G, but can form a base pair with A and G. For example, a stably integrated unnatural nucleic acid may not substantially form a base pair with, G, but can form a base pair with A, T, and, C. For example, a stably integrated unnatural nucleic acid may not substantially form a base pair with, A, but can form a base pair with G, T, and, C. For example, a stably integrated unnatural nucleic acid may not substantially form a base pair with, T, but can form a base pair with G, A, and, C. For example, a stably integrated unnatural nucleic acid may not substantially form a base pair with, C, but can form a base pair with G, T, and, A.

[223] Exemplary unnatural nucleotides capable of forming an unnatural DNA or RNA base pair (UBP) under conditions *in vivo* includes, but is not limited to, 5SICS, d5SICS, NaM, dNaM, dTPT3, dMTMO, dCNMO, TAT1, and combinations thereof. In some embodiments, unnatural nucleotide base pairs include but are not limited to:



IX. Nucleoside Triphosphate Transporters

[224] Nucleosides are hydrophilic molecules which require transport proteins for permeation of cell membranes. Nucleoside transporters (NTs) are a group of membrane transport proteins that facilitate crossing of the nucleosides through cell membranes and vesicles. In some cases, there are two types of nucleoside transporters, concentrative nucleoside transporters which drives a concentrative process by electrochemical gradient, and equilibrative nucleoside transporters which drives an equilibrative bidirectional process by chemical gradient. In some instances, a nucleoside transporter further encompasses a nucleoside triphosphate transporter.

[225] Natural nucleosides comprise adenine, guanine, thymine, uracil, and cytosine; and are recognized by nucleotide transporters for permeation of cell membranes. Unnatural nucleosides, in some cases, are either not recognized by endogenous nucleotide transporters or are recognized but the efficiency of transport is low.

[226] In some embodiments, described herein are nucleotide transporters that recognize and facilitate transport of unnatural nucleic acids into a cell. In some instances, the nucleotide transporter enhances import of unnatural nucleic acids into a cell relative to an endogeneous nucleotide transporter. In some cases, the nucleotide transporter increases unnatural nucleic acid retention within a cell. In additional cases, the nucleotide transporter minimizes toxicity due to its expression, and optionally improves cell doubling time and fitness relative to a cell in the absence of the transporter.

[227] In certain embodiments, described herein are nucleoside triphosphate transporters for transporting unnatural nucleic acids into a cell. In some instances, the nucleoside triphosphate transporter is from *Phaeodactylum tricornutum*(PtNTT2). In some instances, the nucleoside triphosphate transporter further comprises a deletion. In some cases, the deletion is a terminal deletion (e.g., a N-terminal deletion or a C-terminal deletion) or is an internal deletion.

[228] In some embodiments, described herein is an isolated and nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprising a deletion. In some instances, the deletion comprises about 5, 10, 15, 20, 22, 25, 30, 40, 44, 50, 60, 66, 70, 80, 90, or more amino acid residues. In some instances, the deletion comprises about 5, 10, 15, 20, 22, 25, 30, 40, 44, 50, 60, 66, 70, or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a deletion of about 5 or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a deletion of about 10 or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a deletion of about 15 or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a deletion of about 20 or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a deletion of about 22 or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a deletion of about 25 or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a deletion of about 30 or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a deletion of about 40 or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a deletion of about 44 or more amino acid residues. In some cases, the nucleoside

triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a deletion of about 50 or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a deletion of about 60 or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a deletion of about 66 or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a deletion of about 70 or more amino acid residues.

[229] In some embodiments, described herein is an isolated and nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprising a N-terminal deletion. In some instances, the N-terminal deletion comprises about 5, 10, 15, 20, 22, 25, 30, 40, 44, 50, 60, 66, 70, 80, 90, or more amino acid residues. In some instances, the N-terminal deletion comprises about 5, 10, 15, 20, 22, 25, 30, 40, 44, 50, 60, 66, 70, or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a N-terminal deletion of about 5 or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a N-terminal deletion of about 10 or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a N-terminal deletion of about 15 or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a N-terminal deletion of about 20 or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a N-terminal deletion of about 22 or more amino acid residues. In some cases, the isolated and nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a N-terminal deletion of about 25 or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a N-terminal deletion of about 30 or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a N-terminal deletion of about 40 or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a N-terminal deletion of about 44 or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a N-terminal deletion of about 50 or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a N-terminal deletion of about 60 or more

amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a N-terminal deletion of about 66 or more amino acid residues. In some cases, the nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (PtNTT2) comprises a N-terminal deletion of about 70 or more amino acid residues.

[230] In some embodiments, the nucleoside triphosphate transporter comprises at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO: 9. In some instances, the nucleoside triphosphate transporter comprises at least 80% sequence identity to SEQ ID NO: 9. In some instances, the nucleoside triphosphate transporter comprises at least 85% sequence identity to SEQ ID NO: 9. In some instances, the nucleoside triphosphate transporter comprises at least 90% sequence identity to SEQ ID NO: 9. In some instances, the nucleoside triphosphate transporter comprises at least 95% sequence identity to SEQ ID NO: 9. In some instances, the nucleoside triphosphate transporter comprises at least 96% sequence identity to SEQ ID NO: 9. In some instances, the nucleoside triphosphate transporter comprises at least 97% sequence identity to SEQ ID NO: 9. In some instances, the nucleoside triphosphate transporter comprises at least 98% sequence identity to SEQ ID NO: 9. In some instances, the nucleoside triphosphate transporter comprises at least 99% sequence identity to SEQ ID NO: 9. In some instances, the nucleoside triphosphate transporter comprises 100% sequence identity to SEQ ID NO: 9. In some instances, the nucleoside triphosphate transporter consists of 100% sequence identity to SEQ ID NO: 9.

[231] In some embodiments, the nucleoside triphosphate transporter comprises at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO: 10. In some instances, the nucleoside triphosphate transporter comprises at least 80% sequence identity to SEQ ID NO: 10. In some instances, the nucleoside triphosphate transporter comprises at least 85% sequence identity to SEQ ID NO: 10. In some instances, the nucleoside triphosphate transporter comprises at least 90% sequence identity to SEQ ID NO: 10. In some instances, the nucleoside triphosphate transporter comprises at least 95% sequence identity to SEQ ID NO: 10. In some instances, the nucleoside triphosphate transporter comprises at least 96% sequence identity to SEQ ID NO: 10. In some instances, the nucleoside triphosphate transporter comprises at least 97% sequence identity to SEQ ID NO: 10. In some instances, the nucleoside triphosphate transporter comprises at least 98% sequence identity to SEQ ID NO: 10. In some instances, the nucleoside triphosphate transporter comprises at least 99% sequence identity to SEQ ID NO: 10. In some instances, the nucleoside triphosphate transporter

comprises 100% sequence identity to SEQ ID NO: 10. In some instances, the nucleoside triphosphate transporter consists of 100% sequence identity to SEQ ID NO: 10.

[232] In some embodiments, the nucleoside triphosphate transporter described herein is truncated and comprises amino acids 1-22 and 66-575 with reference to SEQ ID NO: 10. In some embodiments, the nucleoside triphosphate transporter described herein is truncated and comprises amino acids 23-575 with reference to SEQ ID NO: 10.

[233] In some embodiments, a nucleoside triphosphate transporter described herein has a specificity for an unnatural nucleic acid that is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% the specificity of the wild type nucleoside triphosphate transporter toward the unnatural nucleic acid. In some embodiments, the nucleoside triphosphate transporter has a specificity for an unnatural nucleic acid comprising a modified sugar that is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% the specificity of the wild type nucleoside triphosphate transporter toward a natural nucleic acid and/or the unnatural nucleic acid without the modified sugar. In some embodiments, the nucleoside triphosphate transporter has a specificity for an unnatural nucleic acid comprising a base that is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% the specificity of the wild type nucleoside triphosphate transporter toward a natural nucleic acid and/or the unnatural nucleic acid without the modified base. In some embodiments, the nucleoside triphosphate transporter has a specificity for an unnatural nucleic acid comprising a triphosphate that is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% the specificity of the wild type nucleoside triphosphate transporter toward a nucleic acid comprising a triphosphate and/or the unnatural nucleic acid without the triphosphate. For example, a nucleoside triphosphate transporter can have a specificity for an unnatural nucleic acid comprising a triphosphate that is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% the specificity of the wild type nucleoside triphosphate transporter toward the unnatural nucleic acid with a diphosphate or monophosphate, or no phosphate, or a combination thereof.

[234] In some embodiments, a nucleoside triphosphate transporter described herein has a specificity for an unnatural nucleic acid and a specificity to a natural nucleic acid that is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% the specificity of the wild type nucleoside triphosphate transporter toward the natural nucleic acid. In some embodiments, the nucleoside triphosphate transporter has a specificity for an unnatural nucleic acid comprising a modified sugar and a specificity to a

natural nucleic acid that is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% the specificity of the wild type nucleoside triphosphate transporter toward the natural nucleic acid. In some embodiments, the nucleoside triphosphate transporter has a specificity for an unnatural nucleic acid comprising a modified base and a specificity to a natural nucleic acid that is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% the specificity of the wild type nucleoside triphosphate transporter toward the natural nucleic acid.

[235] In some embodiments, a sequence of a nucleoside triphosphate transporter is further modified to improve the expression and cellular activity. In some instances, the codon usage is modified to introduce ribosomal pause sites to slow translation and to improve the targeting of the nucleoside triphosphate transporter polypeptide to membrane translocons (Fluman, et al., "mRNA-programmed translation pauses in the targeting of *E. coli* membrane proteins," eLife 2014; 3:e03440). In some instances, modification of one or more transmembrane helices, for example, modification of a first transmembrane helix and/or generating a chimeric transporter comprising a first transmembrane helix of a different protein (e.g., a related transporter) may enhance expression and cellular activities (Marshall, et al., "A link between integral membrane protein expression and simulated integration efficiency," Cell Reports, 16(8): 2169-2177 (2016)). In some instances, an endogenous, a modified, or a heterologous signal peptide is incorporated into the sequence of a nucleoside triphosphate transporter to improve expression and cellular activity. In some cases, the signal peptide is optionally linked in-frame with the sequence of the nucleoside triphosphate transporter through a linker. In some cases, the linker is a non-cleavable linker. In other cases, the linker is a cleavable linker.

[236] In some embodiments, the expression of the nucleoside triphosphate transporter is tuned through modification of the ribosomal binding site to modulate the rate of the nucleoside triphosphate transporter polypeptide's synthesis. See, *e.g.*, Howard, *et al.*, "Automated design of synthetic ribosome binding sites to control protein expression," *Nature Biotechnology* **27**: 946-950 (2009); and Mutalik, *et al.*, "Precise and reliable gene expression via standard transcription and translation initiation elements," *Nature Methods* **10**: 354-360 (2013).

[237] In some embodiments, the expression of the nucleoside triphosphate transporter is modulated by the attachment of a tunable degradation tag. In some instances, a tunable degradation tag comprises a small amino acid sequence that, when fused to a target protein, marks the protein for degradation by a cognate protease in a bacterial cell. Exemplary tunable

degradation tag and cognate protease pairs include, but are not limited to, *E. coli* ssrA (ec-ssrA)/*E. coli* Lon (ec-Lon), and *Mesoplasma florum* ssrA (mf-ssrA)/*Mesoplasma florum* Lon (mf-Lon). In some instances, the tunable degradation tag comprises a modified tag that alters expression and/or degradation dynamis relative to an unmodified degradation tag. In some instances, a tunable degradation tag contemplated herein comprises a degradation tag described in PCT Patent Publication WO2014/160025A2. In some instances, a tunable degradation tag contemplated herein comprises a degradation tag described in Cameron, *et al.*, "Tunable protein degradation in bacteria," *Nature Biotechnology* **32**: 1276-1281 (2014).

[238] In some embodiments, the expression of the nucleoside triphosphate transporter is modulated by the availability of an endogenous or exogenous (e.g. unnatural nucleotide triphosphate or unnatural amino acid) molecule during translation. In some instances, the expression of the nucleoside triphosphate transporter is correlated with the copy number of rare codons, in which the rate of a ribosomal read-through of a rare codon modulates translation of the transporter. See, e.g., Wang, *et al.*, "An engineered rare codon device for optimization of metabolic pathways," *Scientific Reports* **6**:20608 (2016).

[239] In some instances, a nucleoside triphosphate transporter is characterized according to its rate of dissociation from a nucleic acid substrate. In some embodiments, a nucleoside triphosphate transporter has a relatively low dissociation rate for one or more natural and unnatural nucleic acids. In some embodiments, a nucleoside triphosphate transporter has a relatively high dissociation rate for one or more natural and unnatural nucleic acids. The dissociation rate is an activity of a nucleoside triphosphate transporter that can be adjusted to tune reaction rates in methods set forth herein.

[240] Nucleoside triphosphate transporters from native sources or variants thereof can be screened using an assay that detects importation of an unnatural nucleic acid having a particular structure. In one example, the nucleoside triphosphate transporters can be screened for the ability to import an unnatural nucleic acid or UBP; e.g., d5SICSTP, dNaMTP, or d5SICSTP- dNaMTP UBP. A NTT, e.g., a heterologous transporter, can be used that displays a modified property for the unnatural nucleic acid as compared to the wild-type transporter. For example, the modified property can be, *e.g.*, K_m, k_{cat}, V_{max}, NTT importation in the presence of an unnatural nucleic acid (or of a naturally occurring nucleotide), average template read-length by a cell with the nucleoside triphosphate transporter in the presence of an unnatural nucleic acid, specificity of the transporter for an unnatural nucleic acid, rate of binding of an unnatural nucleic acid, or rate of product release, or any combination thereof. In one embodiment, the modified property is a reduced K_m for an unnatural nucleic acid and/or an

increased k_{cat}/K_m or V_{max}/K_m for an unnatural nucleic acid. Similarly, the nucleoside triphosphate transporter optionally has an increased rate of binding of an unnatural nucleic acid, an increased rate of product release, and/or an increased cell importation rate, as compared to a wild-type transporter.

[241] At the same time, a nucleoside triphosphate transporter can import natural nucleic acids, *e.g.*, A, C, G, and T, into cell. For example, a nucleoside triphosphate transporter optionally displays a specific importation activity for a natural nucleic acid that is at least about 5% as high (*e.g.*, 5%, 10%, 25%, 50%, 75%, 100% or higher), as a corresponding wild-type transporter. Optionally, the nucleoside triphosphate transporter displays a k_{cat}/K_m or V_{max}/K_m for a naturally occurring nucleotide that is at least about 5% as high (*e.g.*, about 5%, 10%, 25%, 50%, 75% or 100% or higher) as the wild-type NTT.

[242] Nucleoside triphosphate transporters used herein that can have the ability to import an unnatural nucleic acid of a particular structure can also be produced using a directed evolution approach. A nucleic acid synthesis assay can be used to screen for transporter variants having specificity for any of a variety of unnatural nucleic acids. For example, transporter variants can be screened for the ability to import an unnatural nucleic acid or UBP; e.g., d5SICSTP, dNaMTP, or d5SICSTP- dNaMTP UBP into nucleic acids. In some embodiments, such an assay is an in vitro assay, e.g., using a recombinant transporter variant. In some embodiments, such an assay is an in vivo assay, e.g., expressing a transporter variant in a cell. Such directed evolution techniques can be used to screen variants of any suitable transporter for activity toward any of the unnatural nucleic acids set forth herein.

[243] In some embodiments, described herein is an engineered cell comprising a nucleic acid molecule encoding a nucleoside triphosphate transporter. In some instances, the nucleic acid molecule encodes a nucleoside triphosphate transporter from *Phaeodactylum tricornutum* (*Pt*NTT2). In some instances, the nucleic acid of the nucleoside triphosphate transporter is incorporated in the genomic sequence of the engineered cell. In some embodiments, a *Pt*NTT2 is integrated into a chromosomal *lacZYA* locus, optionally replacing *lacZY*.

[244] In some instances, the engineered cell comprises a plasmid comprising the nucleoside triphosphate transporter. In some cases, the nucleoside triphosphate transporter is a codon optimized nucleoside triphosphate transporter from *Phaeodactylum tricornutum*.

[245] In some embodiments, also described herein is an isolated and purified plasmid comprising a nucleic acid molecule encoding a nucleoside triphosphate transporter from

Phaeodactylum tricornutum (*Pt*NTT2); and a promoter region selected from a pSC plasmid or lacZYA locus.

[246] In some embodiments, the nucleoside triphosphate transporter is under the control of a promoter. In some instances, the promoter is derived from an *E. coli* source. In other instances, the promoter is derived from a phage source. Exemplary promoters, include, but are not limited to, P_{bla}, P_{lac}, P_{lacUV5}, P_{H207}, P_{λ}, P_{tac}, or P_{N25}. In some instances, the promoter replaces the lac operon. In some cases, the nucleoside triphosphate transporter is under the control of a promoter selected from P_{bla}, P_{lac}, P_{lacUV5}, P_{H207}, P_{λ}, P_{tac}, or P_{N25}. In some cases, the nucleoside triphosphate transporter is under the control of promoter P_{lacUV5}. In some embodiments, expression of the nucleoside triphosphate transporter in a cell is modulated and/or increased by integrating a second copy of the expression cassette including the nucleoside triphosphate transporter under an IPTG inducible promoter. In some embodiments, expression of the nucleoside triphosphate transporter in a cell is modulated and/or increased by introducing the nucleoside triphosphate transporter with an expression stabilized promoter, optionally balanced with transcription activator-like effector (TALE) repression. In some embodiments, the nucleoside triphosphate transporter is *Pt*NTT2.

[247] In some instances, the nucleoside triphosphate transporter is encoded within a pSC plasmid.

X. Polymerases

[248] A particularly useful function of a polymerase is to catalyze the polymerization of a nucleic acid strand using an existing nucleic acid as a template. Other functions that are useful are described elsewhere herein. Examples of useful polymerases include DNA polymerases and RNA polymerases.

[249] The ability to improve specificity, processivity, or other features of polymerases unnatural nucleic acids would be highly desirable in a variety of contexts where, e.g., unnatural nucleic acid incorporation is desired, including amplification, sequencing, labeling, detection, cloning, and many others.

[250] In some instances, disclosed herein includes polymerases that incorporate unnatural nucleic acids into a growing template copy, e.g., during DNA amplification. In some embodiments, polymerases can be modified such that the active site of the polymerase is modified to reduce steric entry inhibition of the unnatural nucleic acid into the active site. In some embodiments, polymerases can be modified to provide complementarity with one or more unnatural features of the unnatural nucleic acids. Such polymerases can be expressed or

engineered in cells for stably incorporating a UBP into the cells. Accordingly, the present disclosure includes compositions that include a heterologous or recombinant polymerase and methods of use thereof.

[251] Polymerases can be modified using methods pertaining to protein engineering. For example, molecular modeling can be carried out based on crystal structures to identify the locations of the polymerases where mutations can be made to modify a target activity. A residue identified as a target for replacement can be replaced with a residue selected using energy minimization modeling, homology modeling, and/or conservative amino acid substitutions, such as described in Bordo, et al. J Mol Biol 217: 721-729 (1991) and Hayes, et al. Proc Natl Acad Sci, USA 99: 15926- 15931 (2002).

[252] Any of a variety of polymerases can be used in methods or compositions set forth herein including, for example, protein-based enzymes isolated from biological systems and functional variants thereof. Reference to a particular polymerase, such as those exemplified below, will be understood to include functional variants thereof unless indicated otherwise. In some embodiments, a polymerase is a wild type polymerase. In some embodiments, a polymerase is a modified, or mutant, polymerase.

[253] Polymerases, with features for improving entry of unnatural nucleic acids into active site regions and for coordinating with unnatural nucleotides in the active site region, can also be used. In some embodiments, a modified polymerase has a modified nucleotide binding site.

[254] In some embodiments, a modified polymerase has a specificity for an unnatural nucleic acid that is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% the specificity of the wild type polymerase toward the unnatural nucleic acid. In some embodiments, a modified or wild type polymerase has a specificity for an unnatural nucleic acid comprising a modified sugar that is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% the specificity of the wild type polymerase toward a natural nucleic acid and/or the unnatural nucleic acid without the modified sugar. In some embodiments, a modified or wild type polymerase has a specificity for an unnatural nucleic acid comprising a modified base that is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% the specificity of the wild type polymerase toward a natural nucleic acid and/or the unnatural nucleic acid without the modified base. In some embodiments, a modified or wild type polymerase has a specificity for an unnatural nucleic acid comprising a triphosphate that is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%,

99.5%, 99.99% the specificity of the wild type polymerase toward a nucleic acid comprising a triphosphate and/or the unnatural nucleic acid without the triphosphate. For example, a modified or wild type polymerase can have a specificity for an unnatural nucleic acid comprising a triphosphate that is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% the specificity of the wild type polymerase toward the unnatural nucleic acid with a diphosphate or monophosphate, or no phosphate, or a combination thereof.

[255] In some embodiments, a modified or wild type polymerase has a relaxed specificity for an unnatural nucleic acid. In some embodiments, a modified or wild type polymerase has a specificity for an unnatural nucleic acid and a specificity to a natural nucleic acid that is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% the specificity of the wild type polymerase toward the natural nucleic acid. In some embodiments, a modified or wild type polymerase has a specificity for an unnatural nucleic acid comprising a modified sugar and a specificity to a natural nucleic acid that is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% the specificity of the wild type polymerase toward the natural nucleic acid. In some embodiments, a modified or wild type polymerase has a specificity for an unnatural nucleic acid comprising a modified base and a specificity to a natural nucleic acid that is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% the specificity of the wild type polymerase toward the natural nucleic acid.

[256] Absence of exonuclease activity can be a wild type characteristic or a characteristic imparted by a variant or engineered polymerase. For example, an exo minus Klenow fragment is a mutated version of Klenow fragment that lacks 3’ to 5’ proofreading exonuclease activity.

[257] The methods of the present disclosure can be used to expand the substrate range of any DNA polymerase which lacks an intrinsic 3 to 5' exonuclease proofreading activity or where a 3 to 5' exonuclease proofreading activity has been disabled, e.g. through mutation. Examples of DNA polymerases include polA, polB (see e.g. Parrel & Loeb, Nature Struc Biol 2001) polC, polD, polY, polX and reverse transcriptases (RT) but preferably are processive, high-fidelity polymerases (PCT/GB2004/004643). In some embodiments a modified or wild type polymerase substantially lacks 3’ to 5’ proofreading exonuclease activity. In some embodiments a modified or wild type polymerase substantially lacks 3’ to 5’ proofreading exonuclease activity for an unnatural nucleic acid. In some embodiments, a modified or wild

type polymerase has a 3' to 5' proofreading exonuclease activity. In some embodiments, a modified or wild type polymerase has a 3' to 5' proofreading exonuclease activity for a natural nucleic acid and substantially lacks 3' to 5' proofreading exonuclease activity for an unnatural nucleic acid.

[258] In some embodiments, a modified polymerase has a 3' to 5' proofreading exonuclease activity that is at least about 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% the proofreading exonuclease activity of the wild type polymerase. In some embodiments, a modified polymerase has a 3' to 5' proofreading exonuclease activity for an unnatural nucleic acid that is at least about 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% the proofreading exonuclease activity of the wild type polymerase to a natural nucleic acid. In some embodiments, a modified polymerase has a 3' to 5' proofreading exonuclease activity for an unnatural nucleic acid and a 3' to 5' proofreading exonuclease activity for a natural nucleic acid that is at least about 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% the proofreading exonuclease activity of the wild type polymerase to a natural nucleic acid. In some embodiments, a modified polymerase has a 3' to 5' proofreading exonuclease activity for a natural nucleic acid that is at least about 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99.5%, 99.99% the proofreading exonuclease activity of the wild type polymerase to the natural nucleic acid.

[259] In some embodiments, polymerases are characterized according to their rate of dissociation from nucleic acids. In some embodiments a polymerase has a relatively low dissociation rate for one or more natural and unnatural nucleic acids. In some embodiments a polymerase has a relatively high dissociation rate for one or more natural and unnatural nucleic acids. The dissociation rate is an activity of a polymerase that can be adjusted to tune reaction rates in methods set forth herein.

[260] In some embodiments, polymerases are characterized according to their fidelity when used with a particular natural and/or unnatural nucleic acid or collections of natural and/or unnatural nucleic acid. Fidelity generally refers to the accuracy with which a polymerase incorporates correct nucleic acids into a growing nucleic acid chain when making a copy of a nucleic acid template. DNA polymerase fidelity can be measured as the ratio of correct to incorrect natural and unnatural nucleic acid incorporations when the natural and unnatural nucleic acid are present, e.g., at equal concentrations, to compete for strand synthesis at the same site in the polymerase-strand-template nucleic acid binary complex. DNA polymerase fidelity can be calculated as the ratio of (k_{cat}/K_m) for the natural and unnatural nucleic acid and (k_{cat}/K_m) for the incorrect natural and unnatural nucleic acid; where k_{cat} and

K_m are Michaelis-Menten parameters in steady state enzyme kinetics (Fersht, A. R. (1985) Enzyme Structure and Mechanism, 2nd ed., p 350, W. H. Freeman & Co., New York., incorporated herein by reference). In some embodiments, a polymerase has a fidelity value of at least about 100, 1000, 10,000, 100,000, or $1x10^6$, with or without a proofreading activity.

[261] In some embodiments, polymerases from native sources or variants thereof are screened using an assay that detects incorporation of an unnatural nucleic acid having a particular structure. In one example, polymerases can be screened for the ability to incorporate an unnatural nucleic acid or UBP; e.g., d5SICSTP, dCNMOTP, dTPT3TP, dNaMTP, dCNMOTP-dTPT3TP, or d5SICSTP- dNaMTP UBP. A polymerase, e.g., a heterologous polymerase, can be used that displays a modified property for the unnatural nucleic acid as compared to the wild-type polymerase. For example, the modified property can be, e.g., K_m, k_{cat}, V_{max}, polymerase processivity in the presence of an unnatural nucleic acid (or of a naturally occurring nucleotide), average template read-length by the polymerase in the presence of an unnatural nucleic acid, specificity of the polymerase for an unnatural nucleic acid, rate of binding of an unnatural nucleic acid, rate of product (pyrophosphate, triphosphate, etc.) release, branching rate, or any combination thereof. In one embodiment, the modified property is a reduced K_m for an unnatural nucleic acid and/or an increased k_{cat}/K_m or V_{max}/K_m for an unnatural nucleic acid. Similarly, the polymerase optionally has an increased rate of binding of an unnatural nucleic acid, an increased rate of product release, and/or a decreased branching rate, as compared to a wild-type polymerase.

[262] At the same time, a polymerase can incorporate natural nucleic acids, e.g., A, C, G, and T, into a growing nucleic acid copy. For example, a polymerase optionally displays a specific activity for a natural nucleic acid that is at least about 5% as high (e.g., 5%, 10%, 25%, 50%, 75%, 100% or higher), as a corresponding wild-type polymerase and a processivity with natural nucleic acids in the presence of a template that is at least 5% as high (e.g., 5%, 10%, 25%, 50%, 75%, 100% or higher) as the wild-type polymerase in the presence of the natural nucleic acid. Optionally, the polymerase displays a k_{cat}/K_m or V_{max}/K_m for a naturally occurring nucleotide that is at least about 5% as high (e.g., about 5%, 10%, 25%, 50%, 75% or 100% or higher) as the wild-type polymerase.

[263] Polymerases used herein that can have the ability to incorporate an unnatural nucleic acid of a particular structure can also be produced using a directed evolution approach. A nucleic acid synthesis assay can be used to screen for polymerase variants having specificity for any of a variety of unnatural nucleic acids. For example, polymerase variants can be screened for the ability to incorporate an unnatural nucleoside triphosphate opposite an

unnatural nucleotide in a DNA template; e.g., dTPT3TP opposite dCNMO, dCNMOTP opposite dTPT3, NaMTP opposite dTPT3, or TAT1TP opposite dCNMO or dNaM. In some embodiments, such an assay is an *in vitro* assay, e.g., using a recombinant polymerase variant. In some embodiments, such an assay is an *in vivo* assay, e.g., expressing a polymerase variant in a cell. Such directed evolution techniques can be used to screen variants of any suitable polymerase for activity toward any of the unnatural nucleic acids set forth herein. In some instances, polymerases used herein have the ability to incorporate unnatural ribonucleotides into a nucleic acid, such as RNA. For example, NaM or TAT1 ribonucleotides are incorporated into nucleic acids using the polymerases described herein.

[264] Modified polymerases of the compositions described can optionally be a modified and/or recombinant Φ29-type DNA polymerase. Optionally, the polymerase can be a modified and/or recombinant Φ29, B103, GA-1, PZA, Φ15, BS32, M2Y, Nf, G1, Cp-1, PRD1, PZE, SF5, Cp-5, Cp-7, PR4, PR5, PR722, or L17 polymerase.

[265] Modified polymerases of the compositions described can optionally be modified and/or recombinant prokaryotic DNA polymerase, e.g., DNA polymerase II (Pol II), DNA polymerase III (Pol III), DNA polymerase IV (Pol IV), DNA polymerase V (Pol V). In some embodiments, the modified polymerases comprise polymerases that mediate DNA synthesis across non-instructional damaged nucleotides. In some embodiments, the genes encoding Pol I, Pol II (*polB*), Poll IV (*dinB*), and/or Pol V (*umuCD*) are derepressed, or overexpressed, in the engineered cell, or SSO. In some embodiments, an increase in expression or overexpression of Pol II contributes to an increased retention of unnatural base pairs (UBPs) in an engineered cell, or SSO.

[266] Nucleic acid polymerases generally useful in the present disclosure include DNA polymerases, RNA polymerases, reverse transcriptases, and mutant or altered forms thereof. DNA polymerases and their properties are described in detail in, among other places, DNA Replication 2nd edition, Kornberg and Baker, W. H. Freeman, New York, N. Y. (1991). Known conventional DNA polymerases useful in the present disclosure include, but are not limited to, Pyrococcus furiosus (Pfu) DNA polymerase (Lundberg et al., 1991, Gene, 108: 1, Stratagene), Pyrococcus woesei (Pwo) DNA polymerase (Hinnisdaels et al., 1996, Biotechniques, 20:186-8, Boehringer Mannheim), Thermus thermophilus (Tth) DNA polymerase (Myers and Gelfand 1991, Biochemistry 30:7661), Bacillus stearothermophilus DNA polymerase (Stenesh and McGowan, 1977, Biochim Biophys Acta 475:32), Thermococcus litoralis (TIi) DNA polymerase (also referred to as Vent™ DNA polymerase, Cariello et al, 1991, Polynucleotides Res, 19: 4193, New England Biolabs), 9°Nm™ DNA

polymerase (New England Biolabs), Stoffel fragment, Thermo Sequenase® (Amersham Pharmacia Biotech UK), Therminator™ (New England Biolabs), Thermotoga maritima (Tma) DNA polymerase (Diaz and Sabino, 1998 Braz J Med. Res, 31 :1239), Thermus aquaticus (Taq) DNA polymerase (Chien et al, 1976, J. Bacteoriol, 127: 1550), DNA polymerase, Pyrococcus kodakaraensis KOD DNA polymerase (Takagi et al., 1997, Appl. Environ. Microbiol. 63:4504), JDF-3 DNA polymerase (from thermococcus sp. JDF-3, Patent application WO 0132887), Pyrococcus GB-D (PGB-D) DNA polymerase (also referred as Deep Vent™ DNA polymerase, Juncosa-Ginesta et al., 1994, Biotechniques, 16:820, New England Biolabs), UlTma DNA polymerase (from thermophile Thermotoga maritima; Diaz and Sabino, 1998 Braz J. Med. Res, 31 :1239; PE Applied Biosystems), Tgo DNA polymerase (from thermococcus gorgonarius, Roche Molecular Biochemicals), *E. coli* DNA polymerase I (Lecomte and Doubleday, 1983, Polynucleotides Res. 11 :7505), T7 DNA polymerase (Nordstrom et al, 1981, J Biol. Chem. 256:3112), and archaeal DP1I/DP2 DNA polymerase II (Cann et al, 1998, Proc. Natl. Acad. Sci. USA 95:14250). Both mesophilic polymerases and thermophilic polymerases are contemplated. Thermophilic DNA polymerases include, but are not limited to, ThermoSequenase®, 9°Nm™, Therminator™, Taq, Tne, Tma, Pfu, TfI, Tth, TIi, Stoffel fragment, Vent™ and Deep Vent™ DNA polymerase, KOD DNA polymerase, Tgo, JDF-3, and mutants, variants and derivatives thereof. A polymerase that is a 3 ‘ exonuclease-deficient mutant is also contemplated. Reverse transcriptases useful in the present disclosure include, but are not limited to, reverse transcriptases from HIV, HTLV-I, HTLV-II, FeLV, FIV, SIV, AMV, MMTV, MoMuLV and other retroviruses (see Levin, Cell 88:5-8 (1997); Verma, Biochim Biophys Acta. 473:1-38 (1977); Wu et al, CRC Crit Rev Biochem. 3:289-347(1975)). Further examples of polymerases include, but are not limited to 9°N™ DNA Polymerase, Taq DNA polymerase, Phusion® DNA polymerase, Pfu DNA polymerase, RB69 DNA polymerase, KOD DNA polymerase, and VentR® DNA polymerase Gardner et al. (2004) "Comparative Kinetics of Nucleotide Analog Incorporation by Vent DNA Polymerase (J. Biol. Chem., 279(12), 11834-11842; Gardner and Jack "Determinants of nucleotide sugar recognition in an archaeon DNA polymerase" Nucleic Acids Research, 27(12) 2545-2553.) Polymerases isolated from non-thermophilic organisms can be heat inactivatable. Examples are DNA polymerases from phage. It will be understood that polymerases from any of a variety of sources can be modified to increase or decrease their tolerance to high temperature conditions. In some embodiments, a polymerase can be thermophilic. In some embodiments, a thermophilic polymerase can be heat inactivatable. Thermophilic polymerases are typically

useful for high temperature conditions or in thermocycling conditions such as those employed for polymerase chain reaction (PCR) techniques.

[267] In some embodiments, the polymerase comprises Φ29, B103, GA-1, PZA, Φ15, BS32, M2Y, Nf, G1, Cp-1, PRD1, PZE, SF5, Cp-5, Cp-7, PR4, PR5, PR722, L17, ThermoSequenase®, 9°Nm™, Therminator™ DNA polymerase, Tne, Tma, TfI, Tth, TIi, Stoffel fragment, Vent™ and Deep Vent™ DNA polymerase, KOD DNA polymerase, Tgo, JDF-3, Pfu, Taq, T7 DNA polymerase, T7 RNA polymerase, PGB-D, UlTma DNA polymerase, *E. coli* DNA polymerase I, *E. coli* DNA polymerase III, archaeal DP1I/DP2 DNA polymerase II, 9°N™ DNA Polymerase, Taq DNA polymerase, Phusion® DNA polymerase, Pfu DNA polymerase, SP6 RNA polymerase, RB69 DNA polymerase, Avian Myeloblastosis Virus (AMV) reverse transcriptase, Moloney Murine Leukemia Virus (MMLV) reverse transcriptase, SuperScript® II reverse transcriptase, and SuperScript® III reverse transcriptase.

[268] In some embodiments, the polymerase is DNA polymerase I (or Klenow fragment), Vent polymerase, Phusion® DNA polymerase, KOD DNA polymerase, Taq polymerase, T7 DNA polymerase, T7 RNA polymerase, Therminator™ DNA polymerase, POLB polymerase, SP6 RNA polymerase, *E. coli* DNA polymerase I, *E. coli* DNA polymerase III, Avian Myeloblastosis Virus (AMV) reverse transcriptase, Moloney Murine Leukemia Virus (MMLV) reverse transcriptase, SuperScript® II reverse transcriptase, or SuperScript® III reverse transcriptase.

[269] In some embodiments, the polymerase is T7 RNA polymerase. In some embodiments, the polymerase is a T7 bacteriophage RNA polymerase. In some embodiments, the gene encoding the T7 bacteriophage RNA polymerase is operably linked to an inducible promoter. In some embodiments, the inducible promoter is lac inducible. In some embodiments, the inducible promoter is P_{lacUV5}.

XI. Polypeptides comprising an unnatural amino acid

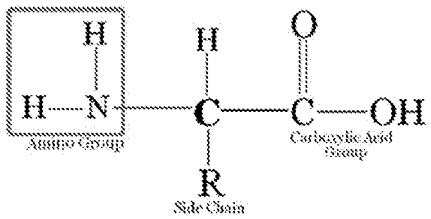
[270] The cells of the present disclosure can produce various polypeptides comprising an unnatural amino acid. In some embodiments, such polypeptides are antibodies. In some embodiments, such polypeptides are therapeutic proteins. In some embodiments, such polypeptides are cytokines. In some embodiments, such cytokines include interleukins. In some embodiments, an interleukin comprising an unnatural amino acid is a modified interleukin, such as those described in WO 2020/163532 A1, WO 2021/050554 A1, WO 2021/041206 A1, and WO 2021/091986 A1. Several of these molecules are referenced herein as “Molecule A,” “Molecule B,” and “Molecule C.” Molecule A is an IL-2 modified with an unnatural amino acid as described in WO 2020/163532 A1. Molecule B is an IL-2 modified

with an unnatural amino acid as described in WO 2021/050554 A1. Molecule C is also an interleukin modified with an unnatural amino acid.

XII. Natural and Unnatural Amino Acids

[271] As used herein, an amino acid residue can refer to a molecule containing both an amino group and a carboxyl group. Suitable amino acids include, without limitation, both the D- and L-isomers of the naturally-occurring amino acids, as well as non-naturally occurring amino acids prepared by organic synthesis or any other methods. The term amino acid, as used herein, includes, without limitation, α-amino acids, natural amino acids, non-natural amino acids, and amino acid analogs.

[272] The term "α-amino acid" can refer to a molecule containing both an amino group and a carboxyl group bound to a carbon which is designated the α-carbon. For example:



[273] The term "β-amino acid" can refer to a molecule containing both an amino group and a carboxyl group in a β configuration.

[274] "Naturally occurring amino acid" can refer to any one of the twenty amino acids commonly found in peptides synthesized in nature, and known by the one letter abbreviations A, R, N, C, D, Q, E, G, H, I, L, K, M, F, P, S, T, W, Y and V.

[275] The following table shows a summary of the properties of natural amino acids:

[276] "Hydrophobic amino acids" include small hydrophobic amino acids and large hydrophobic amino acids. "Small hydrophobic amino acid" can be glycine, alanine, proline, and analogs thereof. "Large hydrophobic amino acids" can be valine, leucine, isoleucine, phenylalanine, methionine, tryptophan, and analogs thereof. "Polar amino acids" can be serine, threonine, asparagine, glutamine, cysteine, tyrosine, and analogs thereof. "Charged amino acids" can be lysine, arginine, histidine, aspartate, glutamate, and analogs thereof.

Table 2

Amino Acid	3-Letter Code	1-Letter Code	Side-chain Polarity	Side-chain charge (pH 7.4)	Hydropathy Index
Alanine	Ala	A	nonpolar	neutral	1.8
Arginine	Arg	R	polar	positive	−4.5

Asparagine	Asn	N	polar	neutral	−3.5
Aspartic acid	Asp	D	polar	negative	−3.5
Cysteine	Cys	C	polar	neutral	2.5
Glutamic acid	Glu	E	polar	negative	−3.5
Glutamine	Gln	Q	polar	neutral	−3.5
Glycine	Gly	G	nonpolar	neutral	−0.4
Histidine	His	H	polar	positive (10%) neutral (90%)	−3.2
Isoleucine	Ile	I	nonpolar	neutral	4.5
Leucine	Leu	L	nonpolar	neutral	3.8
Lysine	Lys	K	polar	positive	−3.9
Methionine	Met	M	nonpolar	neutral	1.9
Phenylalanine	Phe	F	nonpolar	neutral	2.8
Proline	Pro	P	nonpolar	neutral	−1.6
Serine	Ser	S	polar	neutral	−0.8
Threonine	Thr	T	polar	neutral	−0.7
Tryptophan	Trp	W	nonpolar	neutral	−0.9
Tyrosine	Tyr	Y	polar	neutral	−1.3
Valine	Val	V	nonpolar	neutral	4.2

[277] An “amino acid analog” can be a molecule which is structurally similar to an amino acid and which can be substituted for an amino acid in the formation of a peptidomimetic macrocycle Amino acid analogs include, without limitation, β-amino acids and amino acids where the amino or carboxy group is substituted by a similarly reactive group (e.g., substitution of the primary amine with a secondary or tertiary amine, or substitution of the carboxy group with an ester).

[278] Amino acid analogs can include β-amino acid analogs. Examples of β-amino acid analogs include, but are not limited to, the following: cyclic β-amino acid analogs; β-alanine; (R)-β-phenylalanine; (R)-1,2,3,4-tetrahydro-isoquinoline-3-acetic acid; (R)-3-amino-4-(1-naphthyl)-butyric acid; (R)-3-amino-4-(2,4-dichlorophenyl)butyric acid; (R)-3-amino-4-(2-chlorophenyl)-butyric acid; (R)-3-amino-4-(2-cyanophenyl)-butyric acid; (R)-3-amino-4-(2-

fluorophenyl)-butyric acid; (R)-3-amino-4-(2-furyl)-butyric acid; (R)-3-amino-4-(2-methylphenyl)-butyric acid; (R)-3-amino-4-(2-naphthyl)-butyric acid; (R)-3-amino-4-(2-thienyl)-butyric acid; (R)-3-amino-4-(2-trifluoromethylphenyl)-butyric acid; (R)-3-amino-4-(3,4-dichlorophenyl)butyric acid; (R)-3-amino-4-(3,4-difluorophenyl)butyric acid; (R)-3-amino-4-(3-benzothienyl)-butyric acid; (R)-3-amino-4-(3-chlorophenyl)-butyric acid; (R)-3-amino-4-(3-cyanophenyl)-butyric acid; (R)-3-amino-4-(3-fluorophenyl)-butyric acid; (R)-3-amino-4-(3-methylphenyl)-butyric acid; (R)-3-amino-4-(3-pyridyl)-butyric acid; (R)-3-amino-4-(3-thienyl)-butyric acid; (R)-3-amino-4-(3-trifluoromethylphenyl)-butyric acid; (R)-3-amino-4-(4-bromophenyl)-butyric acid; (R)-3-amino-4-(4-chlorophenyl)-butyric acid; (R)-3-amino-4-(4-cyanophenyl)-butyric acid; (R)-3-amino-4-(4-fluorophenyl)-butyric acid; (R)-3-amino-4-(4-iodophenyl)-butyric acid; (R)-3-amino-4-(4-methylphenyl)-butyric acid; (R)-3-amino-4-(4-nitrophenyl)-butyric acid; (R)-3-amino-4-(4-pyridyl)-butyric acid; (R)-3-amino-4-(4-trifluoromethylphenyl)-butyric acid; (R)-3-amino-4-pentafluoro-phenylbutyric acid; (R)-3-amino-5-hexenoic acid; (R)-3-amino-5-hexynoic acid; (R)-3-amino-5-phenylpentanoic acid; (R)-3-amino-6-phenyl-5-hexenoic acid; (S)-1,2,3,4-tetrahydro-isoquinoline-3-acetic acid; (S)-3-amino-4-(1-naphthyl)-butyric acid; (S)-3-amino-4-(2,4-dichlorophenyl)butyric acid; (S)-3-amino-4-(2-chlorophenyl)-butyric acid; (S)-3-amino-4-(2-cyanophenyl)-butyric acid; (S)-3-amino-4-(2-fluorophenyl)-butyric acid; (S)-3-amino-4-(2-furyl)-butyric acid; (S)-3-amino-4-(2-methylphenyl)-butyric acid; (S)-3-amino-4-(2-naphthyl)-butyric acid; (S)-3-amino-4-(2-thienyl)-butyric acid; (S)-3-amino-4-(2-trifluoromethylphenyl)-butyric acid; (S)-3-amino-4-(3,4-dichlorophenyl)butyric acid; (S)-3-amino-4-(3,4-difluorophenyl)butyric acid; (S)-3-amino-4-(3-benzothienyl)-butyric acid; (S)-3-amino-4-(3-chlorophenyl)-butyric acid; (S)-3-amino-4-(3-cyanophenyl)-butyric acid; (S)-3-amino-4-(3-fluorophenyl)-butyric acid; (S)-3-amino-4-(3-methylphenyl)-butyric acid; (S)-3-amino-4-(3-pyridyl)-butyric acid; (S)-3-amino-4-(3-thienyl)-butyric acid; (S)-3-amino-4-(3-trifluoromethylphenyl)-butyric acid; (S)-3-amino-4-(4-bromophenyl)-butyric acid; (S)-3-amino-4-(4-chlorophenyl) butyric acid; (S)-3-amino-4-(4-cyanophenyl)-butyric acid; (S)-3-amino-4-(4-fluorophenyl) butyric acid; (S)-3-amino-4-(4-iodophenyl)-butyric acid; (S)-3-amino-4-(4-methylphenyl)-butyric acid; (S)-3-amino-4-(4-nitrophenyl)-butyric acid; (S)-3-amino-4-(4-pyridyl)-butyric acid; (S)-3-amino-4-(4-trifluoromethylphenyl)-butyric acid; (S)-3-amino-4-pentafluoro-phenylbutyric acid; (S)-3-amino-5-hexenoic acid; (S)-3-amino-5-hexynoic acid; (S)-3-amino-5-phenylpentanoic acid; (S)-3-amino-6-phenyl-5-hexenoic acid; 1,2,5,6-tetrahydropyridine-3-carboxylic acid; 1,2,5,6-tetrahydropyridine-4-carboxylic acid; 3-amino-3-(2-chlorophenyl)-propionic acid; 3-amino-3-(2-thienyl)-propionic acid; 3-amino-3-(3-bromophenyl)-propionic acid; 3-amino-3-(4-

chlorophenyl)-propionic acid; 3-amino-3-(4-methoxyphenyl)-propionic acid; 3-amino-4,4,4-trifluoro-butyric acid; 3-aminoadipic acid; D-β-phenylalanine; β-leucine; L-β-homoalanine; L-β-homoaspartic acid γ-benzyl ester; L-β-homoglutamic acid δ-benzyl ester; L-β-homoisoleucine; L-β-homoleucine; L-β-homomethionine; L-β-homophenylalanine; L-β-homoproline; L-β-homotryptophan; L-β-homovaline; L-Nω-benzyloxycarbonyl-β-homolysine; Nω-L-β-homoarginine; O-benzyl-L-β-homohydroxyproline; O-benzyl-L-β-homoserine; O-benzyl-L-β-homothreonine; O-benzyl-L-β-homotyrosine; γ-trityl-L-β-homoasparagine; (R)-β-phenylalanine; L-β-homoaspartic acid γ-t-butyl ester; L-β-homoglutamic acid δ-t-butyl ester; L-Nω-β-homolysine; Nδ-trityl-L-β-homoglutamine; Nω-2,2,4,6,7-pentamethyl-dihydrobenzofuran-5-sulfonyl-L-β-homoarginine; O-t-butyl-L-β-homohydroxy-proline; O-t-butyl-L-β-homoserine; O-t-butyl-L-β-homothreonine; O-t-butyl-L-β-homotyrosine; 2-aminocyclopentane carboxylic acid; and 2-aminocyclohexane carboxylic acid.

[279] Amino acid analogs can include analogs of alanine, valine, glycine or leucine. Examples of amino acid analogs of alanine, valine, glycine, and leucine include, but are not limited to, the following: α-methoxyglycine; α-allyl-L-alanine; α-aminoisobutyric acid; α-methyl-leucine; β-(1-naphthyl)-D-alanine; β-(1-naphthyl)-L-alanine; β-(2-naphthyl)-D-alanine; β-(2-naphthyl)-L-alanine; β-(2-pyridyl)-D-alanine; β-(2-pyridyl)-L-alanine; β-(2-thienyl)-D-alanine; β-(2-thienyl)-L-alanine; β-(3-benzothienyl)-D-alanine; β-(3-benzothienyl)-L-alanine; β-(3-pyridyl)-D-alanine; β-(3-pyridyl)-L-alanine; β-(4-pyridyl)-D-alanine; β-(4-pyridyl)-L-alanine; β-chloro-L-alanine; β-cyano-L-alanine; β-cyclohexyl-D-alanine; β-cyclohexyl-L-alanine; β-cyclopenten-1-yl-alanine; β-cyclopentyl-alanine; β-cyclopropyl-L-Ala-OH.dicyclohexylammonium salt; β-t-butyl-D-alanine; β-t-butyl-L-alanine; γ-aminobutyric acid; L-α,β-diaminopropionic acid; 2,4-dinitro-phenylglycine; 2,5-dihydro-D-phenylglycine; 2-amino-4,4,4-trifluorobutyric acid; 2-fluoro-phenylglycine; 3-amino-4,4,4-trifluoro-butyric acid; 3-fluoro-valine; 4,4,4-trifluoro-valine; 4,5-dehydro-L-leu-OH.dicyclohexylammonium salt; 4-fluoro-D-phenylglycine; 4-fluoro-L-phenylglycine; 4-hydroxy-D-phenylglycine; 5,5,5-trifluoro-leucine; 6-aminohexanoic acid; cyclopentyl-D-Gly-OH.dicyclohexylammonium salt; cyclopentyl-Gly-OH.dicyclohexylammonium salt; D-α,β-diaminopropionic acid; D-α-aminobutyric acid; D-α-t-butylglycine; D-(2-thienyl)glycine; D-(3-thienyl)glycine; D-2-aminocaproic acid; D-2-indanylglycine; D-allylglycine-dicyclohexylammonium salt; D-cyclohexylglycine; D-norvaline; D-phenylglycine; β-aminobutyric acid; β-aminoisobutyric acid; (2-bromophenyl)glycine; (2-methoxyphenyl)glycine; (2-methylphenyl)glycine; (2-thiazoyl)glycine; (2-thienyl)glycine; 2-amino-3-(dimethylamino)-propionic acid; L-α,β-diaminopropionic acid; L-α-aminobutyric acid; L-α-t-butylglycine; L-(3-thienyl)glycine; L-2-

amino-3-(dimethylamino)-propionic acid; L-2-aminocaproic acid dicyclohexyl-ammonium salt; L-2-indanylglycine; L-allylglycine dicyclohexyl ammonium salt; L-cyclohexylglycine; L-phenylglycine; L-propargylglycine; L-norvaline; N-α-aminomethyl-L-alanine; D-α,γ-diaminobutyric acid; L-α,γ-diaminobutyric acid; β-cyclopropyl-L-alanine; (N-β-(2,4-dinitrophenyl))-L-α,β-diaminopropionic acid; (N-β-1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl)-D-α,β-diaminopropionic acid; (N-β-1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl)-L-α,β-diaminopropionic acid; (N-β-4-methyltrityl)-L-α,β-diaminopropionic acid; (N-β-allyloxycarbonyl)-L-α,β-diaminopropionic acid; (N-γ-1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl)-D-α,γ-diaminobutyric acid; (N-γ-1-(4,4-dimethyl-2,6-dioxocyclohex-1-ylidene)ethyl)-L-α,γ-diaminobutyric acid; (N-γ-4-methyltrityl)-D-α,γ-diaminobutyric acid; (N-γ-4-methyltrityl)-L-α,γ-diaminobutyric acid; (N-γ-allyloxycarbonyl)-L-α,γ-diaminobutyric acid; D-α,γ-diaminobutyric acid; 4,5-dehydro-L-leucine; cyclopentyl-D-Gly-OH; cyclopentyl-Gly-OH; D-allylglycine; D-homocyclohexylalanine; L-1-pyrenylalanine; L-2-aminocaproic acid; L-allylglycine; L-homocyclohexylalanine; and N-(2-hydroxy-4-methoxy-Bzl)-Gly-OH.

[280] Amino acid analogs can include analogs of arginine or lysine. Examples of amino acid analogs of arginine and lysine include, but are not limited to, the following: citrulline; L-2-amino-3-guanidinopropionic acid; L-2-amino-3-ureidopropionic acid; L-citrulline; Lys(Me)$_2$-OH; Lys(N$_3$)—OH; Nδ-benzyloxycarbonyl-L-ornithine; Nω-nitro-D-arginine; Nω-nitro-L-arginine; α-methyl-ornithine; 2,6-diaminoheptanedioic acid; L-ornithine; (Nδ-1-(4,4-dimethyl-2,6-dioxo-cyclohex-1-ylidene)ethyl)-D-ornithine; (Nδ-1-(4,4-dimethyl-2,6-dioxo-cyclohex-1-ylidene)ethyl)-L-ornithine; (Nδ-4-methyltrityl)-D-ornithine; (Nδ-4-methyltrityl)-L-ornithine; D-ornithine; L-ornithine; Arg(Me)(Pbf)-OH; Arg(Me)$_2$-OH (asymmetrical); Arg(Me)2-OH (symmetrical); Lys(ivDde)-OH; Lys(Me)2-OH.HCl; Lys(Me3)-OH chloride; Nω-nitro-D-arginine; and Nω-nitro-L-arginine.

[281] Amino acid analogs can include analogs of aspartic or glutamic acids. Examples of amino acid analogs of aspartic and glutamic acids include, but are not limited to, the following: α-methyl-D-aspartic acid; α-methyl-glutamic acid; α-methyl-L-aspartic acid; γ-methylene-glutamic acid; (N-γ-ethyl)-L-glutamine; [N-α-(4-aminobenzoyl)]-L-glutamic acid; 2,6-diaminopimelic acid; L-α-aminosuberic acid; D-2-aminoadipic acid; D-α-aminosuberic acid; α-aminopimelic acid; iminodiacetic acid; L-2-aminoadipic acid; threo-β-methyl-aspartic acid; γ-carboxy-D-glutamic acid γ,γ-di-t-butyl ester; γ-carboxy-L-glutamic acid γ,γ-di-t-butyl ester; Glu(OAll)-OH; L-Asu(OtBu)—OH; and pyroglutamic acid.

[282] Amino acid analogs can include analogs of cysteine and methionine. Examples of amino acid analogs of cysteine and methionine include, but are not limited to, Cys(farnesyl)-OH, Cys(farnesyl)-OMe, α-methyl-methionine, Cys(2-hydroxyethyl)-OH, Cys(3-aminopropyl)-OH, 2-amino-4-(ethylthio)butyric acid, buthionine, buthioninesulfoximine, ethionine, methionine methylsulfonium chloride, selenomethionine, cysteic acid, [2-(4-pyridyl)ethyl]-DL-penicillamine, [2-(4-pyridyl)ethyl]-L-cysteine, 4-methoxybenzyl-D-penicillamine, 4-methoxybenzyl-L-penicillamine, 4-methylbenzyl-D-penicillamine, 4-methylbenzyl-L-penicillamine, benzyl-D-cysteine, benzyl-L-cysteine, benzyl-DL-homocysteine, carbamoyl-L-cysteine, carboxyethyl-L-cysteine, carboxymethyl-L-cysteine, diphenylmethyl-L-cysteine, ethyl-L-cysteine, methyl-L-cysteine, t-butyl-D-cysteine, trityl-L-homocysteine, trityl-D-penicillamine, cystathionine, homocystine, L-homocystine, (2-aminoethyl)-L-cysteine, seleno-L-cystine, cystathionine, Cys(StBu)—OH, and acetamidomethyl-D-penicillamine.

[283] Amino acid analogs can include analogs of phenylalanine and tyrosine. Examples of amino acid analogs of phenylalanine and tyrosine include β-methyl-phenylalanine, β-hydroxyphenylalanine, α-methyl-3-methoxy-DL-phenylalanine, α-methyl-D-phenylalanine, α-methyl-L-phenylalanine, 1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid, 2,4-dichloro-phenylalanine, 2-(trifluoromethyl)-D-phenylalanine, 2-(trifluoromethyl)-L-phenylalanine, 2-bromo-D-phenylalanine, 2-bromo-L-phenylalanine, 2-chloro-D-phenylalanine, 2-chloro-L-phenylalanine, 2-cyano-D-phenylalanine, 2-cyano-L-phenylalanine, 2-fluoro-D-phenylalanine, 2-fluoro-L-phenylalanine, 2-methyl-D-phenylalanine, 2-methyl-L-phenylalanine, 2-nitro-D-phenylalanine, 2-nitro-L-phenylalanine, 2;4;5-trihydroxy-phenylalanine, 3,4,5-trifluoro-D-phenylalanine, 3,4,5-trifluoro-L-phenylalanine, 3,4-dichloro-D-phenylalanine, 3,4-dichloro-L-phenylalanine, 3,4-difluoro-D-phenylalanine, 3,4-difluoro-L-phenylalanine, 3,4-dihydroxy-L-phenylalanine, 3,4-dimethoxy-L-phenylalanine, 3,5,3′-triiodo-L-thyronine, 3,5-diiodo-D-tyrosine, 3,5-diiodo-L-tyrosine, 3,5-diiodo-L-thyronine, 3-(trifluoromethyl)-D-phenylalanine, 3-(trifluoromethyl)-L-phenylalanine, 3-amino-L-tyrosine, 3-bromo-D-phenylalanine, 3-bromo-L-phenylalanine, 3-chloro-D-phenylalanine, 3-chloro-L-phenylalanine, 3-chloro-L-tyrosine, 3-cyano-D-phenylalanine, 3-cyano-L-phenylalanine, 3-fluoro-D-phenylalanine, 3-fluoro-L-phenylalanine, 3-fluoro-tyrosine, 3-iodo-D-phenylalanine, 3-iodo-L-phenylalanine, 3-iodo-L-tyrosine, 3-methoxy-L-tyrosine, 3-methyl-D-phenylalanine, 3-methyl-L-phenylalanine, 3-nitro-D-phenylalanine, 3-nitro-L-phenylalanine, 3-nitro-L-tyrosine, 4-(trifluoromethyl)-D-phenylalanine, 4-(trifluoromethyl)-L-phenylalanine, 4-amino-D-phenylalanine, 4-amino-L-phenylalanine, 4-benzoyl-D-phenylalanine, 4-benzoyl-L-

phenylalanine, 4-bis(2-chloroethyl)amino-L-phenylalanine, 4-bromo-D-phenylalanine, 4-bromo-L-phenylalanine, 4-chloro-D-phenylalanine, 4-chloro-L-phenylalanine, 4-cyano-D-phenylalanine, 4-cyano-L-phenylalanine, 4-fluoro-D-phenylalanine, 4-fluoro-L-phenylalanine, 4-iodo-D-phenylalanine, 4-iodo-L-phenylalanine, homophenylalanine, thyroxine, 3,3-diphenylalanine, thyronine, ethyl-tyrosine, and methyl-tyrosine.

[284] Amino acid analogs can include analogs of proline. Examples of amino acid analogs of proline include, but are not limited to, 3,4-dehydro-proline, 4-fluoro-proline, cis-4-hydroxy-proline, thiazolidine-2-carboxylic acid, and trans-4-fluoro-proline.

[285] Amino acid analogs can include analogs of serine and threonine. Examples of amino acid analogs of serine and threonine include, but are not limited to, 3-amino-2-hydroxy-5-methylhexanoic acid, 2-amino-3-hydroxy-4-methylpentanoic acid, 2-amino-3-ethoxybutanoic acid, 2-amino-3-methoxybutanoic acid, 4-amino-3-hydroxy-6-methylheptanoic acid, 2-amino-3-benzyloxypropionic acid, 2-amino-3-benzyloxypropionic acid, 2-amino-3-ethoxypropionic acid, 4-amino-3-hydroxybutanoic acid, and α-methylserine.

[286] Amino acid analogs can include analogs of tryptophan. Examples of amino acid analogs of tryptophan include, but are not limited to, the following: α-methyl-tryptophan; β-(3-benzothienyl)-D-alanine; β-(3-benzothienyl)-L-alanine; 1-methyl-tryptophan; 4-methyl-tryptophan; 5-benzyloxy-tryptophan; 5-bromo-tryptophan; 5-chloro-tryptophan; 5-fluoro-tryptophan; 5-hydroxy-tryptophan; 5-hydroxy-L-tryptophan; 5-methoxy-tryptophan; 5-methoxy-L-tryptophan; 5-methyl-tryptophan; 6-bromo-tryptophan; 6-chloro-D-tryptophan; 6-chloro-tryptophan; 6-fluoro-tryptophan; 6-methyl-tryptophan; 7-benzyloxy-tryptophan; 7-bromo-tryptophan; 7-methyl-tryptophan; D-1,2,3,4-tetrahydro-norharman-3-carboxylic acid; 6-methoxy-1,2,3,4-tetrahydronorharman-1-carboxylic acid; 7-azatryptophan; L-1,2,3,4-tetrahydro-norharman-3-carboxylic acid; 5-methoxy-2-methyl-tryptophan; and 6-chloro-L-tryptophan.

[287] Amino acid analogs can be racemic. In some instances, the D isomer of the amino acid analog is used. In some cases, the L isomer of the amino acid analog is used. In some instances, the amino acid analog comprises chiral centers that are in the R or S configuration. Sometimes, the amino group(s) of a β-amino acid analog is substituted with a protecting group, e.g., tert-butyloxycarbonyl (BOC group), 9-fluorenylmethyloxycarbonyl (FMOC), tosyl, and the like. Sometimes, the carboxylic acid functional group of a β-amino acid analog is protected, e.g., as its ester derivative. In some cases, the salt of the amino acid analog is used.

[288] In some embodiments, an unnatural amino acid is an unnatural amino acid described in Liu C.C., Schultz, P.G. *Annu. Rev. Biochem*. 2010, 79, 413. In some embodiments, an unnatural amino acid comprises N6(2-azidoethoxy)-carbonyl-L-lysine.

[289] In some embodiments, an amino acid residue described herein (e.g., within a protein) is mutated to an unnatural amino acid that is suitable for binding to a conjugating moiety. In some cases, the mutation to an unnatural amino acid prevents or minimizes a self-antigen response of the immune system. As used herein, the term "unnatural amino acid" refers to an amino acid other than the 20 amino acids that occur naturally in protein. Non-limiting examples of unnatural amino acids include: p-acetyl-L-phenylalanine, p-iodo-L-phenylalanine, p-methoxyphenylalanine, O-methyl-L-tyrosine, p-propargyloxyphenylalanine, p- propargyl-phenylalanine, L-3-(2-naphthyl)alanine, 3-methyl-phenylalanine, O- 4-allyl-L-tyrosine, 4-propyl-L-tyrosine, tri-O-acetyl-GlcNAcp-serine, L-Dopa, fluorinated phenylalanine, isopropyl-L-phenylalanine, p-azido-L-phenylalanine, p-azido-L-phenylalanine p-azido-phenylalanine, p-benzoyl-L-phenylalanine, *p*-Boronophenylalanine, *O*-propargyltyrosine, L-phosphoserine, phosphonoserine, phosphonotyrosine, p-bromophenylalanine, selenocysteine, p-amino-L-phenylalanine, isopropyl-L-phenylalanine, N6-(propargyloxy)-carbonyl-L-lysine (PrK), azido-lysine (N6-azidoethoxy-carbonyl-L-lysine, AzK), N6-(((2-azidobenzyl)oxy)carbonyl)-L-lysine, N6-(((3-azidobenzyl)oxy)carbonyl)-L-lysine, and N6-(((4-azidobenzyl)oxy)carbonyl)-L-lysine, an unnatural analogue of a tyrosine amino acid; an unnatural analogue of a glutamine amino acid; an unnatural analogue of a phenylalanine amino acid; an unnatural analogue of a serine amino acid; an unnatural analogue of a threonine amino acid; an alkyl, aryl, acyl, azido, cyano, halo, hydrazine, hydrazide, hydroxyl, alkenyl, alkynyl, ether, thiol, sulfonyl, seleno, ester, thioacid, borate, boronate, phospho, phosphono, phosphine, heterocyclic, enone, imine, aldehyde, hydroxylamine, keto, or amino substituted amino acid, or a combination thereof; an amino acid with a photoactivatable cross-linker; a spin-labeled amino acid; a fluorescent amino acid; a metal binding amino acid; a metal-containing amino acid; a radioactive amino acid; a photocaged and/or photoisomerizable amino acid; a biotin or biotin-analogue containing amino acid; a keto containing amino acid; an amino acid comprising polyethylene glycol or polyether; a heavy atom substituted amino acid; a chemically cleavable or photocleavable amino acid; an amino acid with an elongated side chain; an amino acid containing a toxic group; a sugar substituted amino acid; a carbon-linked sugar-containing amino acid; a redox-active amino acid; an a-hydroxy containing acid; an amino thio acid; an α, α disubstituted amino acid; a β-amino acid; a cyclic amino acid other than proline or histidine, and an aromatic amino acid other than phenylalanine, tyrosine or tryptophan.

[290] In some embodiments, the unnatural amino acid comprises a selective reactive group, or a reactive group for site-selective labeling of a target protein or polypeptide. In some instances, the chemistry is a biorthogonal reaction (e.g., biocompatible and selective reactions). In some cases, the chemistry is a Cu(I)-catalyzed or "copper-free" alkyne-azide triazole-forming reaction, the Staudinger ligation, inverse-electron-demand Diels-Alder (IEDDA) reaction, "photo-click" chemistry, or a metal-mediated process such as olefin metathesis and Suzuki-Miyaura or Sonogashira cross-coupling. In some embodiments, the unnatural amino acid comprises a photoreactive group, which crosslinks, upon irradiation with, e.g., UV. In some embodiments, the unnatural amino acid comprises a photo-caged amino acid. In some instances, the unnatural amino acid is a *para*-substituted, *meta*-substituted, or an *ortho*-substituted amino acid derivative.

[291] In some instances, the unnatural amino acid comprises p-acetyl-L-phenylalanine, p-azidomethyl-L-phenylalanine (pAMF), p-iodo-L-phenylalanine, O-methyl-L-tyrosine, p-methoxyphenylalanine, p-propargyloxyphenylalanine, p-propargyl-phenylalanine, L-3-(2-naphthyl)alanine, 3-methyl-phenylalanine, O- 4-allyl-L-tyrosine, 4-propyl-L-tyrosine, tri-O-acetyl-GlcNAcp-serine, L-Dopa, fluorinated phenylalanine, isopropyl-L-phenylalanine, p-azido-L-phenylalanine, p-acyl-L-phenylalanine, p-benzoyl-L-phenylalanine, L-phosphoserine, phosphonoserine, phosphonotyrosine, p-bromophenylalanine, p-amino-L-phenylalanine, or isopropyl-L-phenylalanine.

[292] In some cases, the unnatural amino acid is 3-aminotyrosine, 3-nitrotyrosine, 3,4-dihydroxy-phenylalanine, or 3-iodotyrosine. In some cases, the unnatural amino acid is phenylselenocysteine. In some instances, the unnatural amino acid is a benzophenone, ketone, iodide, methoxy, acetyl, benzoyl, or azide containing phenylalanine derivative. In some instances, the unnatural amino acid is a benzophenone, ketone, iodide, methoxy, acetyl, benzoyl, or azide containing lysine derivative. In some instances, the unnatural amino acid comprises an aromatic side chain. In some instances, the unnatural amino acid does not comprise an aromatic side chain. In some instances, the unnatural amino acid comprises an azido group, e.g., an azidoethoxy group. In some instances, the unnatural amino acid is a lysine modified with a group comprising an azido group, e.g., an azidoethoxy group. In some instances, the unnatural amino acid comprises a Michael-acceptor group. In some instances, Michael-acceptor groups comprise an unsaturated moiety capable of forming a covalent bond through a 1,2-addition reaction. In some instances, Michael-acceptor groups comprise electron-deficient alkenes or alkynes. In some instances, Michael-acceptor groups include but are not limited to alpha,beta unsaturated: ketones, aldehydes, sulfoxides, sulfones, nitriles, imines, or

aromatics. In some instances, the unnatural amino acid is dehydroalanine. In some instances, the unnatural amino acid comprises an aldehyde or ketone group. In some instances, the unnatural amino acid is a lysine derivative comprising an aldehyde or ketone group. In some instances, the unnatural amino acid is a lysine derivative comprising one or more O, N, Se, or S atoms at the beta, gamma, or delta position. In some instances, the unnatural amino acid is a lysine derivative comprising O, N, Se, or S atoms at the gamma position. In some instances, the unnatural amino acid is a lysine derivative wherein the epsilon N atom is replaced with an oxygen atom. In some instances, the unnatural amino acid is a lysine derivative that is not naturally-occurring post-translationally modified lysine.

[293] In some instances, the unnatural amino acid is an amino acid comprising a side chain, wherein the sixth atom from the alpha position comprises a carbonyl group. In some instances, the unnatural amino acid is an amino acid comprising a side chain, wherein the sixth atom from the alpha position comprises a carbonyl group, and the fifth atom from the alpha position is nitrogen. In some instances, the unnatural amino acid is an amino acid comprising a side chain, wherein the seventh atom from the alpha position is an oxygen atom.

[294] In some instances, the unnatural amino acid is a serine derivative comprising selenium. In some instances, the unnatural amino acid is selenoserine (2-amino-3-hydroselenopropanoic acid). In some instances, the unnatural amino acid is 2-amino-3-((2-((3-(benzyloxy)-3-oxopropyl)amino)ethyl)selanyl)propanoic acid. In some instances, the unnatural amino acid is 2-amino-3-(phenylselanyl)propanoic acid. In some instances, the unnatural amino acid comprises selenium, wherein oxidation of the selenium results in the formation of an unnatural amino acid comprising an alkene.

[295] In some instances, the unnatural amino acid comprises a cyclooctynyl group. In some instances, the unnatural amino acid comprises a transcycloctenyl group. In some instances, the unnatural amino acid comprises a norbornenyl group. In some instances, the unnatural amino acid comprises a cyclopropenyl group. In some instances, the unnatural amino acid comprises a diazirine group. In some instances, the unnatural amino acid comprises a tetrazine group.

[296] In some instances, the unnatural amino acid is a lysine derivative, wherein the side-chain nitrogen is carbamylated. In some instances, the unnatural amino acid is a lysine derivative, wherein the side-chain nitrogen is acylated. In some instances, the unnatural amino acid is 2-amino-6-{[(tert-butoxy)carbonyl]amino}hexanoic acid. In some instances, the unnatural amino acid is 2-amino-6-{[(tert-butoxy)carbonyl]amino}hexanoic acid. In some instances, the unnatural amino acid is N6-Boc-N6-methyllysine. In some instances, the

unnatural amino acid is N6-acetyllysine. In some instances, the unnatural amino acid is pyrrolysine. In some instances, the unnatural amino acid is N6-trifluoroacetyllysine. In some instances, the unnatural amino acid is 2-amino-6-{[(benzyloxy)carbonyl]amino}hexanoic acid. In some instances, the unnatural amino acid is 2-amino-6-{[(p-iodobenzyloxy)carbonyl]amino}hexanoic acid. In some instances, the unnatural amino acid is 2-amino-6-{[(p-nitrobenzyloxy)carbonyl]amino}hexanoic acid. In some instances, the unnatural amino acid is N6-prolyllysine. In some instances, the unnatural amino acid is 2-amino-6-{[(cyclopentyloxy)carbonyl]amino}hexanoic acid. In some instances, the unnatural amino acid is N6-(cyclopentanecarbonyl)lysine. In some instances, the unnatural amino acid is N6-(tetrahydrofuran-2-carbonyl)lysine. In some instances, the unnatural amino acid is N6-(3-ethynyltetrahydrofuran-2-carbonyl)lysine. In some instances, the unnatural amino acid is N6-((prop-2-yn-1-yloxy)carbonyl)lysine. In some instances, the unnatural amino acid is 2-amino-6-{[(2-azidocyclopentyloxy)carbonyl]amino}hexanoic acid. In some instances, the unnatural amino acid is N6-((2-azidoethoxy)carbonyl)lysine. In some instances, the unnatural amino acid is 2-amino-6-{[(2-nitrobenzyloxy)carbonyl]amino}hexanoic acid. In some instances, the unnatural amino acid is 2-amino-6-{[(2-cyclooctynyloxy)carbonyl]amino}hexanoic acid. In some instances, the unnatural amino acid is N6-(2-aminobut-3-ynoyl)lysine. In some instances, the unnatural amino acid is 2-amino-6-((2-aminobut-3-ynoyl)oxy)hexanoic acid. In some instances, the unnatural amino acid is N6-(allyloxycarbonyl)lysine. In some instances, the unnatural amino acid is N6-(butenyl-4-oxycarbonyl)lysine. In some instances, the unnatural amino acid is N6-(pentenyl-5-oxycarbonyl)lysine. In some instances, the unnatural amino acid is N6-((but-3-yn-1-yloxy)carbonyl)-lysine. In some instances, the unnatural amino acid is N6-((pent-4-yn-1-yloxy)carbonyl)-lysine. In some instances, the unnatural amino acid is N6-(thiazolidine-4-carbonyl)lysine. In some instances, the unnatural amino acid is 2-amino-8-oxononanoic acid. In some instances, the unnatural amino acid is 2-amino-8-oxooctanoic acid. In some instances, the unnatural amino acid is N6-(2-oxoacetyl)lysine. In some instances, the unnatural amino acid is N6-(((2-azidobenzyl)oxy)carbonyl)-L-lysine. In some instances, the unnatural amino acid is N6-(((3-azidobenzyl)oxy)carbonyl)-L-lysine. In some instances, the unnatural amino acid is N6-(((4-azidobenzyl)oxy)carbonyl)-L-lysine.

[297] In some instances, the unnatural amino acid is N6-propionyllysine. In some instances, the unnatural amino acid is N6-butyryllysine, In some instances, the unnatural amino acid is N6-(but-2-enoyl)lysine, In some instances, the unnatural amino acid is N6-((bicyclo[2.2.1]hept-5-en-2-yloxy)carbonyl)lysine. In some instances, the unnatural amino acid is N6-((spiro[2.3]hex-1-en-5-ylmethoxy)carbonyl)lysine. In some instances, the unnatural

amino acid is N6-(((4-(1-(trifluoromethyl)cycloprop-2-en-1-yl)benzyl)oxy)carbonyl)lysine. In some instances, the unnatural amino acid is N6-((bicyclo[2.2.1]hept-5-en-2-ylmethoxy)carbonyl)lysine. In some instances, the unnatural amino acid is cysteinyllysine. In some instances, the unnatural amino acid is N6-((1-(6-nitrobenzo[d][1,3]dioxol-5-yl)ethoxy)carbonyl)lysine. In some instances, the unnatural amino acid is N6-((2-(3-methyl-3H-diazirin-3-yl)ethoxy)carbonyl)lysine. In some instances, the unnatural amino acid is N6-((3-(3-methyl-3H-diazirin-3-yl)propoxy)carbonyl)lysine. In some instances, the unnatural amino acid is N6-((meta nitrobenyloxy)N6-methylcarbonyl)lysine. In some instances, the unnatural amino acid is N6-((bicyclo[6.1.0]non-4-yn-9-ylmethoxy)carbonyl)-lysine. In some instances, the unnatural amino acid is N6-((cyclohept-3-en-1-yloxy)carbonyl)-L-lysine.

[298] In some instances, the unnatural amino acid is 2-amino-3-(((((benzyloxy)carbonyl)amino)methyl)selanyl)propanoic acid. In some embodiments, the unnatural amino acid is incorporated into an unnatural polypeptide or an unnatural protein by a repurposed amber, opal, or ochre stop codon. In some embodiments, the unnatural amino acid is incorporated into an unnatural polypeptide or an unnatural protein by a 4-base codon. In some embodiments, the unnatural amino acid is incorporated into the protein by a repurposed rare sense codon.

[299] In some embodiments, provided herein is a cell comprising at least one unnatural amino acid. In some embodiments, the unnatural amino acid comprises a lysine analogue; an aromatic side chain; an azido group; an alkyne group; or an aldehyde or ketone group. In some embodiments, the unnatural amino acid does not comprise an aromatic side chain. In some embodiments, the unnatural amino acid is selected from N6-azidoethoxy-carbonyl-L-lysine, N6-propargylethoxy-carbonyl-L-lysine (PraK), N6-(propargyloxy)-carbonyl-L-lysine (PrK), p-azido-phenylalanine(*p*AzF), BCN-L-lysine, norbornene lysine, TCO-lysine, methyltetrazine lysine, allyloxycarbonyllysine, 2-amino-8-oxononanoic acid, 2-amino-8-oxooctanoic acid, p-acetyl-L-phenylalanine, p-azidomethyl-L-phenylalanine (pAMF), p-iodo-L-phenylalanine, m-acetylphenylalanine, 2-amino-8-oxononanoic acid, p-propargyloxyphenylalanine, p-propargyl-phenylalanine, 3-methyl-phenylalanine, L-Dopa, fluorinated phenylalanine, isopropyl-L-phenylalanine, p-azido-L-phenylalanine, p-acyl-L-phenylalanine, p-benzoyl-L-phenylalanine, p-bromophenylalanine, p-amino-L- phenylalanine, isopropyl-L-phenylalanine, O-allyltyrosine, O-methyl-L-tyrosine, O-4-allyl-L-tyrosine, 4-propyl-L-tyrosine, phosphonotyrosine, tri-O-acetyl-GlcNAcp-serine, L-phosphoserine, phosphonoserine, L-3-(2-naphthyl)alanine, 2-amino-3-((2-((3-(benzyloxy)-3-oxopropyl)amino)ethyl)selanyl)propanoic acid, 2-amino-3-(phenylselanyl)propanoic,

selenocysteine, N6-(((2-azidobenzyl)oxy)carbonyl)-L-lysine, N6-(((3-azidobenzyl)oxy)carbonyl)-L-lysine, and N6-(((4-azidobenzyl)oxy)carbonyl)-L-lysine. In some embodiments, the unnatural amino acid is an azidolysine (AzK).

XIII. Cell Types

[300] Taxonomic information for bacteria described herein are interpreted according to the NCBI Taxonomy database as of January 1, 2023, available at http://www.ncbi.nlm.nih.gov/taxonomy; *see also* Schoch CL, et al. NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database (Oxford) (2020) (describing the NCBI Taxonomy database and methodology).

[301] In some embodiments, a cell is a prokaryotic cell.

[302] Any suitable prokaryote may be selected as a host microorganism, engineered microorganism or source for a heterologous polynucleotide. A Gram negative or Gram positive bacteria may be selected. Examples of bacteria include, but are not limited to, Bacillus bacteria (e.g., B. subtilis, B. megaterium), Acinetobacter bacteria, Norcardia baceteria, Xanthobacter bacteria, Escherichia bacteria (e.g., *E. coli* (e.g., strains BL21, BL21(DE3), DH10B, HB101, Stbl2, DH5α, DB3, DB3.1, JM109), DB4, DB5, JDP682 and ccdA-over (e.g., U.S. Application No. 09/518,188))), Streptomyces bacteria, Erwinia bacteria, Klebsiella bacteria, Serratia bacteria (e.g., S. marcessans), Pseudomonas bacteria (e.g., P. aeruginosa), Salmonella bacteria (e.g., S. typhimurium, S. typhi), Megasphaera bacteria (e.g., Megasphaera elsdenii). Bacteria also include, but are not limited to, photosynthetic bacteria (e.g., green non-sulfur bacteria (e.g., Choroflexus bacteria (e.g., C. aurantiacus), Chloronema bacteria (e.g., C. gigateum)), green sulfur bacteria (e.g., Chlorobium bacteria (e.g., C. limicola), Pelodictyon bacteria (e.g., P. luteolum), purple sulfur bacteria (e.g., Chromatium bacteria (e.g., C. okenii)), and purple non-sulfur bacteria (e.g., Rhodospirillum bacteria (e.g., R. rubrum), Rhodobacter bacteria (e.g., R. sphaeroides, R. capsulatus), and Rhodomicrobium bacteria (e.g., R. vanellii)).

[303] Microorganisms or cells used as host organisms or source for a heterologous polynucleotide are commercially available. Microorganisms and cells described herein, and other suitable microorganisms and cells are available, for example, from Invitrogen Corporation, (Carlsbad, CA), American Type Culture Collection (Manassas, Virginia), and Agricultural Research Culture Collection (NRRL; Peoria, Illinois). Host microorganisms and engineered microorganisms may be provided in any suitable form. For example, such microorganisms may be provided in liquid culture or solid culture (e.g., agar-based medium), which may be a primary culture or may have been passaged (e.g., diluted and cultured) one or

more times. Microorganisms also may be provided in frozen form or dry form (e.g., lyophilized). Microorganisms may be provided at any suitable concentration.

[304] In some embodiments, the cell is a gamma proteobacterium. In some embodiments, the cell is a member of order Enterobacterales. In some embodiments, the cell is a member of family *Enterobacteriaceae*. In some embodiments, the cell is a member of genus *Escherichia*. In some embodiments, the cell is an *Escherichia coli* cell.

XIV. Structural Alignment

[305] In general, positions in the constructs described herein can be mapped onto the wild-type sequence by pairwise alignment, e.g., using the Needleman-Wunsch algorithm with standard parameters (EBLOSUM62 matrix, Gap penalty 10, gap extension penalty 0.5). *See also* the discussion below of structural alignment provided herein as an alternative approach for identifying corresponding positions.

[306] As discussed herein, positions of mutations corresponding to those described with respect to a given polypeptide (e.g, *M. barkeri or M. mazei* pyrrolysyl tRNA synthetase) can be identified by pairwise or structural alignment. Structural alignment is relevant to large protein families where the proteins share similar structures despite considerable sequence variation and many members of the family have been structurally characterized, and can also be used to identify corresponding positions in different versions of other polypeptides described herein..

[307] Structural alignment involves identifying corresponding residues across two (or more) polypeptide sequences by (i) modeling the structure of a first sequence using the known structure of the second sequence or (ii) comparing the structures of the first and second sequences where both are known, and identifying the residue in the first sequence most similarly positioned to a residue of interest in the second sequence. Corresponding residues are identified in some algorithms based on alpha-carbon distance minimization in the overlaid structures (e.g., what set of paired alpha carbons provides a minimized root-mean-square deviation for the alignment). There is an extensive literature on structural modeling and alignment; representative disclosures include US 6859736; US 8738343; and those cited in Aslam et al., Electronic Journal of Biotechnology 20 (2016) 9–13. For discussion of modeling a structure based on a known related structure or structures, *see, e.g.*, Bordoli et al., Nature Protocols 4 (2009) 1–13, and references cited therein.

XV. Methods of Preparing Cells To Express A Polypeptide Comprising An Unnatural Amino Acid with Improved Titer and/or Fidelity

[308] In some embodiments, provided herein is a method of preparing a cell that can express a polypeptide comprising an unnatural amino acid with improved titer and/or fidelity, the cell comprising one or more of a DNA sequence encoding LexA, a DNA sequence encoding RecA and RecX, a DNA sequence encoding Pol II, a DNA sequence encoding tRNA synthetase that can charge a tRNA with an unnatural amino acid, and/or a DNA sequence encoding *Pt*NTT2; and the method comprising one or more of: modifying the DNA sequence encoding LexA so that LexA does not undergo RecA-stimulated cleavage; modifying the cell to decrease expression of RecA; modifying the cell to increase expression of Pol II; modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression; and/or modifying the cell to increase expression of the tRNA synthetase. In some embodiments, the method comprises two or more of: modifying the DNA sequence encoding LexA so that LexA does not undergo RecA-stimulated cleavage; modifying the cell to decrease expression of RecA; modifying the cell to increase expression of Pol II; modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression; and/or modifying the cell to increase expression of the tRNA synthetase. In some embodiments, the method comprises three or more of: modifying the DNA sequence encoding LexA so that LexA does not undergo RecA-stimulated cleavage; modifying the cell to decrease expression of RecA; modifying the cell to increase expression of Pol II; modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression; and/or modifying the cell to increase expression of the tRNA synthetase. In some embodiments, the method comprises four or more of: modifying the DNA sequence encoding LexA so that LexA does not undergo RecA-stimulated cleavage; modifying the cell to decrease expression of RecA; modifying the cell to increase expression of Pol II; modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression; and/or modifying the cell to increase expression of the tRNA synthetase. In some embodiments, the method comprises each of: modifying the DNA sequence encoding LexA so that LexA does not undergo RecA-stimulated cleavage; modifying the cell to decrease expression of RecA; modifying the cell to increase expression of Pol II; modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression; and/or modifying the cell to increase expression of the tRNA synthetase.

[309] In some embodiments, modifying the DNA sequence encoding LexA so that LexA does not undergo Rec-A stimulated cleavage comprises a S119 mutation. In some embodiments, the S119 mutation is S119A.

[310] In some embodiments, modifying the cell to decrease expression of RecA results in a *recA* hypomorph. In some embodiments, modifying the cell to decrease expression of RecA comprises modifying the DNA sequence encoding RecA and/or RecX. In some embodiments, modifying the cell to decrease expression of RecA comprises modifying the DNA sequence encoding RecA and RecX so that RecX is overexpressed. In some embodiments, the DNA sequences encoding RecA and RecX are on a single transcript, and modifying the cell to decrease expression of RecA comprises deleting a terminator between the DNA sequence encoding RecA and the DNA sequence encoding RecX.

[311] In some embodiments, modifying the cell to increase expression of Pol II comprises modifying the promoter for Pol II so that the promoter for Pol II is derepressed. In some embodiments, modifying the promoter for Pol II so that it is derepressed comprises deleting or inactivating a LexA binding box.

[312] In some embodiments, modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression comprises modifying the DNA sequence encoding *Pt*NTT2 to be operably linked to a P_{lac} promoter and so that the DNA sequence does not comprise an oppositely oriented promoter within about 1.3-1.5 kb downstream of the 3' end of the *Pt*NTT2 coding sequence. In some embodiments, modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression comprises modifying the DNA sequence encoding *Pt*NTT2 to be operably linked to a P_{lac} promoter and so that the DNA sequence comprises a transcriptional terminator that terminates anterograde PtNTT2 transcription, and further comprises a transcription terminator element downstream of the stop codon of the DNA sequence encoding the *Pt*NTT2 and the transcription terminator element is oriented to terminate retrograde transcription into the DNA sequence encoding the *Pt*NTT2.

[313] In some embodiments, modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to a 5' UTR, wherein the 5' UTR enables increased tRNA synthetase expression compared to SEQ ID NO: 1. In some embodiments, the 5' UTR comprises SEQ ID NO: 2.

[314] In some embodiments, modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to an inducible promoter and providing a DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ} promoter or a P_{lacIQ1} promoter. In some embodiments, the promoter for the repressor of the inducible promoter is a P_{lacIQ} promoter. In some embodiments, the promoter for the repressor of the inducible promoter is a P_{lacIQ1} promoter.

[315] In some embodiments, the method comprises modifying the cell to increase expression of the tRNA synthetase: wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to a 5' UTR, wherein the 5' UTR comprises SEQ ID NO: 2; and wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to an inducible promoter and providing a DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ1} promoter.

[316] In some embodiments, the method comprises modifying the cell to increase expression of the tRNA synthetase and modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression: wherein modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression comprises modifying the DNA sequence encoding *Pt*NTT2 to be operably linked to a P_{lac} promoter and so that the DNA sequence does not comprise an oppositely oriented promoter within about 1.3-1.5 kb downstream of the 3' end of the *Pt*NTT2 coding sequence; wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to a 5' UTR, wherein the 5' UTR comprises SEQ ID NO: 2; and wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to an inducible promoter and providing a DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ1} promoter.

[317] In some embodiments, the method comprises the method comprising modifying the cell to increase expression of the tRNA synthetase, modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression, modifying the DNA sequence encoding LexA so that LexA does not undergo RecA-stimulated cleavage, modifying the cell to increase expression of Pol II, and modifying the cell to decrease expression of RecA: wherein modifying the DNA sequence encoding LexA so that LexA does not undergo Rec-A stimulated cleavage comprises a S119A mutation; wherein modifying the cell to decrease expression of RecA results in a *recA* hypomorph by modifying the DNA sequence encoding RecA and RecX so that RecX is overexpressed; wherein modifying the cell to increase expression of Pol II comprises modifying the promoter for Pol II so that the promoter for Pol II is derepressed, optionally wherein the modification comprises deleting a LexA binding box or inactivating a LexA binding box; wherein modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression comprises modifying the DNA sequence

encoding *Pt*NTT2 to be operably linked to a P_{lac} promoter and so that the DNA sequence does not comprise an oppositely oriented promoter within about 1.3-1.5 kb downstream of the 3' end of the *Pt*NTT2 coding sequence; wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to a 5' UTR, wherein the 5' UTR comprises SEQ ID NO: 2; and wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to an inducible promoter and providing a DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ1} promoter.

[318] In some embodiments, the method comprises modifying the cell to increase expression of the tRNA synthetase: wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to a 5' UTR, wherein the 5' UTR comprises SEQ ID NO: 2; and wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to an inducible promoter and providing a DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ} promoter.

[319] In some embodiments, the method comprises modifying the cell to increase expression of the tRNA synthetase and modifying the cell to decrease expression of RecA: wherein modifying the cell to decrease expression of RecA comprises deleting DNA sequence encoding RecA; wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to a 5' UTR, wherein the 5' UTR comprises SEQ ID NO: 2; and wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to an inducible promoter and providing a DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ1} promoter.

[320] In some embodiments, the method comprises modifying the cell to increase expression of the tRNA synthetase and modifying the cell to decrease expression of RecA: wherein modifying the cell to decrease expression of RecA comprises deleting DNA sequence encoding RecA; wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding the tRNA synthetase to a 5' UTR, wherein the 5' UTR comprises SEQ ID NO: 2; and wherein modifying the cell to increase expression of the tRNA synthetase comprises operably linking the DNA sequence encoding

the tRNA synthetase to an inducible promoter and providing a DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter for the repressor of the inducible promoter is a P_{lacIQ} promoter.

EXAMPLES

[321] The following examples are provided to illustrate certain disclosed embodiments and are not to be construed as limiting the scope of this disclosure in any way.

Example 1. Engineering Improved Pyrrolysyl tRNA Synthetases

[322] A strain of *E. coli* for producing certain polypeptides having an unnatural amino acid that included a pyrrolysyl tRNA synthetase to charge tRNAs with anticodons comprising an unnatural nucleotide is shown in Figure 31.

[323] In some instances, expression of a heterologous tRNA synthetase in cells producing polypeptides comprising an unnatural amino acid can complicate purification of the polypeptides comprising an unnatural amino acid and limit their yields. For example, low solubility and/or high expression levels may cause a significant portion of the heterologous tRNA synthetase to partition with insoluble target polypeptides. Figure 1 shows a Western blot showing two distinct bands: a synthetase band between 37 kDa and 50 kDa and a target interleukin band between 10 kDa and 15 kDa.

[324] Three types of pyrrolysyl tRNA synthetases were investigated to determine whether they could provide improved production of polypeptides comprising an unnatural amino acid relative to *M. barkeri* PylRS (Figure 2). One pyrrolysyl tRNA synthetase was a "Chimeric Synthetase"—*M. barkeri* PylRS NTD (N-terminal domain) fused to *M. mazei* PylRS CTD (C-terminal domain) (Figure 3). Another pyrrolysyl tRNA synthetase was "Split Synthetase"—*M. barkeri* PylRS NTD and, separately, *M. barkeri* PylRS CTD (Figure 4). And another pyrrolysyl tRNA synthetase was "Split Chimeric Synthetase"—*M. barkeri* PylRS NTD and, separately, *M. mazei* PylRS CTD (Figure 5).

A. Split Sites

[325] Experiments in this Example, Example 1A, were conducted using an amber suppression system unless otherwise noted. With respect to the Split Synthetase and Split Chimeric Synthetases, split sites were identified that led to variants displaying higher activity than the parental full-length, wild-type *M. barkeri* synthetase ("Control Synthetase"). An experiment was conducted in which superfolder green fluorescent protein (GFP) fluorescence was used to indicate the activity level of the tested pyrrolysyl tRNA synthetases. A TAG stop codon (1xTAG) at position 151 was recognized by the pyrrolysyl tRNA synthetases as a coding codon, and the pyrrolysyl tRNA synthetases charged tRNAs with a non-canonical

amino acid AzK. The constructs included a T7 promoter. Greater GFP fluorescence indicated greater activity. Four candidate pyrrolysyl tRNA synthetases were tested: Control (wild-type *M. barkeri* synthetase); Split #1 (*M. barkeri* pyrrolysyl-tRNA synthetase having two polypeptide chains, wherein the first polypeptide chain has part of a *M. barkeri* pyrrolysyl-tRNA synthetase with an amino acid sequence of positions 1-110 of SEQ ID NO: 3, and wherein the second polypeptide chain has part of a *M. barkeri* pyrrolysyl-tRNA synthetase with an amino acid sequence of positions 150-419 of SEQ ID NO: 3); Split #2 (*M. barkeri* pyrrolysyl-tRNA synthetase having two polypeptide chains, wherein the first polypeptide chain has part of a *M. barkeri* pyrrolysyl-tRNA synthetase with an amino acid sequence of positions 1-137 of SEQ ID NO: 3, and wherein the second polypeptide chain has part of a *M. barkeri* pyrrolysyl-tRNA synthetase with an amino acid sequence of positions 150-419 of SEQ ID NO: 3); and Split #3 (*M. barkeri* pyrrolysyl-tRNA synthetase having two polypeptide chains, wherein the first polypeptide chain has part of a *M. barkeri* pyrrolysyl-tRNA synthetase with an amino acid sequence of positions 1-149 of SEQ ID NO: 3, and wherein the second polypeptide chain has part of a *M. barkeri* pyrrolysyl-tRNA synthetase with an amino acid sequence of positions 150-419 of SEQ ID NO: 3). The results shown in Figure 6 indicate that the Split #2 pyrrolysyl tRNA synthetase produced the greatest amount of GFP fluorescence, indicating that it supported the highest production of GFP containing the ncAA activity among the four pyrrolysyl tRNA synthetases tested. The Split #1 pyrrolysyl tRNA synthetase produced the second greatest amount of GFP fluorescence. The Control Synthetase produced the third greatest amount of GFP fluorescence. The Split #3 pyrrolysyl tRNA synthetase produced the least amount of GFP fluorescence, lower than the Control Synthetase.

[326] It was also determined that the split pyrrolysyl tRNA synthetases are permissive for C-terminal fusion of a fusion partner to the N-terminal domain. One of skill in the art reading the present disclosure would understand that they can be permissive for N-terminal fusion of a fusion partner to the C-terminal domain. *E. coli* thioredoxin (TrxA) was C-terminally fused to N-terminal domains of split pyrrolysyl synthetases Split #1, Split #2, and Split #3. *See* Figure 7. An experiment was conducted in which superfolder green fluorescent protein (GFP) fluorescence was used to indicate the activity level of the tested pyrrolysyl tRNA synthetases. A TAG stop codon (1xTAG) at position 151 was recognized by the pyrrolysyl tRNA synthetases as a coding codon, and the pyrrolysyl tRNA synthetases charged tRNAs with GFP. Greater GFP fluorescence indicated greater activity. The constructs included a T7 promoter. Six pyrrolysyl tRNA synthetases were tested: Split #1; Split #1 fused to TrxA(N); Split #2, Split #2 fused to TrxA(N), Split #3, and Split #3 fused to TrxA(N). The results shown in

Figure 8 indicate that the Split #2 and Split #2 fused to TrxA(N) tRNA synthetases produced the greatest amount of GFP fluorescence, indicating the highest activity amongst the six pyrrolysyl tRNA synthetases tested. The Split #1 pyrrolysyl tRNA synthetase fused to TrxA(N) produced the next greatest amount of GFP fluorescence, followed by the Split #1 pyrrolysyl tRNA synthetase and then followed by the Split #3 pyrrolysyl tRNA synthetase fused to TrxA(N). The Split #3 pyrrolysyl tRNA synthetase produced the least amount of GFP fluorescence. The results show that fusing E. coli thioredoxin (TrxA) C-terminally to N-terminal domains of split synthetases (#2 and #3) increases GFP production by 25-40%. This tolerance of attached fusion partners demonstrates enablement for attachment of solubilizing proteins, affinity handles, and/or polyanionic/polycationic tags to further aid the removal of contaminating synthetase from preparations of expressed polypeptides containing an ncAA.

[327] An experiment was conducted in which higher-stringency constructs were used to detect improvements in enzymatic activity of the pyrrolysyl tRNA synthetases. Superfolder green fluorescent protein (GFP) fluorescence was used to indicate the activity level of the tested pyrrolysyl tRNA synthetases. Either one, two, or three TAG stop codons (1xTAG, 2xTAG, or 3xTAG) were inserted in the sequence for production of GFP, such that for GFP to be produced and for fluorescence to be measurable, the pyrrolysyl tRNA synthetase would need to recognize as coding (rather than as a stop codon) either one, two, or three TAG stop codons, with three TAG stop codons requiring the highest concentration of charged pyrrolysyl tRNA synthetases. As shown in Figure 9, the 1xTAG construct included a TAG codon at position 151, the 2xTAG construct included TAG codons at positions 151 and 190, and the 3xTAG construct included TAG codons at positions 151, 190, and 200. The constructs all included a T7 promoter ("high expression conditions"). tRNAs for the TAG codon were charged with a non-canonical amino acid AzK. Greater GFP fluorescence indicated greater activity. Four pyrrolysyl tRNA synthetases were tested: Control (wild-type *M. barkeri* synthetase); Split #2; "Chimeric Synthetase"—*M. barkeri* PylRS NTD fused to *M. mazei* PylRS CTD; and "Split Chimeric Synthetase"—*M. barkeri* PylRS NTD and, separately, *M. mazei* PylRS CTD. The results, shown in Figure 10, indicate that under high expression conditions, split and split chimeric pyrrolysyl tRNA synthetases support higher GFP (TAG) production than intact pyrrolysyl tRNA synthetases.

[328] The same four pyrrolysyl tRNA synthetases were tested for insoluble contamination when producing Molecule B with a T7 promoter: "Control Synthetase" (wild-type *M. barkeri* synthetase); Split #2 "Chimeric Synthetase"—*M. barkeri* PylRS NTD fused to *M. mazei* PylRS CTD; and "Split Chimeric Synthetase"—*M. barkeri* PylRS NTD and, separately, *M. mazei*

PylRS CTD. All Experiments were performed in commercial BL21(DE3) *E. coli* strain. Strains prepared with the appropriate plasmids were grown overnight in Power media with antibiotics. In the morning cultures were diluted to an OD of 0.05 with 2xYS media. 400 µL of diluted cultures were added to 48-well plates and cultured at 37°C (250 rpm). After 2 hrs, 40 µL of 1 M AzK was added to each well followed by 40 µL of 1 M IPTG 20 min later. Final measurements were taken 3 hours after induction. Molecule B is an IL-2 modified with an unnatural amino acid as described in WO 2021/050554 A1. The results, a Western blot shown in Figure 11, indicate that under high expression conditions, using split synthetases in place of the Control Synthetase reduces contamination by the tRNA synthetase during purification of the Molecule B product, without affecting specific production and fidelity of Molecule B.

[329] An experiment was conducted in which higher-stringency constructs were used for testing improvement in enzymatic activity in the pyrrolysyl tRNA synthetases under low expression conditions. Low expression conditions are relevant because when the pyrrolysyl tRNA synthetase is moved into the chromosome instead of on plasmids, expression is lower, as there are only one or two copies on the chromosome and many copies when on plasmids. Superfolder green fluorescent protein (GFP) fluorescence was used to indicate the activity level of the tested pyrrolysyl tRNA synthetases. Either one, two, or three TAG stop codons (1xTAG, 2xTAG, or 3xTAG) were inserted in the sequence for production of GFP, such that for GFP to be produced and for fluorescence to be measurable, the pyrrolysyl tRNA synthetase would need to recognize as coding (rather than as a stop codon) either one, two, or three TAG stop codons. As shown in Figure 9, the 1xTAG construct included a TAG codon at position 151, the 2xTAG construct included TAG codons at positions 151 and 190, and the 3xTAG construct included TAG codons at positions 151, 190, and 200. The constructs all included a T7 promoter. tRNAs for the TAG codon were charged with a non-canonical amino acid AzK. Greater GFP fluorescence indicated greater activity. Four pyrrolysyl tRNA synthetases were tested: "Control Synthetase" (wild-type *M. barkeri* synthetase); Split # 2; "Chimeric Synthetase"—*M. barkeri* PylRS NTD fused to *M. mazei* PylRS CTD; and "Split Chimeric Synthetase"—*M. barkeri* PylRS NTD and, separately, *M. mazei* PylRS CTD. The results, shown in Figure 12 and Figure 13, indicate that under low expression conditions, the "Chimeric Synthetase" outperformed the Control Synthetase, "Split #2," and "Split Chimeric Synthetase."

[330] The following experiment was conducted with a system utilizing unnatural nucleotides. The same four pyrrolysyl tRNA synthetases were tested for contamination by the tRNA synthetase during purification of the Molecule B product, when producing Molecule B with a

T7 promoter. Low synthetase expression was achieved with a modified expression cassette which eliminated a SalI recognition site. "Control Synthetase" (wild-type *M. barkeri* synthetase); Split #2 "Chimeric Synthetase"—*M. barkeri* PylRS NTD fused to *M. mazei* PylRS CTD; and "Split Chimeric Synthetase"—*M. barkeri* PylRS NTD and, separately, *M. mazei* PylRS CTD. The results, a Western blot shown in Figure 14, indicate that low expression of the "Chimeric Synthetase" eliminated insoluble contamination while maintaining Molecule B production and/or fidelity.

B. IPYE and IPYEVF Mutations

[331] Experiments in this Example, Example 1B, were conducted using an amber suppression system unless otherwise noted. Mutations to the chimeric pyrrolysyl tRNA synthetases were investigated. One set of mutations investigated was the "IPYE" set of mutations: V31I, T56P, H62Y, and A100E (where the amino acid positions correspond to those of the unmutated chimeric PylRS having SEQ ID NO: 4). Another set of mutations investigated was the "VF" set of mutations: C313V and Y349F (where the amino acid positions correspond to those of the unmutated chimeric PylRS having SEQ ID NO: 4). A chimeric pyrrolysyl tRNA synthetase of SEQ ID NO: 4 having the "IPYE" mutations and the "VF" mutations (collectively "IPYEVF" mutations)("Chimeric (IPYEVF) Synthetase") is the chimeric pyrrolysyl tRNA synthetase of SEQ ID NO: 6.

[332] An experiment was conducted in which higher-stringency constructs were used for testing enzymatic activity in the pyrrolysyl tRNA synthetases under low expression conditions. Superfolder green fluorescent protein (GFP) fluorescence was used to indicate the activity level of the tested pyrrolysyl tRNA synthetases. "GFP (TAG)" constructs were prepared as follows: either one, two, or three TAG stop codons (1xTAG, 2xTAG, or 3xTAG) were inserted in the sequence for production of GFP, such that for GFP to be produced and for fluorescence to be measurable, the pyrrolysyl tRNA synthetase would need to recognize as coding (rather than as a stop codon) either one, two, or three TAG stop codons. As shown in Figure 9, the 1xTAG construct included a TAG codon at position 151, the 2xTAG construct included TAG codons at positions 151 and 190, and the 3xTAG construct included TAG codons at positions 151, 190, and 200. The synthetase constructs all included a modified RBS after ptac with a SalI recognition site removed. tRNAs for the TAG codon were charged with a non-canonical amino acid AzK. Greater GFP fluorescence indicated greater activity. Three pyrrolysyl tRNA synthetases were tested: "Chimeric Synthetase"—*M. barkeri* PylRS NTD fused to *M. mazei* PylRS CTD; "Chimeric (IPYE) Synthetase"—the "Chimeric Synthetase" with the IPYE mutations; and "Chimeric (IPYEVF) Synthetase"—the "Chimeric Synthetase" with the

IPYEVF mutations. The results, shown in Figure 15, indicate that under low expression conditions, the "Chimeric (IPYEVF) Synthetase" outperformed the "Chimeric Synthetase" and the "Chimeric (IPYE) Synthetase" in terms of GFP fluorescence. The results showed that mutations that increase the synthetase's affinity for $tRNA^{Pyl}$ and affinity for AzK enhance GFP (TAG) production.

[333] The following experiment was conducted with a system utilizing unnatural nucleotides. An experiment was conducted in which pyrrolysyl tRNA synthetases were tested and compared in terms of production of Molecule A. Two pyrrolysyl tRNA synthetases were tested: the Control Synthetase and the "Chimeric (IPYEVF) Synthetase." The Control Synthetase was under a high expression promoter, the T7 promoter. The "Chimeric (IPYEVF) Synthetase" was under a low expression promoter, which consisted of the high expression promoter with the deletion of a SalI recognition site. Measurements were taken at 5 hours, 6 hours, and 7 hours post-induction. Both titer and fidelity were tested. The titer results are shown in Figure 16. The fidelity results are shown in Figure 17. The results indicate that "Chimeric (IPYEVF) Synthetase" under low expression conditions was roughly equivalent to the Control Synthetase under high expression conditions for titer (*see* Figure 16) and for fidelity (*see* Figure 17).

[334] The following experiment was conducted with a system utilizing unnatural nucleotides. An experiment was conducted in which pyrrolysyl tRNA synthetases were tested and compared in terms of production of Molecule B. Two pyrrolysyl tRNA synthetases were tested: the Control Synthetase and the "Chimeric (IPYEVF) Synthetase." The Control Synthetase was under a high expression promoter, the standard ptac promoter. The "Chimeric (IPYEVF) Synthetase" was under a low expression promoter, with a SalI recognition site deleted from the original promoter sequence. Measurements were taken at 5 hours and 7 hours post-induction. Both titer and fidelity were tested. The insoluble fraction at the end of the cell culture analyzed chromatographically and the results are shown in Figure 20 and Figure 21. The peaks at approximately 0.5 minutes represent flow-through. The peaks between 5.5 and 6 minutes in Figure 20 and Figure 21 represent Molecule B. Any peak between 4 and 4.5 minutes in Figure 20 and Figure 21 represents synthetase (PylRS) contamination. The results indicate that "Chimeric (IPYEVF) Synthetase" under low expression conditions outperforms the Control Synthetase under high expression conditions for titer. *See* Figure 18. The fidelity results are shown in Figure 19. The results indicate that "Chimeric (IPYEVF) Synthetase" under low expression conditions and the Control Synthetase under high expression conditions perform about the same for fidelity. *See* Figure 19. Results for the same experiment are shown

in Figure 20 and Figure 21. The results show less synthetase (PylRS) contamination for the "Chimeric (IPYEVF) Synthetase" than with the Control Synthetase.

[335] All three of the classes of engineered synthetases in some embodiments exhibited superior solubility and enzymatic activity relative to the Control Synthetase.

C. IPYEVF + TA Mutations

[336] Experiments in this Example, Example 1C, were conducted with a system utilizing unnatural nucleotides unless otherwise noted. The pyrrolysyl tRNA synthetase was designed to charge tRNAs with AzK.

[337] The Strain J with a Chimeric (IPYEVF) Synthetase was susceptible to lowering of AzK concentrations. A Culture Process was used to assess strain performance. Typically, in the Culture Process, media containing unnatural deoxyribonucleotides is first inoculated with cells. Once the culture reaches a certain optical density (OD_{600}), unnatural deoxyribonucleotides are added to the feed, followed by unnatural ribonucleotides and unnatural amino acids. Once the culture reaches an even higher optical density, the cells are induced with IPTG to express a protein. After induction, additional unnatural deoxyribonucleotides, unnatural ribonucleotides, and unnatural amino acids (e.g., AzK) are added to the feed before finally, the proteins are harvested. Even though the fidelity did not substantially drop, lowered AzK concentration dramatically lowered titer. Figure 22 shows Final OD and Titer/OD at different AzK concentrations (mM) (one tenth standard, one quarter standard, one half standard, and standard concentration of AzK). Figure 23 shows Titer (mg/L) and Fidelity (%) at different AzK concentrations (mM) (one tenth standard, one quarter standard, one half standard, and standard concentration of AzK).

[338] Two additional mutations were identified and added to the mutant chimeric pyrrolysyl tRNA sequence of SEQ ID NO: 6 (which is the chimeric pyrrolysyl synthetase of SEQ ID NO: 4 with the IPYEVF mutations described herein). One of these was a mutation at the N-terminal proline residue to either leucine or threonine (P5L or P5T, respectively) within the conserved essential N-terminal domain of the mutant chimeric pyrrolysyl tRNA synthetase of SEQ ID NO: 6. Additionally, a mutation at position 302 within the C-terminal domain of the mutant chimeric pyrrolysyl tRNA synthetase of SEQ ID NO: 6 was identified. Glutamic acid mutated to a smaller non-charged alanine. A new mutant chimeric pyrrolysyl tRNA synthetase of SEQ ID NO: 7 was created (which is the chimeric pyrrolysyl synthetase of SEQ ID NO: 4 with the IPYEVF mutations and TA mutations described herein).

[339] Expression of Molecule B was tested at half the standard AzK concentration with Strain J as a control. The Culture Process described above related to Figure 22 and Figure 23

was used to assess strain performance. Figure 24 and Figure 25 show the titer in mg/L of Strain J and Strain J + 2 new mutations in the pyrrolysyl tRNA synthetase, P5T and E302A ("Strain L"). Titer was also plotted over OD. Strain L had higher titer than the control Strain J.

[340] Another expression experiment was conducted to test Strain L, the strain with a pyrrolysyl tRNA synthetase with the two new mutations, with further limitation of the AzK concentration. One of the tested AzK concentrations was a quarter of the standard AzK concentration. The other tested AzK concentration was one tenth of the standard AzK concentration. The control was Strain J. The Culture Process described above related to Figure 22 and Figure 23 was used to assess strain performance. The results shown in Figures 26-28 show that Strain L, the strain with a pyrrolysyl tRNA synthetase with the two new mutations, had an advantage over Strain J. Strain L grew similarly to, if not better than, Strain J, produced more Molecule B under all tested conditions, and sustained the fidelity over 94% even in extreme limiting conditions, while the fidelity of Molecule B produced by Strain J dropped to the high 80s. Figure 29 shows a plot of titer/OD values for the same experiment, and shows that Strain L outperformed Strain J.

[341] Another expression experiment was conducted in which three different AzK concentrations (standard, one fourth standard, and one tenth standard) were tested for their effect on synthetase levels. The Culture Process described above related to Figure 22 and Figure 23 was used to assess strain performance. A Pierce™ Fast Western Blot Kit, West Pico Substrate, Rabbit – Thermo Scientific was used to measure synthetase levels. OD was normalized to lyse equivalent amounts of cells. The results are shown in Figure 30. The results showed that in the control strain, Strain J, the synthetase expression dropped as AzK became limited. However, in Strain L, with the improved synthetase comprising two new mutations, it was clear that the expression of the synthetase stayed the same even when the AzK was limited, for example, below 7.5 mM.

Example 2. Chromosomal Integration of Improved Pyrrolysyl tRNA Synthetase + Modified 5' UTR

[342] Experiments in this Example, Example 2, were conducted using an amber suppression system unless otherwise noted. The basic components of a strain used for producing certain protein products having an unnatural amino acid are shown in Figure 31, where X and Y each represent an unnatural nucleotide. A pyrrolysyl tRNA synthetase is one component.

[343] The control strain included two plasmids: one encoding Molecule A (or Molecule B), the tRNA, and a zeocin antibiotic resistance marker, and the other encoding the chimeric pyrrolysyl tRNA synthetase and a tetracycline antibiotic resistance marker, as shown in Figure

32. Molecule A is an IL-2 modified with an unnatural amino acid as described in WO 2020/163532 A1. In another strain, the chimeric pyrrolysyl tRNA synthetase was chromosomally integrated and the strain comprised only one plasmid encoding Molecule A (or Molecule B), the tRNA, and a zeocin antibiotic resistance marker, as shown in Figure 33.

[344] An experiment testing expression of the chromosomally integrated chimeric pyrrolysyl tRNA synthetase was performed. In this experiment, the chimeric pyrrolysyl tRNA synthetase included six mutations: V31I, T56P, H62Y, A100E, C313V, and Y349F ("IPYEVF"). In the experiment, superfolder green fluorescent protein (GFP) fluorescence was used to indicate the activity level of the tested pyrrolysyl tRNA synthetases, as shown in Figure 37. A TAG stop codon (1xTAG) at position 151 was recognized by the pyrrolysyl tRNA synthetases as a coding codon, and the pyrrolysyl tRNA synthetases charged tRNAs with a non-canonical amino acid, AzK. If the AzK was not incorporated, the GFP would be truncated, so the greater the AzK incorporation, the greater the GFP fluorescence. Three constructs were tested. The first construct included a plasmid encoding the chimeric pyrrolysyl tRNA synthetase (IPYEVF) with a standard 5' UTR (SEQ ID NO: 1), which includes a lac operator, a tac promoter, and a tetracycline resistance marker, as shown in Figure 34. The second construct included a chromosomally integrated chimeric pyrrolysyl tRNA synthetase (IPYEVF) with a standard 5' UTR (SEQ ID NO: 1), which includes a lac operator, and a tac promoter, as shown in Figure 35. The third construct included a chromosomally integrated chimeric pyrrolysyl tRNA synthetase (IPYEVF) with a 5' UTR that was modified to include SEQ ID NO: 2, which includes a lac operator, and a tac promoter, as shown in Figure 36. Greater GFP fluorescence indicated greater activity. The results shown in Figure 38 indicate that both strains with chromosomally integrated pyrrolysyl tRNA synthetases expressed more GFP than the strain with pyrrolysyl tRNA synthetase on a plasmid. The strain with the chromosomally integrated pyrrolysyl tRNA synthetase with the modified 5' UTR performed the best overall; it produced the greatest amount of GFP fluorescence, indicating the highest activity among the strains tested.

Example 3. Additional LacI Repressor

[345] Experiments in this Example, Example 3, were conducted with a system utilizing unnatural nucleotides unless otherwise noted. For the strains described in Example 2, several options for promoters for LacI repressors were considered. LacI repressor regulates the expression of all induced platform components, including the T7 RNA polymerase, the target mRNA, the tRNA, and the pyrrolysyl tRNA synthetase. More LacI generally means greater

repression and less leakiness, but potentially at the cost of reduced inducibility as assessed by natural GFP expression.

[346] Three options for the promoter for the LacI repressor were considered: (1) the original promoter for the LacI repressor; (2) the LacIQ promoter for the LacI repressor, with about 10-fold promoter strength (shown in Figure 39); and (3) the LacIQ1 promoter for the LacI repressor, with about 100-fold promoter strength (shown in Figure 40).

[347] An experiment was conducted in which three strains having differing characteristics were grown on plates with Power Prime + KPi, dCNMOTP, dTPT3TP, and no antibiotics. Photos were taken of the plates of three different stages: RCB (Research Cell Bank), mMCB (mock Master Cell Bank), and mWCB (mock Working Cell Bank), with a 1:30K dilution with power prime between each stage. When plating, a 1:100 dilution of each glycerol stock was prepared in Power Prime media. Serial dilutions were prepared to plate 100 µL of a 1:5000 dilution onto each agar plate. The first strain, Strain A, included a *M. barkeri* tRNA synthetase encoded on a plasmid. The second strain, Strain E, included a chromosomally integrated chimeric tRNA synthetase with a LacIQ promoter for the LacI repressor. The third strain, Strain F, included a chromosomally integrated chimeric tRNA synthetase with a lacIQ1 promoter for the LacI repressor. The results are shown in Figure 41. Surprisingly, it was observed that the LacIQ1 genotype may confer enhanced resistance to colony size heterogeneity. Colony size heterogeneity is undesirable for a strain because different growth rates between colonies indicate that the strain is genetically unstable.

Example 4. Expression Tests of Strains with Multiple Combinations of Characteristics

[348] Experiments in this Example, Example 4, were conducted with a system utilizing unnatural nucleotides unless otherwise noted. As described in Examples 1-3, several options for characteristics for strains were considered, including options for the pyrrolysyl tRNA synthetase, plasmid-based or chromosomal integration for the pyrrolysyl tRNA synthetase, original or modified 5'UTR for chromosomally integrated pyrrolysyl tRNA synthetases, and different LacI repressor promoters.

[349] In addition, deletion of the *recA* gene was considered. RecA is a master regulator of DNA repair. RecA mediated repair of stalled polymerase forks promotes recombinational repair and is purported to be a mechanism of non-natural base pair loss (mutation to natural). Deletion of the *recA* gene has been previously reported to enhance retention (i.e. fidelity) of the non-natural base pairs at the DNA level, but it comes at the cost of poorer growth. *See* Ledbetter, M. P., et al. *JACS* 140(2): p. 758–765 (2018).

[350] An experiment was performed in which seven strains with different combinations of characteristics were tested for expression of Molecule A precursor. The strains were cultured according to the Culture Process described in Example 1C. Two runs were performed. Characteristics of the seven strains—Strain E, Strain F, Strain G, Strain H, Strain B, Strain C, and Strain A—are shown in Figure 42 and in Table 3 below. In Figure 42 and Table 3, "ch" indicates the strain has a Chimeric (IPYEVF) Synthetase, "Mb" indicates the strain has a Control Synthetase (wild-type *M. barkeri* synthetase), a "C" indicates the strain has a tRNA synthetase that is chromosomally integrated, a "P" indicates the strain has a pyrrolysyl tRNA synthetase that is on a plasmid (not chromosomally integrated), a "Q" indicates the strain comprises a LacI promoter P_{lacIQ} for controlling LacI, a repressor of the inducible promoter for the pyrrolysyl tRNA synthetase; a "Q1" indicates the strain comprises a LacI promoter P_{lacIQ1} for controlling LacI, a repressor of the inducible promoter for the tRNA synthetase; an "N" indicates that LacI expression was not driven by a P_{lacIQ} or P_{lacIQ1} promoter; a "WT" indicates that the *recA* gene in the strain is wild-type; a "Δ" indicates the *recA* gene in the strain has been deleted; a "+" indicates the strain has a modified 5'UTR for chromosomally integrated pyrrolysyl tRNA synthetases; and a "-" indicates that the strain does not have the feature of an modified 5' UTR (the strain has an original 5' UTR).

Table 3							
Strain	Strain E	Strain F	Strain G	Strain H	Strain B	Strain C	Strain A
MbPylRS / chPylRS (IPYE)	ch	ch	ch	ch	ch	ch	Mb
Plasmid / Chromosomal	C	C	C	C	P	P	P
LacI promoter P_{lacIQ} / P_{lacIQ1}	Q	Q1	Q	Q1	N	N	N
recA	WT	WT	Δ	Δ	WT	WT	WT
Met39 RBS opt / 5' UTR opt	+	+	+	+	-	+	-

[351] A Western blot of the results from Run 1 is shown in Figure 42. Molecule A precursor was detected. The asterisk indicates a 3 hour induction instead of 5 hour. Titer (mg/L) results

from Run 1 are shown in Figure 43A. Fidelity (% AzK at P65) results from Run 1 are shown in Figure 43B. Final OD_{600} results from Run 1 are shown in Figure 43C. The asterisk indicates a 3 hour induction instead of 5 hour.

Example 5. Further Improvement of Fidelity and Titer, Including Nucleotide Triphosphate Transporter Expression Improvement

[352] Experiments in this Example, Example 5, were conducted with a system utilizing unnatural nucleotides unless otherwise noted. It was observed that elimination of the tetracycline resistance gene as part of the effort to chromosomally integrate the pyrrolysyl tRNA synthetase conferred enhanced strain stability. The tetracycline resistance gene had negative synergy with the nucleotide triphosphate transporter. An experiment was conducted in which three strains having different properties were tested for expression. The Culture Process described in Example 1C was used to assess strain performance. Strain F had a chromosomally integrated Chimeric (IPYEVF) Synthetase and a zeocin resistance gene. Strain D had a plasmid comprising a Chimeric (IPYEVF) Synthetase and both a zeocin resistance gene and a tetracycline resistance gene. Strain A had a plasmid comprising a MbPylRS (Control Synthetase) and both a zeocin resistance gene and a tetracycline resistance gene. The results, a Western blot shown in Figure 44, indicate that Strain F produced the most Molecule A precursor.

[353] Another experiment was conducted in which two strains, Strain F and Strain A, described above, were tested for fidelity and for colony size morphology. Isolated “big” colonies have low but non-zero fidelity (<20%). Transporter activity is greatly reduced in isolated “big” colonies. No mutations were apparent from shotgun sequencing of a “big” colony. The Culture Process described in Example 1C was used to assess strain performance. Fidelity results from the experiment are shown in Figure 45. Strain F had greater fidelity over time after many doublings than Strain A. Plates were prepared with Power Prime + KPi, dCNMOTP, dTPT3TP, and no antibiotics. Photos were taken of the plates of three different stages: RCB (Research Cell Bank), mMCB (mock Master Cell Bank), and mWCB mock Working Cell Bank), with a 1:30K dilution with power prime between each stage. When plating, a 1:100 dilution of each glycerol stock was prepared in Power Prime media. Serial dilutions were prepared to plate 100 μL of a 1:5000 dilution onto each agar plate. Colony size morphology results from the experiment are shown in Figure 46. Strain F had reduced colony size heterogeneity over time after many doublings compared to Strain A.

[354] Another experiment testing for colony size morphology was conducted with Strain F and Strain A (plus expression plasmid). The Culture Process described in Example 1C was

used to assess strain performance. An experiment was conducted in which strain lineages having differing characteristics were grown on plates with Power Prime + Kpi, dCNMOTP, dTPT3TP, and no antibiotics. Photos were taken of the plates of three different stages: RCB (Research Cell Bank), mMCB (mock Master Cell Bank), and mWCB (mock Working Cell Bank), with a 1:30K dilution with power prime between each stage. When plating, a 1:100 dilution of each glycerol stock was prepared in Power Prime media. Serial dilutions were prepared to plate 100 µL of a 1:5000 dilution onto each agar plate. The results are shown in Figure 47. Again, Strain F had reduced colony size heterogeneity over time after many doublings compared to Strain A.

[355] Two additional changes were made to Strain F to improve *Pt*NTT2 expression: 1) use of a P_{lac} promoter instead of a P_{lacUV5} promoter for *Pt*NTT2; and 2) removal of the oppositely oriented promoter within 1.3-1.5 kb downstream of the 3' end of the *Pt*NTT2 coding sequence. Diagrams showing these changes are shown in Figure 48. Strain F had a P_{lacUV5} promoter for *Pt*NTT2 and an oppositely oriented Cm^R promoter within 1.3-1.5 kb downstream of the 3' end of the *Pt*NTT2 coding sequence. Strain I had a P_{lac} promoter for *Pt*NTT2 and an oppositely oriented Cm^R promoter within 1.3-1.5 kb downstream of the 3' end of the *Pt*NTT2 coding sequence. Strain J had a P_{lac} promoter for *Pt*NTT2 and no oppositely oriented Cm^R promoter within 1.3-1.5 kb downstream of the 3' end of the *Pt*NTT2 coding sequence. An experiment was performed in which ATP export (RLU/s/OD, normalized to Strain F) was conducted with No Transporter, Strain I, Strain J, and Strain F. Overnight 2xYS starter cultures were used to inoculate fresh 2xYS media for each experiment. Cultures were incubated shaking (250 rpm) at 37°C for 2 hours before sampling. Cells were washed and added directly into the luciferase-ATP detection assay buffer. Reaction was initiated by supplementing substrate (CTP) for uptake which is coupled to ATP export. The results are shown in Figure 49. Strain J had an intermediate, "goldilocks" expression level, between that of the lower expression of Strain I and the higher expression level of Strain F. Each strain from this experiment was also plated on power prime agar with dCNMOTP and dTPT3TP for about 15 hours of growth at 37°C. Photographs of plated strains are shown in Figure 50. In addition, Molecule B % fidelity was measured for each of Strain I, Strain J, and Strain F transformed with Molecule B expression plasmid and cultured in 25 mL shake flask expressions with dNaMTP. This data is also shown in Figure 50.

[356] An experiment was performed in which fidelity (%) was measured from 25 mL shake flask over time/doublings. Fidelity was a measure of non-canonical amino acid incorporation at the intended site for each Molecule (versus natural substitutions). Fidelity was tested for

production of Molecule A (Produced according to the Culture Process described in Example 1C with dNaMTP), Molecule B (Produced according to the Culture Process described in Example 1C with dCNMOTP), and Molecule C and molecules related to Molecule C (Produced according to the Culture Process described in Example 1C with dCNMOTP). Molecule C is also an interleukin modified with an unnatural amino acid. The results are shown in Figures 51A, 51B, and 51C. The results show that elimination of the tetracycline resistance gene (by chromosomal integration) and reduction in *Pt*NTT2 expression confers further enhanced strain stability, up to ~60 doublings. Drops in fidelity are correlated with colony phenotype and reduced transporter activity. ATP export (RLU/s/OD) (normalized to Strain F) was measured for Strain F and for Strain J for production of Molecule B. Overnight 2xYS starter cultures were used to inoculate fresh 2xYS media for each experiment. Cultures were incubated shaking (250 rpm) at 37C for 2 hours before sampling. Cells were washed and added directly into the luciferase-ATP detection assay buffer. Reaction was initiated by supplementing substrate (CTP) for uptake which is coupled to ATP export. The results are shown in Figure 52A and Figure 52B. It appears that fidelity does eventually decline for Strain F-based Molecule B banks (worse in 2xYS than Power, consistent with the manifestation of the "big" colony phenotype), while fidelity does not seem to decline for Strain J-based banks generated in either media. It is also noted that Strain J does not have a chloramphenicol marker, so the stability seen (see Figure 52B) is in the absence of chloramphenicol selection. Another experiment comparing Strain F and Strain J during expansion was performed as described above for the experiment with results in Figure 52A and Figure 52B. The results are shown in Figure 53. These results demonstrated that Strain J exhibited stable morphology over propagations as compared to Strain F, which did not.

[357] Cell line populations were consistent with host cell lines. An experiment was conducted in which Strain J (host cell line) and Strain J with an expression plasmid for Molecule B was performed as described above for the experiment with results in Figures 52A and Figure 52B. Strain J (host cell line) and Strain J with an expression plasmid for Molecule B did not exhibit a morphology change across populations over propagations—i.e. it was stable over propagations. *See* Figure 54.

[358] An experiment testing for colony size morphology was conducted with Strain J and Strain F. Photos were taken of the plates of three different stages: RCB (Research Cell Bank), mMCB (mock Master Cell Bank), and mWCB (mock Working Cell Bank), with a 1:30K dilution with power prime between each stage. When plating, a 1:100 dilution of each glycerol stock was prepared in Power Prime media. Serial dilutions were prepared to plate 100 µL of a

1:5000 dilution onto each agar plate. Strain F, with chromosomally integrated pyrrolysyl tRNA synthetase and no tetracycline resistance gene, had more "big" colonies after many doublings compared to Strain J, which has the same characteristics as Strain F except: 1) a P_{lac} promoter instead of a P_{lacUV5} promoter for PtNTT2; and 2) removal of Cm^R terminator for *Pt*NTT2. The results are shown in Figure 55. The results indicated that colony morphology of strain J was consistent across excessive propagations.

[359] Another experiment testing for colony size heterogeneity was conducted with Strain J and Strain F. Photos were taken of the plates of different stages: RCB, mMCB, mWCB, mMCB (propagated in Soytone media; Soytone is the same as 2xYS media), and mWCB (propagated in Soytone media from the mMCB propagated in Soytone media), with a 1:30K dilution between each stage. When plating, a 1:100 dilution of each glycerol stock was prepared in Power Prime media. Serial dilutions were prepared to plate 100 µL of a 1:5000 dilution onto each agar plate. The results are shown in Figure 56. In general, colonies of Strain J look as big as the "big" colonies for Strain F—even the colonies with good transporter expression. The photos in Figure 56 show that Strain J morphology is stable across propagations and regardless of media type. On the other hand, Strain F is stably small only in PowerPrime media, whereas in Soytone/2xYS, "big" colonies (that lost transporter activity) appeared by the time of the WCB propagation. These results demonstrate that Strain J is more stable regardless of media type.

Example 6. Triple Mutant – *recA* Hypomorph, PolB Constitutive Expression, and LexA Cleavability Modification

[360] Experiments in this Example, Example 6, were conducted with a system utilizing unnatural nucleotides unless otherwise noted. Three mutations were made to Strain J to make a "triple mutant" strain, Strain K. First, a *recA* hypomorph was produced via the following modification: *recA-terΔ-recX*. Second, Pol II was derepressed. Third, a modification was made to LexA ($LexA^{S119A}$) to make it uncleavable, and thus, not able to induce the SOS response.

[361] An experiment was performed to compare expression results for the triple mutant, Strain K, with Strain J under limiting deoxyribonucleotide conditions. Molecule B was produced. The experiments were conducted in 25 mL shake flasks according to a modified version of the Culture Process described in Example 1C , modified to limit deoxyribonucleotides. In the unmodified Culture Process, deoxyribonucleotides are added three times in the process—at the beginning before culture, pre-induction (bolus 1), and again post-induction (bolus 2). For these experiments, deoxyribonucleotides were added initially, but one or both of the second (i.e., pre-induction) and third (i.e., post-induction) additions were

skipped. Results are shown in Figure 57A-57D. In Figure 57A-57D, "both" means both boluses were skipped; "none" means none were skipped, "pre-ind" means the pre-induction bolus 1 was skipped; and "post-ind" means that post-induction bolus 2 was skipped. In addition, the starting culture deoxyribonucleotide concentration was about 10-fold lower in the modified Culture Process compared to the unmodified Culture Process. One set of experiments was performed to investigate the effect of substituting dCNMOTP for dNaMTP. Figure 57A shows fidelity (% AzK at H16); Figure 57B shows titer (mg/L), Figure 57C shows Titer / OD (mg/L/OD), and Figure 57D shows final OD.

[362] Titer declined with dNaMTP reduction in Strain J. The titer unexpectedly increased with the triple mutant in the context of using dNaMTP. The triple mutant, Strain K, increased productivity (in terms of titer and/or OD) in every pairwise comparison.

[363] The experiment to compare expression results for the triple mutant, Strain K, with Strain J under the same limiting deoxynucleotide conditions described above was performed again to produce Molecule B, except that induction was at a higher optical density. One set of experiments was performed to investigate the effect of substituting dCNMOTP for dNaMTP. Results are shown in Figures 58A-D: Figure 58A shows fidelity (% AzK at H16); Figure 58B shows titer (mg/L), Figure 58C shows Titer / OD (mg/L/OD), and Figure 58D shows final OD. For dCNMOTP, where the data state "not determined," it means that particular experiment was not run). Even at equimolar concentrations, dCNMOTP was superior to dNaMTP.

Example 7. Further Assessment of Strain Performance

[364] Experiments in this Example, Example 7, were conducted with a system utilizing unnatural nucleotides unless otherwise noted.

[365] In one experiment, Strain J cells were cultured according to the Culture Process described in Example 1C to investigate the effect of substituting dCNMOTP for dNaMTP in producing Molecule B. Fidelity in these experiments was a measure of ncAA incorporation at the intended site (versus natural substitutions). Misincorporation in these experiments was % of species with two ncAA, one at the intended site and one at a site corresponding to a near-cognate natural codon (GTC Val, GCC Ala, GAC Asp, GGC Gly). Misincorporation was not part of the fidelity calculation, i.e., fidelity is not purity. Expression of the protein product, Molecule B, was comparable in terms of titer (mg/L), as shown in Figure 59A, and fidelity, as shown in Figure 59B.

[366] An experiment was conducted to determine whether increasing the expression of pyrrolysyl tRNA synthetase would improve Strain J's tolerance to AzK addition conditions during the culture process. Different promoters were used for the synthetase-TacO, 23119 (also

known as J23119), 0.26x T7, and 3.65x T7. The unnatural deoxyribonucleotide (dCNMOTP) was used. The Culture Process described in Example 1C was used, with modifications to the amount of AzK. Two boluses of AzK were experimented with, at three different possible concentrations for each bolus, relative to standard concentration: standard, half, or quarter AzK additions were not equivolumetric, but were unlikely to be confounding. There was no detectable synthetase in inclusion body (IB), even for the strongest promoter, based on HPLC Titer (data not shown). The results are shown in Figure 60A-60D. The results indicate that the additional pyrrolysyl tRNA synthetase expression did not further improve Strain J's tolerance to AzK addition conditions during the culture process.

EQUIVALENTS

[367] The foregoing written specification is considered to be sufficient to enable one skilled in the art to practice the embodiments. The foregoing description and Examples detail certain embodiments and describes the best mode contemplated by the inventors. It will be appreciated, however, that no matter how detailed the foregoing may appear in text, the embodiment may be practiced in many ways and should be construed in accordance with the appended claims and any equivalents thereof.

[368] As used herein, the term about refers to a numeric value, including, for example, whole numbers, fractions, and percentages, whether or not explicitly indicated. The term about generally refers to a range of numerical values (e.g., +/-5-10% of the recited range) that one of ordinary skill in the art would consider equivalent to the recited value (e.g., having the same function or result). When terms such as at least and about precede a list of numerical values or ranges, the terms modify all of the values or ranges provided in the list. In some instances, the term about may include numerical values that are rounded to the nearest significant figure.

What is Claimed is:

1. A prokaryotic cell comprising:
   a first chromosomally integrated DNA sequence encoding a transfer RNA synthetase (tRNA synthetase) operably linked to an inducible promoter and a 5' UTR;
   a second DNA sequence encoding a promoter for a repressor of an inducible promoter,
   a third DNA sequence encoding at least one unnatural transfer RNA (tRNA) molecule;
   a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising a codon recognized by the unnatural tRNA molecule;
   wherein the 5' UTR enables increased tRNA synthetase expression compared to SEQ ID NO: 1.

2. A prokaryotic cell comprising:
   a DNA sequence encoding a *Pt*NTT2 operably linked to a P_{lac} promoter, wherein (i) the DNA sequence does not comprise an oppositely oriented promoter within about 1.3-1.5 kb downstream of the 3' end of the *Pt*NTT2 coding sequence or (ii) the DNA sequence comprises a transcriptional terminator that terminates anterograde *Pt*NTT2 transcription, and further comprises a transcription terminator element downstream of the stop codon of the DNA sequence encoding the *Pt*NTT2 and the transcription terminator element is oriented to terminate retrograde transcription into the DNA sequence encoding the *Pt*NTT2.

3. A prokaryotic cell comprising:
   a DNA sequence encoding a tRNA synthetase;
   a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed;
   a DNA sequence encoding a non-self-cleaving LexA, optionally wherein the LexA comprises a S119 mutation, optionally wherein the S119 mutation is S119A; and
   a *recA* hypomorph, optionally wherein the *recA* hypomorph results from RecX overexpression.

4. The cell of claim 1, wherein the unnatural tRNA molecule is an unnatural pyrrolysyl tRNA molecule.

5. The cell of claim 1 or claim 4, wherein the 5' UTR comprises the DNA sequence of SEQ ID NO: 2.

6. The cell of any one of claims 1, 4, or 5, wherein the second DNA sequence is chromosomally integrated.

7. The cell of any one of claims 1 or 3-6, wherein the tRNA synthetase comprises a chimeric pyrrolysyl-tRNA synthetase comprising a pyrrolysyl-tRNA synthetase N-terminal domain from a first species and a pyrrolysyl-tRNA synthetase C-terminal domain from a second species, optionally wherein the first species is *Methanosarcina barkeri*, and the second species is *Methanosarcina mazei*, further optionally wherein the tRNA synthetase comprises a chimeric pyrrolysyl-tRNA synthetase comprising a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD).

8. The cell of any one of claims 1 or 3-6, wherein the tRNA synthetase comprises a split chimeric pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the first polypeptide chain comprises a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD), and wherein the second polypeptide chain comprises a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD).

9. The cell of any one of claims 1, 3-6, or 8, wherein the tRNA synthetase comprises a split chimeric pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) comprises an amino acid sequence having at least 80% identity to positions 1-137 of SEQ ID NO: 4, and wherein the *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD) comprises an amino acid sequence having at least 80% identity to positions 150-419 of SEQ ID NO: 4, optionally wherein the C-terminal amino acid of the *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase NTD corresponds to P137 or S149 of SEQ ID NO: 4.

10. The cell of any one of claims 1 or 3-6, wherein the tRNA synthetase comprises a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase.

11. The cell of any one of claims 1 or 3-6, wherein the tRNA synthetase comprises a split *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the first polypeptide chain comprises a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD), and wherein the second polypeptide chain comprises a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase C-terminal domain (CTD).

12. The cell of claim any one of claims 1, 3-6, or 11, wherein the tRNA synthetase comprises a split *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) comprises an amino acid sequence having at least 80% identity to positions 1-110, 1-137, or 1-149 of SEQ ID NO: 3, and wherein the *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase C-terminal domain (CTD) comprises an amino acid sequence having at least 80% identity to positions 150-419 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase NTD corresponds to S110, P137, or S149 of SEQ ID NO: 3.

13. A cell comprising:
a DNA sequence encoding a split *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the first polypeptide chain comprises part of a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 1-110, 1-137, or 1-149 of SEQ ID NO: 3, and wherein the second polypeptide chain comprises part of a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 150-419 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to S110, P137, or S149 of SEQ ID NO: 3.

14. The cell of claim 13, wherein the split *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprises an amino acid sequence having at least 80% identity to positions 1-110 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to S110.

15. The cell of claim 13, wherein the split *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprises an amino acid sequence having at least 80% identity to positions 1-137 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to P137.

16. The cell of claim 13, wherein the split *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprises an amino acid sequence having at least 80% identity to positions 1-149 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to S149.

17. A cell comprising:

a DNA sequence encoding a split chimeric pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the first polypeptide chain comprises part of a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 1-137 of SEQ ID NO: 4, and wherein the second polypeptide chain comprises part of a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 150-419 of SEQ ID NO: 4, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to P137 of SEQ ID NO: 4.

18. A cell comprising:

a DNA sequence encoding a split pyrrolysyl-tRNA synthetase comprising two separate polypeptide chains, wherein the first polypeptide subunit comprises part of a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 1-138 or 1-172 of SEQ ID NO: 5, and wherein the second polypeptide subunit comprises part of a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 185-454 of SEQ ID NO: 5, optionally wherein the C-terminal amino acid of the first polypeptide subunit corresponds to I138 or N172 and of SEQ ID NO: 5.

19. The cell of any one of claims 7-9 or 17, wherein the tRNA synthetase comprises at least the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, and A100E.

20. The cell of any one of claims 7-9, 17, or 19, wherein the tRNA synthetase comprises an amino acid sequence having at least 80% identity to SEQ ID NO: 6.

21. The cell of any one of claims 7-9, 17, 19, or 20, wherein the tRNA synthetase comprises an amino acid sequence having at least 80% identity to SEQ ID NO: 7.

22. The cell of any one of claims 1 or 3-21, wherein the pyrrolysyl-tRNA synthetase is fused to *Escherichia coli* thioredoxin.

23. The cell of any one of claims 1, 4-7, or 19-22, wherein the first sequence comprises a DNA sequence that is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 8.

24. The cell of any one of claims 1, 4-7, 19, or 20, wherein the first sequence comprises a DNA sequence comprising SEQ ID NO: 8.

25. The cell of any one of claims 1, 2, or 4-24, wherein the cell does not comprise a functional *recA* gene.

26. A prokaryotic cell comprising:
a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to an inducible promoter, and a 5' UTR comprising SEQ ID NO: 2;
a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ1} promoter;
a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid; and
a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising a codon recognized by the unnatural pyrrolysyl tRNA, wherein the fourth DNA sequence is on a plasmid; and a fifth DNA sequence encoding a nucleoside triphosphate transporter operably linked to a P_{lacUV5} promoter, optionally wherein the prokaryotic cell further comprises an antibiotic resistance gene.

27. A prokaryotic cell comprising:

a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to an inducible promoter, and a 5' UTR comprising SEQ ID NO: 2;

a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ1} promoter;

a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid; and

a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising an unnatural codon recognized by the unnatural pyrrolysyl tRNA, wherein the fourth DNA sequence is on a plasmid, and a DNA sequence encoding a nucleoside triphosphate transporter, wherein the DNA sequence encoding the nucleoside triphosphate transporter is operably linked to a P_{lac} promoter and does not comprise an oppositely oriented promoter within 1.3-1.5 kb downstream of the 3' end of the nucleoside triphosphate transporter coding sequence.

28. A prokaryotic cell comprising:

a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to an inducible promoter, and a 5' UTR comprising SEQ ID NO: 2;

a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ1} promoter;

a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid;

a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising an unnatural codon recognized by the unnatural pyrrolysyl tRNA, wherein the fourth DNA sequence is on a plasmid;

a fifth DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed;

a sixth DNA sequence encoding a non-self-cleaving LexA, optionally wherein the LexA comprises a S119 mutation, optionally wherein the S119 mutation is S119A;

a DNA sequence encoding a nucleoside triphosphate transporter, wherein the DNA sequence encoding the nucleoside triphosphate transporter is operably linked to a P_{lac} promoter and does not comprise an oppositely oriented promoter within 1.3-1.5 kb downstream of the 3' end of the nucleoside triphosphate transporter coding sequence; and

a *recA* hypomorph, optionally wherein the *recA* hypomorph results from RecX overexpression.

29. A prokaryotic cell comprising:
a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to an inducible promoter, and a 5' UTR comprising SEQ ID NO: 2;
a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ} promoter;
a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid;
a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising an unnatural codon recognized by the unnatural pyrrolysyl tRNA molecule, wherein the fourth DNA sequence is on a plasmid.

30. A prokaryotic cell comprising:
a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to an inducible promoter, and a 5' UTR comprising SEQ ID NO: 2;
a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ1} promoter;
a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid;
a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising a codon recognized by the unnatural pyrrolysyl tRNA molecule, wherein the fourth DNA sequence is on a plasmid;
wherein the cell does not comprise a functional *recA* gene.

31. A prokaryotic cell comprising:
a first chromosomally integrated DNA sequence encoding a tRNA synthetase operably linked to a inducible promoter, and a 5' UTR comprising SEQ ID NO: 2;
a second DNA sequence encoding a promoter for a repressor of an inducible promoter, wherein the promoter is a P_{lacIQ} promoter;
a third DNA sequence encoding at least one unnatural pyrrolysyl tRNA molecule, wherein the third DNA sequence is on a plasmid;
a fourth DNA sequence encoding at least one unnatural mRNA molecule comprising a

codon recognized by the unnatural pyrrolysyl tRNA molecule, wherein the fourth DNA sequence is on a plasmid;
wherein the cell does not comprise a functional *recA* gene.

32. The cell of any one of claims 1-31, wherein the cell comprises a DNA sequence encoding a nucleoside triphosphate transporter, optionally

wherein the nucleoside triphosphate transporter comprises a truncated *Pt*NTT2.

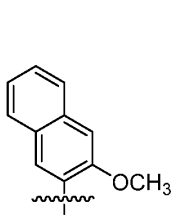
33. The cell of any one of claims 1, 2, 4-27, or 29-32, further comprising a DNA sequence encoding Polymerase II (Pol II) operably linked to a promoter, wherein the promoter is derepressed.

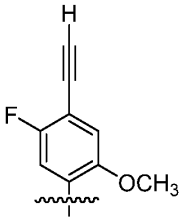
34. The cell of any one of claims 1, 2, 4-27, or 29-33, further comprising a DNA sequence encoding a LexA that does not undergo RecA-stimulated cleavage.

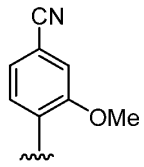
35. The cell of any one of claims 1, 2, 4-24, 26, 27, 29, or 32-34, further comprising a *recA* hypomorph.

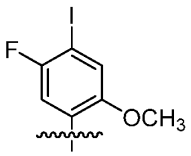
36. The cell of any one of claims 1, 4-12, or 19-35,
wherein the third DNA sequence encoding at least one unnatural tRNA molecule comprises a first unnatural nucleotide comprising a first unnatural base; and
wherein the fourth DNA sequence encoding at least one unnatural mRNA molecule comprises a second unnatural nucleotide comprising a second unnatural base.

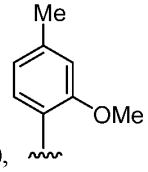
37. The cell of claim 36, wherein the first unnatural nucleotide and the second unnatural nucleotide each comprise an unnatural base independently selected from the group consisting of:

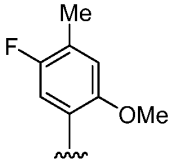


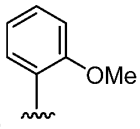
(NaM),


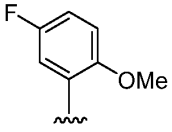
(FEMO),


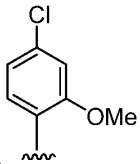
(CNMO),


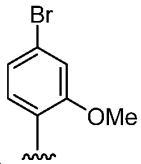


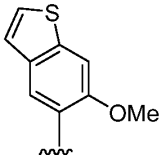
(FIMO),
(MMO2),


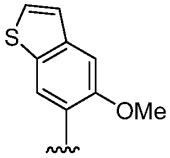


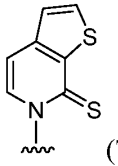
(5FM),
(2OMe),


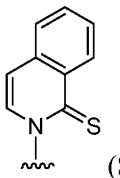


(5F2OMe),
(ClMO),


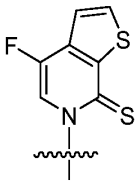
(BrMO),


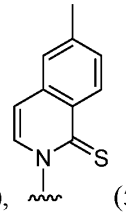
(PTMO),


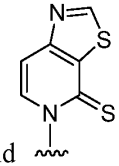
(MTMO),


(TPT3),


(SICS),


(FSICS),


(FTPT3),


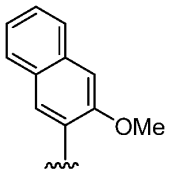
(5SICS), and


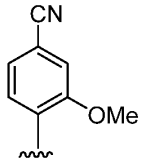
(TAT1).

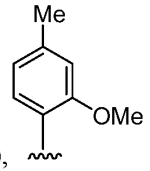
38. The cell of any one of claims 1, 4-12, or 19-35, wherein the codon recognized by the unnatural tRNA, when read from a 5' to 3' direction, comprises NNX or NXN, wherein N is any natural nucleotide, and X is an unnatural nucleotide.

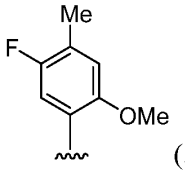
39. The cell of claim 38, wherein the codon recognized by the unnatural tRNA read from a 5' to 3' direction comprises UUX, UGX, CGX, AGX, GAX, CAX, AUX, CUX, GUX, UAX, GGX, GXU, CXU, GXG, AXG, GXC, AXC, GXA, CXC, or UXC.

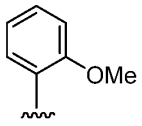
40. The cell of claim 38 or 39, wherein X comprises any one of the following bases:

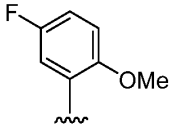


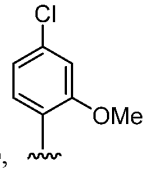
(NaM),


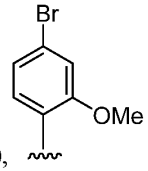
(CNMO),


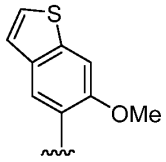
(MMO2),


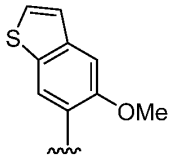
(5FM),


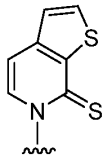
(2OMe),


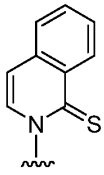
(5F2OMe),


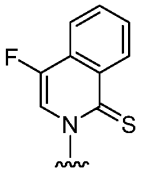
(ClMO),


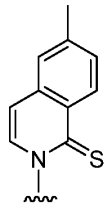
(BrMO),


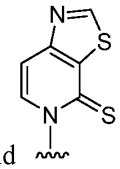
(PTMO),


(MTMO),


(TPT3),


(SICS),


(FSICS),


(5SICS), and


(TAT1).

41. The cell of any one of claims 2, 3, 7-25, or 32-35, wherein the cell comprises unnatural dNTPs.

42. The cell of any one of claims 1-41, wherein the cell comprises at least one unnatural amino acid.

43. The cell of claim 42, wherein the unnatural amino acid comprises
a lysine analogue;
an aromatic side chain;
an azido group;
an alkyne group; or
an aldehyde or ketone group.

44. A pyrrolysyl-tRNA synthetase comprising:

(a) an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: V31I, T56P, H62Y, A100E, C313V, and Y349F;

(b) an amino acid sequence having at least 80% identity to SEQ ID NO: 6, wherein the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 31 of SEQ ID NO: 6 is isoleucine; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 56 of SEQ ID NO: 6 is proline; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 62 of SEQ ID NO: 6 is tyrosine; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 100 of SEQ ID NO: 6 is glutamate; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 313 of SEQ ID NO: 6 is valine; and the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 349 of SEQ ID NO: 6 is phenylalanine;

(c) an *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase N-terminal domain (NTD) fused to an *Methanosarcina mazei* pyrrolysyl-tRNA synthetase C-terminal domain (CTD), wherein the tRNA synthetase comprises the following mutations with reference to SEQ ID NO: 4: P5T or P5L, E302A, V31I, T56P, H62Y, A100E, C313V, and Y349F;

(d) an amino acid sequence having at least 80% identity to SEQ ID NO: 7, wherein the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 5 of SEQ ID NO: 6 is threonine or leucine; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 31 of SEQ ID NO: 6 is isoleucine; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 56 of SEQ ID NO: 6 is proline; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 62 of SEQ ID NO: 6 is tyrosine; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 100 of SEQ ID NO: 6 is glutamate; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 302 of SEQ ID NO: 6 is alanine; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to

position 313 of SEQ ID NO: 6 is valine; the position of the chimeric pyrrolysyl-tRNA synthetase corresponding to position 349 of SEQ ID NO: 6 is phenylalanine;

(e) two separate polypeptide chains, wherein the first polypeptide chain comprises part of a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 1-137 of SEQ ID NO: 4, and wherein the second polypeptide chain comprises part of a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 150-419 of SEQ ID NO: 4, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to P137 of SEQ ID NO: 4;

(f) two separate polypeptide chains, wherein the first polypeptide subunit comprises part of a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 1-110, 1-137, or 1-149 of SEQ ID NO: 3, and wherein the second polypeptide chain comprises part of a *Methanosarcina barkeri* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 150-419 of SEQ ID NO: 3, optionally wherein the C-terminal amino acid of the first polypeptide chain corresponds to S110, P137, or S149 of SEQ ID NO: 3; or

(g) two separate polypeptide chains, wherein the first polypeptide subunit comprises part of a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 1-138 or 1-172 of SEQ ID NO: 5, and wherein the second polypeptide subunit comprises part of a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase comprising an amino acid sequence having at least 80% identity to positions 185-454 of SEQ ID NO: 5, optionally wherein the C-terminal amino acid of the first polypeptide subunit corresponds to I138 or N172 of SEQ ID NO: 5.

45. An isolated nucleic acid encoding the tRNA synthetase of claim 44.

46. A vector comprising the nucleic acid of claim 45.

47. A host cell comprising the vector of claim 46.

48. A host cell that produces the tRNA synthetase of claim 44.

49. A method for making a tRNA synthetase, comprising culturing the host cell of claim 47 or 48 under conditions suitable for expression of the tRNA synthetase.

50. A method of preparing a cell that can express a polypeptide comprising an unnatural amino acid with improved titer and/or fidelity, the cell comprising one or more of a DNA sequence encoding LexA, a DNA sequence encoding RecA and RecX, a DNA sequence encoding Pol II, a DNA sequence encoding tRNA synthetase that can charge a tRNA with an unnatural amino acid, and/or a DNA sequence encoding *Pt*NTT2; and the method comprising one or more of:
modifying the DNA sequence encoding LexA so that LexA does not undergo RecA-stimulated cleavage;
modifying the cell to decrease expression of RecA;
modifying the cell to increase expression of Pol II;
modifying the cell to decrease expression of *Pt*NTT2 without eliminating *Pt*NTT2 expression; and/or
modifying the cell to increase expression of the tRNA synthetase.