

# COMPOSITIONS OF tRNA AND USES THEREOF

## CROSS-REFERENCE TO RELATED APPLICATIONS

[01] This application claims priority to U.S. provisional patent application Serial No. 60/709,364, filed August 18, 2005, the specification of which is incorporated herein in its entirety.

## FIELD OF THE PRESENT INVENTION

[02] The invention pertains to the field of translation biochemistry. The invention relates to methods for producing and compositions of tRNA and uses thereof. The invention also relates to methods of producing proteins in cells using such tRNA and related compositions.

## BACKGROUND OF THE PRESENT INVENTION

[03] The genetic code of every known organism, from bacteria to humans, encodes the same twenty common amino acids. Different combinations of the same twenty natural amino acids form proteins that carry out virtually all the complex processes of life, from photosynthesis to signal transduction and the immune response. In order to study and modify protein structure and function, scientists have attempted to manipulate both the genetic code and the amino acid sequence of protein. However, it has been difficult to remove the constraints imposed by the genetic code that limit proteins to twenty genetically encoded standard building blocks (with the rare exception of selenocysteine (*see, e.g.*, A. Bock et al., (1991), Molecular Microbiology 5:515-20) and pyrrolysine (*see, e.g.*, G. Srinivasan, et al., (2002), Science 296:1459-62).

[04] Some progress has been made to remove these constraints, although this progress has been limited and the ability to rationally control protein structure and function is still in its infancy. For example, chemists have developed methods and strategies to synthesize and manipulate the structures of small molecules (*see, e.g.*, E. J. Corey, & X.-M. Cheng, The Logic of Chemical Synthesis (Wiley-Interscience, New York, 1995)). Total synthesis (*see, e.g.*, B. Merrifield, (1986), Science 232:341-7 (1986)), and semi-synthetic methodologies (*see, e.g.*, D. Y. Jackson et al., (1994) Science 266:243-7; and, P. E. Dawson, & S. B. Kent, (2000), Annual Review of Biochemistry 69:923-60), have made it possible to synthesize peptides and small proteins, but these methodologies have limited utility with proteins over 10 kilo Daltons (kDa). Mutagenesis methods, though powerful, are restricted to a limited

number of structural changes. In a number of cases, it has been possible to competitively incorporate close structural analogues of common amino acids throughout proteins. *See, e.g.*, R. Furter, (1998), Protein Science 7:419-26; K. Kirshenbaum, et al., (2002), ChemBioChem 3:235-7; and, V. Doring et al., (2001), Science 292:501-4. Chemical peptide ligation and native chemical ligation are described in U.S. Patent No. 6,184,344, U.S. Patent Publication No. 2004/0138412, U.S. Patent Publication No. 2003/0208046, WO 02/098902, and WO 03/042235, which are incorporated by reference herein. Lu et al. in *Mol Cell*. 2001 Oct;8(4):759-69 describe a method in which a protein is chemically ligated to a synthetic peptide containing unnatural amino acids (expressed protein ligation).

[05] Early work demonstrated that the translational machinery of *E. coli* would accommodate amino acids similar in structure to the common twenty. *See*, Hortin, G., and Boime, I. (1983) Methods Enzymol. 96:777-784. This work was further extended by relaxing the specificity of endogenous *E. coli* synthetases so that they activate unnatural amino acids as well as their cognate natural amino acid. Moreover, it was shown that mutations in editing domains could also be used to extend the substrate scope of the endogenous synthetase. *See*, Doring, V., et al., (2001) Science 292:501-504. However, these strategies are limited to *recoding* the genetic code rather than *expanding* the genetic code and lead to varying degrees of substitution of one of the common twenty amino acids with an unnatural amino acid.

[06] Later it was shown that unnatural amino acids could be site-specifically incorporated into proteins *in vitro* by the addition of chemically aminoacylated orthogonal amber suppressor tRNA's to an *in vitro* transcription/translation reaction. *See, e.g.*, Noren, C. J., et al. (1989) Science 244:182-188; Bain, J. D., et al., (1989) J. Am. Chem. Soc. 111:8013-8014; Dougherty, D. A. (2000) Curr. Opin. Chem. Biol. 4, 645-652; Cornish, V. W., et al. (1995) Angew. Chem., Int. Ed. 34:621-633; J. A. Ellman, et al., (1992), Science 255:197-200; and, D. Mendel, et al., (1995), Annual Review of Biophysics and Biomolecular Structure 24:435-462. These studies show that the ribosome and translation factors are compatible with a large number of unnatural amino acids, even those with unusual structures. Unfortunately, the chemical aminoacylation of tRNA's is difficult, and the stoichiometric nature of this process severely limited the amount of protein that could be generated.

[07] Unnatural amino acids have been microinjected into cells. For example, unnatural amino acids were introduced into the nicotinic acetylcholine receptor in *Xenopus* oocytes (e.g., M.W. Nowak, et al. (1998), *In vivo incorporation of unnatural amino acids into ion channels in Xenopus oocyte expression system*, Method Enzymol. 293:504-529) by

microinjection of a chemically misacylated *Tetrahymena thermophila* tRNA (e.g., M.E. Saks, et al. (1996), *An engineered Tetrahymena tRNA<sup>Gln</sup> for in vivo incorporation of unnatural amino acids into proteins by nonsense suppression*, J. Biol. Chem. 271:23169-23175), and the relevant mRNA. See, also, D.A. Dougherty (2000), *Unnatural amino acids as probes of protein structure and function*, Curr. Opin. Chem. Biol. 4:645-652 and M. W. Nowak, P. C. Kearney, J. R. Sampson, M. E. Saks, C. G. Labarca, S. K. Silverman, W. G. Zhong, J. Thorson, J. N. Abelson, N. Davidson, P. G. Schultz, D. A. Dougherty and H. A. Lester, Science, 268:439 (1995). A *Xenopus* oocyte was coinjected with two RNA species made in vitro: an mRNA encoding the target protein with a UAG stop codon at the amino acid position of interest and an amber suppressor tRNA aminoacylated with the desired unnatural amino acid. The translational machinery of the oocyte then inserts the unnatural amino acid at the position specified by UAG. Unfortunately, this methodology is limited to proteins in cells that can be microinjected, and because the relevant tRNA is chemically acylated in vitro, and cannot be re-acylated, the yields of protein are very low.

[08] To overcome these limitations, new components, e.g., orthogonal tRNA's, orthogonal aminoacyl-tRNA synthetases and pairs thereof, were added to the protein biosynthetic machinery of the prokaryote *Escherichia coli* (*E. coli*) (see e.g., L. Wang, et al., (2001), Science 292:498-500) and the eukaryote *Saccharomyces cerevisiae* (*S. cerevisiae*) (e.g., J. Chin et al., Science 301:964-7 (2003)) which has enabled the incorporation of non-genetically encoded amino acids to proteins *in vivo*. A number of new amino acids with novel chemical, physical or biological properties, including photoaffinity labels and photoisomerizable amino acids, photocrosslinking amino acids (see, e.g., Chin, J. W., et al. (2002) Proc. Natl. Acad. Sci. U. S. A. 99:11020-11024; and, Chin, J. W., et al., (2002) J. Am. Chem. Soc. 124:9026-9027), keto amino acids (see, e.g., Wang, L., et al., (2003) Proc. Natl. Acad. Sci. U. S. A. 100:56-61 and Zhang, Z. et al., Biochem. 42(22):6735-6746 (2003)), heavy atom containing amino acids, and glycosylated amino acids have been incorporated efficiently and with high fidelity into proteins in *E. coli* and in yeast in response to, e.g., the amber codon (TAG), using this methodology. See, e.g., J. W. Chin, & P. G. Schultz, (2002), ChemBioChem 3(11):1135-1137 and, L. Wang, & P. G. Schultz, (2002), Chem. Comm., 1:1-11.

[09] Several other orthogonal pairs have been reported. Glutaminyl (see, e.g., Liu, D. R., and Schultz, P. G. (1999) Proc. Natl. Acad. Sci. U. S. A. 96:4780-4785), aspartyl (see, e.g., Pasternak, M., et al., (2000) Helv. Chim. Acta 83:2277-2286), and tyrosyl (see, e.g.,

Ohno, S., et al., (1998) J. Biochem. (Tokyo, Jpn.) 124:1065-1068; and, Kowal, A. K., et al., (2001) Proc. Natl. Acad. Sci. U. S. A. 98:2268-2273) systems derived from *S. cerevisiae* tRNA's and synthetases have been described for the potential incorporation of unnatural amino acids in *E. coli*. Systems derived from the *E. coli* glutaminy (see, e.g., Kowal, A. K., et al., (2001) Proc. Natl. Acad. Sci. U. S. A. 98:2268-2273) and tyrosyl (see, e.g., Edwards, H., and Schimmel, P. (1990) Mol. Cell. Biol. 10:1633-1641) synthetases have been described for use in *S. cerevisiae*. The *E. coli* tyrosyl system has been used for the incorporation of 3-iodo-L-tyrosine *in vivo*, in mammalian cells. See, Sakamoto, K., et al., (2002) Nucleic Acids Res. 30:4692-4699. Typically, these systems have made use of the amber stop codon. To further expand the genetic code, there is a need to develop improved and/or additional components of the biosynthetic machinery, e.g., tRNA's. This invention fulfills these and other needs, as will be apparent upon review of the following disclosure.

#### **SUMMARY OF THE PRESENT INVENTION**

[10] To expand the genetic code, the invention provides compositions of and methods of producing orthogonal tRNA's. Aminoacyl-tRNA synthetases aminoacylate tRNA's of the present invention with a non-naturally encoded amino acid. These translational components can be used to incorporate a selected amino acid in a specific position in a growing polypeptide chain (during nucleic acid translation) in response to a selector codon that is recognized by the tRNA.

[11] Methods of producing a protein in a cell with a selected amino acid at a specified position are also a feature of the present invention. For example, a method includes growing, in an appropriate medium, a cell, where the cell comprises a nucleic acid that comprises at least one selector codon and encodes a protein; and, providing the selected amino acid. The cell further comprises: an orthogonal tRNA (O-tRNA) that functions in the cell and recognizes the selector codon; and, an orthogonal aminoacyl-tRNA synthetase (O-RS) that preferentially aminoacylates the O-tRNA with the selected amino acid. Typically, the O-tRNA comprises suppression activity in the presence of a cognate synthetase. A protein produced by this method is also a feature of the present invention.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

[12] Figure 1 – The cloverleaf structure of J17 tRNA with TΨC stem mutation sites is shown.

[13] Figure 2 – Supression of an amber mutation in human growth hormone is shown using J17 or J17 mutants (F12, F13, F14). Total cell lysate for each sample was analyzed by SDS PAGE.

[14] Figure 3 – Supression of an amber mutation in human growth hormone is shown in different cell lines using F13.

#### DEFINITIONS

[15] Before describing the invention in detail, it is to be understood that this invention is not limited to particular biological systems, which can, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention, which will be limited only by the appended claims. As used herein and in the appended claims, the singular forms "a", "an" and "the" include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to "a cell" includes a combination of two or more cells and includes equivalents thereof known to those of ordinary skill in the art, and so forth. Reference to "bacteria" includes mixtures of bacteria, and the like.

[16] Unless defined herein and below in the remainder of the specification, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains.

[17] All publications and patents mentioned herein are incorporated herein by reference for the purpose of describing and disclosing, for example, the constructs and methodologies that are described in the publications, which might be used in connection with the presently described invention. The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention or for any other reason.

[18] Homologous: Proteins and/or protein sequences are "homologous" when they are derived, naturally or artificially, from a common ancestral protein or protein sequence. Similarly, nucleic acids and/or nucleic acid sequences are homologous when they are derived, naturally or artificially, from a common ancestral nucleic acid or nucleic acid sequence. For example, any naturally occurring nucleic acid can be modified by any available mutagenesis method to include one or more selector codon. When expressed, this mutagenized nucleic acid encodes a polypeptide comprising one or more selected amino acid, e.g. unnatural amino

acid. The mutation process can, of course, additionally alter one or more standard codon, thereby changing one or more standard amino acid in the resulting mutant protein as well. The one or more standard amino acid may be changed to an unnatural amino acid or a natural amino acid. Homology is generally inferred from sequence similarity between two or more nucleic acids or proteins (or sequences thereof). The precise percentage of similarity between sequences that is useful in establishing homology varies with the nucleic acid and protein at issue, but as little as 25% sequence similarity is routinely used to establish homology. Higher levels of sequence similarity, e.g., 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more, can also be used to establish homology. Methods for determining sequence similarity percentages (e.g., BLASTP and BLASTN using default parameters) are described herein and are generally available.

[19]        Orthogonal: As used herein, the term "orthogonal" refers to a molecule (e.g., an orthogonal tRNA (O-tRNA) and/or an orthogonal aminoacyl tRNA synthetase (O-RS)) that is used with reduced efficiency by a system of interest (e.g., a translational system, e.g., a cell). Orthogonal refers to the inability or reduced efficiency, e.g., less than 20% efficient, less than 10% efficient, less than 5% efficient, or e.g., less than 1% efficient, of an orthogonal tRNA and/or orthogonal RS to function in the translation system of interest. For example, an orthogonal tRNA in a translation system of interest is aminoacylated by any endogenous RS of a translation system of interest with reduced or even zero efficiency, when compared to aminoacylation of an endogenous tRNA by an endogenous RS. In another example, an orthogonal RS aminoacylates any endogenous tRNA in the translation system of interest with reduced or even zero efficiency, as compared to aminoacylation of the endogenous tRNA by an endogenous RS. A second orthogonal molecule can be introduced into the cell that functions with the first orthogonal molecule. For example, an orthogonal tRNA/RS pair includes introduced complementary components that function together in the cell with an efficiency (e.g., about 50% efficiency, about 60% efficiency, about 70% efficiency, about 75% efficiency, about 80% efficiency, about 85% efficiency, about 90% efficiency, about 95% efficiency, or about 99% or more efficiency) to that of a corresponding tRNA/RS endogenous pair. "Improvement in orthogonality" refers to enhanced orthogonality compared to a starting material or a naturally occurring tRNA or RS.

[20] Cognate: The term “cognate” refers to components that function together, e.g., a tRNA and an aminoacyl-tRNA synthetase. The components can also be referred to as being complementary.

[21] Preferentially aminoacylates: The term “preferentially aminoacylates” refers to an efficiency, e.g., about 70% efficient, about 75% efficient, about 80% efficient, about 85% efficient, about 90% efficient, about 95 % efficient, or about 99% or more efficient, at which an O-RS aminoacylates an O-tRNA with a selected amino acid, e.g., an unnatural amino acid, compared to the O-RS aminoacylating a naturally occurring tRNA or a starting material used to generate the O-tRNA. The unnatural amino acid is then incorporated into a growing polypeptide chain with high fidelity, e.g., at greater than about 70% efficiency for a given selector codon, at greater than about 75% efficiency for a given selector codon, at greater than about 80% efficiency for a given selector codon, at greater than about 85% efficiency for a given selector codon, at greater than about 90% efficiency for a given selector codon, greater than about 95% efficiency for a given selector codon, or greater than about 99% efficiency for a given selector codon.

[22] Selector codon: The term “selector codon” refers to codons recognized by the O-tRNA in the translation process and not recognized by an endogenous tRNA. The O-tRNA anticodon loop recognizes the selector codon on the mRNA and incorporates its amino acid, e.g., a selected amino acid, such as an unnatural amino acid, at this site in the polypeptide. Selector codons can include but are not limited to, e.g., nonsense codons, such as, stop codons, including but not limited to, amber, ochre, and opal codons; four or more base codons; rare codons; codons derived from natural or unnatural base pairs and/or the like. For a given system, a selector codon can also include one of the natural three base codons, wherein the endogenous system does not use (or rarely uses) said natural three base codon. For example, this includes a system that is lacking a tRNA that recognizes the natural three base codon, and/or a system wherein the natural three base codon is a rare codon.

[23] Suppressor tRNA: A suppressor tRNA is a tRNA that alters the reading of a messenger RNA (mRNA) in a given translation system, e.g., by providing a mechanism for incorporating an amino acid into a polypeptide chain in response to a selector codon. For example, a suppressor tRNA can read through a codon including but not limited to, a stop codon, a four base codon, or a rare codon.

[24]        Suppression activity: The term “suppression activity” refers to the ability of a tRNA, e.g., a suppressor tRNA, to read through a selector codon. Activity can be expressed as a percentage of activity observed as compared to a control (e.g., lacking a cognate synthetase).

[25]        Translation system: The term “translation system” refers to the components necessary to incorporate a naturally occurring amino acid into a growing polypeptide chain (protein). Components of a translation system can include, e.g., ribosomes, tRNA's, synthetases, mRNA and the like. The components of the present invention can be added to an in vitro or in vivo translation system. Examples of translation systems include but are not limited to, a non-eukaryotic cell, e.g., a bacterium (such as *E. coli*), a eukaryotic cell, e.g., a yeast cell, a mammalian cell, a plant cell, an algae cell, a fungus cell, an insect cell, a cell-free translational system e.g., a cell lysate, and/or the like.

[26]        Translation systems may be cellular or cell-free, and may be prokaryotic or eukaryotic. Cellular translation systems include, but are not limited to, whole cell preparations such as permeabilized cells or cell cultures wherein a desired nucleic acid sequence can be transcribed to mRNA and the mRNA translated. Cell-free translation systems are commercially available and many different types and systems are well-known. Examples of cell-free systems include, but are not limited to, prokaryotic lysates such as *Escherichia coli* lysates, and eukaryotic lysates such as wheat germ extracts, insect cell lysates, rabbit reticulocyte lysates, rabbit oocyte lysates and human cell lysates. Eukaryotic extracts or lysates may be preferred when the resulting protein is glycosylated, phosphorylated or otherwise modified because many such modifications are only possible in eukaryotic systems. Some of these extracts and lysates are available commercially (Promega; Madison, Wis.; Stratagene; La Jolla, Calif.; Amersham; Arlington Heights, Ill.; GIBCO/BRL; Grand Island, N.Y.). Membranous extracts, such as the canine pancreatic extracts containing microsomal membranes, are also available which are useful for translating secretory proteins.

[27]        Reconstituted translation systems may also be used. Mixtures of purified translation factors have also been used successfully to translate mRNA into protein as well as combinations of lysates or lysates supplemented with purified translation factors such as initiation factor-1 (IF-1), IF-2, IF-3 ( $\alpha$  or  $\beta$ ), elongation factor T (EF-Tu), or termination factors. Cell-free systems may also be coupled transcription/translation systems wherein DNA is introduced to the system, transcribed into mRNA and the mRNA translated as described in *Current Protocols in Molecular Biology* (F. M. Ausubel et al. editors, Wiley

Interscience, 1993), which is hereby specifically incorporated by reference. RNA transcribed in eukaryotic transcription system may be in the form of heteronuclear RNA (hnRNA) or 5'-end caps (7-methyl guanosine) and 3'-end poly A tailed mature mRNA, which can be an advantage in certain translation systems. For example, capped mRNAs are translated with high efficiency in the reticulocyte lysate system.

[28]       Selected amino acid: The term “selected amino acid” refers to any desired naturally occurring amino acid or unnatural amino acid. As used herein, the term “unnatural amino acid” or “non-naturally encoded amino acid” refers to any amino acid, modified amino acid, and/or amino acid analogue that is not one of the 20 common naturally occurring amino acids or selenocysteine or pyrrolysine. Other terms that may be used synonymously with the term “non-naturally encoded amino acid” and “unnatural amino acid” are “non-natural amino acid,” “non-naturally-occurring amino acid,” and variously hyphenated and non-hyphenated versions thereof. The term “non-naturally encoded amino acid” also includes, but is not limited to, amino acids that occur by modification (e.g. post-translational modifications) of a naturally encoded amino acid (including but not limited to, the 20 common amino acids or pyrrolysine and selenocysteine) but are not themselves naturally incorporated into a growing polypeptide chain by the translation complex. Examples of such non-naturally-occurring amino acids include, but are not limited to, *N*-acetylglucosaminyl-L-serine, *N*-acetylglucosaminyl-L-threonine, and O-phosphotyrosine.

[29]       Derived from: As used herein, the term “derived from” refers to a component that is isolated from or made using information from a specified molecule or organism.

[30]       Positive selection or screening marker: As used herein, the term “positive selection or screening marker” refers to a marker that when present, e.g., expressed, activated or the like, results in identification of a cell with the positive selection marker from those without the positive selection marker.

[31]       Negative selection or screening marker: As used herein, the term “negative selection or screening marker” refers to a marker that when present, e.g., expressed, activated or the like, allows identification of a cell that does not possess the desired property (e.g., as compared to a cell that does possess the desired property).

[32]       Reporter: As used herein, the term “reporter” refers to a component that can be used to select target components of a system of interest. For example, a reporter can include a protein, e.g., an enzyme, that confers antibiotic resistance or sensitivity (including, but not

limited to,  $\beta$ -lactamase, chloramphenicol acetyltransferase (CAT), and the like), a fluorescent screening marker (including, but not limited to, green fluorescent protein (e.g., GFP), YFP, EGFP, RFP, a luminescent marker (including but not limited to, a firefly luciferase protein), an affinity based screening marker, or positive or negative selectable marker genes such as lacZ,  $\beta$ -gal/lacZ ( $\beta$ -galactosidase), ADH (alcohol dehydrogenase), his3, ura3, leu2, lys2, or the like.

[33] Eukaryote: As used herein, the term “eukaryote” refers to organisms belonging to the phylogenetic domain Eucarya such as animals (including but not limited to, mammals, insects, reptiles, birds, etc.), ciliates, plants (including but not limited to, monocots, dicots, algae, etc.), fungi, yeasts, flagellates, microsporidia, protists, etc.

[34] Non-eukaryote: As used herein, the term “non-eukaryote” refers to non-eukaryotic organisms. For example, a non-eukaryotic organism can belong to the Eubacteria (including but not limited to, *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, *Pseudomonas fluorescens*, *Pseudomonas aeruginosa*, *Pseudomonas putida*, etc.) phylogenetic domain, or the Archaea (e.g., *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* such as *Haloferax volcanii* and *Halobacterium* species NRC-1, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeuropyrum pernix*, etc.) phylogenetic domain.

[35] Conservative variant: The term “conservative variant” refers to a translation component, e.g., a conservative variant O-tRNA or a conservative variant O-RS, that functionally performs like the component from which the conservative variant is based, e.g., an O-tRNA or O-RS, but has variations in the sequence. For example, an O-RS will aminoacylate a complementary O-tRNA or a conservative variant O-tRNA with a selected amino acid, e.g., an unnatural amino acid, although the O-tRNA and the conservative variant O-tRNA do not have the same sequence. Similarly, a tRNA will be aminoacylated with a selected amino acid, e.g., an unnatural amino acid, by a complementary O-RS or a conservative variant O-RS, although the O-RS and the conservative variant O-RS do not have the same sequence. The conservative variant can have, e.g., one variation, two variations, three variations, four variations, or five or more variations in sequence, as long as the conservative variant is complementary to the corresponding O-tRNA or O-RS.

[36] Selection or screening agent: As used herein, the term “selection or screening agent” refers to an agent that, when present, allows for a selection/screening of certain

components from a population. For example, a selection or screening agent includes, but is not limited to, e.g., a nutrient, an antibiotic, a wavelength of light, an antibody, an expressed polynucleotide, or the like. The selection agent can be varied, e.g., by concentration, intensity, etc.

[37] The term "not efficiently recognized" refers to an efficiency, e.g., less than about 10%, less than about 5%, or less than about 1%, at which a RS from one organism aminoacylates O-tRNA.

#### **DETAILED DESCRIPTION**

[38] Translation systems that are suitable for making proteins that include one or more selected amino acids, e.g., an unnatural amino acid, are described in U.S. patent applications 10/126,931, entitled "METHODS AND COMPOSITION FOR THE PRODUCTION OF ORTHOGONAL tRNA-AMINOACYL tRNA SYNTHETASE PAIRS" and 10/126,927, entitled "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS." In addition, see USSN 10/825,867 entitled "EXPANDING THE EUKARYOTIC GENETIC CODE." Each of these applications is incorporated herein by reference in its entirety. Such translation systems generally comprise cells that include an orthogonal tRNA (O-tRNA), an orthogonal aminoacyl tRNA synthetase (O-RS), and a selected amino acid, e.g., an unnatural amino acid, where the O-RS aminoacylates the O-tRNA with the selected amino acid. An orthogonal pair of the present invention is composed of an O-tRNA, e.g., a suppressor tRNA, a frameshift tRNA, or the like, and an O-RS. The O-tRNA recognizes a first selector codon and has suppression activity in presence of a cognate synthetase in response to a selector codon. The cell uses the components to incorporate the selected amino acid into a growing polypeptide chain. For example, a nucleic acid that comprises a polynucleotide that encodes a polypeptide of interest can also be present, where the polynucleotide comprises a selector codon that is recognized by the O-tRNA. The translation system can also be an in vitro system. tRNA molecules of the present invention are useful in any translational system, including systems that utilize ribosomes in translation.

[39] The translation system may also be a cell-free (in-vitro) translational system. In these systems, which can include either mRNA as a template (in-vitro translation) or DNA as a template (combined in-vitro transcription and translation), the in vitro synthesis is directed by the ribosomes. Considerable effort has been applied to the development of cell-free protein expression systems. See, e.g., Kim, D.M. and J.R. Swartz, *Biotechnology and Bioengineering*, 74 :309-316 (2001); Kim, D.M. and J.R. Swartz, *Biotechnology Letters*, 22,

1537-1542, (2000); Kim, D.M., and J.R. Swartz, *Biotechnology Progress*, 16, 385-390, (2000); Kim, D.M., and J.R. Swartz, *Biotechnology and Bioengineering*, 66, 180-188, (1999); and Patnaik, R. and J.R. Swartz, *Biotechniques* 24, 862-868, (1998); U.S. Patent No. 6,337,191; U.S. Patent Publication No. 2002/0081660; WO 00/55353; WO 90/05785, which are incorporated by reference herein. Another approach that may be applied includes the mRNA-peptide fusion technique. *See, e.g.*, R. Roberts and J. Szostak, *Proc. Natl Acad. Sci. (USA)* 94:12297-12302 (1997); A. Frankel, *et al.*, *Chemistry & Biology* 10:1043-1050 (2003). In this approach, an mRNA template linked to puromycin is translated into peptide on the ribosome. If one or more tRNA molecules have been modified, non-natural amino acids can be incorporated into the peptide as well. After the last mRNA codon has been read, puromycin captures the C-terminus of the peptide. If the resulting mRNA-peptide conjugate is found to have interesting properties in an in vitro assay, its identity can be easily revealed from the mRNA sequence. In this way, one may screen libraries of polypeptides comprising one or more non-naturally encoded amino acids to identify polypeptides having desired properties. More recently, in vitro ribosome translations with purified components have been reported that permit the synthesis of peptides substituted with non-naturally encoded amino acids. *See, e.g.*, A. Forster *et al.*, *Proc. Natl Acad. Sci. (USA)* 100:6353 (2003).

[40] In certain embodiments, an *E. coli* cell comprising the tRNA of the present invention includes such a translation system. For example, the *E. coli* cell of the present invention includes an orthogonal tRNA (O-tRNA), where the O-tRNA comprises suppression activity in presence of a cognate synthetase in response to a selector codon; an orthogonal aminoacyl-tRNA synthetase (O-RS); a selected amino acid; and, a nucleic acid that comprises a polynucleotide that encodes a polypeptide of interest, where the polynucleotide comprises a selector codon that is recognized by the O-tRNA.

[41] The invention also features multiple O-tRNA/O-RS pairs in a cell, which allows incorporation of more than one selected amino acid. In certain embodiments, the cell can further include an additional different O-tRNA/O-RS pair and a second selected amino acid, where the O-tRNA recognizes a second selector codon and the O-RS preferentially aminoacylates the O-tRNA with the second selected amino acid. For example, a cell can further comprise, e.g., an amber suppressor tRNA-aminoacyl tRNA synthetase pair derived from the tyrosyl-tRNA synthetase of *Methanococcus jannaschii*.

[42] The O-tRNA and/or the O-RS can be naturally occurring or can be derived by mutation of a naturally occurring tRNA and/or RS, e.g., which generates libraries of tRNA's

and/or libraries of RSs, from a variety of organisms. For example, one strategy of producing an orthogonal tRNA/aminoacyl-tRNA synthetase pair involves importing a heterologous tRNA/synthetase pair from, e.g., a source other than the host cell, or multiple sources, into the host cell. The properties of the heterologous synthetase candidate include, e.g., that it does not charge any host cell tRNA, and the properties of the heterologous tRNA candidate include, e.g., that it is not aminoacylated by any host cell synthetase. In addition, the heterologous tRNA is orthogonal to all host cell synthetases.

[43] A second strategy for generating an orthogonal pair involves generating mutant libraries from which to screen and/or select an O-tRNA or O-RS. These strategies can also be combined.

[44] In various embodiments, the O-tRNA and O-RS are derived from at least one organism. In another embodiment, the O-tRNA is derived from a naturally occurring or mutated naturally occurring tRNA from a first organism and the O-RS is derived from naturally occurring or mutated naturally occurring RS from a second organism. In one embodiment, the first and second organisms are different. For example, an orthogonal pair may include a tRNA synthetase derived from *Methanobacterium thermoautotrophicum*, and a tRNA derived from an archaeal tRNA (e.g., from *Halobacterium sp. NRC-1*). Alternatively, the first and second organisms are the same. See the section entitled "Sources and Host Organisms" herein for additional information.

[45] In certain embodiments of the present invention, an O-tRNA of the present invention comprises or is encoded by a polynucleotide sequence as set forth in SEQ ID NO.: 1, 2, or 3, or a complementary polynucleotide sequence thereof, or a conservative variation thereof. See also the section entitled "Nucleic Acid and Polypeptide Sequence and Variants" herein.

#### Orthogonal tRNA (O-tRNA)

[46] An orthogonal tRNA (O-tRNA) mediates incorporation of a selected amino acid into a protein that is encoded by a polynucleotide that comprises a selector codon that is recognized by the O-tRNA, e.g., in vivo or in vitro. An O-tRNA of the present invention may be aminoacylated with a desired amino acid by any method or technique, including but not limited to, chemical or enzymatic aminoacylation. The aminoacylated O-tRNA of the present invention may be added directly to a translation system. An O-tRNA of the present invention may be aminoacylated by an RS with a selected amino acid in vitro or in vivo. In addition,

the RS may be an O-RS. An O-tRNA of the present invention can be provided to the translation system (e.g., in vitro translation components, or a cell) directly, or by providing a polynucleotide that encodes an O-tRNA or a portion thereof. For example, an O-tRNA, or a portion thereof, is encoded by a polynucleotide sequence as set forth in SEQ ID NO.: 1, 2, 3, or a complementary polynucleotide sequence thereof, or a conservative variation thereof.

[47] An O-tRNA of the present invention comprises suppression activity in the presence of a cognate synthetase in response to a selector codon. Suppression activity can be determined by any of a number of assays known in the art. For example, a  $\beta$ -galactosidase reporter assay can be used. A derivative of a plasmid that expresses *lacZ* gene under the control of promoter is used, e.g., where the Leu-25 of the peptide VVLQRRDWEN of *lacZ* is replaced by a selector codon, e.g., TAG, TGA, AGGA, etc. codons, or sense codons (as a control) for tyrosine, serine, leucine, etc. The derivatived *lacZ* plasmid is introduced into cells from an appropriate organism (e.g., an organism where the orthogonal components can be used) along with plasmid comprising an O-tRNA of the present invention. A cognate synthetase can also be introduced (either as a polypeptide or a polynucleotide that encodes the cognate synthetase when expressed). The cells are grown in media to a desired density, e.g., to an OD<sub>600</sub> of about 0.5., and  $\beta$ -galactosidase assays are performed, e.g., using the BetaFluor™  $\beta$ -Galactosidase Assay Kit (Novagen). Percent suppression is calculated as the percentage of activity for a sample relative to a comparable control, e.g., the value observed from the derivatived *lacZ* construct, where the construct has a corresponding sense codon at desired position rather than a selector codon.

[48] In the tRNA molecule, Thymine (T) is replaced with Uracil (U). In addition, additional modifications to the bases can be present. The invention also includes conservative variations of O-tRNA. For example, conservative variations of O-tRNA include those molecules that function like the O-tRNA and maintain the tRNA L-shaped structure, but do not have the same sequence (and are other than wild type tRNA molecules). *See also* the section herein entitled "Nucleic Acid and Polypeptide Sequence and Variants."

[49] The composition comprising an O-tRNA can further include an orthogonal aminoacyl-tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O-tRNA with a selected amino acid (e.g., an unnatural amino acid). In certain embodiments, a composition including an O-tRNA can further include a translation system (e.g., an in vitro or an in vivo translation system). A nucleic acid comprising a polynucleotide encoding a

polypeptide of interest, wherein the polynucleotide comprises one or more selector codons recognized by the O-tRNA, or a combination of one or more of these, can also be present in the cell or other translation system. *See also*, the section herein entitled “Orthogonal Aminoacyl-tRNA Synthetases (O-RS).”

[50] Methods of producing an orthogonal tRNA (O-tRNA), e.g., an O-tRNA, are also a feature of the present invention. A tRNA, e.g., an O-tRNA, produced by the method is also a feature of the present invention.

[51] Methods of producing an orthogonal tRNA include mutating the anticodon loop of each of a pool of tRNA's to allow recognition of a selector codon (e.g., an amber codon, an opal codon, a four base codon, etc.), thereby providing a plurality of potential O-tRNA's; and analyzing secondary structure of a member of the plurality potential O-tRNA to identify non-canonical base pairs in the secondary structure, and optionally mutating the non-canonical base pairs (e.g., the non-canonical base pairs are mutated to canonical base pairs). The non-canonical base pairs can be located in the stem region of the secondary structure. An O-tRNA may possess an improvement of one or more characteristics or activities, such as improvement in orthogonality for a desired organism compared to the starting material, e.g., the plurality of tRNA sequences, while preserving its affinity towards a desired RS.

[52] Alternatively, O-tRNA's may be developed by mutating a known tRNA to modulate its interaction with or binding affinity to one or more molecules that influence translation or are components of translation machinery. Such components include, but are not limited to, elongation factors. Bacterial elongation factor EF-Tu plays a key role in the elongation step in protein synthesis. Following aminoacylation of the tRNA by tRNA synthetase, EF-Tu binds the aminoacylated tRNA and brings it to the A site of the ribosome. The ester bond between the charged amino acid and the tRNA is protected from spontaneous hydrolysis due to the binding between EF-Tu and aminoacylated tRNA. Stortchevoi et al. investigated mutants of the *E. coli* initiation tRNA<sup>Met</sup> U50:G64 wobble base pair in the TΨC stem, since this base pair was found to be a secondary negative determinant blocking the tRNA's activity in elongation, presumably due to a weakened interaction between the EF-Tu.GTP and aminoacylated tRNA (JBC 2003 278(20):17672-17679). Also, LaRiviere et al. described in Science 2001 Oct 5;294(5540):165-8 the thermodynamic contributions of the amino acid and the tRNA body to the overall binding affinity to EF-Tu. They indicated that the contributions of the tRNA body and the amino acid are independent of each other and that they compensate for one another when the tRNAs are correctly acylated. Alterations to the

interaction between EF-Tu.GTP and the tRNA aminoacylated with the unnatural amino acid may affect the efficiency of the loading of the tRNA to the A site of the ribosome. Potential mutation sites may also be found by analyzing crystal structures of complexes between tRNA and other components of translational machinery such as EF-Tu. For example, Nissen et al. have indicated that EF-Tu.GTP binds directly to the phosphate backbone of the TΨC stem of yeast phenylalanyl-transfer RNA (Phe-tRNA) (Science 1995 270(5241):1464-1472).

[53] The methods optionally include analyzing the homology of sequences of tRNA's and/or aminoacyl-tRNA synthetases to determine potential candidates for an O-tRNA, O-RS and/or pairs thereof, that appear to be orthogonal for a specific organism. Computer programs known in the art and described herein can be used for the analysis. In one example, to choose potential orthogonal translational components for use in a prokaryotic organism, a synthetase and/or a tRNA is chosen that does not display unusual homology to prokaryotic organisms.

[54] A pool of tRNA's can also be produced by a consensus strategy. For example, the pool of tRNA's is produced by aligning a plurality of tRNA sequences; determining a consensus sequence; and generating a library of tRNA's using at least a portion, most of, or the entire consensus sequence. For example, a consensus sequence can be compiled with a computer program, e.g., the GCG program *pileup*. Optionally, degenerate positions determined by the program are changed to the most frequent base at those positions. A library is synthesized by techniques known in the art using the consensus sequence. For example, overlap extension of oligonucleotides in which each site of the tRNA gene can be synthesized as a doped mixture of 90% the consensus sequence and 10% a mixture of the other 3 bases can be used to provide the library based on the consensus sequence. Other mixtures can also be used, e.g., 75% the consensus sequence and 25% a mixture of the other 3 bases, 80% the consensus sequence and 20% a mixture of the other 3 bases, 95% the consensus sequence and 5% a mixture of the other 3 bases, etc.

[55] Libraries of mutant tRNA's can be generated using various mutagenesis techniques known in the art. For example, the mutant tRNA's can be generated by site-specific mutations, random point mutations, homologous recombination, DNA shuffling or other recursive mutagenesis methods, chimeric construction or any combination thereof.

[56] Additional mutations can be introduced at a specific position(s), e.g., at a non-conservative position(s), or at a conservative position(s), at a randomized position(s), or a

combination thereof in a desired loop or region of a tRNA, e.g., an anticodon loop, the acceptor stem, D arm or loop, variable loop, TΨC arm or loop, other regions of the tRNA molecule, or a combination thereof. Mutations may include matched base pairs in the stem region.

[57] Typically, an O-tRNA is obtained by subjecting to negative selection a population of cells of a first species, where the cells comprise a member of the plurality of potential O-tRNA's. The negative selection eliminates cells that comprise a member of the plurality of potential O-tRNA's that is aminoacylated by an aminoacyl-tRNA synthetase (RS) that is endogenous to the cells. This provides a pool of tRNA's that are orthogonal to the cell of the first species.

[58] In certain embodiments of the negative selection, a selector codon(s) is introduced into polynucleotide that encodes a negative selection marker, e.g., an enzyme that confers antibiotic resistance, e.g.,  $\beta$ -lactamase, an enzyme that confers a detectable product, e.g.,  $\beta$ -galactosidase, chloramphenicol acetyltransferase (CAT), e.g., a toxic product, such as barnase, at a non-essential position, etc. Screening/selection can be done by growing the population of cells in the presence of a selection agent (e.g., an antibiotic, such as ampicillin). In one embodiment, the concentration of the selection agent is varied.

[59] For example, to measure the activity of suppressor tRNA's, a selection system is used that is based on the *in vivo* suppression of selector codon, e.g., nonsense or frameshift mutations introduced into a polynucleotide that encodes a negative selection marker, e.g., a gene for  $\beta$ -lactamase (*bla*). For example, polynucleotide variants, e.g., *bla* variants, with, e.g., TAG, AGGA, and TGA, at position a certain position, are constructed. Cells, e.g., bacteria, are transformed with these polynucleotides. In the case of an orthogonal tRNA, which cannot be efficiently charged by endogenous *E. coli* synthetases, antibiotic resistance, e.g., ampicillin resistance, should be about or less than that for a bacteria transformed with no plasmid. If the tRNA is not orthogonal, or if a heterologous synthetase capable of charging the tRNA is co-expressed in the system, a higher level of antibiotic, e.g., ampicillin, resistance is observed. Cells, e.g., bacteria, are chosen that are unable to grow on LB agar plates with antibiotic concentrations about equal to cells transformed with no plasmids.

[60] In the case of a toxic product (e.g., ribonuclease barnase), when a member of the plurality of potential tRNA's is aminoacylated by endogenous host, e.g., *Escherichia coli* synthetases (i.e., it is not orthogonal to the host, e.g., *Escherichia coli* synthetases), the

selector codon is suppressed and the toxic polynucleotide product produced leads to cell death. Cells harboring orthogonal tRNA or non-functional tRNA's survive.

[61] In one embodiment, the pool of tRNA's that are orthogonal to a desired organism are then subjected to a positive selection in which a selector codon is placed in a positive selection marker, e.g., encoded by a drug resistance gene, such a  $\beta$ -lactamase gene. The positive selection is performed on cell comprising a polynucleotide encoding or comprising a member of the pool of tRNA's, a polynucleotide encoding a positive selection marker, and a polynucleotide encoding cognate RS. These polynucleotides are expressed in the cell and the cell is grown in the presence of a selection agent, e.g., ampicillin. tRNA's are then selected for their ability to be aminoacylated by the coexpressed cognate synthetase and to insert an amino acid in response to this selector codon. Typically, these cells show an enhancement in suppression efficiency compared to cells harboring non-functional tRNA's, or tRNA's that cannot efficiently be recognized by the synthetase of interest. The cell harboring the non-functional or tRNA's that are not efficiently recognized by the synthetase of interest are sensitive to the antibiotic. Therefore, tRNA's that: (i) are not substrates for endogenous host, e.g., *Escherichia coli*, synthetases; (ii) can be aminoacylated by the synthetase of interest; and (iii) are functional in translation survive both selections.

[62] The stringency of the selection, e.g., the positive selection, the negative selection or both the positive and negative selection, in the above described-methods, optionally may be varied. For example, because barnase is an extremely toxic protein, the stringency of the negative selection can be controlled by introducing different numbers of selector codons into the barnase gene and/or by using an inducible promoter. In another example, the concentration of the selection or screening agent is varied (e.g., ampicillin). In one aspect, the stringency is varied because the desired activity can be low during early rounds. Thus, less stringent selection criteria are applied in early rounds and more stringent criteria are applied in later rounds of selection. In certain embodiments, the negative selection, the positive selection, or both the negative and positive selection can be repeated multiple times. Multiple different negative selection markers, positive selection markers or both negative and positive selection markers can be used. In certain embodiments, the positive and negative selection marker can be the same.

[63] Other types of selections/screening can be used in the invention for producing orthogonal translational components, e.g., an O-tRNA, an O-RS, and an O-tRNA/O-RS pair. For example, the negative selection marker, the positive selection marker or both the positive

and negative selection markers can include a marker that fluoresces or catalyzes a luminescent reaction in the presence of a suitable reactant. In another embodiment, a product of the marker is detected by fluorescence-activated cell sorting (FACS) or by luminescence. Optionally, the marker includes an affinity based screening marker. *See*, Francisco, J. A., et al., (1993) *Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragment on the external surface*. Proc Natl Acad Sci U S A. 90:10444-8.

[64] Additional methods for producing a recombinant orthogonal tRNA can be found, e.g., in U.S. patent applications 10/126,931, entitled "Methods and Compositions for the Production of Orthogonal tRNA-Aminoacyl tRNA Synthetase Pairs" and 10/126,127, entitled "In vivo Incorporation of Unnatural Amino Acids," and USSN 10/825,867 entitled "EXPANDING THE EUKARYOTIC GENETIC CODE." *See also*, Forster et al., (2003) *Programming peptidomimetic synthetases by translating genetic codes designed de novo*. PNAS 100(11):6353-6357; and, Feng et al., (2003), *Expanding tRNA recognition of a tRNA synthetase by a single amino acid change*, PNAS 100(10): 5676-5681.

[65] A tRNA of the present invention may be aminoacylated with a desired amino acid by any method or technique, including but not limited to, chemical or enzymatic aminoacylation.

[66] Aminoacylation may be accomplished by aminoacyl tRNA synthetases or by other enzymatic molecules, including but not limited to, ribozymes. The term "ribozyme" is interchangeable with "catalytic RNA." Cech and coworkers (Cech, 1987, *Science*, 236:1532-1539; McCorkle et al., 1987, *Concepts Biochem.* 64:221-226) demonstrated the presence of naturally occurring RNAs that can act as catalysts (ribozymes). However, although these natural RNA catalysts have only been shown to act on ribonucleic acid substrates for cleavage and splicing, the recent development of artificial evolution of ribozymes has expanded the repertoire of catalysis to various chemical reactions. Studies have identified RNA molecules that can catalyze aminoacyl-RNA bonds on their own (2')3'-termini (Illangakekare et al., 1995 *Science* 267:643-647), and an RNA molecule which can transfer an amino acid from one RNA molecule to another (Lohse et al., 1996, *Nature* 381:442-444).

[67] U.S. Patent Application Publication 2003/0228593, which is incorporated by reference herein, describes methods to construct ribozymes and their use in aminoacylation of tRNAs with naturally encoded and non-naturally encoded amino acids. Substrate-

immobilized forms of enzymatic molecules that can aminoacylate tRNAs, including but not limited to, ribozymes, may enable efficient affinity purification of the aminoacylated products. Examples of suitable substrates include agarose, sepharose, and magnetic beads. The production and use of a substrate-immobilized form of ribozyme for aminoacylation is described in Chemistry and Biology 2003, 10:1077-1084 and U.S. Patent Application Publication 2003/0228593, which is incorporated by reference herein.

[68] Chemical aminoacylation methods include, but are not limited to, those introduced by Hecht and coworkers (Hecht, S. M. Acc. Chem. Res. 1992, 25, 545; Heckler, T. G.; Roesser, J. R.; Xu, C.; Chang, P.; Hecht, S. M. Biochemistry 1988, 27, 7254; Hecht, S. M.; Alford, B. L.; Kuroda, Y.; Kitano, S. J. Biol. Chem. 1978, 253, 4517) and by Schultz, Chamberlin, Dougherty and others (Cornish, V. W.; Mendel, D.; Schultz, P. G. Angew. Chem. Int. Ed. Engl. 1995, 34, 621; Robertson, S. A.; Ellman, J. A.; Schultz, P. G. J. Am. Chem. Soc. 1991, 113, 2722; Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. Science 1989, 244, 182; Bain, J. D.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R. J. Am. Chem. Soc. 1989, 111, 8013; Bain, J. D. et al. Nature 1992, 356, 537; Gallivan, J. P.; Lester, H. A.; Dougherty, D. A. Chem. Biol. 1997, 4, 740; Turcatti, et al. J. Biol. Chem. 1996, 271, 19991; Nowak, M. W. et al. Science, 1995, 268, 439; Saks, M. E. et al. J. Biol. Chem. 1996, 271, 23169; Hohsaka, T. et al. J. Am. Chem. Soc. 1999, 121, 34), to avoid the use of synthetases in aminoacylation. Such methods or other chemical aminoacylation methods may be used to aminoacylate tRNA molecules of the invention.

[69] Biosynthetic methods that employ chemically modified aminoacyl-tRNAs have been used to incorporate several biophysical probes into proteins synthesized in vitro. See the following publications and references cited within: Brunner, J. *New Photolabeling and crosslinking methods*, Annu. Rev Biochem, 62:483-514 (1993); and, Krieg, U.C., Walter, P., Hohnson, A.E. *Photocrosslinking of the signal sequence of nascent preprolactin of the 54-kilodalton polypeptide of the signal recognition particle*, Proc. Natl. Acad. Sci, 83(22):8604-8608 (1986).

[70] Previously, it has been shown that unnatural amino acids can be site-specifically incorporated into proteins in vitro by the addition of chemically aminoacylated suppressor tRNAs to protein synthesis reactions programmed with a gene containing a desired amber nonsense mutation. Using these approaches, one can substitute a number of the common twenty amino acids with close structural homologues, e.g., fluorophenylalanine for phenylalanine, using strains auxotrophic for a particular amino acid. See, e.g., Noren, C.J.,

Anthony-Cahill, Griffith, M.C., Schultz, P.G. *A general method for site-specific incorporation of unnatural amino acids into proteins*, Science, 244: 182-188 (1989); M.W. Nowak, et al., Science 268:439-42 (1995); Bain, J.D., Glabe, C.G., Dix, T.A., Chamberlin, A.R., Diala, E.S. *Biosynthetic site-specific Incorporation of a non-natural amino acid into a polypeptide*, J. Am Chem Soc, 111:8013-8014 (1989); N. Budisa et al., FASEB J. 13:41-51 (1999); Ellman, J.A., Mendel, D., Anthony-Cahill, S., Noren, C.J., Schultz, P.G. *Biosynthetic method for introducing unnatural amino acids site-specifically into proteins*, Methods in Enz., vol. 202, 301-336 (1992); and, Mendel, D., Cornish, V.W. & Schultz, P.G. *Site-Directed Mutagenesis with an Expanded Genetic Code*, Annu Rev Biophys. Biomol Struct. 24, 435-62 (1995).

[71] For example, a suppressor tRNA was prepared that recognized the stop codon UAG and was chemically aminoacylated with an unnatural amino acid. Conventional site-directed mutagenesis was used to introduce the stop codon TAG, at the site of interest in the protein gene. *See, e.g.*, Sayers, J.R., Schmidt, W. Eckstein, F. *5'-3' Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis*, Nucleic Acids Res, 16(3):791-802 (1988). When the acylated suppressor tRNA and the mutant gene were combined in an in vitro transcription/translation system, the unnatural amino acid was incorporated in response to the UAG codon which gave a protein containing that amino acid at the specified position. Experiments using [<sup>3</sup>H]-Phe and experiments with  $\alpha$ -hydroxy acids demonstrated that only the desired amino acid is incorporated at the position specified by the UAG codon and that this amino acid is not incorporated at any other site in the protein. *See, e.g.*, Noren, et al, *supra*; Kobayashi et al., (2003) *Nature Structural Biology* 10(6):425-432; and, Ellman, J.A., Mendel, D., Schultz, P.G. *Site-specific incorporation of novel backbone structures into proteins*, Science, 255(5041):197-200 (1992).

[72] Methods for generating catalytic RNA may involve generating separate pools of randomized ribozyme sequences, performing directed evolution on the pools, screening the pools for desirable aminoacylation activity, and selecting sequences of those ribozymes exhibiting desired aminoacylation activity.

[73] Ribozymes can comprise motifs and/or regions that facilitate acylation activity, such as a GGU motif and a U-rich region. For example, it has been reported that U-rich regions can facilitate recognition of an amino acid substrate, and a GGU-motif can form base pairs with the 3' termini of a tRNA. In combination, the GGU and motif and U-rich region

facilitate simultaneous recognition of both the amino acid and tRNA simultaneously, and thereby facilitate aminoacylation of the 3' terminus of the tRNA.

[74] Ribozymes can be generated by *in vitro* selection using a partially randomized r24mini conjugated with tRNA<sup>Asn</sup> cccg, followed by systematic engineering of a consensus sequence found in the active clones. An exemplary ribozyme obtained by this method is termed "Fx3 ribozyme" and is described in U.S. Pub. App. No. 2003/0228593, the contents of which is incorporated by reference herein, acts as a versatile catalyst for the synthesis of various aminoacyl-tRNAs charged with cognate non-natural amino acids.

[75] Immobilization on a substrate may be used to enable efficient affinity purification of the aminoacylated tRNAs. Examples of suitable substrates include, but are not limited to, agarose, sepharose, and magnetic beads. Ribozymes can be immobilized on resins by taking advantage of the chemical structure of RNA, such as the 3'-cis-diol on the ribose of RNA can be oxidized with periodate to yield the corresponding dialdehyde to facilitate immobilization of the RNA on the resin. Various types of resins can be used including inexpensive hydrazide resins wherein reductive amination makes the interaction between the resin and the ribozyme an irreversible linkage. Synthesis of aminoacyl-tRNAs can be significantly facilitated by this on-column aminoacylation technique. Kourouklis et al. Methods 2005; 36:239-4 describe a column-based aminoacylation system.

[76] Isolation of the aminoacylated tRNAs can be accomplished in a variety of ways. One suitable method is to elute the aminoacylated tRNAs from a column with a buffer such as a sodium acetate solution with 10 mM EDTA, a buffer containing 50 mM N-(2-hydroxyethyl)piperazine-N'-(3-propanesulfonic acid), 12.5 mM KCl, pH 7.0, 10 mM EDTA, or simply an EDTA buffered water (pH 7.0).

[77] The aminoacylated tRNAs of the present invention can be added to translation reactions in order to incorporate the amino acid with which the tRNA was aminoacylated in a position of choice in a polypeptide made by the translation reaction. Examples of translation systems in which the aminoacylated tRNAs of the present invention may be used include, but are not limited to cell lysates. Cell lysates provide reaction components necessary for *in vitro* translation of a polypeptide from an input mRNA. Examples of such reaction components include but are not limited to ribosomal proteins, rRNA, amino acids, tRNAs, GTP, ATP, translation initiation and elongation factors and additional factors associated with translation. Additionally, translation systems may be batch translations or compartmentalized translation.

Batch translation systems combine reaction components in a single compartment while compartmentalized translation systems separate the translation reaction components from reaction products that can inhibit the translation efficiency. Such translation systems are available commercially.

[78] Further, a coupled transcription/translation system may be used. Coupled transcription/translation systems allow for both transcription of an input DNA into a corresponding mRNA, which is in turn translated by the reaction components. An example of a commercially available coupled transcription/translation is the Rapid Translation System (RTS, Roche Inc.). The system includes a mixture containing E. coli lysate for providing translational components such as ribosomes and translation factors. Additionally, an RNA polymerase is included for the transcription of the input DNA into an mRNA template for use in translation. RTS can use compartmentalization of the reaction components by way of a membrane interposed between reaction compartments, including a supply/waste compartment and a transcription/translation compartment.

[79] Aminoacylation of tRNA may be performed by other agents, including but not limited to, transferases, polymerases, catalytic antibodies, multi-functional proteins, and the like.

#### Orthogonal Aminoacyl-tRNA Synthetases (O-RS)

[80] An O-RS preferentially aminoacylates an O-tRNA of the present invention with a selected amino acid in vitro or in vivo. An O-RS of the present invention can be provided to the translation system (e.g., in vitro translation components, or a cell) by a polypeptide that includes an O-RS and/or by a polynucleotide that encodes an O-RS or a portion thereof. An O-RS, or a portion thereof, is encoded by a polynucleotide sequence or a complementary polynucleotide sequence thereof, or a conservative variation thereof. An O-tRNA of the present invention may be aminoacylated by a number of different O-RS molecules, including but not limited to, those disclosed herein.

[81] Methods for identifying an orthogonal aminoacyl-tRNA synthetase (O-RS), e.g., an O-RS, for use with an O-tRNA, e.g., an O-tRNA, are also a feature of the present invention. For example a method includes subjecting to positive selection a population of cells of a first species, where the cells each comprise: 1) a member of a plurality of aminoacyl-tRNA synthetases (RSs), where the plurality of RSs comprise mutant RSs, RSs derived from a species other than the first species or both mutant RSs and RSs derived from a

species other than the first species; 2) the orthogonal tRNA (O-tRNA) from a second species; and 3) a polynucleotide that encodes a positive selection marker and comprises at least one selector codon. Cells are selected or screened for those that show an enhancement in suppression efficiency compared to cells lacking or with a reduced amount of the member of the plurality of RSs. Cells having an enhancement in suppression efficiency comprise an active RS that aminoacylates the O-tRNA. A level of aminoacylation (in vitro or in vivo) by the active RS of a first set of tRNA's from the first species is compared to the level of aminoacylation (in vitro or in vivo) by the active RS of a second set of tRNA's from the second species. The level of aminoacylation can be determined by a detectable substance (e.g., a labeled amino acid or unnatural amino acid). The active RS that more efficiently aminoacylates the second set of tRNA's compared to the first set of tRNA's is selected, thereby providing the orthogonal aminoacyl-tRNA synthetase for use with the O-tRNA. An O-RS, e.g., an O-RS, identified by the method is also a feature of the present invention.

[82] Any of a number of assays can be used to determine aminoacylation. These assays can be performed in vitro or in vivo. For example, in vitro aminoacylation assays are described in, e.g., Hoben, P., and Soll, D. (1985) Methods Enzymol. 113:55-59 and in U.S. Patent Application Publication No. 2003/0228593. Aminoacylation can also be determined by using a reporter along with orthogonal translation components and detecting the reporter in a cell expressing a polynucleotide comprising at least one selector codon that encodes a protein. See also, U.S. patent application 10/126,927, entitled "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS;" and, USSN 10/825,867 entitled "EXPANDING THE EUKARYOTIC GENETIC CODE."

[83] An identified O-RS can be further manipulated to alter the substrate specificity of the synthetase so that only a desired unnatural amino acid, but not any of the common 20 amino acids, are charged to the O-tRNA. Methods to generate an orthogonal aminoacyl tRNA synthetases with a substrate specificity for an unnatural amino acid include mutating the synthetase, e.g., at the active site in the synthetase, at the editing mechanism site in the synthetase, at different sites by combining different domains of synthetases, or the like, and applying a selection process. A strategy is used that is based on the combination of a positive selection followed by a negative selection. In the positive selection, suppression of the selector codon introduced at a non-essential position(s) of a positive marker allows cells to survive under positive selection pressure. In the presence of both natural and unnatural amino acids, survivors thus encode active synthetases charging the orthogonal suppressor

tRNA with either a natural or unnatural amino acid. In the negative selection, suppression of a selector codon introduced at a non-essential position(s) of a negative marker removes synthetases with natural amino acid specificities. Survivors of the negative and positive selection encode synthetases that aminoacylate (charge) the orthogonal suppressor tRNA with unnatural amino acids only. These synthetases can then be subjected to further mutagenesis, e.g., DNA shuffling or other recursive mutagenesis methods.

[84] The library of mutant O-RSs can be generated using various mutagenesis techniques known in the art. For example, the mutant RSs can be generated by site-specific mutations, random point mutations, homologous recombination, DNA shuffling or other recursive mutagenesis methods, chimeric construction or any combination thereof. For example, a library of mutant RSs can be produced from two or more other, e.g., smaller, less diverse "sub-libraries." Chimeric libraries of RSs are also included in the invention. It should be noted that libraries of tRNA synthetases from various organisms (e.g., microorganisms such as eubacteria or archaeobacteria) such as libraries that comprise natural diversity (see, e.g., U.S. Patent No. 6,238,884 to Short et al; U.S. Patent No. 5,756,316 to Schallenger et al; U.S. Patent No. 5,783,431 to Petersen et al; U.S. Patent No. 5,824,485 to Thompson et al; U.S. Patent No. 5,958,672 to Short et al), are optionally constructed and screened for orthogonal pairs.

[85] Once the synthetases are subject to the positive and negative selection/screening strategy, these synthetases can then be subjected to further mutagenesis. For example, a nucleic acid that encodes the O-RS can be isolated; a set of polynucleotides that encode mutated O-RSs (e.g., by random mutagenesis, site-specific mutagenesis, recombination or any combination thereof) can be generated from the nucleic acid; and, these individual steps or a combination of these steps can be repeated until a mutated O-RS is obtained that preferentially aminoacylates the O-tRNA with the unnatural amino acid. In one aspect of the present invention, the steps are performed multiple times, e.g., at least two times.

[86] Additional levels of selection/screening stringency can also be used in the methods of the present invention, for producing O-tRNA, O-RS, or pairs thereof. The selection or screening stringency can be varied on one or both steps of the method to produce an O-RS. This could include, e.g., varying the amount of selection/screening agent that is used, etc. Additional rounds of positive and/or negative selections can also be performed. Selecting or screening can also comprise one or more positive or negative selection or screening that includes, e.g., a change in amino acid permeability, a change in translation

efficiency, a change in translational fidelity, etc. Typically, the one or more change is based upon a mutation in one or more gene in an organism in which an orthogonal tRNA-tRNA synthetase pair is used to produce protein.

[87] Other types of selections can be used in the present invention for, e.g., O-RS, O-tRNA, and O-tRNA/O-RS pair. The positive selection marker can be any of a variety of molecules including, but not limited to, a product that provides a nutritional supplement for growth and the selection is performed on a medium that lacks the nutritional supplement. Examples of polynucleotides that encode positive selection markers include, but are not limited to, e.g., a reporter gene based on complementing the amino acid auxotrophy of a cell, a *his3* gene (e.g., where the *his3* gene encodes an imidazole glycerol phosphate dehydratase, detected by providing 3-aminotriazole (3-AT)), *ura3* gene, *leu2* gene, *lys2* gene, *lacZ* gene, *adh* gene, etc. See, e.g., G. M. Kishore, & D. M. Shah, (1988), Amino acid biosynthesis inhibitors as herbicides, *Annual Review of Biochemistry* 57:627-663. In one embodiment, *lacZ* production is detected by ortho-nitrophenyl- $\beta$ -D-galactopyranoside (ONPG) hydrolysis. See, e.g., I. G. Serebriiskii, & E. A. Golemis, (2000), Uses of *lacZ* to study gene function: evaluation of beta-galactosidase assays employed in the yeast two-hybrid system, *Analytical Biochemistry* 285:1-15. Additional positive selection markers include, e.g., luciferase, green fluorescent protein (GFP), YFP, EGFP, RFP, the product of an antibiotic resistant gene (e.g., chloramphenicol acetyltransferase (CAT)), a transcriptional modulator protein (e.g., GAL4), etc. Optionally, a polynucleotide that encodes a positive selection marker comprises a selector codon.

[88] A polynucleotide that encodes the positive selection marker can be operably linked to a response element. An additional polynucleotide that encodes a transcriptional modulator protein that modulates transcription from the response element, and comprises at least one selector codon, can also be present. The incorporation of the unnatural amino acid into the transcriptional modulator protein by the O-tRNA aminoacylated with the unnatural amino acid results in transcription of the polynucleotide (e.g., reporter gene) encoding the positive selection marker. Optionally, the selector codon is located in or substantially near a portion of the polynucleotide that encodes a DNA binding domain of the transcriptional modulator protein.

[89] A polynucleotide that encodes the negative selection marker can also be operably linked to a response element from which transcription is mediated by the transcriptional modulator protein. See, e.g., A. J. DeMaggio, et al., (2000), The yeast split-hybrid system,

Method Enzymol. 328:128-137; H. M. Shih, et al., (1996), A positive genetic selection for disrupting protein-protein interactions: identification of CREB mutations that prevent association with the coactivator CBP, Proc. Natl. Acad. Sci. U. S. A. 93:13896-13901; M. Vidal, et al., (1996), Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system, Proc. Natl. Acad. Sci. U. S. A. 93:10321-10326; and, M. Vidal, et al., (1996), Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions (Proc. Natl. Acad. Sci. U. S. A. 93:10315-10320). The incorporation of a natural amino acid into the transcriptional modulator protein by the O-tRNA aminoacylated with a natural amino acid results in transcription of the negative selection marker. Optionally, the negative selection marker comprises a selector codon. The positive selection marker and/or negative selection marker of the invention can comprise at least two selector codons, which each or both can comprise at least two different selector codons or at least two of the same selector codons.

[90] The transcriptional modulator protein is a molecule that binds (directly or indirectly) to a nucleic acid sequence (e.g., a response element) and modulates transcription of a sequence that is operably linked to the response element. A transcriptional modulator protein can be a transcriptional activator protein (e.g., GAL4, nuclear hormone receptors, AP1, CREB, LEF/tcf family members, SMADs, VP16, SP1, etc.), a transcriptional repressor protein (e.g., nuclear hormone receptors, Groucho/tle family, Engrailed family, etc), or a protein that can have both activities depending on the environment (e.g., LEF/tcf, homobox proteins, etc.). A response element is typically a nucleic acid sequence that is recognized by the transcriptional modulator protein or an additional agent that acts in concert with the transcriptional modulator protein.

[91] Another example of a transcriptional modulator protein is the transcriptional activator protein, GAL4. See, e.g., A. Laughon, et al., (1984), Identification of two proteins encoded by the *Saccharomyces cerevisiae* GAL4 gene, Molecular & Cellular Biology 4:268-275; A. Laughon, & R. F. Gesteland, (1984), Primary structure of the *Saccharomyces cerevisiae* GAL4 gene, Molecular & Cellular Biology 4:260-267; L. Keegan, et al., (1986), Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein, Science 231:699-704; and, M. Ptashne, (1988), How eukaryotic transcriptional activators work, Nature 335:683-689. The N-terminal 147 amino acids of this 881 amino acid protein form a DNA binding domain (DBD) that binds DNA sequence specifically. See, e.g., M. Carey, et al., (1989), An amino-terminal fragment of GAL4 binds

DNA as a dimer, *J. Mol. Biol.* 209:423-432; and, E. Giniger, et al., (1985), Specific DNA binding of GAL4, a positive regulatory protein of yeast, *Cell* 40:767-774. The DBD is linked, by an intervening protein sequence, to a C-terminal 113 amino acid activation domain (AD) that can activate transcription when bound to DNA. See, e.g., J. Ma, & M. Ptashne, (1987), Deletion analysis of GAL4 defines two transcriptional activating segments, *Cell* 48:847-853; and, J. Ma, & M. Ptashne, (1987), The carboxy-terminal 30 amino acids of GAL4 are recognized by GAL80, *Cell* 50:137-142. By placing amber codons towards, e.g., the N-terminal DBD of a single polypeptide that contains both the N-terminal DBD of GAL4 and its C-terminal AD, amber suppression by the O-tRNA/O-RS pair can be linked to transcriptional activation by GAL4. GAL4 activated reporter genes can be used to perform both positive and negative selections with the gene.

[92] The medium used for negative selection can comprise a selecting or screening agent that is converted to a detectable substance by the negative selection marker. In one aspect of the invention, the detectable substance is a toxic substance. A polynucleotide that encodes a negative selection marker can be, e.g., an *ura3* gene. For example, the *URA3* reporter can be placed under control of a promoter that contains GAL4 DNA binding sites. When the negative selection marker is produced, e.g., by translation of a polynucleotide encoding the GAL4 with selector codons, GAL4 activates transcription of *URA3*. The negative selection is accomplished on a medium that comprises 5-fluoroorotic acid (5-FOA), which is converted into a detectable substance (e.g., a toxic substance which kills the cell) by the gene product of the *ura3* gene. See, e.g., J. D. Boeke, et al., (1984), A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoroorotic acid resistance, *Molecular & General Genetics* 197:345-346; M. Vidal, et al., (1996), Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system., *Proc. Natl. Acad. Sci. U. S. A.* 93:10321-10326; and, M. Vidal, et al., (1996), Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions., *Proc. Natl. Acad. Sci. U. S. A.* 93:10315-10320.

[93] As with the positive selection marker, the negative selection marker can also be any of a variety of molecules. The positive selection marker and/or the negative selection marker may be a polypeptide that fluoresces or catalyzes a luminescent reaction in the presence of a suitable reactant. For example, negative selection markers include, but are not limited to, e.g., luciferase, green fluorescent protein (GFP), YFP, EGFP, RFP, the product of an antibiotic resistant gene (e.g., chloramphenicol acetyltransferase (CAT)), the product of a

lacZ gene, transcriptional modulator protein, etc. The positive selection marker and/or the negative selection marker may be detected by fluorescence-activated cell sorting (FACS) or by luminescence. The positive selection marker and/or negative selection marker may comprise an affinity based screening marker. The same polynucleotide can encode both the positive selection marker and the negative selection marker. For example, the positive selection step, the negative selection step or both the positive and negative selection steps and can include using a reporter, wherein the reporter is detected by fluorescence-activated cell sorting (FACS). For example, a positive selection can be done first with a positive selection marker, e.g., chloramphenicol acetyltransferase (CAT) gene, where the CAT gene comprises a selector codon, e.g., an amber stop codon, in the CAT gene, which followed by a negative selection screen, that is based on the inability to suppress a selector codon(s), e.g., two or more, at positions within a negative marker, e.g., T7 RNA polymerase gene. The positive selection marker and the negative selection marker can be found on the same vector, e.g., plasmid. Expression of the negative marker drives expression of the reporter, e.g., green fluorescent protein (GFP). The stringency of the selection and screen can be varied, e.g., the intensity of the light need to fluorescence the reporter can be varied. A positive selection can be done with a reporter as a positive selection marker, which is screened by FACS, followed by a negative selection screen, that is based on the inability to suppress a selector codon(s), e.g., two or more, at positions within a negative marker, e.g., barnase gene.

[94]        Optionally, the reporter is displayed on a cell surface, e.g., on a phage display or the like. Cell-surface display, e.g., the OmpA-based cell-surface display system, relies on the expression of a particular epitope, e.g., a poliovirus C3 peptide fused to an outer membrane porin OmpA, on the surface of the *Escherichia coli* cell. The epitope is displayed on the cell surface only when a selector codon in the protein message is suppressed during translation. The displayed peptide then contains the amino acid recognized by one of the mutant aminoacyl-tRNA synthetases in the library, and the cell containing the corresponding synthetase gene can be isolated with antibodies raised against peptides containing specific unnatural amino acids. The OmpA-based cell-surface display system was developed and optimized by Georgiou et al. as an alternative to phage display. See, Francisco, J. A., Campbell, R., Iverson, B. L. & Georgiou, G. Production and fluorescence-activated cell sorting of *Escherichia coli* expressing a functional antibody fragment on the external surface. *Proc. Natl. Acad. Sci. U S A* 90:10444-8 (1993).

[95] Other embodiments of the present invention include carrying one or more of the selection steps in vitro. The selected component, e.g., synthetase and/or tRNA, can then be introduced into a cell for use in in vivo incorporation of an unnatural amino acid.

[96] Additional details for producing O-RS, and altering the substrate specificity of the synthetase can be found in U.S. patent application 10/126,931 entitled "Methods and Compositions for the Production of Orthogonal tRNA-Aminoacyl tRNA Synthetase Pairs;" and, USSN 10/825,867 entitled "EXPANDING THE EUKARYOTIC GENETIC CODE," which are incorporated by reference herein. Additional details for producing O-RS can be found in Hamano-Takaku et al., (2000) A mutant *Escherichia coli* Tyrosyl-tRNA Synthetase Utilizes the Unnatural Amino Acid Azatyrosine More Efficiently than Tyrosine, *Journal of Biological Chemistry*, 275(51):40324-40328; Kiga et al. (2002), An engineered *Escherichia coli* tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cell-free system, *PNAS* 99(15): 9715-9723; and, Francklyn et al., (2002), Aminoacyl-tRNA synthetases: Versatile players in the changing theater of translation; *RNA*, 8:1363-1372, each of which are incorporated by reference herein.

#### SOURCE AND HOST ORGANISMS

[97] The translational components of the present invention are typically derived from non-eukaryotic organisms. For example, the orthogonal O-tRNA can be derived from a non-eukaryotic organism, e.g., an archaeobacterium, such as *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* such as *Haloferax volcanii* and *Halobacterium* species *NRC-1*, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeuropyrum pernix*, or the like, or a eubacterium, such as *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, or the like, while the orthogonal O-RS can be derived from a non-eukaryotic organism, e.g., *Methanobacterium thermoautotrophicum*, *Halobacterium* such as *Haloferax volcanii* and *Halobacterium* species *NRC-1*, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeuropyrum pernix*, or the like, or a eubacterium, such as *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, or the like. In one embodiment, eukaryotic sources can also be used, including but not limited to, plants, algae, protists, fungi, yeasts, animals (e.g., mammals, insects, arthropods, etc.), or the like.

[98] The individual components of an O-tRNA/O-RS pair can be derived from the same organism or different organisms. In one embodiment, the O-tRNA/O-RS pair is from the same organism. Alternatively, the O-tRNA and the O-RS of the O-tRNA/O-RS pair are from different organisms. For example, the O-tRNA can be derived from, e.g., a *Halobacterium sp NRC-1*, and the O-RS can be derived from, e.g., a *Methanobacterium thermoautotrophicum*.

[99] The O-tRNA, O-RS or O-tRNA/O-RS pair can be selected or screened in vivo or in vitro and/or used in a cell, e.g., a non-eukaryotic cells (such as *E. coli* cell), or a eukaryotic cell, to produce a polypeptide with a selected amino acid (e.g., an unnatural amino acid). A non-eukaryotic cell can be from a variety of sources, such as the Archaea phylogenetic domain, including but not limited to, *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* such as *Haloferax volcanii* and *Halobacterium* species *NRC-1*, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeuiropyrum pernix*, or the like, or can belong to the Eubacteria phylogenetic domain (including but not limited to, *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, *Pseudomonas fluorescens*, *Pseudomonas aeruginosa*, *Pseudomonas putida*, etc.), or the like. A eukaryotic cell can be from a variety of sources, including but not limited to, a plant (e.g., complex plant such as monocots, or dicots), an algae, a protist, a fungus, a yeast (including but not limited to, *Saccharomyces cerevisiae*), an animal (including but not limited to, a mammal, an insect, an arthropod, etc.), or the like. Compositions of cells with translational components of the present invention are also a feature of the present invention. See also USSN 10/825,867 entitled "Expanding the Eukaryotic Genetic Code" for screening O-tRNA and/or O-RS in one species for use in another species.

[100] To express a polypeptide of interest with a selected amino acid in a host cell, one may subclone polynucleotides encoding a polypeptide of interest into an expression vector that contains a promoter to direct transcription, a transcription/translation terminator, and if for a nucleic acid encoding a protein, a ribosome binding site for translational initiation. Suitable bacterial promoters are well known in the art and described, e.g., in Sambrook *et al.* and Ausubel *et al.*

[101] Bacterial expression systems for expressing a polypeptide of interest are available in, including but not limited to, *E. coli*, *Bacillus sp.*, *Pseudomonas fluorescens*, *Pseudomonas aeruginosa*, *Pseudomonas putida*, and *Salmonella* (Palva *et al.*, *Gene* 22:229-235 (1983); Mosbach *et al.*, *Nature* 302:543-545 (1983)). Kits for such expression systems are

commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known in the art and are also commercially available.

[102] A tRNA and/or RS of the present invention and/or a polypeptide of interest may be utilized and/or expressed in any number of suitable expression systems including, for example, yeast, insect cells, mammalian cells, and bacteria. A description of exemplary expression systems is provided below.

[103] Yeast As used herein, the term “yeast” includes any of the various yeasts capable of expressing a polypeptide of interest. Such yeasts include, but are not limited to, ascosporogenous yeasts (*Endomycetales*), basidiosporogenous yeasts and yeasts belonging to the Fungi imperfecti (*Blastomycetes*) group. The ascosporogenous yeasts are divided into two families, *Spermophthoraceae* and *Saccharomycetaceae*. The latter is comprised of four subfamilies, *Schizosaccharomycoideae* (e.g., genus *Schizosaccharomyces*), *Nadsonioideae*, *Lipomycoideae* and *Saccharomycoideae* (e.g., genera *Pichia*, *Kluyveromyces* and *Saccharomyces*). The basidiosporogenous yeasts include the genera *Leucosporidium*, *Rhodosporeidium*, *Sporidiobolus*, *Filobasidium*, and *Filobasidiella*. Yeasts belonging to the Fungi Imperfecti (*Blastomycetes*) group are divided into two families, *Sporobolomycetaceae* (e.g., genera *Sporobolomyces* and *Bullera*) and *Cryptococcaceae* (e.g., genus *Candida*).

[104] Of particular interest for use with the present invention are species within the genera *Pichia*, *Kluyveromyces*, *Saccharomyces*, *Schizosaccharomyces*, *Hansenula*, *Torulopsis*, and *Candida*, including, but not limited to, *P. pastoris*, *P. guilliermondii*, *S. cerevisiae*, *S. carlsbergensis*, *S. diastaticus*, *S. douglasii*, *S. kluyveri*, *S. norbensis*, *S. oviformis*, *K. lactis*, *K. fragilis*, *C. albicans*, *C. maltosa*, and *H. polymorpha*. Yeast are generally available from a variety of sources including, but not limited to, the Yeast Genetic Stock Center, Department of Biophysics and Medical Physics, University of California (Berkeley, CA), and the American Type Culture Collection (“ATCC”) (Manassas, VA).

[105] The term “yeast host” or “yeast host cell” includes yeast that can be, or has been, used as a recipient for recombinant vectors or other transfer DNA. The term includes the progeny of the original yeast host cell that has received the recombinant vectors or other transfer DNA. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement to the original parent, due to accidental or deliberate mutation. Progeny of the parental cell that are sufficiently similar to the parent to be characterized by the relevant property, such as the

presence of a nucleotide sequence encoding a polypeptide of interest, are included in the progeny intended by this definition.

[106] Expression and transformation vectors, including extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeast hosts. For example, expression vectors have been developed for *S. cerevisiae* (Sikorski et al., GENETICS (1989) 122:19; Ito et al., J. BACTERIOL. (1983) 153:163; Hinnen et al., PROC. NATL. ACAD. SCI. USA (1978) 75:1929); *C. albicans* (Kurtz et al., MOL. CELL. BIOL. (1986) 6:142); *C. maltosa* (Kunze et al., J. BASIC MICROBIOL. (1985) 25:141); *H. polymorpha* (Gleeson et al., J. GEN. MICROBIOL. (1986) 132:3459; Roggenkamp et al., MOL. GENETICS AND GENOMICS (1986) 202:302); *K. fragilis* (Das et al., J. BACTERIOL. (1984) 158:1165); *K. lactis* (De Louvencourt et al., J. BACTERIOL. (1983) 154:737; Van den Berg et al., BIOTECHNOLOGY (NY) (1990) 8:135); *P. guillermondii* (Kunze et al., J. BASIC MICROBIOL. (1985) 25:141); *P. pastoris* (U.S. Patent Nos. 5,324,639; 4,929,555; and 4,837,148; Cregg et al., MOL. CELL. BIOL. (1985) 5:3376); *Schizosaccharomyces pombe* (Beach et al., NATURE (1982) 300:706); and *Y. lipolytica*; *A. nidulans* (Ballance et al., BIOCHEM. BIOPHYS. RES. COMMUN. (1983) 112:284-89; Tilburn et al., GENE (1983) 26:205-221; and Yelton et al., PROC. NATL. ACAD. SCI. USA (1984) 81:1470-74); *A. niger* (Kelly and Hynes, EMBO J. (1985) 4:475-479); *T. reesia* (EP 0 244 234); and filamentous fungi such as, e.g., *Neurospora*, *Penicillium*, *Tolypocladium* (WO 91/00357), each incorporated by reference herein.

[107] Control sequences for yeast vectors are known to those of ordinary skill in the art and include, but are not limited to, promoter regions from genes such as alcohol dehydrogenase (ADH) (EP 0 284 044); enolase; glucokinase; glucose-6-phosphate isomerase; glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH); hexokinase; phosphofructokinase; 3-phosphoglycerate mutase; and pyruvate kinase (PyK) (EP 0 329 203). The yeast PHO5 gene, encoding acid phosphatase, also may provide useful promoter sequences (Miyanojara et al., PROC. NATL. ACAD. SCI. USA (1983) 80:1). Other suitable promoter sequences for use with yeast hosts may include the promoters for 3-phosphoglycerate kinase (Hitzeman et al., J. BIOL. CHEM. (1980) 255:12073); and other glycolytic enzymes, such as pyruvate decarboxylase, triosephosphate isomerase, and phosphoglucose isomerase (Holland et al., BIOCHEMISTRY (1978) 17:4900; Hess et al., J. ADV. ENZYME REG. (1969) 7:149). Inducible yeast promoters having the additional advantage of transcription controlled by growth conditions may include the promoter regions for alcohol dehydrogenase 2; isocytochrome C; acid phosphatase; metallothionein;

glyceraldehyde-3-phosphate dehydrogenase; degradative enzymes associated with nitrogen metabolism; and enzymes responsible for maltose and galactose utilization. Suitable vectors and promoters for use in yeast expression are further described in EP 0 073 657.

[108] Yeast enhancers also may be used with yeast promoters. In addition, synthetic promoters may also function as yeast promoters. For example, the upstream activating sequences (UAS) of a yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region. See U.S. Patent Nos. 4,880,734 and 4,876,197, which are incorporated by reference herein. Other examples of hybrid promoters include promoters that consist of the regulatory sequences of the ADH2, GAL4, GAL10, or PHO5 genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK. See EP 0 164 556. Furthermore, a yeast promoter may include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription.

[109] Other control elements that may comprise part of the yeast expression vectors include terminators, for example, from GAPDH or the enolase genes (Holland et al., J. BIOL. CHEM. (1981) 256:1385). In addition, the origin of replication from the 2 $\mu$  plasmid origin is suitable for yeast. A suitable selection gene for use in yeast is the *trp1* gene present in the yeast plasmid. See Tschumper et al., GENE (1980) 10:157; Kingsman et al., GENE (1979) 7:141. The *trp1* gene provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan. Similarly, Leu2-deficient yeast strains (ATCC 20,622 or 38,626) are complemented by known plasmids bearing the *Leu2* gene.

[110] Methods of introducing exogenous DNA into yeast hosts are known to those of ordinary skill in the art, and typically include, but are not limited to, either the transformation of spheroplasts or of intact yeast host cells treated with alkali cations. For example, transformation of yeast can be carried out according to the method described in Hsiao et al., PROC. NATL. ACAD. SCI. USA (1979) 76:3829 and Van Solingen et al., J. BACT. (1977) 130:946. However, other methods for introducing DNA into cells such as by nuclear injection, electroporation, or protoplast fusion may also be used as described generally in SAMBROOK ET AL., MOLECULAR CLONING: A LAB. MANUAL (2001). Yeast host cells may then be cultured using standard techniques known to those of ordinary skill in the art.

[111] Other methods for expressing heterologous proteins in yeast host cells are known to those of ordinary skill in the art. *See generally* U.S. Patent Publication No. 20020055169, U.S. Patent Nos. 6,361,969; 6,312,923; 6,183,985; 6,083,723; 6,017,731; 5,674,706; 5,629,203; 5,602,034; and 5,089,398; U.S. Reexamined Patent Nos. RE37,343 and RE35,749; PCT Published Patent Applications WO 99/07862; WO 98/37208; and WO 98/26080; European Patent Applications EP 0 946 736; EP 0 732 403; EP 0 480 480; WO 90/10277; EP 0 340 986; EP 0 329 203; EP 0 324 274; and EP 0 164 556. *See also* Gellissen et al., *ANTONIE VAN LEEUWENHOEK* (1992) 62(1-2):79-93; Romanos et al., *YEAST* (1992) 8(6):423-488; Goeddel, *METHODS IN ENZYMOLOGY* (1990) 185:3-7, each incorporated by reference herein.

[112] The yeast host strains may be grown in fermentors during the amplification stage using standard feed batch fermentation methods known to those of ordinary skill in the art. The fermentation methods may be adapted to account for differences in a particular yeast host's carbon utilization pathway or mode of expression control. For example, fermentation of a *Saccharomyces* yeast host may require a single glucose feed, complex nitrogen source (e.g., casein hydrolysates), and multiple vitamin supplementation. In contrast, the methylotrophic yeast *P. pastoris* may require glycerol, methanol, and trace mineral feeds, but only simple ammonium (nitrogen) salts for optimal growth and expression. *See, e.g.*, U.S. Patent No. 5,324,639; Elliott et al., *J. PROTEIN CHEM.* (1990) 9:95; and Fieschko et al., *BIOTECH. BIOENG.* (1987) 29:1113, incorporated by reference herein.

[113] Such fermentation methods, however, may have certain common features independent of the yeast host strain employed. For example, a growth limiting nutrient, typically carbon, may be added to the fermentor during the amplification phase to allow maximal growth. In addition, fermentation methods generally employ a fermentation medium designed to contain adequate amounts of carbon, nitrogen, basal salts, phosphorus, and other minor nutrients (vitamins, trace minerals and salts, etc.). Examples of fermentation media suitable for use with *Pichia* are described in U.S. Patent Nos. 5,324,639 and 5,231,178, which are incorporated by reference herein.

[114] Baculovirus-Infected Insect Cells The term "insect host" or "insect host cell" refers to an insect that can be, or has been, used as a recipient for recombinant vectors or other transfer DNA. The term includes the progeny of the original insect host cell that has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement to the original

parent, due to accidental or deliberate mutation. Progeny of the parental cell that are sufficiently similar to the parent to be characterized by the relevant property, such as the presence of a nucleotide sequence encoding a polypeptide of interest, are included in the progeny intended by this definition.

[115] The selection of suitable insect cells for expression of a polypeptide of interest is known to those of ordinary skill in the art. Several insect species are well described in the art and are commercially available including *Aedes aegypti*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni*. In selecting insect hosts for expression, suitable hosts may include those shown to have, *inter alia*, good secretion capacity, low proteolytic activity, and overall robustness. Insect are generally available from a variety of sources including, but not limited to, the Insect Genetic Stock Center, Department of Biophysics and Medical Physics, University of California (Berkeley, CA); and the American Type Culture Collection ("ATCC") (Manassas, VA).

[116] Generally, the components of a baculovirus-infected insect expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene to be expressed; a wild type baculovirus with sequences homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media. The materials, methods and techniques used in constructing vectors, transfecting cells, picking plaques, growing cells in culture, and the like are known in the art and manuals are available describing these techniques.

[117] After inserting the heterologous gene into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome recombine. The packaged recombinant virus is expressed and recombinant plaques are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, for example, Invitrogen Corp. (Carlsbad, CA). These techniques are generally known to those of ordinary skill in the art and fully described in SUMMERS AND SMITH, TEXAS AGRICULTURAL EXPERIMENT STATION BULLETIN NO. 1555 (1987), herein incorporated by reference. *See also*, RICHARDSON, 39 METHODS IN MOLECULAR BIOLOGY: BACULOVIRUS EXPRESSION PROTOCOLS (1995); AUSUBEL ET AL., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY 16.9-16.11 (1994); KING

AND POSSEE, THE BACULOVIRUS SYSTEM: A LABORATORY GUIDE (1992); and O'REILLY ET AL., BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL (1992).

[118] Indeed, the production of various heterologous proteins using baculovirus/insect cell expression systems is known to those of ordinary skill in the art. *See, e.g.*, U.S. Patent Nos. 6,368,825; 6,342,216; 6,338,846; 6,261,805; 6,245,528, 6,225,060; 6,183,987; 6,168,932; 6,126,944; 6,096,304; 6,013,433; 5,965,393; 5,939,285; 5,891,676; 5,871,986; 5,861,279; 5,858,368; 5,843,733; 5,762,939; 5,753,220; 5,605,827; 5,583,023; 5,571,709; 5,516,657; 5,290,686; WO 02/06305; WO 01/90390; WO 01/27301; WO 01/05956; WO 00/55345; WO 00/20032; WO 99/51721; WO 99/45130; WO 99/31257; WO 99/10515; WO 99/09193; WO 97/26332; WO 96/29400; WO 96/25496; WO 96/06161; WO 95/20672; WO 93/03173; WO 92/16619; WO 92/02628; WO 92/01801; WO 90/14428; WO 90/10078; WO 90/02566; WO 90/02186; WO 90/01556; WO 89/01038; WO 89/01037; WO 88/07082, which are incorporated by reference herein.

[119] Vectors that are useful in baculovirus/insect cell expression systems are known in the art and include, for example, insect expression and transfer vectors derived from the baculovirus *Autographa californica* nuclear polyhedrosis virus (AcNPV), which is a helper-independent, viral expression vector. Viral expression vectors derived from this system usually use the strong viral polyhedrin gene promoter to drive expression of heterologous genes. *See generally*, O'Reilly ET AL., BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL (1992).

[120] Prior to inserting the foreign gene into the baculovirus genome, the above-described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are typically assembled into an intermediate transplacement construct (transfer vector). Intermediate transplacement constructs are often maintained in a replicon, such as an extra chromosomal element (e.g., plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification. More specifically, the plasmid may contain the polyhedrin polyadenylation signal (Miller, ANN. REV. MICROBIOL. (1988) 42:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli*.

[121] One commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed

including, for example, pVL985, which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 base pairs downstream from the ATT. *See* Luckow and Summers, *VIROLOGY* 170:31 (1989). Other commercially available vectors include, for example, PBlueBac4.5/V5-His; pBlueBacHis2; pMelBac; pBlueBac4.5 (Invitrogen Corp., Carlsbad, CA).

[122] After insertion of the heterologous gene, the transfer vector and wild type baculoviral genome are co-transfected into an insect cell host. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. *See* SUMMERS AND SMITH, TEXAS AGRICULTURAL EXPERIMENT STATION BULLETIN No. 1555 (1987); Smith et al., *MOL. CELL. BIOL.* (1983) 3:2156; Luckow and Summers, *VIROLOGY* (1989) 170:31. For example, the insertion can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. *See* Miller et al., *BIOESSAYS* (1989) 11(4):91.

[123] Transfection may be accomplished by electroporation. *See* TROTTER AND WOOD, 39 *METHODS IN MOLECULAR BIOLOGY* (1995); Mann and King, *J. GEN. VIROL.* (1989) 70:3501. Alternatively, liposomes may be used to transfect the insect cells with the recombinant expression vector and the baculovirus. *See, e.g.,* Liebman et al., *BIOTECHNIQUES* (1999) 26(1):36; Graves et al., *BIOCHEMISTRY* (1998) 37:6050; Nomura et al., *J. BIOL. CHEM.* (1998) 273(22):13570; Schmidt et al., *PROTEIN EXPRESSION AND PURIFICATION* (1998) 12:323; Siffert et al., *NATURE GENETICS* (1998) 18:45; TILKINS ET AL., *CELL BIOLOGY: A LABORATORY HANDBOOK* 145-154 (1998); Cai et al., *PROTEIN EXPRESSION AND PURIFICATION* (1997) 10:263; Dolphin et al., *NATURE GENETICS* (1997) 17:491; Kost et al., *GENE* (1997) 190:139; Jakobsson et al., *J. BIOL. CHEM.* (1996) 271:22203; Rowles et al., *J. BIOL. CHEM.* (1996) 271(37):22376; Reverey et al., *J. BIOL. CHEM.* (1996) 271(39):23607-10; Stanley et al., *J. BIOL. CHEM.* (1995) 270:4121; Sisk et al., *J. VIROL.* (1994) 68(2):766; and Peng et al., *BIOTECHNIQUES* (1993) 14(2):274. Commercially available liposomes include, for example, Cellfectin® and Lipofectin® (Invitrogen, Corp., Carlsbad, CA). In addition, calcium phosphate transfection may be used. *See* TROTTER AND WOOD, 39 *METHODS IN MOLECULAR BIOLOGY* (1995); Kitts, *NAR* (1990) 18(19):5667; and Mann and King, *J. GEN. VIROL.* (1989) 70:3501.

[124] Baculovirus expression vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA

polymerase and initiating the downstream (3') transcription of a coding sequence (e.g., structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region typically includes an RNA polymerase binding site and a transcription initiation site. A baculovirus promoter may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Moreover, expression may be either regulated or constitutive.

[125] Structural genes, abundantly transcribed at late times in the infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein (FRIESEN ET AL., *The Regulation of Baculovirus Gene Expression* in THE MOLECULAR BIOLOGY OF BACULOVIRUSES (1986); EP 0 127 839 and 0 155 476) and the gene encoding the p10 protein (Vlak et al., J. GEN. VIROL. (1988) 69:765).

[126] The newly formed baculovirus expression vector is packaged into an infectious recombinant baculovirus and subsequently grown plaques may be purified by techniques known to those of ordinary skill in the art. See Miller et al., BIOESSAYS (1989) 11(4):91; SUMMERS AND SMITH, TEXAS AGRICULTURAL EXPERIMENT STATION BULLETIN No. 1555 (1987).

[127] Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*, *Aedes aegypti* (ATCC No. CCL-125), *Bombyx mori* (ATCC No. CRL-8910), *Drosophila melanogaster* (ATCC No. 1963), *Spodoptera frugiperda*, and *Trichoplusia ni*. See Wright, NATURE (1986) 321:718; Carbonell et al., J. VIROL. (1985) 56:153; Smith et al., MOL. CELL. BIOL. (1983) 3:2156. See generally, Fraser et al., IN VITRO CELL. DEV. BIOL. (1989) 25:225. More specifically, the cell lines used for baculovirus expression vector systems commonly include, but are not limited to, Sf9 (*Spodoptera frugiperda*) (ATCC No. CRL-1711), Sf21 (*Spodoptera frugiperda*) (Invitrogen Corp., Cat. No. 11497-013 (Carlsbad, CA)), Tri-368 (*Trichoplusia ni*), and High-Five™ BTI-TN-5B1-4 (*Trichoplusia ni*).

[128] Cells and culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression, and cell culture technology is generally known to those of ordinary skill in the art.

[129] *E. Coli, Pseudomonas species, and other Prokaryotes* Bacterial expression techniques are known to those of ordinary skill in the art. A wide variety of vectors are available for use in bacterial hosts. The vectors may be single copy or low or high multicopy vectors. Vectors may serve for cloning and/or expression. In view of the ample literature concerning vectors, commercial availability of many vectors, and even manuals describing vectors and their restriction maps and characteristics, no extensive discussion is required here. As is well-known, the vectors normally involve markers allowing for selection, which markers may provide for cytotoxic agent resistance, prototrophy or immunity. Frequently, a plurality of markers is present, which provide for different characteristics.

[130] A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (e.g. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region typically includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in *Escherichia coli* (*E. coli*) [Raibaud et al., ANNU. REV. GENET. (1984) 18:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

[131] Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (lac) [Chang et al., NATURE (1977) 198:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (trp) [Goeddel et al., NUC. ACIDS RES. (1980) 8:4057; Yelverton et al., NUCL. ACIDS RES. (1981) 9:731; U.S. Pat. No. 4,738,921; EP Pub. Nos. 036 776 and 121 775, which are incorporated by reference herein]. The  $\beta$ -galactosidase (bla) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In Interferon 3

(Ed. I. Gresser)], bacteriophage lambda PL [Shimatake et al., NATURE (1981) 292:128] and T5 [U.S. Pat. No. 4,689,406, which are incorporated by reference herein] promoter systems also provide useful promoter sequences. Strong promoters, such as the T7 promoter may be used to induce the polypeptide of interest at high levels. Examples of such vectors are known to those of ordinary skill in the art and include the pET29 series from Novagen, and the pPOP vectors described in WO99/05297, which is incorporated by reference herein. Such expression systems produce high levels of polypeptide in the host without compromising host cell viability or growth parameters. pET19 (Novagen) is another vector known in the art.

[132] In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [U.S. Pat. No. 4,551,433, which is incorporated by reference herein]. For example, the tac promoter is a hybrid trp-lac promoter comprised of both trp promoter and lac operon sequences that is regulated by the lac repressor [Amann et al., GENE (1983) 25:167; de Boer et al., PROC. NATL. ACAD. SCI. (1983) 80:21]. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier et al., J. MOL. BIOL. (1986) 189:113; Tabor et al., Proc Natl. Acad. Sci. (1985) 82:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an E. coli operator region (EP Pub. No. 267 851).

[133] In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In E. coli, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine et al., NATURE (1975) 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' end of E. coli 16S rRNA [Steitz et al. "Genetic signals and nucleotide sequences in messenger RNA", In Biological Regulation and Development: Gene Expression (Ed. R. F. Goldberger, 1979)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-

binding site [Sambrook et al. "Expression of cloned genes in *Escherichia coli*", Molecular Cloning: A Laboratory Manual, 1989].

[134] The term "bacterial host" or "bacterial host cell" refers to a bacterial that can be, or has been, used as a recipient for recombinant vectors or other transfer DNA. The term includes the progeny of the original bacterial host cell that has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement to the original parent, due to accidental or deliberate mutation. Progeny of the parental cell that are sufficiently similar to the parent to be characterized by the relevant property, such as the presence of a nucleotide sequence encoding a polypeptide of interest, are included in the progeny intended by this definition.

[135] The selection of suitable host bacteria for expression of polypeptides is known to those of ordinary skill in the art. In selecting bacterial hosts for expression, suitable hosts may include those shown to have, *inter alia*, good inclusion body formation capacity, low proteolytic activity, and overall robustness. Bacterial hosts are generally available from a variety of sources including, but not limited to, the Bacterial Genetic Stock Center, Department of Biophysics and Medical Physics, University of California (Berkeley, CA); and the American Type Culture Collection ("ATCC") (Manassas, VA). Industrial/pharmaceutical fermentation generally use bacterial derived from K strains (e.g. W3110) or from bacteria derived from B strains (e.g. BL21). These strains are particularly useful because their growth parameters are extremely well known and robust. In addition, these strains are non-pathogenic, which is commercially important for safety and environmental reasons. Other examples of suitable *E. coli* hosts include, but are not limited to, strains of BL21, DH10B, or derivatives thereof. In another embodiment of the methods of the present invention, the *E. coli* host is a protease minus strain including, but not limited to, OMP- and LON-. The host cell strain may be a species of *Pseudomonas*, including but not limited to, *Pseudomonas fluorescens*, *Pseudomonas aeruginosa*, and *Pseudomonas putida*. *Pseudomonas fluorescens* biovar 1, designated strain MB101, is known to be useful for recombinant production and is available for therapeutic protein production processes. Examples of a *Pseudomonas* expression system include the system available from The Dow Chemical Company as a host strain (Midland, MI available on the World Wide Web at [dow.com](http://dow.com)). U.S. Patent Nos. 4,755,465 and 4,859,600, which are incorporated by reference herein, describe the use of *Pseudomonas* strains as a host cell for hGH production.

[136] Once a recombinant host cell strain has been established (i.e., the expression construct has been introduced into the host cell and host cells with the proper expression construct are isolated), the recombinant host cell strain is cultured under conditions appropriate for production of the polypeptide of interest. As will be apparent to one of skill in the art, the method of culture of the recombinant host cell strain will be dependent on the nature of the expression construct utilized and the identity of the host cell. Recombinant host strains are normally cultured using methods that are well known to the art. Recombinant host cells are typically cultured in liquid medium containing assimilatable sources of carbon, nitrogen, and inorganic salts and, optionally, containing vitamins, amino acids, growth factors, and other proteinaceous culture supplements known to those of ordinary skill in the art. Liquid media for culture of host cells may optionally contain antibiotics or anti-fungals to prevent the growth of undesirable microorganisms and/or compounds including, but not limited to, antibiotics to select for host cells containing the expression vector.

[137] Recombinant host cells may be cultured in batch or continuous formats, with either cell harvesting (in the case where the polypeptide of interest accumulates intracellularly) or harvesting of culture supernatant in either batch or continuous formats. For production in prokaryotic host cells, batch culture and cell harvest are preferred.

### SELECTOR CODONS

[138] Selector codons of the present invention expand the genetic codon framework of protein biosynthetic machinery. For example, a selector codon includes, e.g., a unique three base codon, a nonsense codon, such as a stop codon, including but not limited to, an amber codon (UAG), an ochre codon, or an opal codon (UGA), an unnatural codon, a four base (or more) codon, a rare codon, or the like. A number of selector codons can be introduced into a desired gene or polynucleotide, e.g., one or more, two or more, three or more, etc.

[139] In one embodiment, the methods involve the use of a selector codon that is a stop codon for the incorporation of a selected amino acid, e.g., an unnatural amino acid, *in vivo*. For example, an O-tRNA is produced that recognizes the stop codon and is aminoacylated by an O-RS with a selected amino acid. This O-tRNA is not recognized by the naturally occurring host's aminoacyl-tRNA synthetases. Conventional site-directed mutagenesis can be used to introduce the stop codon at the site of interest in a polypeptide of interest. *See, e.g., Sayers, J.R., et al. (1988), 5'-3' Exonucleases in phosphorothioate-based*

*oligonucleotide-directed mutagenesis*. Nucleic Acids Res, 16:791-802. When the O-RS, O-tRNA and the nucleic acid that encodes a polypeptide of interest are combined, e.g., in vivo, the selected amino acid is incorporated in response to the stop codon to give a polypeptide containing the selected amino acid, e.g., an unnatural amino acid, at the specified position. In one embodiment of the present invention, a stop codon used as a selector codon is an amber codon, UAG, and/or an opal codon, UGA. For example, see SEQ ID NO.: 6 for an example of an O-tRNA that recognizes an amber codon, and see SEQ ID NO.: 7 for an example of an O-tRNA that recognizes an opal codon. A genetic code in which UAG and UGA are both used as a selector codon can encode 22 amino acids while preserving the ochre nonsense codon, UAA, which is the most abundant termination signal.

[140] The incorporation of selected amino acids, e.g., unnatural amino acids, in vivo can be done without significant perturbation of the host cell. For example in non-eukaryotic cells, such as *Escherichia coli*, because the suppression efficiency for the UAG codon depends upon the competition between the O-tRNA, e.g., the amber suppressor tRNA, and the release factor 1 (RF1) (which binds to the UAG codon and initiates release of the growing peptide from the ribosome), the suppression efficiency can be modulated by, e.g., either increasing the expression level of O-tRNA, e.g., the suppressor tRNA, or using an RF1 deficient strain. In eukaryotic cells, because the suppression efficiency for the UAG codon depends upon the competition between the O-tRNA, e.g., the amber suppressor tRNA, and a eukaryotic release factor (e.g., eRF) (which binds to a stop codon and initiates release of the growing peptide from the ribosome), the suppression efficiency can be modulated by, e.g., increasing the expression level of O-tRNA, e.g., the suppressor tRNA.

[141] Unnatural amino acids can also be encoded with rare codons. For example, when the arginine concentration in an in vitro protein synthesis reaction is reduced, the rare arginine codon, AGG, has proven to be efficient for insertion of Ala by a synthetic tRNA acylated with alanine. See, e.g., Ma et al., Biochemistry, 32:7939 (1993). In this case, the synthetic tRNA competes with the naturally occurring tRNA<sup>Arg</sup>, which exists as a minor species in *Escherichia coli*. Some organisms do not use all triplet codons. An unassigned codon AGA in *Micrococcus luteus* has been utilized for insertion of amino acids in an in vitro transcription/translation extract. See, e.g., Kowal and Oliver, Nucl. Acid. Res., 25:4685 (1997). Components of the present invention can be generated to use these rare codons in vivo.

[142] Selector codons also comprise extended codons, e.g., four or more base codons, such as, four, five, six or more base codons. Examples of four base codons include but are not limited to, AGGA, CUAG, UAGA, CCCU, and the like. Examples of five base codons include but are not limited to, AGGAC, CCCCUC, CUAGA, CUACU, UAGGC and the like. A feature may include using extended codons based on frameshift suppression. Four or more base codons can insert, e.g., one or multiple selected amino acids, including but not limited to, unnatural amino acids, into the same protein. For example, in the presence of mutated O-tRNA's, e.g., a special frameshift suppressor tRNA's, with anticodon loops, e.g., with a CU(X)<sub>n</sub> XXXAA sequence (where n=1), the four or more base codon is read as single amino acid. For example, see SEQ ID NOs.: 6, 12 from PCT/US04/22061 for O-tRNA's that recognize a four base codon. In other embodiments, the anticodon loops can decode, e.g., at least a four-base codon, at least a five-base codon, or at least a six-base codon or more. Since there are 256 possible four-base codons, multiple unnatural amino acids can be encoded in the same cell using a four or more base codon. See, Anderson et al., (2002) *Exploring the Limits of Codon and Anticodon Size*, Chemistry and Biology, 9:237-244; Magliery, (2001) *Expanding the Genetic Code: Selection of Efficient Suppressors of Four-base Codons and Identification of "Shifty" Four-base Codons with a Library Approach in Escherichia coli*, J. Mol. Biol. 307: 755-769.

[143] For example, four-base codons have been used to incorporate unnatural amino acids into proteins using in vitro biosynthetic methods. See, e.g., Ma et al., (1993) Biochemistry, 32:7939; and Hohsaka et al., (1999) J. Am. Chem. Soc., 121:34. CGGG and AGGU were used to simultaneously incorporate 2-naphthylalanine and an NBD derivative of lysine into streptavidin in vitro with two chemically acylated frameshift suppressor tRNA's. See, e.g., Hohsaka et al., (1999) J. Am. Chem. Soc., 121:12194. In an in vivo study, Moore et al. examined the ability of tRNA<sup>Leu</sup> derivatives with NCUA anticodons to suppress UAGN codons (N can be U, A, G, or C), and found that the quadruplet UAGA can be decoded by a tRNA<sup>Leu</sup> with a UCUA anticodon with an efficiency of 13 to 26% with little decoding in the 0 or -1 frame. See, Moore et al., (2000) J. Mol. Biol., 298:195. In one embodiment, extended codons based on rare codons or nonsense codons can be used in invention, which can reduce missense readthrough and frameshift suppression at other unwanted sites.

[144] For a given system, a selector codon can also include one of the natural three base codons, where the endogenous system does not use (or rarely uses) the natural base

codon. For example, this includes a system that is lacking a tRNA that recognizes the natural three base codon, and/or a system where the three base codon is a rare codon.

[145] Selector codons optionally include unnatural base pairs. These unnatural base pairs further expand the existing genetic alphabet. One extra base pair increases the number of triplet codons from 64 to 125. Properties of third base pairs include stable and selective base pairing, efficient enzymatic incorporation into DNA with high fidelity by a polymerase, and the efficient continued primer extension after synthesis of the nascent unnatural base pair. Descriptions of unnatural base pairs which can be adapted for methods and compositions include, e.g., Hirao, et al., (2002) *An unnatural base pair for incorporating amino acid analogues into protein*, Nature Biotechnology, 20:177-182. See, also, Wu, Y., et al., (2002) J. Am. Chem. Soc. 124:14626-14630. Other relevant publications are listed below.

[146] For in vivo usage, the unnatural nucleoside is membrane permeable and is phosphorylated to form the corresponding triphosphate. In addition, the increased genetic information is stable and not destroyed by cellular enzymes. Previous efforts by Benner and others took advantage of hydrogen bonding patterns that are different from those in canonical Watson-Crick pairs, the most noteworthy example of which is the iso-C:iso-G pair. See, e.g., Switzer et al., (1989) J. Am. Chem. Soc., 111:8322; and Piccirilli et al., (1990) Nature, 343:33; Kool, (2000) Curr. Opin. Chem. Biol., 4:602. These bases in general mispair to some degree with natural bases and cannot be enzymatically replicated. Kool and co-workers demonstrated that hydrophobic packing interactions between bases can replace hydrogen bonding to drive the formation of base pair. See, Kool, (2000) Curr. Opin. Chem. Biol., 4:602; and Guckian and Kool, (1998) Angew. Chem. Int. Ed. Engl., 36, 2825. In an effort to develop an unnatural base pair satisfying all the above requirements, Schultz, Romesberg and co-workers have systematically synthesized and studied a series of unnatural hydrophobic bases. A PICS:PICS self-pair is found to be more stable than natural base pairs, and can be efficiently incorporated into DNA by Klenow fragment of Escherichia coli DNA polymerase I (KF). See, e.g., McMinn et al., (1999) J. Am. Chem. Soc., 121: 11585-6; and Ogawa et al., (2000) J. Am. Chem. Soc., 122:3274. A 3MN:3MN self-pair can be synthesized by KF with efficiency and selectivity sufficient for biological function. See, e.g., Ogawa et al., (2000) J. Am. Chem. Soc., 122:8803. However, both bases act as a chain terminator for further replication. A mutant DNA polymerase has been recently evolved that can be used to replicate the PICS self pair. In addition, a 7AI self pair can be replicated. See, e.g., Tae et al., (2001) J. Am. Chem. Soc., 123:7439. A novel metallobase pair, Dipic:Py, has also been

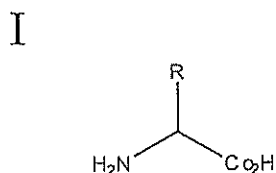
developed, which forms a stable pair upon binding Cu(II). See, Meggers et al., (2000) J. Am. Chem. Soc., 122:10714. Because extended codons and unnatural codons are intrinsically orthogonal to natural codons, the methods of the present invention can take advantage of this property to generate orthogonal tRNAs for them.

[147] A translational bypassing system can also be used to incorporate a selected amino acid, e.g., an unnatural amino acid, in a desired polypeptide. In a translational bypassing system, a large sequence is inserted into a gene but is not translated into protein. The sequence contains a structure that serves as a cue to induce the ribosome to hop over the sequence and resume translation downstream of the insertion.

#### SELECTED AND UNNATURAL AMINO ACIDS

[148] As used herein, a selected amino acid refers to any desired naturally occurring amino acid or unnatural amino acid. A naturally occurring amino acid includes any one of the twenty genetically encoded alpha-amino acids: alanine, arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, valine. In one embodiment, the selected amino acid is incorporated into a growing polypeptide chain with high fidelity, e.g., at greater than about 70% efficiency for given selector codon, at greater than 75% efficiency for a given selector codon, at greater than about 80% efficiency for a given selector codon, at greater than about 85% efficiency for a given selector codon, at greater than about 90% efficiency for a given selector codon, at greater than about 95% efficiency for a given selector codon, or at greater than about 99% or more efficiency for a given selector codon.

[149] As used herein, an unnatural amino acid refers to any amino acid, modified amino acid, or amino acid analogue other than selenocysteine and/or pyrrolysine and the following twenty genetically encoded alpha-amino acids: alanine, arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, valine. The generic structure of an alpha-amino acid is illustrated by Formula I:

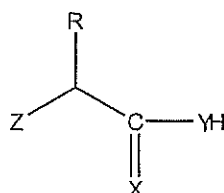


[150] An unnatural amino acid is typically any structure having Formula I wherein the R group is any substituent other than one used in the twenty natural amino acids. See, e.g., *Biochemistry* by L. Stryer, 3<sup>rd</sup> ed. 1988, Freeman and Company, New York, for structures of the twenty natural amino acids. Note that, the unnatural amino acids of the present invention can be naturally occurring compounds other than the twenty alpha-amino acids above.

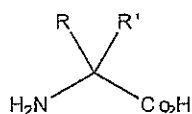
[151] Because the unnatural amino acids of the present invention typically differ from the natural amino acids only in the structure of the side chain, the unnatural amino acids form amide bonds with other amino acids, including but not limited to, natural or unnatural, in the same manner in which they are formed in naturally occurring proteins. However, the unnatural amino acids have side chain groups that distinguish them from the natural amino acids. For example, R in Formula I may comprise an alkyl-, aryl-, acyl-, keto-, azido-, hydroxyl-, hydrazine, cyano-, halo-, hydrazide, alkenyl, alkynyl, ether, thiol, seleno-, sulfonyl-, borate, boronate, phospho, phosphono, phosphine, heterocyclic, enone, imine, aldehyde, ester, thioacid, hydroxylamine, amine, and the like, or any combination thereof. Other non-naturally occurring amino acids include, but are not limited to, amino acids comprising a photoactivatable cross-linker, spin-labeled amino acids, fluorescent amino acids, metal binding amino acids, metal-containing amino acids, radioactive amino acids, amino acids with novel functional groups, amino acids that covalently or noncovalently interact with other molecules, photocaged and/or photoisomerizable amino acids, amino acids comprising biotin or a biotin analogue, glycosylated amino acids such as a sugar substituted serine, other carbohydrate modified amino acids, keto-containing amino acids, amino acids comprising polyethylene glycol or polyether, heavy atom substituted amino acids, chemically cleavable and/or photocleavable amino acids, amino acids with an elongated side chains as compared to natural amino acids, including but not limited to, polyethers or long chain hydrocarbons, including but not limited to, greater than about 5 or greater than about 10 carbons, carbon-linked sugar-containing amino acids, redox-active amino acids, amino thioacid containing amino acids, and amino acids comprising one or more toxic moiety. See, also, U.S. Patent Application Publications 2003/0082575 and 2003/0108885, which are incorporated by reference herein. Unnatural amino acids may have a photoactivatable cross-linker that is used, e.g., to link a protein to a solid support. Unnatural amino acids may have a saccharide moiety attached to the amino acid side chain.

[152] In addition to unnatural amino acids that contain novel side chains, unnatural amino acids also optionally comprise modified backbone structures, e.g., as illustrated by the structures of Formula II and III:

II



III



wherein Z typically comprises OH, NH<sub>2</sub>, SH, NH-R', or S-R'; X and Y, which can be the same or different, typically comprise S or O, and R and R', which are optionally the same or different, are typically selected from the same list of constituents for the R group described above for the unnatural amino acids having Formula I as well as hydrogen. For example, unnatural amino acids may comprise substitutions in the amino or carboxyl group as illustrated by Formulas II and III. Unnatural amino acids of this type include, but are not limited to, α-hydroxy acids, α-thioacids, α-aminothiocarboxylates, e.g., with side chains corresponding to the common twenty natural amino acids or unnatural side chains. In addition, substitutions at the α-carbon optionally include L, D, or α-α-disubstituted amino acids such as D-glutamate, D-alanine, D-methyl-O-tyrosine, aminobutyric acid, and the like. Other structural alternatives include cyclic amino acids, such as proline analogues as well as 3, 4, 6, 7, 8, and 9 membered ring proline analogues, β and γ amino acids such as substituted β-alanine and γ-amino butyric acid.

[153] Many unnatural amino acids are based on natural amino acids, such as tyrosine, glutamine, phenylalanine, and the like. Tyrosine analogs include para-substituted tyrosines, ortho-substituted tyrosines, and meta substituted tyrosines, wherein the substituted tyrosine comprises a keto group (including but not limited to, an acetyl group), a benzoyl group, an

amino group, a hydrazine, an hydroxyamine, a thiol group, a carboxy group, an isopropyl group, a methyl group, a C<sub>6</sub> - C<sub>20</sub> straight chain or branched hydrocarbon, a saturated or unsaturated hydrocarbon, an O-methyl group, a polyether group, a nitro group, or the like. In addition, multiply substituted aryl rings are also contemplated. Glutamine analogs include, but are not limited to,  $\alpha$ -hydroxy derivatives,  $\gamma$ -substituted derivatives, cyclic derivatives, and amide substituted glutamine derivatives. Example phenylalanine analogs include, but are not limited to, para-substituted phenylalanines, ortho-substituted phenylalanines, and meta-substituted phenylalanines, wherein the substituent comprises a hydroxy group, a methoxy group, a methyl group, an allyl group, an aldehyde, an azido, an iodo, a bromo, a keto group (including but not limited to, an acetyl group), or the like. Specific examples of unnatural amino acids include, but are not limited to, a *p*-acetyl-L-phenylalanine, a *p*-propargyl-phenylalanine, O-methyl-L-tyrosine, an L-3-(2-naphthyl)alanine, a 3-methyl-phenylalanine, an O-4-allyl-L-tyrosine, a 4-propyl-L-tyrosine, a tri-O-acetyl-GlcNAc $\beta$ -serine, an L-Dopa, a fluorinated phenylalanine, an isopropyl-L-phenylalanine, a *p*-azido-L-phenylalanine, a *p*-acyl-L-phenylalanine, a *p*-benzoyl-L-phenylalanine, an L-phosphoserine, a phosphoserine, a phosphonotyrosine, a *p*-iodo-phenylalanine, a *p*-bromophenylalanine, a *p*-amino-L-phenylalanine, an isopropyl-L-phenylalanine, and a *p*-propargyloxy-phenylalanine, and the like. Examples of structures of a variety of unnatural amino acids are provided in, for example, WO 2002/085923 entitled "In vivo incorporation of unnatural amino acids," which is incorporated by reference herein. See also Kiick et al., (2002) *Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation*, PNAS 99:19-24, which is incorporated by reference herein, for additional methionine analogs.

[154] A non-natural amino acid incorporated into a polypeptide at the amino terminus can be composed of an R group that is any substituent other than one used in the twenty natural amino acids and a 2<sup>nd</sup> reactive group different from the NH<sub>2</sub> group normally present in  $\alpha$ -amino acids (see Formula I). A similar non-natural amino acid can be incorporated at the carboxyl terminus with a 2<sup>nd</sup> reactive group different from the COOH group normally present in  $\alpha$ -amino acids (see Formula I).

[155] The unnatural amino acids of the invention may be selected or designed to provide additional characteristics unavailable in the twenty natural amino acids. For example, unnatural amino acid may be optionally designed or selected to modify the biological properties of a protein, e.g., into which they are incorporated. For example, the following properties may be optionally modified by inclusion of an unnatural amino acid into a protein:

toxicity, biodistribution, solubility, stability, e.g., thermal, hydrolytic, oxidative, resistance to enzymatic degradation, and the like, facility of purification and processing, structural properties, spectroscopic properties, chemical and/or photochemical properties, catalytic activity, redox potential, half-life, ability to react with other molecules, e.g., covalently or noncovalently, and the like.

[156] The structures of a variety of unnatural amino acids are provided in, for example, Figures 16, 17, 18, 19, 26, and 29 of WO 2002/085923 entitled “In vivo incorporation of unnatural amino acids,” which is incorporated by reference herein. The examples are not meant to be limiting in any way of amino acids that may be attached to a tRNA of the present invention.

[157] One advantage of an unnatural amino acid is that it presents additional chemical moieties that can be used to add additional molecules. These modifications can be made in vivo in a eukaryotic or non-eukaryotic cell, or in vitro. Thus, in certain embodiments, the post-translational modification is through the unnatural amino acid. An unnatural amino acid in a polypeptide may be used to attach another molecule to the polypeptide, including but not limited to, a label, a dye, a polymer, a water-soluble polymer, a derivative of polyethylene glycol, a photocrosslinker, a radionuclide, a cytotoxic compound, a drug, an affinity label, a photoaffinity label, a reactive compound, a resin, a second protein or polypeptide or polypeptide analog, an antibody or antibody fragment, a metal chelator, a cofactor, a fatty acid, a carbohydrate, a polynucleotide, a DNA, a RNA, an antisense polynucleotide, a saccharide, a water-soluble dendrimer, a cyclodextrin, an inhibitory ribonucleic acid, a biomaterial, a nanoparticle, a spin label, a fluorophore, a metal-containing moiety, a radioactive moiety, a novel functional group, a group that covalently or noncovalently interacts with other molecules, a photocaged moiety, an actinic radiation excitable moiety, a photoisomerizable moiety, biotin, a derivative of biotin, a biotin analogue, a moiety incorporating a heavy atom, a chemically cleavable group, a photocleavable group, an elongated side chain, a carbon-linked sugar, a redox-active agent, an amino thioacid, a toxic moiety, an isotopically labeled moiety, a biophysical probe, a phosphorescent group, a chemiluminescent group, an electron dense group, a magnetic group, an intercalating group, a chromophore, an energy transfer agent, a biologically active agent, a detectable label, a small molecule, a quantum dot, a nanotransmitter, or any combination of the above or any other desirable compound or substance, comprising a second reactive group to at least one

unnatural amino acid comprising a first reactive group utilizing chemistry methodology that is known to one of ordinary skill in the art to be suitable for the particular reactive groups.

[158] For example, the post-translational modification can be through a nucleophilic-electrophilic reaction. Most reactions currently used for the selective modification of proteins involve covalent bond formation between nucleophilic and electrophilic reaction partners, including but not limited to the reaction of  $\alpha$ -haloketones with histidine or cysteine side chains. Selectivity in these cases is determined by the number and accessibility of the nucleophilic residues in the protein. In proteins of the invention, other more selective reactions can be used such as the reaction of an unnatural keto-amino acid with hydrazides or aminooxy compounds, in vitro and in vivo. See, e.g., Cornish, et al., (1996) *J. Am. Chem. Soc.*, 118:8150-8151; Mahal, et al., (1997) *Science*, 276:1125-1128; Wang, et al., (2001) *Science* 292:498-500; Chin, et al., (2002) *J. Am. Chem. Soc.* 124:9026-9027; Chin, et al., (2002) *Proc. Natl. Acad. Sci.*, 99:11020-11024; Wang, et al., (2003) *Proc. Natl. Acad. Sci.*, 100:56-61; Zhang, et al., (2003) *Biochemistry*, 42:6735-6746; and, Chin, et al., (2003) *Science*, 301:964-7, all of which are incorporated by reference herein. This allows the selective labeling of virtually any protein with a host of reagents including fluorophores, crosslinking agents, saccharide derivatives and cytotoxic molecules. See also, U.S. Patent No. 6,927,042 entitled "Glycoprotein synthesis," which is incorporated by reference herein. Post-translational modifications, including but not limited to, through an azido amino acid, can also made through the Staudinger ligation (including but not limited to, with triarylphosphine reagents). See, e.g., Kiick et al., (2002) *Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation*, *PNAS* 99:19-24.

#### Chemical Synthesis of Unnatural Amino Acids

[159] Many unnatural amino acids are commercially available, e.g., from Sigma-Aldrich (St. Louis, MO, USA), Novabiochem (a division of EMD Biosciences, Darmstadt, Germany), or Peptech (Burlington, MA, USA). Those that are not commercially available are optionally synthesized as provided herein or using standard methods known to those of ordinary skill in the art. For organic synthesis techniques, see, e.g., *Organic Chemistry* by Fessenden and Fessenden, (1982, Second Edition, Willard Grant Press, Boston Mass.); *Advanced Organic Chemistry* by March (Third Edition, 1985, Wiley and Sons, New York); and *Advanced Organic Chemistry* by Carey and Sundberg (Third Edition, Parts A and B, 1990, Plenum Press, New York). Additional publications describing the synthesis of

unnatural amino acids include, e.g., WO 2002/085923 entitled "In vivo incorporation of Unnatural Amino Acids;" Matsoukas et al., (1995) *J. Med. Chem.*, 38, 4660-4669; King, F.E. & Kidd, D.A.A. (1949) *A New Synthesis of Glutamine and of  $\gamma$ -Dipeptides of Glutamic Acid from Phthylated Intermediates*. *J. Chem. Soc.*, 3315-3319; Friedman, O.M. & Chatterji, R. (1959) *Synthesis of Derivatives of Glutamine as Model Substrates for Anti-Tumor Agents*. *J. Am. Chem. Soc.* 81, 3750-3752; Craig, J.C. et al. (1988) *Absolute Configuration of the Enantiomers of 7-Chloro-4 [[4-(diethylamino)-1-methylbutyl]amino]quinoline (Chloroquine)*. *J. Org. Chem.* 53, 1167-1170; Azoulay, M., Vilmont, M. & Frappier, F. (1991) *Glutamine analogues as Potential Antimalarials*, *Eur. J. Med. Chem.* 26, 201-5; Koskinen, A.M.P. & Rapoport, H. (1989) *Synthesis of 4-Substituted Prolines as Conformationally Constrained Amino Acid Analogues*. *J. Org. Chem.* 54, 1859-1866; Christie, B.D. & Rapoport, H. (1985) *Synthesis of Optically Pure Pípecolates from L-Asparagine. Application to the Total Synthesis of (+)-Apovincamine through Amino Acid Decarbonylation and Iminium Ion Cyclization*. *J. Org. Chem.* 50:1239-1246; Barton et al., (1987) *Synthesis of Novel  $\alpha$ -Amino-Acids and Derivatives Using Radical Chemistry: Synthesis of L- and D- $\alpha$ -Amino-Adipic Acids, L- $\alpha$ -aminopimelic Acid and Appropriate Unsaturated Derivatives*. *Tetrahedron* 43:4297-4308; and, Subasinghe et al., (1992) *Quisqualic acid analogues: synthesis of beta-heterocyclic 2-aminopropanoic acid derivatives and their activity at a novel quisqualate-sensitized site*. *J. Med. Chem.* 35:4602-7. See also, U.S. Patent Publication No. US 2004/0198637 entitled "Protein Arrays," which is incorporated by reference.

#### Cellular uptake of unnatural amino acids

[160] Unnatural amino acid uptake by a cell is one issue that is typically considered when designing and selecting unnatural amino acids, e.g., for incorporation into a protein. For example, the high charge density of  $\alpha$ -amino acids suggests that these compounds are unlikely to be cell permeable. Natural amino acids are taken up into the cell via a collection of protein-based transport systems. A rapid screen can be done which assesses which unnatural amino acids, if any, are taken up by cells. See, e.g., the toxicity assays in, e.g., U.S. Patent Publication No. US 2004/0198637 entitled "Protein Arrays" which is incorporated by reference herein, and Liu, D.R. & Schultz, P.G. (1999) *Progress toward the evolution of an organism with an expanded genetic code*. *PNAS United States* 96:4780-4785. Although uptake is easily analyzed with various assays, an alternative to designing unnatural

amino acids that are amenable to cellular uptake pathways is to provide biosynthetic pathways to create amino acids *in vivo*.

#### Biosynthesis of Unnatural Amino Acids

[161] Many biosynthetic pathways already exist in cells for the production of amino acids and other compounds. While a biosynthetic method for a particular unnatural amino acid may not exist in nature, including but not limited to, in a cell, the invention provides such methods. For example, biosynthetic pathways for unnatural amino acids are optionally generated in host cell by adding new enzymes or modifying existing host cell pathways. Additional new enzymes are optionally naturally occurring enzymes or artificially evolved enzymes. For example, the biosynthesis of *p*-aminophenylalanine (as presented in an example in WO 2002/085923 entitled "In vivo incorporation of unnatural amino acids") relies on the addition of a combination of known enzymes from other organisms. The genes for these enzymes can be introduced into a cell by transforming the cell with a plasmid comprising the genes. The genes, when expressed in the cell, provide an enzymatic pathway to synthesize the desired compound. Examples of the types of enzymes that are optionally added are provided in the examples below. Additional enzymes sequences are found, for example, in Genbank. Artificially evolved enzymes are also optionally added into a cell in the same manner. In this manner, the cellular machinery and resources of a cell are manipulated to produce unnatural amino acids.

[162] A variety of methods are available for producing novel enzymes for use in biosynthetic pathways or for evolution of existing pathways. For example, recursive recombination, e.g., as developed by Maxygen, Inc. (available on the World Wide Web at maxygen.com), is optionally used to develop novel enzymes and pathways. *See, e.g.,* Stemmer (1994), *Rapid evolution of a protein in vitro by DNA shuffling*, Nature 370(4):389-391; and, Stemmer, (1994), *DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution*, Proc. Natl. Acad. Sci. USA., 91:10747-10751. Similarly DesignPath™, developed by Genencor (available on the World Wide Web at genencor.com) is optionally used for metabolic pathway engineering, e.g., to engineer a pathway to create O-methyl-L-tyrosine in a cell. This technology reconstructs existing pathways in host organisms using a combination of new genes, including but not limited to, those identified through functional genomics, and molecular evolution and design. Diversa Corporation (available on the World Wide Web at diversa.com) also provides technology for

rapidly screening libraries of genes and gene pathways, including but not limited to, to create new pathways.

[163] Typically, the unnatural amino acid produced with an engineered biosynthetic pathway of the present invention is produced in a concentration sufficient for efficient protein biosynthesis, e.g., a natural cellular amount, but not to such a degree as to affect the concentration of the other amino acids or exhaust cellular resources. Typical concentrations produced *in vivo* in this manner are about 10 mM to about 0.05 mM. Once a cell is transformed with a plasmid comprising the genes used to produce enzymes desired for a specific pathway and an unnatural amino acid is generated, *in vivo* selections are optionally used to further optimize the production of the unnatural amino acid for both ribosomal protein synthesis and cell growth.

#### NUCLEIC ACID AND POLYPEPTIDE SEQUENCE AND VARIANTS

[164] As described above and below, the invention provides for nucleic acid polynucleotide sequences and polypeptide amino acid sequences, e.g., tRNA's and RSs, and, e.g., compositions and methods comprising said sequences. Examples of said sequences, e.g., tRNA's and RSs are disclosed herein. However, one of skill in the art will appreciate that the invention is not limited to those sequences disclosed herein, e.g., the Examples. One of skill will appreciate that the invention also provides many related and unrelated sequences with the functions described herein, e.g., encoding an O-tRNA or an O-RS.

[165] The invention provides polypeptides (O-RSs) and polynucleotides, e.g., O-tRNA, polynucleotides that encode O-RSs or portions thereof, oligonucleotides used to isolate aminoacyl-tRNA synthetase clones, etc. Polynucleotides of the present invention include those that encode proteins or polypeptides of interests of the present invention with one or more selector codon. In addition, polynucleotides of the present invention include, e.g., a polynucleotide comprising a nucleotide sequence as set forth in any one of SEQ ID NO.: 1, 2, 3; a polynucleotide that is complementary to, or a conservative variation thereof. Similarly, a nucleic acid that hybridizes to a polynucleotide indicated above under highly stringent conditions over substantially the entire length of the nucleic acid is a polynucleotide of the present invention.

[166] In certain embodiments, a vector (e.g., a plasmid, a cosmid, a phage, a bacterium, a virus, a naked polynucleotide, a conjugated polynucleotide, etc.) comprises a polynucleotide of the present invention. In one embodiment, the vector is an expression

vector. In another embodiment, the expression vector includes a promoter operably linked to one or more of the polynucleotides of the present invention. In another embodiment, a cell comprises a vector that includes a polynucleotide of the present invention.

[167] One of skill will also appreciate that many variants of the disclosed sequences are included in the invention. For example, conservative variations of the disclosed sequences that yield a functionally identical sequence are included in the invention. Variants of the nucleic acid polynucleotide sequences, wherein the variants hybridize to at least one disclosed sequence, are considered to be included in the invention. Unique subsequences of the sequences disclosed herein, as determined by, e.g., standard sequence comparison techniques, are also included in the invention.

#### Conservative variations

[168] Owing to the degeneracy of the genetic code, “silent substitutions” (*i.e.*, substitutions in a nucleic acid sequence which do not result in an alteration in an encoded polypeptide) are an implied feature of *every* nucleic acid sequence which encodes an amino acid. Similarly, “conservative amino acid substitutions,” in one or a few amino acids in an amino acid sequence are substituted with different amino acids with highly similar properties, are also readily identified as being highly similar to a disclosed construct. Such conservative variations of each disclosed sequence are a feature of the present invention.

[169] “Conservative variations” of a particular nucleic acid sequence refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or, where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. One of ordinary skill in the art will recognize that individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence are “conservatively modified variations” or “conservatively modified variants” where the alterations result in the deletion of an amino acid, addition of an amino acid, or substitution of an amino acid with a chemically similar amino acid. Thus, “conservative variations” of a listed polypeptide sequence of the present invention include substitutions of a small percentage, typically less than 5%, more typically less than 4%, 2% or 1%, of the amino acids of the polypeptide sequence, with a conservatively selected amino acid of the same conservative substitution group. The addition of sequences which do not alter the encoded activity of a nucleic acid molecule, such as the addition of a non-functional sequence, is a conservative variation of the basic nucleic acid.

[170] Conservative substitution tables providing functionally similar amino acids are known to those of ordinary skill in the art. The following eight groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Glycine (G);
- 2) Aspartic acid (D), Glutamic acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V);
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W);
- 7) Serine (S), Threonine (T); and
- 8) Cysteine (C), Methionine (M)

(see, e.g., Creighton, *Proteins: Structures and Molecular Properties* (W H Freeman & Co.; 2nd edition (December 1993)

#### Nucleic Acid Hybridization

[171] Comparative hybridization can be used to identify nucleic acids of the present invention, such as SEQ ID NO.: 1- 3, including conservative variations of nucleic acids of the present invention, and this comparative hybridization method is a preferred method of distinguishing nucleic acids of the present invention. In addition, target nucleic acids which hybridize to the nucleic acids represented by SEQ ID NO: 1-3 under high, ultra-high, and/or ultra-ultra high stringency conditions are a feature of the present invention. Examples of such nucleic acids include those with one or a few silent or conservative nucleic acid substitutions as compared to a given nucleic acid sequence.

[172] A test nucleic acid is said to specifically hybridize to a probe nucleic acid when it hybridizes at least  $\frac{1}{2}$  as well to the probe as to the perfectly matched complementary target, i.e., with a signal to noise ratio at least  $\frac{1}{2}$  as high as hybridization of the probe to the target under conditions in which the perfectly matched probe binds to the perfectly matched complementary target with a signal to noise ratio that is at least about 5x-10x as high as that observed for hybridization to any of the unmatched target nucleic acids.

[173] Nucleic acids “hybridize” when they associate, typically in solution. Nucleic acids hybridize due to a variety of well characterized physico-chemical forces, such as hydrogen bonding, solvent exclusion, base stacking and the like. The phrase “stringent hybridization conditions” refers to conditions of low ionic strength and high temperature as is

known in the art. Typically, under stringent conditions a probe will hybridize to its target subsequence in a complex mixture of nucleic acid (including but not limited to, total cellular or library DNA or RNA) but does not hybridize to other sequences in the complex mixture. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes* part I chapter 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays," (Elsevier, New York), as well as in Ausubel *et al.*, *Current Protocols in Molecular Biology* (1995). Hames and Higgins (1995) Gene Probes 1 IRL Press at Oxford University Press, Oxford, England, (Hames and Higgins 1) and Hames and Higgins (1995) Gene Probes 2 IRL Press at Oxford University Press, Oxford, England (Hames and Higgins 2) provide details on the synthesis, labeling, detection and quantification of DNA and RNA, including oligonucleotides. Generally, stringent conditions are selected to be about 5-10° C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength pH. The  $T_m$  is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at  $T_m$ , 50% of the probes are occupied at equilibrium). Stringent conditions may be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (including but not limited to, 10 to 50 nucleotides) and at least about 60°C for long probes (including but not limited to, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal may be at least two times background, optionally 10 times background hybridization. Exemplary stringent hybridization conditions can be as following: 50% formamide, 5X SSC, and 1% SDS, incubating at 42°C, or 5X SSC, 1% SDS, incubating at 65°C, with wash in 0.2X SSC, and 0.1% SDS at 65°C. Such washes can be performed for 5, 15, 30, 60, 120, or more minutes.

[174] An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or Northern blot is 50% formalin with 1 mg of heparin at 42°C, with the hybridization being carried out overnight. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see*, Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (3rd ed. 2001) for a description of SSC buffer). Often the high stringency wash is

preceded by a low stringency wash to remove background probe signal. An example low stringency wash is 2x SSC at 40°C for 15 minutes. In general, a signal to noise ratio of 5x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

[175] “Stringent hybridization wash conditions” in the context of nucleic acid hybridization experiments such as Southern and Northern hybridizations are sequence dependent, and are different under different environmental parameters. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993), *supra*. and in Hames and Higgins, 1 and 2, *supra*. Stringent hybridization and wash conditions can easily be determined empirically for any test nucleic acid. For example, in determining highly stringent hybridization and wash conditions, the hybridization and wash conditions are gradually increased (e.g., by increasing temperature, decreasing salt concentration, increasing detergent concentration and/or increasing the concentration of organic solvents such as formalin in the hybridization or wash), until a selected set of criteria are met. For example, the hybridization and wash conditions are gradually increased until a probe binds to a perfectly matched complementary target with a signal to noise ratio that is at least 5x as high as that observed for hybridization of the probe to an unmatched target.

[176] “Very stringent” conditions are selected to be equal to the thermal melting point ( $T_m$ ) for a particular probe. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the test sequence hybridizes to a perfectly matched probe. For the purposes of the present invention, generally, “highly stringent” hybridization and wash conditions are selected to be about 5°C lower than the  $T_m$  for the specific sequence at a defined ionic strength and pH.

[177] “Ultra high-stringency” hybridization and wash conditions are those in which the stringency of hybridization and wash conditions are increased until the signal to noise ratio for binding of the probe to the perfectly matched complementary target nucleic acid is at least 10x as high as that observed for hybridization to any of the unmatched target nucleic acids. A target nucleic acid which hybridizes to a probe under such conditions, with a signal to noise ratio of at least  $\frac{1}{2}$  that of the perfectly matched complementary target nucleic acid is said to bind to the probe under ultra-high stringency conditions.

[178] Similarly, even higher levels of stringency can be determined by gradually increasing the hybridization and/or wash conditions of the relevant hybridization assay. For example, those in which the stringency of hybridization and wash conditions are increased until the signal to noise ratio for binding of the probe to the perfectly matched complementary target nucleic acid is at least 10X, 20X, 50X, 100X, or 500X or more as high as that observed for hybridization to any of the unmatched target nucleic acids. A target nucleic acid which hybridizes to a probe under such conditions, with a signal to noise ratio of at least ½ that of the perfectly matched complementary target nucleic acid is said to bind to the probe under ultra-ultra-high stringency conditions.

[179] Nucleic acids which do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, *e.g.*, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

#### Unique subsequences

[180] In one aspect, the invention provides a nucleic acid that comprises a unique subsequence in a nucleic acid selected from the sequences of O-tRNA's and O-RSs disclosed herein. The unique subsequence is unique as compared to a nucleic acid corresponding to any known O-tRNA or O-RS nucleic acid sequence. Alignment can be performed using, *e.g.*, BLAST set to default parameters. Any unique subsequence is useful, *e.g.*, as a probe to identify the nucleic acids of the present invention.

[181] Similarly, the invention includes a polypeptide which comprises a unique subsequence in a polypeptide selected from the sequences of O-RSs disclosed herein. Here, the unique subsequence is unique as compared to a polypeptide corresponding to any of known polypeptide sequence.

[182] The invention also provides for target nucleic acids which hybridizes under stringent conditions to a unique coding oligonucleotide which encodes a unique subsequence in a polypeptide selected from the sequences of O-RSs wherein the unique subsequence is unique as compared to a polypeptide corresponding to any of the control polypeptides (*e.g.*, parental sequences from which synthetases of the present invention were derived, *e.g.*, by mutation). Unique sequences are determined as noted above.

### Sequence comparison, identity, and homology

[183] The terms “identical” or percent “identity,” in the context of two or more nucleic acid or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a comparison window, or designated region, as measured using one of the sequence comparison algorithms described below (or other algorithms available to persons of ordinary skill in the art) or by manual alignment and visual inspection.

[184] The phrase “substantially identical,” in the context of two nucleic acids or polypeptides (*e.g.*, DNAs encoding an O-tRNA or O-RS, or the amino acid sequence of an O-RS) refers to two or more sequences or subsequences that have at least about 60%, about 80%, about 90-95%, about 98%, about 99% or more nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence over a comparison window, or designated region, as measured using a sequence comparison algorithm (or other algorithms available to persons of ordinary skill in the art) or by manual alignment and visual inspection. Such “substantially identical” sequences are typically considered to be “homologous,” without reference to actual ancestry. “Substantial identity” may exist over a region of the sequences that is at least about 50 residues in length, a region of at least about 100 residues, or a region of at least about 150 residues, or over the full length of the two sequences to be compared.

[185] For sequence comparison and homology determination, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

[186] Methods of alignment of sequences for comparison are known to those of ordinary skill in the art. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, Adv. Appl. Math. 2:482c (1981), by the homology alignment algorithm of Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), by the search for similarity method of Pearson & Lipman, Proc. Nat’l. Acad. Sci. USA 85:2444 (1988), by computerized implementations of these algorithms (GAP,

BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (*see e.g., Ausubel et al., Current Protocols in Molecular Biology* (1995 supplement)).

[187] One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm and BLAST 2.0 algorithms, which are described in Altschul *et al.* (1997) *Nuc. Acids Res.* 25:3389-3402, and Altschul *et al., J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information available at the World Wide Web at [ncbi.nlm.nih.gov](http://ncbi.nlm.nih.gov). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length  $W$  in the query sequence, which either match or satisfy some positive-valued threshold score  $T$  when aligned with a word of the same length in a database sequence.  $T$  is referred to as the neighborhood word score threshold (Altschul *et al., supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters  $M$  (reward score for a pair of matching residues; always  $> 0$ ) and  $N$  (penalty score for mismatching residues; always  $< 0$ ). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity  $X$  from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters  $W$ ,  $T$ , and  $X$  determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength ( $W$ ) of 11, an expectation ( $E$ ) of 10, a cutoff of 100,  $M=5$ ,  $N=-4$ , and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength ( $W$ ) of 3, an expectation ( $E$ ) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff (1992) *Proc. Natl. Acad. Sci. USA* 89:10915) alignments ( $B$ ) of 50, expectation ( $E$ ) of 10,  $M=5$ ,  $N=-4$ , and a comparison of both strands. The BLAST algorithm is typically performed with the "low complexity" filter turned off.

[188] In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g., Karlin &*

Altschul, Proc. Nat'l. Acad. Sci. USA 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid may be considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, less than about 0.01, or less than about 0.001.

#### Mutagenesis and Other Molecular Biology Techniques

[189] Polynucleotide and polypeptides of the present invention and used in the invention can be manipulated using molecular biological techniques. A nucleotide sequence may be conveniently modified by site-directed mutagenesis in accordance with conventional methods. Alternatively, the nucleotide sequence may be prepared by chemical synthesis, including but not limited to, by using an oligonucleotide synthesizer, wherein oligonucleotides are designed based on the amino acid sequence of the desired polypeptide, and preferably selecting those codons that are favored in the host cell in which the recombinant polypeptide will be produced. For example, several small oligonucleotides coding for portions of the desired polypeptide may be synthesized and assembled by PCR, ligation or ligation chain reaction. *See, e.g., Barany, et al., Proc. Natl. Acad. Sci.* 88: 189-193 (1991); U.S. Patent 6,521,427 which are incorporated by reference herein.

[190] This invention utilizes routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this invention include Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (3rd ed. 2001); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel *et al.*, eds., 1994)).

[191] General texts which describe molecular biological techniques include Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook *et al.*, Molecular Cloning - A Laboratory Manual (2nd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 ("Sambrook") and Current Protocols in Molecular Biology, F.M. Ausubel *et al.*, eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 1999) ("Ausubel"). These texts describe mutagenesis, the use of vectors, promoters and many other relevant topics related to, e.g., the generation of genes or polynucleotides that include selector codons for production of proteins

that include selected amino acids (e.g., unnatural amino acids), orthogonal tRNA's, orthogonal synthetases, and pairs thereof.

[192] Various types of mutagenesis are used in the invention for a variety of purposes, including but not limited to, to produce novel synthetases or tRNAs, to mutate tRNA molecules, to produce libraries of tRNAs, to mutate RS molecules, to produce libraries of synthetases, to produce selector codons, to insert selector codons that encode a selected amino acid in a protein or polypeptide of interest. They include but are not limited to, site-directed, random point mutagenesis, homologous recombination, DNA shuffling or other recursive mutagenesis methods, chimeric construction, mutagenesis using uracil containing templates, oligonucleotide-directed mutagenesis, phosphorothioate-modified DNA mutagenesis, mutagenesis using gapped duplex DNA or the like, or any combination thereof. Additional suitable methods include point mismatch repair, mutagenesis using repair-deficient host strains, restriction-selection and restriction-purification, deletion mutagenesis, mutagenesis by total gene synthesis, double-strand break repair, and the like. Mutagenesis, including but not limited to, involving chimeric constructs, is also included in the present invention. In one embodiment, mutagenesis can be guided by known information of the naturally occurring molecule or altered or mutated naturally occurring molecule, including but not limited to, sequence, sequence comparisons, physical properties, secondary, tertiary, or quaternary structure, crystal structure or the like.

[193] The texts and examples found herein describe these procedures. Additional information is found in the following publications and references cited within: Ling et al., *Approaches to DNA mutagenesis: an overview*, Anal Biochem. 254(2): 157-178 (1997); Dale et al., *Oligonucleotide-directed random mutagenesis using the phosphorothioate method*, Methods Mol. Biol. 57:369-374 (1996); Smith, *In vitro mutagenesis*, Ann. Rev. Genet. 19:423-462 (1985); Botstein & Shortle, *Strategies and applications of in vitro mutagenesis*, Science 229:1193-1201 (1985); Carter, *Site-directed mutagenesis*, Biochem. J. 237:1-7 (1986); Kunkel, *The efficiency of oligonucleotide directed mutagenesis*, in Nucleic Acids & Molecular Biology (Eckstein, F. and Lilley, D.M.J. eds., Springer Verlag, Berlin) (1987); Kunkel, *Rapid and efficient site-specific mutagenesis without phenotypic selection*, Proc. Natl. Acad. Sci. USA 82:488-492 (1985); Kunkel et al., *Rapid and efficient site-specific mutagenesis without phenotypic selection*, Methods in Enzymol. 154, 367-382 (1987); Bass et al., *Mutant Trp repressors with new DNA-binding specificities*, Science 242:240-245 (1988); Zoller & Smith, *Oligonucleotide-directed mutagenesis using M13-derived vectors: an*

efficient and general procedure for the production of point mutations in any DNA fragment, Nucleic Acids Res. 10:6487-6500 (1982); Zoller & Smith, *Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors*, Methods in Enzymol. 100:468-500 (1983); Zoller & Smith, *Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template*, Methods in Enzymol. 154:329-350 (1987); Taylor et al., *The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA*, Nucl. Acids Res. 13: 8749-8764 (1985); Taylor et al., *The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA*, Nucl. Acids Res. 13: 8765-8785 (1985); Nakamaye & Eckstein, *Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis*, Nucl. Acids Res. 14: 9679-9698 (1986); Sayers et al., *5'-3' Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis*, Nucl. Acids Res. 16:791-802 (1988); Sayers et al., *Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide*, (1988) Nucl. Acids Res. 16: 803-814; Kramer et al., *The gapped duplex DNA approach to oligonucleotide-directed mutation construction*, Nucl. Acids Res. 12: 9441-9456 (1984); Kramer & Fritz *Oligonucleotide-directed construction of mutations via gapped duplex DNA*, Methods in Enzymol. 154:350-367 (1987); Kramer et al., *Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations*, Nucl. Acids Res. 16: 7207 (1988); Fritz et al., *Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro*, Nucl. Acids Res. 16: 6987-6999 (1988); Kramer et al., *Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of E. coli*, Cell 38:879-887 (1984); Carter et al., *Improved oligonucleotide site-directed mutagenesis using M13 vectors*, Nucl. Acids Res. 13: 4431-4443 (1985); Carter, *Improved oligonucleotide-directed mutagenesis using M13 vectors*, Methods in Enzymol. 154: 382-403 (1987); Eghtedarzadeh & Henikoff, *Use of oligonucleotides to generate large deletions*, Nucl. Acids Res. 14: 5115 (1986); Wells et al., *Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin*, Phil. Trans. R. Soc. Lond. A 317: 415-423 (1986); Nambiar et al., *Total synthesis and cloning of a gene coding for the ribonuclease S protein*, Science 223: 1299-1301 (1984); Sakmar and Khorana, *Total synthesis and expression of a gene for the alpha-subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin)*, Nucl. Acids Res. 14: 6361-6372 (1988); Wells et al., *Cassette mutagenesis: an efficient method for generation of multiple mutations at defined*

sites, Gene 34:315-323 (1985); Grundström et al., *Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesis*, Nucl. Acids Res. 13: 3305-3316 (1985); Mandecki, *Oligonucleotide-directed double-strand break repair in plasmids of Escherichia coli: a method for site-specific mutagenesis*, Proc. Natl. Acad. Sci. USA, 83:7177-7181 (1986); Arnold, *Protein engineering for unusual environments*, Current Opinion in Biotechnology 4:450-455 (1993); Sieber, et al., *Nature Biotechnology*, 19:456-460 (2001); W. P. C. Stemmer, Nature 370, 389-91 (1994); and, I. A. Lorimer, I. Pastan, Nucleic Acids Res. 23, 3067-8 (1995). Additional details on many of the above methods can be found in Methods in Enzymology Volume 154, which also describes useful controls for trouble-shooting problems with various mutagenesis methods.

[194] Oligonucleotides, e.g., for use in mutagenesis of the present invention, e.g., mutating libraries of synthetases, or altering tRNAs, are typically synthesized chemically according to the solid phase phosphoramidite triester method described by Beaucage and Caruthers, *Tetrahedron Letts.* 22(20):1859-1862, (1981) e.g., using an automated synthesizer, as described in Needham-VanDevanter et al., *Nucleic Acids Res.*, 12:6159-6168 (1984).

[195] In addition, essentially any nucleic acid can be custom or standard ordered from any of a variety of commercial sources, such as The Midland Certified Reagent Company (mcrc@oligos.com), The Great American Gene Company (www.genco.com), ExpressGen Inc. (www.expressgen.com), Operon Technologies Inc. (Alameda, Calif.) and many others.

[196] The invention also relates to eukaryotic host cells, non-eukaryotic host cells, and organisms for the in vivo incorporation of an unnatural amino acid via orthogonal tRNA/RS pairs. Host cells are genetically engineered (including but not limited to, transformed, transduced or transfected) with the polynucleotides of the present invention or constructs which include a polynucleotide of the present invention, including but not limited to, a vector of the present invention, which can be, for example, a cloning vector or an expression vector. For example, the coding regions for the orthogonal tRNA, the orthogonal tRNA synthetase, and the protein to be derivatized are operably linked to gene expression control elements that are functional in the desired host cell. The vector can be, for example, in the form of a plasmid, a cosmid, a phage, a bacterium, a virus, a naked polynucleotide, or a conjugated polynucleotide. The vectors are introduced into cells and/or microorganisms by standard methods including electroporation (Fromm et al., Proc. Natl. Acad. Sci. USA 82, 5824 (1985)), infection by viral vectors, high velocity ballistic penetration by small particles with

the nucleic acid either within the matrix of small beads or particles, or on the surface (Klein et al., Nature 327, 70-73 (1987)), and/or the like.

[197] Several well-known methods of introducing target nucleic acids into cells are available, any of which can be used in the invention. These include: fusion of the recipient cells with bacterial protoplasts containing the DNA, electroporation, projectile bombardment, and infection with viral vectors (discussed further, below), etc. Bacterial cells can be used to amplify the number of plasmids containing DNA constructs of this invention. The bacteria are grown to log phase and the plasmids within the bacteria can be isolated by a variety of methods known in the art (*see*, for instance, Sambrook). In addition, kits are commercially available for the purification of plasmids from bacteria, (*see*, e.g., EasyPrep™, FlexiPrep™, both from Pharmacia Biotech; StrataClean™ from Stratagene; and, QIAprep™ from Qiagen). The isolated and purified plasmids are then further manipulated to produce other plasmids, used to transfect cells or incorporated into related vectors to infect organisms. Typical vectors contain transcription and translation terminators, transcription and translation initiation sequences, and promoters useful for regulation of the expression of the particular target nucleic acid. The vectors optionally comprise generic expression cassettes containing at least one independent terminator sequence, sequences permitting replication of the cassette in eukaryotes, or prokaryotes, or both, (including but not limited to, shuttle vectors) and selection markers for both prokaryotic and eukaryotic systems. Vectors are suitable for replication and/or integration in prokaryotes, eukaryotes, or both. *See*, Gillam & Smith, Gene 8:81 (1979); Roberts, *et al.*, Nature, 328:731 (1987); Schneider, E., *et al.*, Protein Expr. Purif. 6(1)10-14 (1995); Ausubel, Sambrook, Berger (*all supra*). A catalogue of bacteria and bacteriophages useful for cloning is provided, e.g., by the ATCC, e.g., *The ATCC Catalogue of Bacteria and Bacteriophage* (1992) Gherna *et al.* (eds) published by the ATCC. Additional basic procedures for sequencing, cloning and other aspects of molecular biology and underlying theoretical considerations are also found in Watson *et al.* (1992) *Recombinant DNA Second Edition Scientific American Books, NY*. In addition, essentially any nucleic acid (and virtually any labeled nucleic acid, whether standard or non-standard) can be custom or standard ordered from any of a variety of commercial sources, such as the Midland Certified Reagent Company (Midland, TX available on the World Wide Web at [mcrc.com](http://mcrc.com)), The Great American Gene Company (Ramona, CA available on the World Wide Web at [genco.com](http://genco.com)), ExpressGen Inc. (Chicago, IL available on the World Wide Web at [expressgen.com](http://expressgen.com)), Operon Technologies Inc. (Alameda, CA) and many others.

[198] The engineered host cells can be cultured in conventional nutrient media modified as appropriate for such activities as, for example, screening steps, activating promoters or selecting transformants. These cells can optionally be cultured into transgenic organisms. Other useful references, e.g. for cell isolation and culture (e.g., for subsequent nucleic acid isolation) include Freshney (1994) Culture of Animal Cells, a Manual of Basic Technique, third edition, Wiley- Liss, New York and the references cited therein; Payne *et al.* (1992) Plant Cell and Tissue Culture in Liquid Systems John Wiley & Sons, Inc. New York, NY; Gamborg and Phillips (eds.) (1995) Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York) and Atlas and Parks (eds.) The Handbook of Microbiological Media (1993) CRC Press, Boca Raton, FL.

[199] The ability to incorporate unnatural amino acids directly into proteins in vivo offers a wide variety of advantages including but not limited to high yields of mutant proteins, technical ease, the potential to study the mutant proteins in cells or possibly in living organisms and the use of these mutant proteins in therapeutic treatments and diagnostic uses. The ability to include unnatural amino acids with various sizes, acidities, nucleophilicities, hydrophobicities, and other properties into proteins can greatly expand our ability to rationally and systematically manipulate the structures of proteins, both to probe protein function and create new proteins or organisms with novel properties.

#### PROTEINS AND POLYPEPTIDES OF INTEREST

[200] The incorporation of an unnatural amino acid can be done for a variety of purposes, including but not limited to, tailoring changes in protein structure and/or function, changing size, acidity, nucleophilicity, hydrogen bonding, hydrophobicity, accessibility of protease target sites, targeting to a moiety (including but not limited to, for a protein array), adding a biologically active molecule, attaching a polymer, attaching a radionuclide, modulating serum half-life, modulating tissue penetration (e.g. tumors), modulating active transport, modulating tissue, cell or organ specificity or distribution, modulating immunogenicity, modulating protease resistance, etc. Proteins that include an unnatural amino acid can have enhanced or even entirely new catalytic or biophysical properties. For example, the following properties are optionally modified by inclusion of an unnatural amino acid into a protein: toxicity, biodistribution, structural properties, spectroscopic properties, chemical and/or photochemical properties, catalytic ability, half-life (including but not limited to, serum half-life), ability to react with other molecules, including but not limited to,

covalently or noncovalently, and the like. The compositions including proteins that include at least one unnatural amino acid are useful for, including but not limited to, novel therapeutics, diagnostics, catalytic enzymes, industrial enzymes, binding proteins (including but not limited to, antibodies), and including but not limited to, the study of protein structure and function. See, e.g., Dougherty, (2000) *Unnatural Amino Acids as Probes of Protein Structure and Function*, Current Opinion in Chemical Biology, 4:645-652.

[201] A protein may have at least one, including but not limited to, at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, or at least ten or more unnatural amino acids. The unnatural amino acids can be the same or different, including but not limited to, there can be 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more different sites in the protein that comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more different unnatural amino acids. A protein may have at least one, but fewer than all, of a particular amino acid present in the protein is substituted with the unnatural amino acid. For a given protein with more than one unnatural amino acids, the unnatural amino acids can be identical or different (including but not limited to, the protein can include two or more different types of unnatural amino acids, or can include two of the same unnatural amino acid). For a given protein with more than two unnatural amino acids, the unnatural amino acids can be the same, different or a combination of a multiple unnatural amino acid of the same kind with at least one different unnatural amino acid.

[202] By producing proteins or polypeptides of interest with at least one unnatural amino acid in eukaryotic cells, proteins or polypeptides will typically include eukaryotic post-translational modifications. In certain embodiments, a protein includes at least one unnatural amino acid and at least one post-translational modification that is made in vivo by a eukaryotic cell, where the post-translational modification is not made by a prokaryotic cell. For example, the post-translation modification includes, including but not limited to, acetylation, acylation, lipid-modification, palmitoylation, palmitate addition, phosphorylation, glycolipid-linkage modification, glycosylation, and the like. In yet another aspect, the post-translation modification includes proteolytic processing of precursors (including but not limited to, calcitonin precursor, calcitonin gene-related peptide precursor, preproparathyroid hormone, preproinsulin, proinsulin, prepro-opiomelanocortin, pro-opiomelanocortin and the like), assembly into a multisubunit protein or macromolecular assembly, translation to another site in the cell (including but not limited to, to organelles, such as the endoplasmic reticulum, the Golgi apparatus, the nucleus, lysosomes, peroxisomes,

mitochondria, chloroplasts, vacuoles, etc., or through the secretory pathway). In certain embodiments, the protein comprises a secretion or localization sequence, an epitope tag, a FLAG tag, a polyhistidine tag, a GST fusion, or the like.

[203] Methods of producing a protein in a cell with a selected amino acid at a specified position are also a feature of the present invention. For example, a method includes growing, in an appropriate medium, the cell, where the cell comprises a nucleic acid that comprises at least one selector codon and encodes a protein; and, providing the selected amino acid; where the cell further comprises: an orthogonal tRNA (O-tRNA) that functions in the cell and recognizes the selector codon; and, an orthogonal aminoacyl-tRNA synthetase (O-RS) that preferentially aminoacylates the O-tRNA with the selected amino acid. Typically, the O-tRNA comprises suppression activity in presence of a cognate synthetase in response to a selector codon. A protein produced by this method is also a feature of the present invention.

[204] The compositions of the present invention and compositions made by the methods of the present invention optionally are in a cell. The O-tRNA/O-RS pairs or individual components of the present invention can then be used in a host system's translation machinery, which results in a selected amino acid, e.g., unnatural amino acid, being incorporated into a protein. Patent applications USSN 10/825,867, entitled "Expanding the Eukaryotic Genetic Code;" and 10/126,927, entitled "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS", describe this process and are incorporated herein by reference. For example, when an O-tRNA/O-RS pair is introduced into a host, e.g., *Escherichia coli*, the pair leads to the in vivo incorporation of selected amino acid, such as an unnatural amino acid, e.g., a synthetic amino acid, such as derivative of a leucine amino acid, which can be exogenously added to the growth medium, into a protein, in response to a selector codon. Optionally, the compositions of the present invention can be in an in vitro translation system, or in an in vivo system(s).

[205] Any protein (or portion thereof) that includes a selected amino acid, e.g., an unnatural amino acid, (and any corresponding coding nucleic acid, e.g., which includes one or more selector codons) can be produced using the compositions and methods herein. Any polypeptide is suitable for incorporation of one or more selected amino acids. No attempt is made to identify the hundreds of thousands of known proteins, any of which can be modified to include one or more unnatural amino acid, e.g., by tailoring any available mutation methods to include one or more appropriate selector codon in a relevant translation system.

Common sequence repositories for known proteins include GenBank EMBL, DDBJ and the NCBI. Other repositories can easily be identified by searching the internet.

[206] Typically, the proteins are, e.g., at least 60%, at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, or at least 99% or more identical to any available protein (e.g., a therapeutic protein, a diagnostic protein, an industrial enzyme, or portion thereof, and the like), and they comprise one or more selected amino acid. Examples of therapeutic, diagnostic, and other proteins that can be modified to comprise one or more selected amino acid, e.g., an unnatural amino acid, can be found, but not limited to, those in USSN 10/825,867 entitled "Expanding the Eukaryotic Genetic Code;" and, U. S. Patent Application Serial Number 10/126,927, entitled "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS."

[207] In certain embodiments, the protein or polypeptide of interest (or portion thereof) in the methods and/or compositions of the present invention is encoded by a nucleic acid. Typically, the nucleic acid comprises at least one selector codon, at least two selector codons, at least three selector codons, at least four selector codons, at least five selector codons, at least six selector codons, at least seven selector codons, at least eight selector codons, at least nine selector codons, ten or more selector codons.

[208] Genes coding for proteins or polypeptides of interest can be mutagenized using methods well-known to one of skill in the art and described herein under "Mutagenesis and Other Molecular Biology Techniques" to include, e.g., one or more selector codon for the incorporation of a selected amino acid, e.g., an unnatural amino acid. For example, a nucleic acid for a protein of interest is mutagenized to include one or more selector codon, providing for the insertion of the one or more selected amino acids, e.g., unnatural amino acid. The invention includes any such variant, e.g., mutant, versions of any protein, e.g., including at least one selected amino acid. Similarly, the invention also includes corresponding nucleic acids, i.e., any nucleic acid with one or more selector codon that encodes one or more selected amino acid.

[209] To make a protein that includes a selected amino acid, one can use host cells and organisms that are adapted for the in vivo incorporation of the selected amino acid via orthogonal tRNA/RS pairs. Host cells are genetically engineered (e.g., transformed, transduced or transfected) with one or more vectors that express the orthogonal tRNA, the orthogonal tRNA synthetase, and a vector that encodes the protein to be derivatized. Each of

these components can be on the same vector, or each can be on a separate vector, two components can be on one vector and the third component on a second vector. The vector can be, for example, in the form of a plasmid, a cosmid, a phage, a bacterium, a virus, a naked polynucleotide, or a conjugated polynucleotide.

#### ALTERNATE SYSTEMS

[210] Several strategies have been employed to introduce unnatural amino acids into proteins in non-recombinant host cells, mutagenized host cells, or in cell-free systems. Derivatization of amino acids with reactive side-chains such as Lys, Cys and Tyr resulted in the conversion of lysine to N<sup>2</sup>-acetyl-lysine. Chemical synthesis also provides a straightforward method to incorporate unnatural amino acids. With the recent development of enzymatic ligation and native chemical ligation of peptide fragments, it is possible to make larger proteins. *See, e.g.,* P. E. Dawson and S. B. H. Kent, Ann. Rev. Biochem. 69:923 (2000). Chemical peptide ligation and native chemical ligation are described in U.S. Patent No. 6,184,344, U.S. Patent Publication No. 2004/0138412, U.S. Patent Publication No. 2003/0208046, WO 02/098902, and WO 03/042235, which are incorporated by reference herein. A general in vitro biosynthetic method in which a suppressor tRNA chemically acylated with the desired unnatural amino acid is added to an in vitro extract capable of supporting protein biosynthesis, has been used to site-specifically incorporate over 100 unnatural amino acids into a variety of proteins of virtually any size. *See, e.g.,* V. W. Cornish, D. Mendel and P. G. Schultz, Angew. Chem. Int. Ed. Engl., 1995, 34:621 (1995); C.J. Noren, S.J. Anthony-Cahill, M.C. Griffith, P.G. Schultz, *A general method for site-specific incorporation of unnatural amino acids into proteins*, Science 244:182-188 (1989); and, J.D. Bain, C.G. Glabe, T.A. Dix, A.R. Chamberlin, E.S. Diala, *Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide*, J. Am. Chem. Soc. 111:8013-8014 (1989). A broad range of functional groups has been introduced into proteins for studies of protein stability, protein folding, enzyme mechanism, and signal transduction.

[211] An in vivo method, termed selective pressure incorporation, was developed to exploit the promiscuity of wild-type synthetases. *See, e.g.,* N. Budisa, C. Minks, S. Alefelder, W. Wenger, F. M. Dong, L. Moroder and R. Huber, FASEB J., 13:41 (1999). An auxotrophic strain, in which the relevant metabolic pathway supplying the cell with a particular natural amino acid is switched off, is grown in minimal media containing limited concentrations of the natural amino acid, while transcription of the target gene is repressed. At the onset of a stationary growth phase, the natural amino acid is depleted and replaced

with the unnatural amino acid analog. Induction of expression of the recombinant protein results in the accumulation of a protein containing the unnatural analog. For example, using this strategy, o, m and p-fluorophenylalanines have been incorporated into proteins, and exhibit two characteristic shoulders in the UV spectrum which can be easily identified, *see, e.g.*, C. Minks, R. Huber, L. Moroder and N. Budisa, Anal. Biochem., 284:29 (2000); trifluoromethionine has been used to replace methionine in bacteriophage T4 lysozyme to study its interaction with chitooligosaccharide ligands by  $^{19}\text{F}$  NMR, *see, e.g.*, H. Duewel, E. Daub, V. Robinson and J. F. Honek, Biochemistry, 36:3404 (1997); and trifluoroleucine has been incorporated in place of leucine, resulting in increased thermal and chemical stability of a leucine-zipper protein. *See, e.g.*, Y. Tang, G. Ghirlanda, W. A. Petka, T. Nakajima, W. F. DeGrado and D. A. Tirrell, Angew. Chem. Int. Ed. Engl., 40:1494 (2001). Moreover, selenomethionine and telluromethionine are incorporated into various recombinant proteins to facilitate the solution of phases in X-ray crystallography. *See, e.g.*, W. A. Hendrickson, J. R. Horton and D. M. Lemaster, EMBO J., 9:1665 (1990); J. O. Boles, K. Lewinski, M. Kunkle, J. D. Odom, B. Dunlap, L. Lebioda and M. Hatada, Nat. Struct. Biol., 1:283 (1994); N. Budisa, B. Steipe, P. Demange, C. Eckerskorn, J. Kellermann and R. Huber, Eur. J. Biochem., 230:788 (1995); and, N. Budisa, W. Karnbrock, S. Steinbacher, A. Humm, L. Prade, T. Neuefeind, L. Moroder and R. Huber, J. Mol. Biol., 270:616 (1997). Methionine analogs with alkene or alkyne functionalities have also been incorporated efficiently, allowing for additional modification of proteins by chemical means. *See, e.g.*, J. C. van Hest and D. A. Tirrell, FEBS Lett., 428:68 (1998); J. C. van Hest, K. L. Kiick and D. A. Tirrell, J. Am. Chem. Soc., 122:1282 (2000); and, K. L. Kiick and D. A. Tirrell, Tetrahedron, 56:9487 (2000); U.S. Patent No. 6,586,207; U.S. Patent Publication 2002/0042097, which are incorporated by reference herein.

[212] The success of this method depends on the recognition of the unnatural amino acid analogs by aminoacyl-tRNA synthetases, which, in general, require high selectivity to insure the fidelity of protein translation. One way to expand the scope of this method is to relax the substrate specificity of aminoacyl-tRNA synthetases, which has been achieved in a limited number of cases. For example, replacement of Ala<sup>294</sup> by Gly in *Escherichia coli* phenylalanyl-tRNA synthetase (PheRS) increases the size of substrate binding pocket, and results in the acylation of tRNA<sup>Phe</sup> by p-Cl-phenylalanine (p-Cl-Phe). *See*, M. Ibba, P. Kast and H. Hennecke, Biochemistry, 33:7107 (1994). An *Escherichia coli* strain harboring this mutant PheRS allows the incorporation of p-Cl-phenylalanine or p-Br-phenylalanine in place of phenylalanine. *See, e.g.*, M. Ibba and H. Hennecke, FEBS Lett., 364:272 (1995); and, N.

Sharma, R. Furter, P. Kast and D. A. Tirrell, FEBS Lett., 467:37 (2000). Similarly, a point mutation Phe130Ser near the amino acid binding site of *Escherichia coli* tyrosyl-tRNA synthetase was shown to allow azatyrosine to be incorporated more efficiently than tyrosine. See, F. Hamano-Takaku, T. Iwama, S. Saito-Yano, K. Takaku, Y. Monden, M. Kitabatake, D. Soll and S. Nishimura, J. Biol. Chem., 275:40324 (2000).

[213] Another strategy to incorporate unnatural amino acids into proteins in vivo is to modify synthetases that have proofreading mechanisms. These synthetases cannot discriminate and therefore activate amino acids that are structurally similar to the cognate natural amino acids. This error is corrected at a separate site, which deacylates the mischarged amino acid from the tRNA to maintain the fidelity of protein translation. If the proofreading activity of the synthetase is disabled, structural analogs that are misactivated may escape the editing function and be incorporated. This approach has been demonstrated recently with the valyl-tRNA synthetase (ValRS). See, V. Doring, H. D. Mootz, L. A. Nangle, T. L. Hendrickson, V. de Crecy-Lagard, P. Schimmel and P. Marliere, Science, 292:501 (2001). ValRS can misaminoacylate tRNA<sup>Val</sup> with Cys, Thr, or aminobutyrate (Abu); these noncognate amino acids are subsequently hydrolyzed by the editing domain. After random mutagenesis of the *Escherichia coli* chromosome, a mutant *Escherichia coli* strain was selected that has a mutation in the editing site of ValRS. This edit-defective ValRS incorrectly charges tRNA<sup>Val</sup> with Cys. Because Abu sterically resembles Cys (–SH group of Cys is replaced with –CH<sub>3</sub> in Abu), the mutant ValRS also incorporates Abu into proteins when this mutant *Escherichia coli* strain is grown in the presence of Abu. Mass spectrometric analysis shows that about 24% of valines are replaced by Abu at each valine position in the native protein.

[214] Solid-phase synthesis and semisynthetic methods have also allowed for the synthesis of a number of proteins containing novel amino acids. For example, see the following publications and references cited within, which are as follows: Crick, F.H.C., Barrett, L. Brenner, S. Watts-Tobin, R. *General nature of the genetic code for proteins*. Nature, 192:1227-1232 (1961); Hofmann, K., Bohn, H. *Studies on polypeptides*. XXXVI. *The effect of pyrazole-imidazole replacements on the S-protein activating potency of an S-peptide fragment*, J. Am Chem, 88(24):5914-5919 (1966); Kaiser, E.T. *Synthetic approaches to biologically active peptides and proteins including enzymes*, Acc Chem Res, 22:47-54 (1989); Nakatsuka, T., Sasaki, T., Kaiser, E.T. *Peptide segment coupling catalyzed by the semisynthetic enzyme thiosubtilisin*, J Am Chem Soc, 109:3808-3810 (1987); Schmolzer, M., Kent, S B H. *Constructing proteins by dovetailing unprotected synthetic peptides: backbone-*

*engineered HIV protease*, Science, 256(5054):221-225 (1992); Chaiken, I.M. *Semisynthetic peptides and proteins*, CRC Crit Rev Biochem, 11(3):255-301 (1981); Offord, R.E. *Protein engineering by chemical means?* Protein Eng., 1(3):151-157 (1987); and, Jackson, D.Y., Burnier, J., Quan, C., Stanley, M., Tom, J., Wells, J.A. *A Designed Peptide Ligase for Total Synthesis of Ribonuclease A with Unnatural Catalytic Residues*, Science, 266(5183):243 (1994).

[215] Chemical modification has been used to introduce a variety of unnatural side chains, including cofactors, spin labels and oligonucleotides into proteins in vitro. *See, e.g.*, Corey, D.R., Schultz, P.G. *Generation of a hybrid sequence-specific single-stranded deoxyribonuclease*, Science, 238(4832):1401-1403 (1987); Kaiser, E.T., Lawrence D.S., Rokita, S.E. *The chemical modification of enzymatic specificity*, Annu Rev Biochem, 54:565-595 (1985); Kaiser, E.T., Lawrence, D.S. *Chemical mutation of enzyme active sites*, Science, 226(4674):505-511 (1984); Neet, K.E., Nanci A, Koshland, D.E. *Properties of thiol-subtilisin*, J Biol. Chem, 243(24):6392-6401 (1968); Polgar, L. et M.L. Bender. *A new enzyme containing a synthetically formed active site. Thiol-subtilisin*. J. Am Chem Soc, 88:3153-3154 (1966); and, Pollack, S.J., Nakayama, G. Schultz, P.G. *Introduction of nucleophiles and spectroscopic probes into antibody combining sites*, Science, 242(4881):1038-1040 (1988).

#### Defining Polypeptides by Immunoreactivity

[216] Because the polypeptides of the present invention provide a variety of new polypeptide sequences (e.g., comprising selected amino acids (e.g., unnatural amino acids) in the case of proteins synthesized in the translation systems herein, or, e.g., in the case of the novel synthetases, novel sequences of standard amino acids), the polypeptides also provide new structural features which can be recognized, e.g., in immunological assays. The generation of antisera, which specifically bind the polypeptides of the present invention, as well as the polypeptides which are bound by such antisera, are a feature of the present invention. The term "antibody," as used herein, includes, but is not limited to a polypeptide substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof which specifically bind and recognize an analyte (antigen). Examples include polyclonal, monoclonal, chimeric, and single chain antibodies, and the like. Fragments of immunoglobulins, including Fab fragments and fragments produced by an expression library, including phage display, are also included in the term "antibody" as used herein. *See, e.g.*,

Paul, *Fundamental Immunology*, 4th Ed., 1999, Raven Press, New York, for antibody structure and terminology.

[217] For example, the invention includes proteins made utilizing the tRNAs and/or RSs of the present invention that specifically bind to or that are specifically immunoreactive with an antibody or antisera generated against an immunogen comprising an amino acid sequence. To eliminate cross-reactivity with other homologues, the antibody or antisera is subtracted with available protein, such as the wild-type polypeptide, e.g., the "control" polypeptides. Where the wild-type protein corresponds to a nucleic acid, a polypeptide encoded by the nucleic acid is generated and used for antibody/antisera subtraction purposes.

[218] In one typical format, the immunoassay uses a polyclonal antiserum which was raised against one or more polypeptide or a substantial subsequence thereof (i.e., at least about 30% of the full length sequence provided). The set of potential polypeptide immunogens derived from the protein are collectively referred to below as "the immunogenic polypeptides." The resulting antisera is optionally selected to have low cross-reactivity against the control synthetase homologues and any such cross-reactivity is removed, e.g., by immunoabsorbtion, with one or more of the control homologues, prior to use of the polyclonal antiserum in the immunoassay.

[219] In order to produce antisera for use in an immunoassay, one or more of the immunogenic polypeptides is produced and purified as described herein. For example, recombinant protein can be produced in a recombinant cell. An inbred strain of mice (used in this assay because results are more reproducible due to the virtual genetic identity of the mice) is immunized with the immunogenic protein(s) in combination with a standard adjuvant, such as Freund's adjuvant, and a standard mouse immunization protocol (*see, e.g.,* Harlow and Lane (1988) Antibodies, A Laboratory Manual, Cold Spring Harbor Publications, New York, for a standard description of antibody generation, immunoassay formats and conditions that can be used to determine specific immunoreactivity).

[220] Additional references and discussion of antibodies is also found herein and can be applied here to defining polypeptides by immunoreactivity). Alternatively, one or more synthetic or recombinant polypeptide derived from the sequences disclosed herein is conjugated to a carrier protein and used as an immunogen.

[221] Polyclonal sera are collected and titered against the immunogenic polypeptide in an immunoassay, for example, a solid phase immunoassay with one or more of the

immunogenic proteins immobilized on a solid support. Polyclonal antisera with a titer of  $10^6$  or greater are selected, pooled and subtracted with the control synthetase polypeptides to produce subtracted pooled titrated polyclonal antisera.

[222] The subtracted pooled titrated polyclonal antisera are tested for cross reactivity against the control homologues in a comparative immunoassay. In this comparative assay, discriminatory binding conditions are determined for the subtracted titrated polyclonal antisera which result in at least about a 5-10 fold higher signal to noise ratio for binding of the titrated polyclonal antisera to the immunogenic protein as compared to binding to the control synthetase homologues. That is, the stringency of the binding reaction is adjusted by the addition of non-specific competitors such as albumin or non-fat dry milk, and/or by adjusting salt conditions, temperature, and/or the like. These binding conditions are used in subsequent assays for determining whether a test polypeptide (a polypeptide being compared to the immunogenic polypeptides and/or the control polypeptides) is specifically bound by the pooled subtracted polyclonal antisera.

[223] In another example, immunoassays in the competitive binding format are used for detection of a test polypeptide. For example, as noted, cross-reacting antibodies are removed from the pooled antisera mixture by immunoabsorbtion with the control polypeptides. The immunogenic polypeptide(s) are then immobilized to a solid support which is exposed to the subtracted pooled antisera. Test proteins are added to the assay to compete for binding to the pooled subtracted antisera. The ability of the test protein(s) to compete for binding to the pooled subtracted antisera as compared to the immobilized protein(s) is compared to the ability of the immunogenic polypeptide(s) added to the assay to compete for binding (the immunogenic polypeptides compete effectively with the immobilized immunogenic polypeptides for binding to the pooled antisera). The percent cross-reactivity for the test proteins is calculated, using standard calculations.

[224] In a parallel assay, the ability of the control proteins to compete for binding to the pooled subtracted antisera is optionally determined as compared to the ability of the immunogenic polypeptide(s) to compete for binding to the antisera. Again, the percent cross-reactivity for the control polypeptides is calculated, using standard calculations. Where the percent cross-reactivity is at least 5-10x as high for the test polypeptides as compared to the control polypeptides and or where the binding of the test polypeptides is approximately in the range of the binding of the immunogenic polypeptides, the test polypeptides are said to specifically bind the pooled subtracted antisera.

[225] In general, the immunoabsorbed and pooled antisera can be used in a competitive binding immunoassay as described herein to compare any test polypeptide to the immunogenic and/or control polypeptide(s). In order to make this comparison, the immunogenic, test and control polypeptides are each assayed at a wide range of concentrations and the amount of each polypeptide required to inhibit 50% of the binding of the subtracted antisera to, e.g., an immobilized control, test or immunogenic protein is determined using standard techniques. If the amount of the test polypeptide required for binding in the competitive assay is less than twice the amount of the immunogenic polypeptide that is required, then the test polypeptide is said to specifically bind to an antibody generated to the immunogenic protein, provided the amount is at least about 5-10x as high as for the control polypeptide.

[226] As an additional determination of specificity, the pooled antisera is optionally fully immunosorbed with the immunogenic polypeptide(s) (rather than the control polypeptides) until little or no binding of the resulting immunogenic polypeptide subtracted pooled antisera to the immunogenic polypeptide(s) used in the immunosorbition is detectable. This fully immunosorbed antisera is then tested for reactivity with the test polypeptide. If little or no reactivity is observed (i.e., no more than 2x the signal to noise ratio observed for binding of the fully immunosorbed antisera to the immunogenic polypeptide), then the test polypeptide is specifically bound by the antisera elicited by the immunogenic protein.

[227] Additional details on proteins, antibodies, antisera, etc. can be found in USSN 10/825,867 entitled "Expanding the Eukaryotic Genetic Code;" WO 2002/085923, entitled "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS;" U.S. Patent No. 6,927,042 entitled "Glycoprotein synthesis"; and U.S. Patent Publication No. US 2004/0198637 entitled "Protein Arrays," which is incorporated by reference.

#### KITS

[228] Kits are also a feature of the present invention. For example, a kit for producing a protein that comprises at least one selected amino acid, e.g., an unnatural amino acid, in a cell is provided, where the kit includes a container containing a polynucleotide sequence encoding an O-tRNA, and/or an O-tRNA, and/or a polynucleotide sequence encoding an O-RS, and/or an O-RS. In one embodiment, the kit further includes at least selected amino acid. In another embodiment, the kit includes an aminoacylated tRNA of the invention. In another embodiment, the kit further comprises instructional materials for producing the protein.

[229] An additional example is a kit for producing a protein that comprises at least one selected amino acid, e.g., an unnatural amino acid, in a cell-free translation system, where the kit includes a container containing a polynucleotide sequence encoding an O-tRNA, and/or an O-tRNA, and/or a polynucleotide sequence encoding an O-RS, and/or an O-RS. In one embodiment, the kit further includes a selected amino acid. In another embodiment, the kit includes an aminoacylated tRNA of the invention. In another embodiment, the kit further comprises instructional materials for producing the protein.

## EXAMPLES

[230] The following examples are offered to illustrate, but not to limit the claimed invention. One of skill will recognize a variety of non-critical parameters that may be altered without departing from the scope of the claimed invention.

### Example 1:

#### Aminoacyl-tRNA Synthetase Selection against Para-Acetyl Phenylalanine

[231] Two DNA libraries were screened for aminoacyl-tRNA synthetases against para-acetyl phenylalanine, a non-naturally encoded amino acid. These libraries consisted of six mutations in the tyrosyl tRNA synthetase gene from *Methanococcus janneschii* in the pBK plasmid.

[232] The selection procedure was performed which consisted of five alternating rounds of selection, three positive, two negative. The libraries were combined in a 1:1 ratio and electroporated into the positive selection cell line (GeneHog with positive selection plasmid, pREP) and plated on minimal media plates (GMML) with appropriate antibiotics and the non-naturally encoded amino acid para-acetyl phenylalanine (pAF). The plates were incubated at 37°C for about 40 hours at which point the cells were harvested by scraping. The DNA was extracted using a Qiagen Mini-Prep procedure, and then was agarose gel purified to isolate the library plasmid DNA.

[233] This DNA was then electroporated into the negative selection cell line (GeneHog with negative selection plasmid pBAD derivative). These transformants were plated on LB plates with appropriate antibiotic without the non-naturally encoded amino acid (pAF). After about 17 hours these cells were harvested by scraping and the plasmid DNA was purified using the Qiagen Mini-Prep procedure and agarose gel purification.

[234] The subsequent rounds of selection were done utilizing the same method of electroporation, plating, harvesting, and DNA purification. In the last (fifth) round of

selection, serial dilutions were made of the transformed positive selection cells which were plated on minimal media plates. Individual colonies were then picked and grown in a 96 well block overnight. This block was then replica plated on minimal media plates with varying concentrations of chloramphenicol (the positive selection antibiotic) with and without unnatural amino acid pAF. After about 40 hours of growth at 37°C, the plates were visually compared to determine which colonies grew on the highest chloroamphenicol concentration but did not grow or grew poorly in the absence of the non-naturally encoded amino acid pAF. The colonies which met these criteria were grown overnight. The DNA was isolated from the cultures by Mini-Prep and agarose gel purification and were sequenced.

[235] From this selection for pAF, 13 clones were found to have unique amino acid sequences and were subjected to further characterization to determine the fidelity and processivity of the pAF-tRNA synthetase.

[236] To characterize these synthetases, small scale amber suppressions were performed to show that the non-naturally encoded amino acid pAF was incorporated into a polypeptide, and the results were visualized by SDS-PAGE. A single colony was picked and grown overnight in LB broth, which was then used to inoculate 50 mL of LB. The cells were grown to an OD of 0.3-0.4, at which point 1.5 mL aliquots were taken as pre-induction points and the culture was split into two flasks. 1 mM pAF was added to one split and both were grown for 30 minutes. Following the 30 minute growth, both cultures (+/- pAF) were induced with 0.2% L-Arabinose and grown 4.5 hours and the OD<sub>600</sub> was recorded. 1.5 mL aliquots were then taken of the +/- pAF flasks for SDS-PAGE analysis.

[237] The 1.5 mL aliquots (Preinduction, + pAF, -pAF) were centrifuged at 10,000 xg for 10 minutes to pellet the cells. The cells were then suspended in proportional Bacterial Protein Extraction Reagent (BPER, Pierce) amounts relative to their OD<sub>600</sub> at the time of harvest. DNase I was added to the lysed cells and incubated at 4°C for 20 minutes. The samples were then combined with a reducing agent and loading dye and run on a 4-12% Bis-TRIS gel in MES buffer for 30 minutes. The gel was washed in DI H<sub>2</sub>O twice for 10 minutes and stained with coomassie blue dye. The +/- pAF bands were compared for the fidelity of the pAF-tRNA RS to result in incorporation of pAF, and the + pAF band was compared to the previously selected pAF-tRNA RS.

[238] To check for the processivity of the RSs the same procedure was performed with a plasmid containing C-H6 S4am myoglobin (S4am-Myo). The S4am Myo was then purified by IMAC and sent for protein sequencing to determine the amount of pAF incorporation.

[239] Of the pAF-tRNA RSs identified from this selection, one synthetase (E9) was found to incorporate pAF efficiently, with greater than 95% efficiency of incorporation of pAF into S4am-Myo. The incorporation was determined by amino acid sequencing while the processivity was shown by comparing protein bands on a SDS-PAGE gels. The nucleotide sequence for E9 is shown in SEQ ID NO: 4, and the amino acid sequence of E9 is shown in SEQ ID NO : 5.

[240] An additional mutant with similar activity to E9 was identified, and has the amino acid sequence shown in SEQ ID NO: 17.

Example 2:

#### **tRNA Mutagenesis**

[241] Three mutants were generated of tRNA J17. The DNA sequence of wild-type J17 is shown as SEQ ID NO: 8 and in U.S. Patent Publication Nos. 2003/0108885 as SEQ ID NO: 1 and US 2003/0082575 as SEQ ID NO: 1 (U.S. Patent Application Serial Nos. 10/126,931 and 10/126,927, respectively), both of which are incorporated by reference herein. J17 tRNA has a U51:G63 wobble pair in the T $\Psi$ C stem as shown in Figure 1.

[242] Three J17 mutants (F12, F13, and F14) were generated to produce Watson-Crick base pairs at positions 51 and 63 of the T $\Psi$ C stem. Mutagenesis was performed by overlapping PCR, and the final constructs were cloned into EcoRI and NdeI sites in a pET19 plasmid comprising the polynucleotide sequence encoding the aminoacyl tRNA synthetase E9 (SEQ ID NO: 4) and the polynucleotide sequence encoding human growth hormone (hGH) with an amber codon substitution (SEQ ID NO: 16). The expression of hGH was under the control of the T7 promoter.

[243] Two fragments were generated for overlapping PCR. The first fragment was obtained by primer extension. The sequence of the forward primer used to generate each of the three mutants was:

GTAACGCTGAATTCCCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAAT  
CCGCATGGCGC (FTam11; SEQ ID NO: 9).

[244] To generate the F12 mutant (51C:63G), the following reverse primer was used:  
GATCTGCAGTGGTCCGGCGGGCCGGATTTGAACCGGCGCCATGCGGATTAGAG  
TCCGCCGTTCTGC (FTam12; SEQ ID NO: 10).

[245] To generate the F13 mutant (51U:63A), the following reverse primer was used:  
GATCTGCAGTGGTCCGGCGGGCTGGATTTGAACCGGCGCCATGCGGATTAGAGT  
CCGCCGTTCTGC (FTam13; SEQ ID NO: 11).

[246] To generate the F14 mutant (51A:63U), the following reverse primer was used:

GATCTGCAGTGGTCCGGCGGGCAGGATTTGAACCTGCGCCATGCGGATTTAGAGT  
CCGCCGTTCTGC (FTam14; SEQ ID NO: 12).

[247] To generate the second fragment, plasmid pET19 J17 E9 hGH comprising the polynucleotide sequence for J17 tRNA (SEQ ID NO: 8), the polynucleotide sequence encoding the tRNA synthetase E9 (SEQ ID NO: 4) and the polynucleotide sequence encoding human growth hormone with an amber codon substitution (SEQ ID NO: 16) was used as a template for amplification with the following set of primers:

CGCCGGACCACTGCAGATCCTTAGCGAAAGCTAAGGATTTTTTTTAAGC (forward primer; FTam15; SEQ ID NO: 13) and CAAATTCGTCCATATGGGATTCC (FTam 16; SEQ ID NO: 14). The forward primer was used to extend the sequence from the 3' end of tRNA to the Nde I site of the plasmid. The resulting product was gel purified.

[248] The final step of overlapping PCR involved forward primer GTAACGCTGAATTCCCGGCG (FTam17, SEQ ID NO: 15), reverse primer FTam16 (SEQ ID NO: 14), the first fragment and the second fragment. The assembled products were digested with EcoR I and Nde I and ligated into the plasmid pET19 J17 E9 hGH digested with EcoR I and Nde I. The sequence of each construct was confirmed by sequencing, and the DNA sequences for each of the J17 mutant tRNAs are shown as SEQ ID NO: 1 (F12), SEQ ID NO: 2 (F13), and SEQ ID NO: 3 (F14). The tRNAs were named after their corresponding reverse primers.

### **Protein expression**

[249] Plasmids encoding the tRNAs (J17, F12, F13 or F14) were each transformed into E. coli strain 1 and strain 2 bacterial host cells by chemical means and plated onto LB agar plates with 50 ug/ml carbenicillin. The plates were incubated at 37°C overnight. For each tRNA, a single colony was picked to start an overnight culture at 37°C in 1 ml 2xYT with 50 ug/ml carbenicillin. This 1 ml culture was used to inoculate two 10 ml 2xYT cultures with 50 ug/ml carbenicillin at 37°C. One 10 ml culture was supplemented with 4 mM *para*-acetylphenylalanine. At OD<sub>600</sub>=0.7, the hGH expression was induced with 0.4 mM IPTG. After culturing the cells at 37°C for 4 hours with 250 rpm, the cells were harvested by centrifugation at 5000 x g for 5 minutes. The cells were lysed with B-PER Reagent (Pierce, Rockford, IL) supplemented with 5 ug/ml DNase I. The total cell lysate was analyzed by 4-12% SDS PAGE.

[250] Figure 2 shows an analysis of E. coli strain 1 total cell lysates by SDS PAGE. Suppression of a selector codon in human growth hormone was performed using J17 or J17

mutant (F12, F13, F14) tRNA and the aminoacyl tRNA synthetase E9. Cells harboring J17 mutants grew slightly slower than cells harboring J17. No full length hGH product was observed by SDS-PAGE for the tRNA mutants in the absence of 4 mM *para*-acetylphenylalanine. In the presence of 4 mM *para*-acetylphenylalanine, full length product was produced with each of the tRNA mutants, demonstrating that these tRNA mutant-RS E9 pairs are orthogonal to *E. coli* machinery. Based on SDS-PAGE, the suppressed hGH yield of the J17 mutants was approximately 1.5~2 fold higher than that of J17 in *E. coli* strain 1.

[251] One J17 mutant, F13, was further tested in *E. coli* strain 2 bacterial cell line for amber suppression as shown in Figure 3. In *E. coli* strain 2, the expression as well as amber suppression yields were reduced relative to that in *E. coli* strain 1. In the absence of *para*-acetylphenylalanine, no full length hGH product was observed by SDS-PAGE. In the presence 4 mM *para*-acetylphenylalanine, full length hGH was observed for both tRNAs. Based on SDS-PAGE, the suppressed hGH yield of F13 was about three fold higher than that of J17.

[252] A fermentation run comparing J17 and F13 was performed with a final volume of approximately 1.5 L. The plasmid encoding the J17 tRNA and the plasmid encoding F13 tRNA were each transformed into *E. coli* strain 1. The final cell density for each was approximately 190 g wet cells/l. The hGH titer was 347 mg/L for the J17 clone and 542 mg/L for the F13 clone.

[253] While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the present invention. For example, all the techniques and apparatus described above can be used in various combinations. All publications, patents, patent applications, and/or other documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent, patent application, and/or other document were individually indicated to be incorporated by reference for all purposes.

TABLE 1

| SEQ ID NO: | Label                                                            | SEQUENCE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|------------|------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1          | F12 DNA                                                          | CCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCCGCA<br>TGGCGCCGGTTCAAATCCGGCCCCGGGACCA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| 2          | F13 DNA                                                          | CCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCCGCA<br>TGGCGCTGGTTCAAATCCAGCCCCGGGACCA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| 3          | F14 DNA                                                          | CCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCCGCA<br>TGGCGCAGGTTCAAATCCTGCCCCGGGACCA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| 4          | E9 RS nucleic acid                                               | ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTATC<br>AGCGAGGAAGAGTTAAGAGAGGTTTTAAAAAAGATGAAAAATC<br>TGCTGTTATAGGTTTTGAACCAAGTGGTAAAATACATTTAGGGCAT<br>TATCTCCAAATAAAAAAGATGATTGATTTACAAAATGCTGGATTTG<br>ATATAATTATATATTTGGCTGATTTACACGCCTATTTAAACCAGAA<br>AGGAGAGTTGGATGAGATTAGAAAAATAGGAGATTATAACAAAA<br>AAGTTTTGAAGCAATGGGGTTAAAGGCAAAATATGTTTATGGAA<br>GTGAACATGGTCTTGATAAGGATTATACACTGAATGTCTATAGATT<br>GGCTTTAAAAACTACCTTAAAAAGAGCAAGAAGGAGTATGGAAC<br>TATAGCAAGAGAGGATGAAAATCCAAAGGTTGCTGAAGTTATCTA<br>TCCAATAATGCAGGTTAATGGGATTCATTATGAGGGCGTTGATGTT<br>GCAGTTGGAGGGATGGAGCAGAGAAAAATACACATGTTAGCAAG<br>GGAGCTTTTACAAAAAAGGTTGTTTGTATTACAAACCCTGTCTTA<br>ACGGGTTTGGATGGAGAAGGAAAGATGAGTTCTTCAAAGGGAAT<br>TTTATAGCTGTTGATGACTCTCCAGAAGAGATTAGGGCTAAGATAA<br>AGAAAGCATACTGCCCAGCTGGAGTTGTTGAAGGAAATCCAATAA<br>TGGAGATAGCTAAATACTTCCTTGAATATCCTTTAACCATAAAAAAG<br>GCCAGAAAAATTTGGTGGAGATTGACAGTTAATAGCTATGAGGA<br>GTTAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTTA<br>AAAAATGCTGTAGCTGAAGAACTTATAAAGATTTTAGAGCCAATT<br>AGAAAGAGATTATAA |
| 5          | E9 RS Amino Acid                                                 | MDEFEMIKRNTSEIIEEELREVLKKDEKSAVIGFEPGKIHLGHYLQIK<br>KMIDLQNAGFDIITYLADLHAYLNQKGELDEIRKIGDYNKKVFEAMGL<br>KAKYVYGSEHGLDKDYTLNVYRLALKTTLKRRRSMELIAREDENP<br>KVAEVIYPIMQVNGIHYEGVDVAVGGMEQRKIHMLARELLPKKVCCI<br>HNPVLTGLDGEGKMSSSKGNFIAVDDSPPEIRAKIKAYCPAGVVEG<br>NPIMEIAKYFLEYPLTIKRPEKFGGDLTVNSYEELESFLKNKELHPMDL<br>KNAVAEELIKILEPIRKRL                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| 6          | HL(TAG)3 tRNA DNA                                                | CCCAGGGTAGCCAAGCTCGGCCAACGGCGACGGACTCTAAATCCG<br>TTCTCGTAGGAGTTCGAGGGTTCGAATCCCTTCCCTGGGACCA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| 7          | HL(TGA)1 tRNA DNA                                                | GCGGGGGTTGCCGAGCCTGGCCAAAGGCGCCGGACTTCAAATCCG<br>GTCCCGTAGGGGTTCCGGGGTTCAAATCCCCGCCCCCGCACCA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| 8          | J17 <i>M. jannaschii</i> mtRNA <sup>Tyr</sup> <sub>CUA</sub> DNA | CCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCCGCA<br>TGGCGCTGGTTCAAATCCGGCCCCGGGACCA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| 9          | FTam11 primer                                                    | GTAACGCTGAATTCCCGGCGGTAGTTCAGCAGGGCAGAACGGCGG<br>ACTCTAAATCCGCATGGCGC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |

|    |                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|----|--------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 10 | FTam12 primer      | GATCTGCAGTGGTCCGGCGGGCCGGATTGTAACCGGCGCCATGCG<br>GATTTAGAGTCCGCCGTTCTGC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| 11 | FTam13 primer      | GATCTGCAGTGGTCCGGCGGGCTGGATTGTAACAGCGCCATGCG<br>GATTTAGAGTCCGCCGTTCTGC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| 12 | FTam14 primer      | GATCTGCAGTGGTCCGGCGGGCAGGATTGTAACCTGCGCCATGCG<br>GATTTAGAGTCCGCCGTTCTGC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| 13 | FTam15 primer      | CGCCGGACCACTGCAGATCCTTAGCGAAAGCTAAGGATTTTTTTTA<br>AGC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| 14 | FTam16 primer      | CAAATTCGTCCATATGGGATTCC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| 15 | FTam17 primer      | GTAACGCTGAATTCCCGGCG                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| 16 | hGH (DNA)          | ATGGGGCCACCACCACCACCACCTTCCCAACCATTCCTTATCCA<br>GGCTTTTTGACAACGCTATGCTCCGCGCCCATCGTCTGCACCAGCT<br>GGCCTTTGACACCTACCAGGAGTTTGAAGAAGCCTAGATCCCAA<br>GGAACAGAAGTATTCATTCCTGCAGAACCCCCAGACCTCCCTCTGT<br>TTCTCAGAGTCTATTCCGACACCCTCCAACAGGGAGGAAACACAA<br>CAGAAATCCAACCTAGAGCTGCTCCGCATCTCCCTGCTGCTCATCC<br>AGTCGTGGCTGGAGCCCGTGCAGTTCCTCAGGAGTGTCTTCGCCAA<br>CAGCCTGGTGTACGGCGCCTCTGACAGCAACGTCTATGACCTCCTA<br>AAGGACCTAGAGGAAGGCATCCAAACGCTGATGGGGAGGCTGGA<br>AGATGGCAGCCCCCGGACTGGGCAGATCTTCAAGCAGACCTACAG<br>CAAGTTCGACACAACTCACACAACGATGACGCACTACTCAAGAA<br>CTACGGGCTGCTCTACTGCTTCAGGAAGGACATGGACAAGGTCGA<br>GACATTCCTGCGCATCGTGCAGTGCCGCTCTGTGGAGGGCAGCTGT<br>GGCTTCTAA |
| 17 | D286R mutant of E9 | MDEFEMIKRN TSEIISEEEL REVLKKDEKS AVIGFEPSGK<br>IHLGHYLQIK KMIDLQNAGF DIITYLADLH AYLNQKGELD<br>EIRKIGDYNK KVFEAMGLKA KYVYGSEHGL DKDYTLNVYR<br>LALKTTLKRA RRSMEIARE DENPKVAEVI YPIMQVNGIH<br>YEGVDVAVGG MEQRKIHMLA RELPKKVVC IHNPVLTGLD<br>GEGKMSSSKG NFIAVDDSP EIRAKIKKAY CPAGVVEGNP<br>IMEIAKYFLE YPLTIKRPEK FGD LTVNSY EELES LFKNK<br>ELHPMRLKNA VAEELIKILE PIRKRL                                                                                                                                                                                                                                                                                                      |

What is claimed is:

1. A composition comprising a tRNA having the nucleic acid sequence selected from the group consisting of SEQ ID NOs: 1, 2, and 3, and the full length complementary polynucleotide sequence that encodes SEQ ID NO: 1, 2 and 3 in a buffer.
2. The composition of claim 1, wherein said tRNA is aminoacylated.
3. The composition of claim 2, wherein said tRNA is aminoacylated with a non-naturally encoded amino acid.
4. The composition of claim 3, wherein said non-naturally encoded amino acid is para-acetylphenylalanine.
5. The composition of claim 2, wherein said tRNA is chemically aminoacylated.
6. The composition of claim 2, wherein said tRNA is enzymatically aminoacylated.
7. The composition of claim 6, wherein said tRNA is enzymatically aminoacylated by an aminoacyl tRNA synthetase.
8. The composition of claim 6, wherein said tRNA is enzymatically aminoacylated by a ribozyme.
9. The composition of claim 1, further comprising an aminoacyl tRNA synthetase, wherein the tRNA is aminoacylated with an amino acid.
10. The composition of claim 9, wherein said amino acid is a non-naturally encoded amino acid.
11. The composition of claim 10, wherein said non-naturally encoded amino acid is para-acetylphenylalanine.
12. The composition of claim 1, wherein the tRNA is derived from an archaeal tRNA.
13. The composition of claim 1, wherein the tRNA is derived from *M. janneschii*.
14. The composition of claim 1, further comprising a translation system.
15. The composition of claim 14, wherein said translation system is a cell-free translation system.
16. The composition of claim 14, wherein said translation system is a cell lysate.
17. The composition of claim 14, wherein said translation system is a reconstituted system.
18. The composition of claim 14, wherein said translation system is a cellular translation system.
19. A cell comprising a translation system, wherein the translation system comprises the tRNA of claim 1.
20. The cell of claim 19, wherein the cell is a eukaryotic cell.

21. The cell of claim 20, wherein the eukaryotic cell is a yeast cell.

22. The cell of claim 20, wherein the eukaryotic cell is a fungal cell.

23. The cell of claim 20, wherein the eukaryotic cell is a mammalian cell.

24. The cell of claim 20, wherein the eukaryotic cell is an insect cell.

25. The cell of claim 20, wherein the eukaryotic cell is a plant cell.

26. The cell of claim 19, wherein the cell is a non-eukaryotic cell.

27. The cell of claim 26, wherein the non-eukaryotic cell is an *E. coli* cell.

28. The cell of claim 19, further comprising a polynucleotide encoding a polypeptide of interest, wherein the polynucleotide comprises a selector codon that is recognized by the tRNA.

29. The cell of claim 28, wherein said polypeptide of interest is human growth hormone.

30. An *E. coli* cell comprising the tRNA of claim 1.

31. A yeast cell comprising the tRNA of claim 1.

32. A fungal cell comprising a tRNA of claim 1.

33. A mammalian cell comprising a tRNA of claim 1.

34. An insect cell comprising a tRNA of claim 1.

35. A plant cell comprising a tRNA of claim 1.

36. A vector comprising a polynucleotide encoding a tRNA having the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1, 2, and 3, and the full-length complementary polynucleotide sequence that encodes SEQ ID NO: 1, 2 and 3.

37. The vector of claim 36, wherein the vector comprises a plasmid, a cosmid, a phage, or a virus.

38. The vector of claim 36, wherein the vector is an expression vector.

39. A cell comprising the vector of claim 36.

40. A method of producing a polypeptide in a cell with a selected amino acid at a specified position, the method comprising: growing, in an appropriate medium, the cell, where the cell comprises a nucleic acid that comprises at least one selector codon and encodes a polypeptide; and, providing the selected amino acid; wherein the cell further comprises: an orthogonal tRNA (O-tRNA) that functions in the cell and recognizes the selector codon having the amino acid sequence selected from the group consisting of SEQ ID NO: 1, 2, 3, and the full-length complementary polynucleotide sequence that encodes SEQ ID NO: 1, 2 and 3; and, an orthogonal aminoacyl-tRNA synthetase (O-RS), wherein said O-RS aminoacylates the O-tRNA with the selected amino acid.

41. The method of claim 40 wherein said selected amino acid is para-acetyl phenylalanine.

42. The method of claim 40 wherein said polypeptide is human growth hormone.

43. A polynucleotide having a nucleic acid sequence selected from a group consisting of SEQ ID Nos: 1, 2, and 3, and the full-length complementary polynucleotide sequence thereof.

## SEQUENCE LISTING

<110> Tian, Feng

<120> Compositions of tRNA and Uses Thereof

<130> AMBX-0069.00PCT

<150> US 60/709,364

<151> 2005-08-18

<160> 17

<170> PatentIn version 3.3

<210> 1

<211> 77

<212> DNA

<213> Artificial Sequence

<220>

<223> Artificial Sequence

<400> 1

ccggcggtag ttcagcaggg cagaacggcg gactctaaat ccgcatggcg ccggttcaaa 60

tccggcccgc cggacca 77

<210> 2

<211> 77

<212> DNA

<213> Artificial Sequence

<220>

<223> Artificial Sequence

<400> 2

ccggcggtag ttcagcaggg cagaacggcg gactctaaat ccgcatggcg ctggttcaaa 60

tccagcccgc cggacca 77

<210> 3

<211> 77

<212> DNA

<213> Artificial Sequence

<220>

<223> Artificial Sequence

<400> 3

ccggcggtag ttcagcaggg cagaacggcg gactctaaat ccgcatggcg caggttcaaa 60

tcctgcccgc cggacca 77

<210> 4  
 <211> 921  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Artificial Sequence

<400> 4  
 atggacgaat ttgaaatgat aaagagaaac acatctgaaa ttatcagcga ggaagagtta 60  
 agagagggttt taaaaaaaga tgaaaaatct gctgttatag gttttgaacc aagtggtaaa 120  
 atacatttag ggcattatct ccaaataaaa aagatgattg atttacaaa tgctggattt 180  
 gatataatta tatatttggc tgatttacac gcctatttaa accagaaagg agagttggat 240  
 gagattagaa aaataggaga ttataacaaa aaagtttttg aagcaatggg gttaaaggca 300  
 aaatatgttt atggaagtga acatgggtctt gataaggatt atacactgaa tgtctataga 360  
 ttggctttta aaactacctt aaaaagagca agaaggagta tggaacttat agcaagagag 420  
 gatgaaaatc caaagggttc tgaagttatc tatccaataa tgcagggtta tgggattcat 480  
 tatgagggcg ttgatgttgc agttggaggg atggagcaga gaaaaatata catgttagca 540  
 agggagcttt taccaaaaaa ggttggttgc attcacaacc ctgtcttaac ggggttggat 600  
 ggagaaggaa agatgagttc ttcaaaaggg aattttatag ctgttgatga ctctocagaa 660  
 gagattaggg ctaagataaa gaaagcatac tgcccagctg gagttgttga aggaaatcca 720  
 ataattggaga tagctaaata cttccttgaa tatcctttta ccataaaaag gccagaaaaa 780  
 tttggtggag atttgacagt taatagctat gaggagttag agagtttatt taaaaataag 840  
 gaattgcac caatggattt aaaaaatgct gtagctgaag aacttataaa gatttttagag 900  
 ccaattagaa agagattata a 921

<210> 5  
 <211> 306  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Artificial Sequence

<400> 5

Met Asp Glu Phe Glu Met Ile Lys Arg Asn Thr Ser Glu Ile Ile Ser  
 1 5 10 15

Glu Glu Glu Leu Arg Glu Val Leu Lys Lys Asp Glu Lys Ser Ala Val

20 25 30  
 Ile Gly Phe Glu Pro Ser Gly Lys Ile His Leu Gly His Tyr Leu Gln  
 35 40 45  
 Ile Lys Lys Met Ile Asp Leu Gln Asn Ala Gly Phe Asp Ile Ile Ile  
 50 55 60  
 Tyr Leu Ala Asp Leu His Ala Tyr Leu Asn Gln Lys Gly Glu Leu Asp  
 65 70 75 80  
 Glu Ile Arg Lys Ile Gly Asp Tyr Asn Lys Lys Val Phe Glu Ala Met  
 85 90 95  
 Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Glu His Gly Leu Asp Lys  
 100 105 110  
 Asp Tyr Thr Leu Asn Val Tyr Arg Leu Ala Leu Lys Thr Thr Leu Lys  
 115 120 125  
 Arg Ala Arg Arg Ser Met Glu Leu Ile Ala Arg Glu Asp Glu Asn Pro  
 130 135 140  
 Lys Val Ala Glu Val Ile Tyr Pro Ile Met Gln Val Asn Gly Ile His  
 145 150 155 160  
 Tyr Glu Gly Val Asp Val Ala Val Gly Gly Met Glu Gln Arg Lys Ile  
 165 170 175  
 His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His  
 180 185 190  
 Asn Pro Val Leu Thr Gly Leu Asp Gly Glu Gly Lys Met Ser Ser Ser  
 195 200 205  
 Lys Gly Asn Phe Ile Ala Val Asp Asp Ser Pro Glu Glu Ile Arg Ala  
 210 215 220  
 Lys Ile Lys Lys Ala Tyr Cys Pro Ala Gly Val Val Glu Gly Asn Pro  
 225 230 235 240  
 Ile Met Glu Ile Ala Lys Tyr Phe Leu Glu Tyr Pro Leu Thr Ile Lys  
 245 250 255

Arg Pro Glu Lys Phe Gly Gly Asp Leu Thr Val Asn Ser Tyr Glu Glu  
260 265 270

Leu Glu Ser Leu Phe Lys Asn Lys Glu Leu His Pro Met Asp Leu Lys  
275 280 285

Asn Ala Val Ala Glu Glu Leu Ile Lys Ile Leu Glu Pro Ile Arg Lys  
290 295 300

Arg Leu  
305

<210> 6  
<211> 88  
<212> DNA  
<213> Halobacterium sp. NRC-1

<400> 6  
cccagggtag ccaagctcgg ccaacggcga cggactctaa atccgttctc gtaggagttc 60  
gagggttcga atcccttccc tgggacca 88

<210> 7  
<211> 88  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Artificial Sequence

<400> 7  
cgggggggttg ccgagcctgg ccaaaggcgc cggacttcaa atccggtccc gtaggggttc 60  
cggggttcaa atccccgccc ccgcacca 88

<210> 8  
<211> 77  
<212> DNA  
<213> Methanococcus jannaschii

<400> 8  
ccggcggttag ttcagcaggg cagaacggcg gactctaaat ccgcatggcg ctggttcaaa 60  
tccggcccgc cggacca 77

<210> 9  
<211> 65  
<212> DNA

<213> Artificial Sequence

<220>

<223> Artificial Sequence

<400> 9

gtaacgctga attcccgcg gtagttcagc agggcagaac ggcggactct aaatccgcat 60

ggcgc 65

<210> 10

<211> 67

<212> DNA

<213> Artificial Sequence

<220>

<223> Artificial Sequence

<400> 10

gatctgcagt ggtccggcgg gccggatttg aaccggcgcc atgcggattt agagtccgcc 60

gttctgc 67

<210> 11

<211> 67

<212> DNA

<213> Artificial Sequence

<220>

<223> Artificial Sequence

<400> 11

gatctgcagt ggtccggcgg gctggatttg aaccagcgcc atgcggattt agagtccgcc 60

gttctgc 67

<210> 12

<211> 67

<212> DNA

<213> Artificial Sequence

<220>

<223> Artificial Sequence

<400> 12

gatctgcagt ggtccggcgg gcaggatttg aacctgcgcc atgcggattt agagtccgcc 60

gttctgc 67

<210> 13

<211> 49

<212> DNA

<213> Artificial Sequence

<220>  
<223> Artificial Sequence

<400> 13  
cgccggacca ctgcagatcc ttagcgaaag ctaaggattt tttttaagc 49

<210> 14  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Artificial Sequence

<400> 14  
caaattcgtc catatgggat tcc 23

<210> 15  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Artificial Sequence

<400> 15  
gtaacgctga attcccggcg 20

<210> 16  
<211> 600  
<212> DNA  
<213> Homo sapiens

<400> 16  
atggggccacc accaccacca ccacttccca accattccct tatccaggct ttttgacaac 60  
gctatgctcc gcgcccacg tctgcaccag ctggcctttg acacctacca ggagtttgaa 120  
gaagcctaga tcccaaagga acagaagtat tcattcctgc agaaccacca gacctccctc 180  
tgtttctcag agtctattcc gacaccctcc aacagggagg aaacacaaca gaaatccaac 240  
ctagagctgc tccgcatctc cctgctgctc atccagtcgt ggctggagcc cgtgcagttc 300  
ctcaggagtg ttttcgcaa cagcctgggtg tacggcgct ctgacagcaa cgtctatgac 360  
ctcctaaagg acctagagga aggcaccaa acgctgatgg ggaggctgga agatggcagc 420  
ccccggactg ggcagatctt caagcagacc tacagcaagt tcgacacaaa ctcacacaac 480  
gatgacgcac tactcaagaa ctacgggctg ctctaactgt tcaggaagga catggacaag 540  
gtcgagacat tcctgcgcat cgtgcagtgc cgctctgtgg agggcagctg tggcttctaa 600

<210> 17  
 <211> 306  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Artificial Sequence

<400> 17

Met Asp Glu Phe Glu Met Ile Lys Arg Asn Thr Ser Glu Ile Ile Ser  
 1 5 10 15

Glu Glu Glu Leu Arg Glu Val Leu Lys Lys Asp Glu Lys Ser Ala Val  
 20 25 30

Ile Gly Phe Glu Pro Ser Gly Lys Ile His Leu Gly His Tyr Leu Gln  
 35 40 45

Ile Lys Lys Met Ile Asp Leu Gln Asn Ala Gly Phe Asp Ile Ile Ile  
 50 55 60

Tyr Leu Ala Asp Leu His Ala Tyr Leu Asn Gln Lys Gly Glu Leu Asp  
 65 70 75 80

Glu Ile Arg Lys Ile Gly Asp Tyr Asn Lys Lys Val Phe Glu Ala Met  
 85 90 95

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Glu His Gly Leu Asp Lys  
 100 105 110

Asp Tyr Thr Leu Asn Val Tyr Arg Leu Ala Leu Lys Thr Thr Leu Lys  
 115 120 125

Arg Ala Arg Arg Ser Met Glu Leu Ile Ala Arg Glu Asp Glu Asn Pro  
 130 135 140

Lys Val Ala Glu Val Ile Tyr Pro Ile Met Gln Val Asn Gly Ile His  
 145 150 155 160

Tyr Glu Gly Val Asp Val Ala Val Gly Gly Met Glu Gln Arg Lys Ile  
 165 170 175

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His

180

185

190

Asn Pro Val Leu Thr Gly Leu Asp Gly Glu Gly Lys Met Ser Ser Ser  
195 200 205

Lys Gly Asn Phe Ile Ala Val Asp Asp Ser Pro Glu Glu Ile Arg Ala  
210 215 220

Lys Ile Lys Lys Ala Tyr Cys Pro Ala Gly Val Val Glu Gly Asn Pro  
225 230 235 240

Ile Met Glu Ile Ala Lys Tyr Phe Leu Glu Tyr Pro Leu Thr Ile Lys  
245 250 255

Arg Pro Glu Lys Phe Gly Gly Asp Leu Thr Val Asn Ser Tyr Glu Glu  
260 265 270

Leu Glu Ser Leu Phe Lys Asn Lys Glu Leu His Pro Met Arg Leu Lys  
275 280 285

Asn Ala Val Ala Glu Glu Leu Ile Lys Ile Leu Glu Pro Ile Arg Lys  
290 295 300

Arg Leu  
305