SITE SPECIFIC INCORPORATION OF NON-NATURAL AMINO ACIDS BY VERTEBRATE CELLS

FIELD OF THE INVENTION

[01] The invention pertains to the field of translation biochemistry in vertebrate cells. The invention relates to methods for producing and compositions of orthogonal tRNA's, orthogonal synthetases and pairs thereof, in vertebrate cells. The invention also relates to compositions of unnatural amino acids, proteins and methods of producing proteins in vertebrate cells that include unnatural amino acids.

BACKGROUND OF THE INVENTION

[02] 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 proteins. 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).

[03] 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.

[04] In an attempt to expand the ability to manipulate protein structure and function, *in vitro* methods using chemically acylated orthogonal tRNA's were developed that allowed unnatural amino acids to be selectively incorporated in response to a nonsense codon, *in vitro (see, e.g.*, J. A. Ellman, et al., (1992), Science 255:197-200). Amino acids with novel structures and physical properties were selectively incorporated into proteins to study protein folding and stability and biomolecular recognition and catalysis. *See, e.g.*, D. Mendel, et al., (1995), Annual Review of Biophysics and Biomolecular Structure 24:435-462; and, V. W. Cornish, et al. (Mar. 31, 1995), Angewandte Chemie-International Edition in English 34:621-633. However, the stoichiometric nature of this process severely limited the amount of protein that could be generated.

[05] 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 tRNAGln for in vivo incorporation of unnatural amino acids into proteins by nonsense suppression*, J. Biol. Chem. 271:23169-23175), and the relevant mRNA. This has allowed detailed biophysical studies of the receptor in oocytes by the introduction of amino acids containing side chains with unique physical or chemical properties. *See, e.g.*, D.A. Dougherty (2000), *Unnatural amino acids as probes of protein structure and function*, Curr. Opin. Chem. Biol. 4:645-652. 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.

[06] To overcome these limitations, new components were added to the protein biosynthetic machinery of the prokaryote *Escherichia coli* (*E. coli*) (e.g., L. Wang, et al., (2001), Science 292:498-500), which allowed genetic encoding of unnatural amino acids *in vivo*. A number of new amino acids with novel chemical, physical or biological properties, including photoaffinity labels and photoisomerizable amino acids, keto amino acids, and

glycosylated amino acids have been incorporated efficiently and with high fidelity into proteins in *E. coli* in response to the amber codon, TAG, using this methodology. *See, e.g.*, J. W. Chin et al., (2002), Journal of the American Chemical Society 124:9026-9027; J. W. Chin, & P. G. Schultz, (2002), ChemBioChem 11:1135-1137; J. W. Chin, et al., (2002), PNAS United States of America 99:11020-11024: and, L. Wang, & P. G. Schultz, (2002), Chem. Comm., 1-10. However, the translational machinery of prokaryotes and eukaryotes are not highly conserved; thus, components of the biosynthetic machinery added to *E.coli* cannot often be used to site-specifically incorporate unnatural amino acids into proteins in vertebrate cells. For example, the *Methanococcus jannaschii* tyrosyl-tRNA synthetase/tRNA pair that was used in *E.coli* is not orthogonal in vertebrate cells. In addition, the transcription of tRNA in eukaryotes, but not in prokaryotes, is carried out by RNA Polymerase III and this places restrictions on the primary sequence of the tRNA structural genes that can be transcribed in vertebrate cells. Moreover, in contrast to prokaryotic cells, tRNA's in vertebrate cells need to be exported from the nucleus, where they are transcribed, to the cytoplasm, to function in translation. Finally, the vertebrate 80S ribosome is distinct from the 70S prokaryotic ribosome. Thus, there is a need to develop improved components of the biosynthetic machinery to expand the vertebrate genetic code. This invention fulfills these and other needs, as will be apparent upon review of the following disclosure.

[07] SUMMARY OF THE INVENTION

[08] The invention provides vertebrate cells with translation components, e.g., pairs of orthogonal aminoacyl-tRNA synthetases (O-RSs) and orthogonal tRNA's (O-tRNA's) and individual components thereof, that are used in vertebrate protein biosynthetic machinery to incorporate an unnatural amino acid in a growing polypeptide chain, in a vertebrate cell.

[09] Compositions of the invention include a vertebrate cell (e.g., a mammalian cell, an avian cell, a fish cell, a reptile cell, an amphibian cell, cells derived from non-mammalian animals, etc.) comprising an orthogonal aminoacyl-tRNA synthetase (O-RS) (e.g., derived from a non-vertebrate organism, such as *Escherichia coli, Bacillus stearothermophilus,* etc.), where the O-RS preferentially aminoacylates an orthogonal tRNA (O-tRNA) with at least one unnatural amino acid in the vertebrate cell. Optionally, two or more OtRNA's can be aminoacylated in a given vertebrate cell. In one aspect, an O-RS

aminoacylates an O-tRNA with the unnatural amino acid, e.g., at least 40%, at least 45%, at least 50%, at least 60%, at least 75%, at least 80%, or even 90% or more as efficiently as does an O-RS having an amino acid sequence, e.g., as set forth in SEQ ID NO.: 86 or 45. In one embodiment, an O-RS of the invention aminoacylates the O-tRNA with the unnatural amino acid, e.g., at least 10-fold, at least 20-fold, at least 30-fold, etc., more efficiently than the O-RS aminoacylates the O-tRNA with a natural amino acid.

[10] In one embodiment, the O-RS or a portion thereof is encoded by a polynucleotide sequence as set forth in any one of SEQ ID NO.: 3-35, or a complementary polynucleotide sequence thereof. In another embodiment, the O-RS comprises an amino acid sequence as set forth in any one of SEQ ID NO.: 36-63, and/or 86, or a conservative variation thereof. In yet another embodiment, the O-RS comprises an amino acid sequence that is, e.g., at least 90%, at least 95%, at least 98%, at least 99%, or at least 99.5 % or more, identical to that of a naturally occurring tyrosyl aminoacyl-tRNA synthetase (TyrRS) and comprises two or more amino acids from groups A-E. Group A includes valine, isoleucine, leucine, glycine, serine, alanine, or threonine at a position corresponding to Tyr37 of an *E. coli* TyrRS. Group B includes aspartate at a position corresponding to Asn126 of an *E. coli* TyrRS. Group C includes threonine, serine, arginine, asparagine or glycine at a position corresponding to Asp182 of an *E. coli* TyrRS. Group D includes methionine, alanine, valine, or tyrosine at a position corresponding to Phe183 of an *E. coli* TyrRS; and, group E includes serine, methionine, valine, cysteine, threonine, or alanine at a position corresponding to Leu186 of an *E. coli* TyrRS.

[11] In another embodiment, the O-RS has one or more improved or enhanced enzymatic properties for the unnatural amino acid as compared to a natural amino acid. For example, the improved or enhanced properties for the unnatural amino acid as compared to a natural amino acid include any of, e.g., a higher Km, a lower Km, a higher kcat, a lower kcat, a lower kcat/km, a higher kcat/km, etc.

[12] The vertebrate cell also optionally includes an unnatural amino acid(s). The vertebrate cell optionally includes an orthogonal tRNA (O-tRNA) (e.g., derived from a non-vertebrate organism, such as *Escherichia coli*, *Bacillus stearothermophilus*, and/or the like), where the O-tRNA recognizes a selector codon and is preferentially aminoacylated with the unnatural amino acid by the O-RS. In one aspect, the O-tRNA mediates the incorporation of the unnatural amino acid into a protein with, e.g., at least 45%, at least 50%, at least 60%,

at least 75%, at least 80%, at least 90%, at least 95%, or 99% or the efficiency of a tRNA that comprises or is processed in a cell from a polynucleotide sequence as set forth in SEQ ID NO.: 65. In another aspect, the O-tRNA comprises the sequence of SEQ ID NO.:65, and the O-RS comprises a polypeptide sequence selected from an amino acid sequence set forth in any one of SEQ ID NO.: 36-63, and/or 86, and/or a conservative variation thereof.

[13] In another embodiment, the vertebrate cell comprises 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. In one aspect, the yield of the polypeptide of interest comprising the unnatural amino acid is, e.g., at least 2.5%, at least 5%, at least 10%, at least 25%, at least 30%, at least 40%, 50% or more, of that obtained for the naturally occurring polypeptide of interest from a cell in which the polynucleotide lacks the selector codon. In another aspect, the cell produces the polypeptide of interest in the absence of the unnatural amino acid, with a yield that is, e.g., less than 35%, less than 30%, less than 20%, less than 15%, less than 10%, less than 5%, less than 2.5%, etc., of the yield of the polypeptide in the presence of the unnatural amino acid.

[14] The invention also provides a vertebrate cell comprising an orthogonal aminoacyl-tRNA synthetase (O-RS), an orthogonal tRNA (O-tRNA), an unnatural amino acid, and a nucleic acid that comprises a polynucleotide that encodes a polypeptide of interest. The polynucleotide comprises a selector codon that is recognized by the O-tRNA. In addition, the O-RS preferentially aminoacylates the orthogonal tRNA (O-tRNA) with the unnatural amino acid in the vertebrate cell, and the cell produces the polypeptide of interest in the absence of the unnatural amino acid, with a yield that is, e.g., less than 30%, less than 20%, less than 15%, less than 10%, less than 5%, less than 2.5%, etc., of the yield of the polypeptide in the presence of the unnatural amino acid.

[15] Compositions that include a vertebrate cell comprising an orthogonal tRNA (O-tRNA) are also a feature of the invention. Typically, the O-tRNA mediates incorporation of an unnatural amino acid into a protein that is encoded by a polynucleotide that comprises a selection codon that is recognized by the O-tRNA in vivo. In one embodiment, the O-tRNA mediates the incorporation of the unnatural amino acid into the protein with, e.g., at least 45%, at least 50%, at least 60%, at least 75%, at least 80%, at least 90%, at least 95%, or even 99% or more the efficiency of a tRNA that comprises or is

processed in a cell from a polynucleotide sequence as set forth in SEQ ID NO.: 65. In another embodiment, the O-tRNA comprises or is processed from a polynucleotide sequence as set forth in SEQ ID NO.: 65, or a conservative variation thereof. In yet another embodiment, the O-tRNA comprises a recyclable O-tRNA.

[16] In one aspect of the invention, the O-tRNA is post-transcriptionally modified. The invention also provides a nucleic acid that encodes an O-tRNA in a vertebrate cell, or a complementary polynucleotide thereof. In one embodiment, the nucleic acid comprises an A box and a B box.

[17] The invention also features methods of producing translational components, e.g., O-RSs or O-tRNA/O-RS pairs (and translational components produced by these methods). For example, the invention provides methods of producing an orthogonal aminoacyl-tRNA synthetase (O-RS) that preferentially aminoacylates an orthogonal tRNA with an unnatural amino acid in a vertebrate cell. The method includes, e.g., (a) subjecting to positive selection, in the presence of an unnatural amino acid, a population of vertebrate cells of a first species, where the vertebrate cells each comprise: i) a member of a library of aminoacyl-tRNA synthetases (RSs), ii) an orthogonal tRNA (O-tRNA), iii) a polynucleotide that encodes a positive selection marker, and iv) a polynucleotide that encodes a negative selection marker; where cells that survive the positive selection comprise an active RS that aminoacylates the orthogonal tRNA (O-tRNA) in the presence of an unnatural amino acid. The cells that survive the positive selection are subjected to negative selection in the absence of the unnatural amino acid to eliminate active RSs that aminoacylate the O-tRNA with a natural amino acid. This provides the O-RS that preferentially aminoacylates the O-tRNA with the unnatural amino acid.

[18] In certain embodiments, the polynucleotide that encodes the positive selection marker is operably linked to a response element and the cells further comprise a polynucleotide that: a) encodes a transcriptional modulator protein (e.g., a vertebrate transcriptional modulator protein, etc.) that modulates transcription from the response element, and b) comprises at least one selector codon. 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 positive selection marker. In one embodiment, the transcriptional modulator protein is a transcriptional activator protein (e.g., GAL4, etc.), and the selector codon is an amber stop codon, e.g., where the amber stop

codon is located in or substantially near a portion of the polynucleotide that encodes a DNA binding domain of the transcriptional activator protein.

[19] The positive selection marker can be any of a variety of molecules. In one embodiment, the positive selection marker comprises a nutritional supplement for growth and the selection is performed on a medium that lacks the nutritional supplement. In another embodiment, the polynucleotide that encodes the positive selection marker is, e.g., an ura3, leu2, lys2, lacZ gene, his3 (e.g., where the his3 gene encodes an imidazole glycerol phosphate dehydratase, detected by providing 3-aminotriazole (3-AT)), and/or the like. In yet another embodiment, the polynucleotide that encodes the positive selection marker comprises a selector codon.

[20] As with the positive selection marker, the negative selection marker can also be any of a variety of molecules. In certain embodiments, the polynucleotide that encodes the negative selection marker is operably linked to a response element from which transcription is mediated by the transcriptional modulator protein. 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. In one embodiment, the polynucleotide that encodes the negative selection marker is, e.g., an ura3 gene and the negative selection is accomplished on a medium that comprises 5-fluroorotic acid (5-FOA). In another embodiment, the medium used for negative selection comprises 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. In one embodiment, the polynucleotide that encodes the negative selection marker comprises a selector codon.

[21] In certain embodiments, the positive selection marker and/or the negative selection marker comprises a polypeptide that fluoresces or catalyzes a luminescent reaction in the presence of a suitable reactant. In one aspect of the invention, the positive selection marker and/or the negative selection marker is detected by fluorescence-activated cell sorting (FACS), or by luminescence. In certain embodiments, the positive selection marker and/or negative selection marker comprises an affinity based screening marker, or a transcriptional modulator protein. In one embodiment, the same polynucleotide encodes both the positive selection marker and the negative selection marker.

[22] In one embodiment, the polynucleotide that encodes the positive selection marker and/or negative selection marker of the invention can comprises 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.

[23] Additional levels of selection/screening stringency can also be used in the methods of the invention. In one embodiment, the methods can comprise, e.g., providing a varying amount of an inactive synthetase in step (a), (b) or both (a) and (b), where the varying amount of the inactive synthetase provides an additional level of selection or screening stringency. In one embodiment, step (a), (b) or both steps (a) and (b) of the method for producing an O-RS includes varying a selection or screening stringency, e.g., of the positive and/or negative selection marker. The method optionally includes subjecting the O-RS that preferentially aminoacylates the O-tRNA with the unnatural amino acid to an additional selection round, e.g., an additional positive selection round(s), an additional negative selection round(s) or combinations of both additional positive and negative selection rounds.

[24] In one embodiment, the selecting/screening comprises one or more positive or negative selection/screening chosen from, e.g., a change in amino acid permeability, a change in translation efficiency, a change in translational fidelity, etc. The one or more change is based upon a mutation in one or more polynucleotide that encodes a component of orthogonal tRNA-tRNA synthetase pair is used to produce protein.

[25] Typically, the library of RSs (e.g., a library of mutant RSs) comprises RSs derived from at least one aminoacyl-tRNA synthetase (RS), e.g., from a non-vertebrate organism. In one embodiment, the library of RSs is derived from an inactive RS, e.g., where the inactive RS is generated by mutating an active RS. In another embodiment, the inactive RS comprises an amino acid binding pocket and one or more amino acids that comprise the binding pocket are substituted with one or more different amino acids, e.g., the substituted amino acids are substituted with alanines.

[26] In certain embodiments, the method of producing an O-RS further includes performing random mutation, site-specific mutation, recombination, chimeric construction, or any combination thereof, on a nucleic acid that encodes an RS, thereby producing the library of mutant RSs. In certain embodiments, the method further includes, e.g., (c) isolating a nucleic acid that encodes the O-RS; (d) generating from the nucleic acid a set of

polynucleotides that encode mutated O-RSs (e.g., by random mutagenesis, site-specific mutagenesis, chimeric construction, recombination or any combination thereof); and, (e) repeating steps (a) and/or (b) until a mutated O-RS is obtained that preferentially aminoacylates the O-tRNA with the unnatural amino acid. In one aspect of the invention, steps (c)-(e) are performed at least two times.

[27] Methods of producing O-tRNA/O-RS pairs are also a feature of the invention. In one embodiment, the O-RS is obtained as described above and the O-tRNA is obtained by subjecting to negative selection a population of vertebrate cells of a first species, where the vertebrate cells comprise a member of a library of tRNA's, to eliminate cells that comprise a member of the library of tRNA's that is aminoacylated by an aminoacyl-tRNA synthetase (RS) that is endogenous to the vertebrate cells. This provides a pool of tRNA's that are orthogonal to the vertebrate cell of the first species. In one aspect of the invention, the library of tRNA's comprises tRNA's derived from at least one tRNA, e.g., from a non-vertebrate organism. In another aspect of the invention, the library of aminoacyl-tRNA synthetases (RSs) comprises RSs derived from at least one aminoacyl-tRNA synthetase (RS), e.g., from a non-vertebrate organism. In yet another aspect of the invention, the library of tRNA's comprises tRNA's derived from at least one tRNA from a first non-vertebrate organism. The library of aminoacyl-tRNA synthetases (RSs) optionally comprises RSs derived from at least one aminoacyl-tRNA synthetase (RS) from a second non-vertebrate organism. In one embodiment, the first and second non-vertebrate organisms are the same. Alternatively, the first and second non-vertebrate organisms can be different. Specific O-tRNA/O-RS pairs produced by the methods of the invention are also a feature of the invention.

[28] Another feature of the invention is a method for producing translational components in one species and introducing the selected/screened translational components into a second species. For example, the method of producing a O-tRNA/O-RS pair in a first species (e.g., a vertebrate species, such as a yeast and the like) further includes introducing a nucleic acid that encodes the O-tRNA and a nucleic acid that encodes the O-RS into a vertebrate cell of a second species (e.g., a mammal, an insect, a fungus, an algae, a plant and the like). The second species can use the introduced translational components to incorporate an unnatural amino acid into a growing polypeptide chain in vivo, e.g., during translation.

[29] In another example, a method of producing an orthogonal aminoacyl-tRNA synthetase (O-RS) that preferentially aminoacylates an orthogonal tRNA with an unnatural amino acid in a vertebrate cell includes: (a) subjecting to positive selection, in the presence of an unnatural amino acid, a population of vertebrate cells of a first species (e.g., a vertebrate species, such as a yeast or the like). The vertebrate cells of the first species each comprise: i) a member of a library of aminoacyl-tRNA synthetases (RSs), ii) an orthogonal tRNA (O-tRNA), iii) a polynucleotide that encodes a positive selection marker, and iv) a polynucleotide that encodes a negative selection marker. The cells that survive the positive selection comprise an active RS that aminoacylates the orthogonal tRNA (O-tRNA) in the presence of an unnatural amino acid. The cells that survive the positive selection are subjected to negative selection in the absence of the unnatural amino acid to eliminate active RSs that aminoacylate the O-tRNA with a natural amino acid, thereby providing an O-RS that preferentially aminoacylates the O-tRNA with the unnatural amino acid. A nucleic acid that encodes the O-tRNA and a nucleic acid that encodes the O-RS are introduced into a vertebrate cell of a second species (e.g., mammal, an insect, a fungus, an algae, a plant and/or the like). These components, when translated in the second species, can be used to incorporate unnatural amino acids into a protein or polypeptide of interest in the second species. In one embodiment, the O-tRNA and/or the O-RS are introduced into a vertebrate cell of a second species.

[30] In certain embodiments, the O-tRNA is obtained by subjecting to negative selection a population of vertebrate cells of a first species, where the vertebrate cells comprise a member of a library of tRNA's, to eliminate cells that comprise a member of the library of tRNA's that is aminoacylated by an aminoacyl-tRNA synthetase (RS) that is endogenous to the vertebrate cells. This provides a pool of tRNA's that are orthogonal to the vertebrate cell of the first species and the second species.

[31] Proteins (or polypeptides of interest) with at least one unnatural amino acid are also a feature of the invention. In certain embodiments of the invention, a protein with at least one unnatural amino acid includes at least one post-translational modification. In one embodiment, the at least one post-translational modification comprises attachment of a molecule (e.g., a dye, a polymer, e.g., a derivative of polyethylene glycol, a photocrosslinker, a cytotoxic compound, an affinity label, a derivative of biotin, a resin, a second protein or polypeptide, a metal chelator, a cofactor, a fatty acid, a carbohydrate, a polynucleotide (e.g., DNA, RNA, etc.), etc.) comprising a second reactive group by a [3+2]

cycloaddition to the at least one unnatural amino acid comprising a first reactive group. For example, the first reactive group is an alkynyl moiety (e.g., in the unnatural amino acid *p*-propargyloxyphenylalanine) (this group is also sometimes refer to as an acetylene moiety) and the second reactive group is an azido moiety. In another example, the first reactive group is the azido moiety (e.g., in the unnatural amino acid *p*-azido-L-phenylalanine) and the second reactive group is the alkynyl moiety. In certain embodiments, a protein of the invention includes at least one unnatural amino acid (e.g., a keto unnatural amino acid) comprising at least one post-translational modification, where the at least one post-translational modification comprises a saccharide moiety. In certain embodiments, the post-translational modification is made in vivo in a vertebrate cell.

[32] In certain embodiments, the protein includes at least one post-translational modification that is made in vivo by a vertebrate cell, where the post-translational modification is not made by a prokaryotic cell. Examples of post-translational modifications include, but are not limited to, acetylation, acylation, lipid-modification, palmitoylation, palmitate addition, phosphorylation, glycolipid-linkage modification, and the like. In one embodiment, the post-translational modification comprises attachment of an oligosaccharide to an asparagine by a GlcNAc-asparagine linkage (e.g., where the oligosaccharide comprises $(GlcNAc\text{-}Man)_2\text{-}Man\text{-}GlcNAc\text{-}GlcNAc$, and the like). In another embodiment, the post-translational modification comprises attachment of an oligosaccharide (e.g., Gal-GalNAc, Gal-GlcNAc, etc.) to a serine or threonine by a GalNAc-serine, a GalNAc-threonine, a GlcNAc-serine, or a GlcNAc-threonine linkage. In certain embodiments, a protein or polypeptide of the invention can comprise a secretion or localization sequence, an epitope tag, a FLAG tag, a polyhistidine tag, a GST fusion, and/or the like.

[33] 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 even 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/or the like), and they comprise one or more unnatural amino acid. In one embodiment, a composition of the invention includes a protein or polypeptide of interest and an excipient (e.g., a buffer, a pharmaceutically acceptable excipient, etc.).

[34] The protein or polypeptide of interest can contain at least one, 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 ten or more unnatural amino acids. The unnatural amino acids can be the same or different, e.g., there can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more different sites in the protein that comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more different unnatural amino acids. In certain embodiments, at least one, but fewer than all, of a particular amino acid present in a naturally occurring version of the protein is substituted with an unnatural amino acid.

[35] Examples of a protein (or polypeptide of interest) include, but are not limited to, e.g., a cytokine, a growth factor, a growth factor receptor, an interferon, an interleukin, an inflammatory molecule, an oncogene product, a peptide hormone, a signal transduction molecule, a steroid hormone receptor, erythropoietin (EPO), insulin, human growth hormone, an Alpha-1 antitrypsin, an Angiostatin, an Antihemolytic factor, an antibody, an Apolipoprotein, an Apoprotein, an Atrial natriuretic factor, an Atrial natriuretic polypeptide, an Atrial peptide, a C-X-C chemokine, T39765, NAP-2, ENA-78, a Gro-a, a Gro-b, a Gro-c, an IP-10, a GCP-2, an NAP-4, an SDF-1, a PF4, a MIG, a Calcitonin, a c-kit ligand, a cytokine, a CC chemokine, a Monocyte chemoattractant protein-1, a Monocyte chemoattractant protein-2, a Monocyte chemoattractant protein-3, a Monocyte inflammatory protein-1 alpha, a Monocyte inflammatory protein-1 beta, RANTES, I309, R83915, R91733, HCC1, T58847, D31065, T64262, a CD40, a CD40 ligand, a C-kit Ligand, a Collagen, a Colony stimulating factor (CSF), a Complement factor 5a, a Complement inhibitor, a Complement receptor 1, a cytokine, DHFR, an epithelial Neutrophil Activating Peptide-78, a GROα/MGSA, a GROβ, a GROγ a MIP-1α, a MIP-1δ, a MCP-1, an Epidermal Growth Factor (EGF), an epithelial Neutrophil Activating Peptide, an Erythropoietin (EPO), an Exfoliating toxin, a Factor IX, a Factor VII, a Factor VIII, a Factor X, a Fibroblast Growth Factor (FGF), a Fibrinogen, a Fibronectin, a G-CSF, a GM-CSF, a Glucocerebrosidase, a Gonadotropin, a growth factor, a growth factor receptor, a Hedgehog protein, a Hemoglobin, a Hepatocyte Growth Factor (HGF), a Hirudin, a Human serum albumin, an ICAM-1, an ICAM-1 receptor, an LFA-1, an LFA-1 receptor, an Insulin, an Insulin-like Growth Factor (IGF), an IGF-I, an IGF-II, an interferon, an IFN-α, an IFN-β, an IFN-γ, an interleukin, an IL-1, an IL-2, an IL-3, an IL-4, an IL-5, an IL-6, an IL-7, an IL-8, an IL-9, an IL-10, an IL-11, an IL-12, a Keratinocyte Growth Factor (KGF), a Lactoferrin, a leukemia inhibitory factor, a Luciferase, a Neurturin, a Neutrophil inhibitory factor (NIF), an oncostatin M, an Osteogenic protein, an oncogene product, a Parathyroid hormone, a PD-ECSF, a PDGF, a peptide hormone, a Human Growth Hormone, a Pleiotropin, a Protein A, a Protein G, a Pyrogenic exotoxins A, B, or C, a Relaxin, a Renin,

an SCF, a Soluble complement receptor I, a Soluble I-CAM 1, a Soluble interleukin receptors, a Soluble TNF receptor, a Somatomedin, a Somatostatin, a Somatotropin, a Streptokinase, a Superantigens, a Staphylococcal enterotoxins, an SEA, an SEB, an SEC1, an SEC2, an SEC3, an SED, an SEE, a steroid hormone receptor, a Superoxide dismutase (SOD), a Toxic shock syndrome toxin, a Thymosin alpha 1, a Tissue plasminogen activator, a tumor growth factor (TGF), a TGF-α, a TGF-β, a Tumor Necrosis Factor, a Tumor Necrosis Factor alpha, a Tumor necrosis factor beta, a Tumor necrosis factor receptor (TNFR), a VLA-4 protein, a VCAM-1 protein, a Vascular Endothelial Growth Factor (VEGEF), a Urokinase, a Mos, a Ras, a Raf, a Met; a p53, a Tat, a Fos, a Myc, a Jun, a Myb, a Rel, an estrogen receptor, a progesterone receptor, a testosterone receptor, an aldosterone receptor, an LDL receptor, a SCF/c-Kit, a CD40L/CD40, a VLA-4/VCAM-1, an ICAM-1/LFA-1, a hyalurin/CD44, a corticosterone, a protein present in Genebank or other available databases, and the like, and/or a portion thereof. In one embodiment, the polypeptide of interest includes a transcriptional modulator protein (e.g., a transcriptional activator protein (such as GAL4), or a transcriptional repressor protein, etc.) or a portion thereof.

[36] A vertebrate cell of the invention provides the ability to synthesize proteins that comprise unnatural amino acids in large useful quantities. For example, proteins comprising an unnatural amino acid can be produced at a concentration of, e.g., at least 10 μg/liter, at least 50 μg/liter, at least 75 μg/liter, at least 100 μg/liter, at least 200 μg/liter, at least 250 μg/liter, or at least 500 μg/liter or more of protein in a cell extract, a buffer, a pharmaceutically acceptable excipient, and/or the like. In certain embodiments, a composition of the invention includes, e.g., at least 10 μg, at least 50μg, at least 75 μg, at least 100 μg, at least 200 μg, at least 250 μg, or at least 500μg or more of protein that comprises a unnatural amino acid.

[37] In certain embodiments, the protein or polypeptide of interest (or portion thereof) 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, or even ten or more selector codons.

[38] The invention also provides methods for producing, in a vertebrate cell, at least one protein comprising at least one unnatural amino acid (as well as proteins produced by such methods). The methods include, e.g., growing, in an appropriate medium, a vertebrate cell that comprises a nucleic acid that comprises at least one selector codon and encodes the protein. The vertebrate cell also 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 unnatural amino acid, and the medium comprises an unnatural amino acid. In one embodiment, the O-RS aminoacylates the O-tRNA with the unnatural amino acid e.g., at least 45%, at least 50%, at least 60%, at least 75%, at least 80%, at least 90%, at least 95%, or even 99% or more as efficiently as does an O-RS having an amino acid sequence, e.g., as set forth in SEQ ID NO.: 86 or 45. In another embodiment, the O-tRNA comprises, is processed from, or is encoded by SEQ ID NO.: 64 or 65, or a complementary polynucleotide sequence thereof. In yet another embodiment, the O-RS comprises an amino acid sequence as set forth in any one of SEQ ID NO.: 36-63, and/or 86.

[39] In one embodiment, the method further includes incorporating into the protein the unnatural amino acid, where the unnatural amino acid comprises a first reactive group; and contacting the protein with a molecule (e.g., a dye, a polymer, e.g., a derivative of polyethylene glycol, a photocrosslinker, a cytotoxic compound, an affinity label, a derivative of biotin, a resin, a second protein or polypeptide, a metal chelator, a cofactor, a fatty acid, a carbohydrate, a polynucleotide (e.g., DNA, RNA, etc.), etc.) that comprises a second reactive group. The first reactive group reacts with the second reactive group to attach the molecule to the unnatural amino acid through a [3+2] cycloaddition. In one embodiment, the first reactive group is an alkynyl or azido moiety and the second reactive group is an azido or alkynyl moiety. For example, the first reactive group is the alkynyl moiety (e.g., in unnatural amino acid p-propargyloxyphenylalanine) and the second reactive group is the azido moiety. In another example, the first reactive group is the azido moiety (e.g., in the unnatural amino acid p-azido-L-phenylalanine) and the second reactive group is the alkynyl moiety.

[40] In certain embodiments, the encoded protein comprises a therapeutic protein, a diagnostic protein, an industrial enzyme, or portion thereof. In one embodiment, the protein that is produced by the method is further modified through the unnatural amino acid. For example, the unnatural amino acid is modified through, e.g., a nucleophilic-electrophilic

reaction, through a [3+2] cycloaddition, etc. In another embodiment, the protein produced by the method is modified by at least one post-translational modification (e.g., N-glycosylation, O-glycosylation, acetylation, acylation, lipid-modification, palmitoylation, palmitate addition, phosphorylation, glycolipid-linkage modification, and the like) in vivo.

[41] Methods of producing a screening or selecting transcriptional modulator protein are also provided (as are screening or selecting transcriptional modulator proteins produced by such methods). The methods include, e.g., selecting a first polynucleotide sequence, where the polynucleotide sequence encodes a nucleic acid binding domain; and mutating the first polynucleotide sequence to include at least one selector codon. This provides a screening or selecting polynucleotide sequence. The methods also include, e.g., selecting a second polynucleotide sequence, where the second polynucleotide sequence encodes a transcriptional activation domain; providing a construct that comprises the screening or selecting polynucleotide sequence operably linked to the second polynucleotide sequence; and, introducing the construct, an unnatural amino acid, an orthogonal tRNA synthetase (O-RS) and an orthogonal tRNA (O-tRNA), into a cell. With these components, the O-RS preferentially aminoacylates the O-tRNA with the unnatural amino acid and the O-tRNA recognizes the selector codon and incorporates the unnatural amino acid into the nucleic acid binding domain, in response to the selector codon in the screening or selecting polynucleotide sequence. This provides the screening or selecting transcriptional modulator protein.

[42] In certain embodiments, the compositions and the methods of the invention include vertebrate cells. A vertebrate cell of the invention includes any of, e.g., a mammalian cell, a yeast cell, a fungus cell, a plant cell, an insect cell, etc. The translation components of the invention can be derived from a variety of organisms, e.g., non-vertebrate organisms, such as a prokaryotic organism (e.g., *E. coli, Bacillus stearothermophilus,* or the like), or an archaebacterium, or e.g., a vertebrate organism.

[43] A selector codon of the invention expands the genetic codon framework of vertebrate protein biosynthetic machinery. Any of a variety of selector codons can be used in the invention, including stop codons (e.g., an amber codon, an ochre codon, or an opal stop codon), nonsense codons, rare codons, four (or more) base codons, and/or the like.

[44] Examples of unnatural amino acids that can be used in the compositions and methods described herein include (but are not limited to): a *p*-acetyl-L-phenylalanine, a *p*-

iodo-L-phenylalanine, an O-methyl-L-tyrosine, a *p*-propargyloxyphenylalanine, a *p*-propargyl-phenylalanine, 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-GlcNAcβ-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 phosphonoserine, a phosphonotyrosine, a *p*-bromophenylalanine, a *p*-amino-L-phenylalanine, an isopropyl-L-phenylalanine, an unnatural analogue of a tyrosine amino acid; an unnatural analogue of a glutamine amino acid; an unnatural analogue of a phenylalanine amino acid; an unnatural analogue of a serine amino acid; an unnatural analogue of a threonine amino acid; an alkyl, aryl, acyl, azido, cyano, halo, hydrazine, hydrazide, hydroxyl, alkenyl, alkynl, ether, thiol, sulfonyl, seleno, ester, thioacid, borate, boronate, phospho, phosphono, phosphine, heterocyclic, enone, imine, aldehyde, hydroxylamine, keto, or amino substituted amino acid, or any combination thereof; an amino acid with a photoactivatable cross-linker; a spin-labeled amino acid; a fluorescent amino acid; a metal binding amino acid; a metal-containing amino acid; a radioactive amino acid; a photocaged and/or photoisomerizable amino acid; a biotin or biotin-analogue containing amino acid; a keto containing amino acid; an amino acid comprising polyethylene glycol or polyether; a heavy atom substituted amino acid; a chemically cleavable or photocleavable amino acid; an amino acid with an elongated side chain; an amino acid containing a toxic group; a sugar substituted amino acid; a carbon-linked sugar-containing amino acid; a redox-active amino acid; an α-hydroxy containing acid; an amino thio acid; an α,α disubstituted amino acid; a β-amino acid; a cyclic amino acid other than proline or histidine, an aromatic amino acid other than phenylalanine, tyrosine or tryptophan, and/or the like.

[45] The invention also provides polypeptides (O-RSs) and polynucleotides, e.g., O-tRNA's, polynucleotides that encode O-RSs or portions thereof (e.g., the active site of the synthetase), oligonucleotides used to construct aminoacyl-tRNA synthetase mutants, polynucleotides that encode a protein or polypeptide of interest that comprise one or more selector codon, etc. For example, a polypeptide of the invention includes a polypeptide that comprises an amino acid sequence as set forth in any one of SEQ ID NO.: 36-63, and/or 86, a polypeptide that comprises an amino acid sequence encoded by a polynucleotide sequence as set forth in any one of SEQ ID NO.: 3-35, and a polypeptide that is specifically immunoreactive with an antibody specific for a polypeptide that comprises an amino acid sequence as shown in any one of SEQ ID NO.: 36-63, and/or 86, or a polypeptide that

comprises an amino acid sequence encoded by a polynucleotide sequence as shown in any one of SEQ ID NO.: 3-35.

[46] Also included among the polypeptides of the invention is a polypeptide that comprises an amino acid sequence that is at least 90% identical to that of a naturally occurring tyrosyl aminoacyl-tRNA synthetase (TyrRS) (e.g., SEQ ID NO.:2) and comprises two or more amino acids of groups A-E (noted above). Similarly, polypeptides of the invention also optionally include a polypeptide that comprises at least 20 contiguous amino acids of any one of SEQ ID NO.: 36-63, and/or 86, and two or more amino acid substitutions as indicated above in groups A-E. An amino acid sequence comprising a conservative variation of any of the above polypeptides is also included as a polypeptide of the invention.

[47] In one embodiment, a composition includes a polypeptide of the invention and an excipient (e.g., buffer, water, pharmaceutically acceptable excipient, etc.). The invention also provides an antibody or antisera specifically immunoreactive with a polypeptide of the invention.

[48] Polynucleotides are also provided in the invention. Polynucleotides of the invention include those that encode proteins or polypeptides of interests of the invention with one or more selector codon. In addition, polynucleotides of the invention include, e.g., a polynucleotide comprising a nucleotide sequence as set forth in any one of SEQ ID NO.: 3-35, 64-85; a polynucleotide that is complementary to or that encodes a polynucleotide sequence thereof; and/or a polynucleotide encoding a polypeptide that comprises an amino acid sequence as set forth in any one of SEQ ID NO.: 36-63, and/or 86, or a conservative variation thereof. A polynucleotide of the invention also includes a polynucleotide that encodes a polypeptide of the invention. 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 invention.

[49] A polynucleotide of the invention also includes a polynucleotide that encodes a polypeptide that comprises an amino acid sequence that is at least 90% identical to that of a naturally occurring tyrosyl aminoacyl-tRNA synthetase (TyrRS) (e.g., SEQ ID NO.: 2) and comprises two or more mutations as indicated above in groups A-E (noted above). A polynucleotide that is that is at least 70%, (or at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or least 99% or more) identical to a

polynucleotide indicated above and/or a polynucleotide comprising a conservative variation of any of the polynucleotides indicated above are also included among the polynucleotides of the invention.

[50] In certain embodiments, a vector (e.g., a plasmid, a cosmid, a phage, a virus, etc.) comprises a polynucleotide of the 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 invention. In another embodiment, a cell comprises a vector that includes a polynucleotide of the invention.

[51] In another aspect, the invention provides compositions of compounds and methods of producing such compounds. For example, compounds include, e.g., an unnatural amino acid (such as *p*-(propargyloxy)-phenyalanine (e.g., 1 in **Figure 11**), azido dyes (such as shown in chemical structure 4 and chemical structure 6), an alkynyl polyethylene glycol (e.g., as shown in chemical structure 7), where n is an integer between, e.g., 50 and 10,000, 75 and 5,000, 100 and 2,000, 100 and 1,000, etc., and the like. In embodiment of the invention, the alkynyl polyethylene glycol has a molecular weight of, e.g., about 5,000 to about 100,000 Da, about 20,000 to about 50, 000 Da, about 20,000 to about 10,000 Da (e.g., 20,000 Da).

[52] Various compositions comprising these compounds, e.g., with proteins and cells, are also provided. In one aspect, the composition that includes the *p*-(propargyloxy)-phenyalanine unnatural amino acid, further includes an orthogonal tRNA. The unnatural amino acid can be bonded (e.g., covalently) to the orthogonal tRNA, e.g., covalently bonded to the orthogonal tRNA though an amino-acyl bond, covalently bonded to a 3'OH or a 2'OH of a terminal ribose sugar of the orthogonal tRNA, etc.

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

[54] BRIEF DESCRIPTION OF THE DRAWINGS

[55] Figure 1 shows incorporation of para-acetyl-phenylalanine into hGH.

[56] Figure 2 shows incorporation of para-acetyl-phenylalanine at various concentrations into hGH.

[56.1] Figure 3: A bar graph of the results from Example 3 showing the Fc titer in micrograms per mL along the Y-axis from CHO cells with different ratios of pAFRS/I21/4xhtRNA shown along the x-axis.

[57] DETAILED DESCRIPTION

[58] Before describing the present invention in detail, it is to be understood that this invention is not limited to particular devices or 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 be limiting. As used in this specification and 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; reference to "bacteria" includes mixtures of bacteria, and the like.

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

[60] 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 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. 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%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 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.

[61] 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 functions with endogenous components of a cell with reduced efficiency as compared to a corresponding molecule that is endogenous to the cell or translation system, or that fails to function with endogenous components of the cell. In the context of tRNA's and aminoacyl-tRNA synthetases, orthogonal refers to an inability or reduced efficiency, e.g., less than 20 % efficient, less than 10 % efficient, less than 5 % efficient, or less than 1% efficient, of an orthogonal tRNA to function with an endogenous tRNA synthetase compared to an endogenous tRNA to function with the endogenous tRNA synthetase, or of an orthogonal aminoacyl-tRNA synthetase to function with an endogenous tRNA compared to an endogenous tRNA synthetase to function with the endogenous tRNA. The orthogonal molecule lacks a functional endogenous complementary molecule in the cell. For example, an orthogonal tRNA in a cell is aminoacylated by any endogenous RS of the cell with reduced or even zero efficiency, when compared to aminoacylation of an endogenous tRNA by the endogenous RS. In another example, an orthogonal RS aminoacylates any endogenous tRNA in a cell 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., 50% efficiency, 60% efficiency, 70% efficiency, 75% efficiency, 80% efficiency, 90% efficiency, 95% efficiency, or 99% or more efficiency) to that of a corresponding tRNA/RS endogenous pair.

[62] Complementary: The term "complementary" refers to components of an orthogonal pair, O-tRNA and O-RS that can function together, e.g., where the O-RS aminoacylates the O-tRNA.

[63] Preferentially aminoacylates: The term "preferentially aminoacylates" refers to an efficiency, e.g., 70 % efficient, 75 % efficient, 85% efficient, 90% efficient, 95 % efficient, or 99% or more efficient, at which an O-RS aminoacylates an O-tRNA with an unnatural amino acid as 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 incorporated into a growing polypeptide chain with high fidelity, e.g., 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 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.

[64] 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., an unnatural amino acid, at this site in the polypeptide. Selector codons can include, e.g., nonsense codons, such as, stop codons, e.g., amber, ochre, and opal codons; four or more base codons; rare codons; codons derived from natural or unnatural base pairs and/or the like.

[65] 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, e.g., a stop codon, a four base codon, a rare codon, and/or the like.

[66] Recyclable tRNA: The term "recyclable tRNA" refers to a tRNA that is aminoacylated and can be repeatedly reaminoacylated with an amino acid (e.g., an unnatural amino acid) for the incorporation of the amino acid (e.g., the unnatural amino acid) into one or more polypeptide chains during translation.

[67] Translation system: The term "translation system" refers to the collective set of components that 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, amino acids, and the like. The components of the invention (e.g., ORS, OtRNA's, unnatural amino acids, etc.) can be added to an in vitro or in vivo translation system, e.g., a vertebrate cell, e.g., a yeast cell, a mammalian cell, a plant cell, an algae cell, a fungus cell, an insect cell, and/or the like.

[68] Unnatural amino acid: As used herein, the term "unnatural 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, seleno cysteine or pyrrolysine.

[69] 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.

[70] Inactive RS: As used herein, the term "inactive RS" refers to a synthetase that has been mutated so that it no longer can aminoacylate its natural cognate tRNA with an amino acid.

[71] 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.

[72] 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).

[73] 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 fluorescent screening marker (e.g., green fluorescent protein), a luminescent marker (e.g., a firefly luciferase protein), an affinity based screening marker, or selectable marker genes such as his3, ura3, leu2, lys2, lacZ, β-gal/lacZ (β-galactosidase), Adh (alcohol dehydrogenase), or the like.

[74] Vertebrate: As used herein, the term "vertebrate" refers to organisms belonging to the phylogenetic domain Eucarya such as animals e.g., mammals, reptiles, birds, etc.

[75] Non-eukaryote: As used herein, the term "non-eukaryote" refers to non-vertebrate organisms. For example, a non-vertebrate organism can belong to the Eubacteria (e.g., *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, 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.

[76] Antibody: 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.

[77] 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 an unnatural amino acid, although the O-tRNA and the conservative variant O-tRNA 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.

[78] 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 (e.g., a transcriptional modulator protein), or the like. The selection agent can be varied, e.g., by concentration, intensity, etc.

[79] Detectable substance: The term "detectable substance," as used herein, refers to an agent that, when activated, altered, expressed or the like, allows for the selection/screening of certain components from a population. For example, the detectable substance can be a chemical agent, e.g., 5-fluroorotic acid (5-FOA), which under certain conditions, e.g., expression of a URA3 reporter, becomes detectable, e.g., a toxic product that kills cells that express the URA3 reporter.

[80] The ability to genetically modify the structures of proteins directly in vertebrate cells, beyond the chemical constraints imposed by the genetic code, would provides a powerful molecular tool to both probe and manipulate cellular processes. The invention provides translational components that expand the number of genetically encoded amino acids in vertebrate cells. These include tRNA's (e.g., orthogonal tRNA's (O-tRNA's)), aminoacyl-tRNA synthetases (e.g., orthogonal synthetase (O-RS)), pairs of O-tRNA/O-RSs, and unnatural amino acids.

[81] Typically, O-tRNA's of the invention are expressed and processed efficiently, and function in translation in a vertebrate cell, but are not significantly aminoacylated by the host's aminoacyl-tRNA synthetases. In response to a selector codon, an O-tRNA of the invention delivers an unnatural amino acid, which does not encode any of the common twenty amino acids, to a growing polypeptide chain during mRNA translation.

[82] An O-RS of the invention preferentially aminoacylates an O-tRNA of the invention with an unnatural amino acid in a vertebrate cell, but does not aminoacylate any of the cytoplasmic host's tRNA's. Moreover, the specificity of an aminoacyl-tRNA synthetase of the invention provides acceptance of an unnatural amino acid while excluding any endogenous amino acids. Polypeptides that include amino acid sequences of example O-RSs, or portions thereof, are also a feature of the invention. In addition, polynucleotides that encode translational components, O-tRNA's, O-RSs and portions thereof, are features of the invention.

[83] The invention also provides methods of producing the desired translational components, e.g., O-RS, and or an orthogonal pair (orthogonal tRNA and orthogonal aminoacyl-tRNA synthetase), that utilizes an unnatural amino acid for use in a vertebrate cell (and translational components produced by such methods). For example, a tyrosyl-tRNA synthetase/$tRNA_{CUA}$ pair from *E. coli* is an O-tRNA/O-RS pair of the invention. In addition, the invention also features methods of selecting/screening translational components in one vertebrate cell, and once selected/screened, using those components in a different vertebrate cell (a vertebrate cell that was not used for selection/screening). For example, the selection/screening methods to produce the translation components for vertebrate cells can be done in yeast, e.g., *Saccharomyces cerevisiae,* and then those selected components can be used in another vertebrate cell, e.g., another yeast cell, a mammalian cell, an insect cell, a plant cell, a fungus cell, etc.

[84] The invention further provides methods for producing a protein in a vertebrate cell, where the protein comprises an unnatural amino acid. The protein is produced using the translation components of the invention. The invention also provides proteins (and proteins produced by the methods of the invention), which include unnatural amino acids. The protein or polypeptide of interest can also include a post-translational modification, e.g., that is added through a [3+2] cycloaddition, or a nucleophilic-electrophilic reaction, that is not made by a prokaryotic cell, etc. In certain embodiments,

methods of producing a transcriptional modulator protein with an unnatural amino acid (and proteins produced by such methods) are also included in the invention. Compositions, which include proteins that include an unnatural amino acid is also a feature of the invention.

[85] Kits for producing a protein or polypeptide with an unnatural amino acid are also a feature of the invention.

[86] Orthogonal aminoacyl-TRNA synthetases (O-RS)

[87] In order to specifically incorporate an unnatural amino acid in to a protein or polypeptide of interest, in a vertebrate cell, the substrate specificity of the synthetase is altered so that only the desired unnatural amino acid, but not any of the common 20 amino acids are charged to the tRNA. If the orthogonal synthetase is promiscuous, it will result in mutant proteins with a mixture of natural and unnatural amino acids at the target position. The invention provides compositions of, and methods of, producing orthogonal aminoacyl-tRNA synthetases that have modified substrate specificity for a specific unnatural amino acid.

[88] A vertebrate cell that includes an orthogonal aminoacyl-tRNA synthetase (O-RS) is a feature of the invention. The O-RS preferentially aminoacylates an orthogonal tRNA (O-tRNA) with an unnatural amino acid in the vertebrate cell. In certain embodiments, the O-RS utilizes more than one unnatural amino acid, e.g., two or more, three or more, etc. Thus, an O-RS of the invention can have the capability to preferentially aminoacylate an O-tRNA with different unnatural amino acids. This allows an additional level of control by selecting which unnatural amino acid or combination of unnatural amino acids are put with the cell and/or by selecting the different amounts of unnatural amino acids that are put with the cell for their incorporation.

[89] An O-RS of the invention optionally has one or more improved or enhanced enzymatic properties for the unnatural amino acid as compared to a natural amino acid. These properties include, e.g., higher Km, lower Km, higher kcat, lower kcat, lower kcat/km, higher kcat/km, etc., for the unnatural amino acid, as compared to a naturally occurring amino acid, e.g., one of the 20 known common amino acids.

[90] Optionally, the O-RS can be provided to the vertebrate cell by a polypeptide that includes an O-RS and/or by a polynucleotide that encodes an O-RS or a portion thereof.

For example, an O-RS, or a portion thereof, is encoded by a polynucleotide sequence as set forth in any one of SEQ ID NO.: 3-35, or a complementary polynucleotide sequence thereof. In another example, an O-RS comprises an amino acid sequence as set forth in any one of SEQ ID NO.: 36-63, and/or 86, or a conservative variation thereof. *See, e.g.,* Tables 5, 6 and 8, and Example 6 herein for sequences of exemplary O-RS molecules.

[91] An O-RS can also comprise an amino acid sequence that is, e.g., at least 90%, at least 95%, at least 98%, at least 99%, or even at least 99.5 % identical to that of a naturally occurring tyrosyl aminoacyl-tRNA synthetase (TyrRS) (e.g., as set forth in SEQ ID NO.:2) and comprises two or more amino acids of group A-E. Group A includes valine, isoleucine, leucine, glycine, serine, alanine, or threonine at a position corresponding to Tyr37 of *E. coli* TyrRS; group B includes aspartate at a position corresponding to Asn126 of *E. coli* TyrRS; group C includes threonine, serine, arginine, asparagine or glycine at a position corresponding to Asp182 of *E. coli* TyrRS; group D includes methionine, alanine, valine, or tyrosine at a position corresponding to Phe183 of *E. coli* TyrRS; and, group E includes serine, methionine, valine, cysteine, threonine, or alanine at a position corresponding to Leu186 of *E. coli* TyrRS. *See also, e.g.,* Table 4, Table 6 and Table 8, herein.

[92] Besides the O-RS, a vertebrate cell of the invention can include additional components, e.g., an unnatural amino acid(s). The vertebrate cell also includes an orthogonal tRNA (O-tRNA) (e.g., derived from a non-vertebrate organism, such as *Escherichia coli, Bacillus stearothermophilus,* and/or the like), where the O-tRNA recognizes a selector codon and is preferentially aminoacylated with the unnatural amino acid by the O-RS. A nucleic acid that comprises a polynucleotide that encodes a polypeptide of interest, wherein the polynucleotide comprises a selector codon that is recognized by the O-tRNA, or a combination of one or more of these, can also be present in the cell.

[93] In one aspect, the O-tRNA mediates the incorporation of the unnatural amino acid into a protein with, e.g., at least 45%, at least 50%, at least 60%, at least 75%, at least 80%, at least 90%, at least 95%, or 99% or the efficiency of as a tRNA that comprises or is processed from a polynucleotide sequence as set forth in SEQ ID NO.: 65. In another aspect, the O-tRNA comprises SEQ ID NO.:65, and the O-RS comprises a polypeptide sequence set forth in any one of SEQ ID NO.: 36-63, and/or 86, and/or a conservative

variation thereof. *See also*, e.g., Table 5 and Example 6, herein, for sequences of exemplary O-RS and O-tRNA molecules.

[94] In one example, a vertebrate cell comprises an orthogonal aminoacyl-tRNA synthetase (O-RS), an orthogonal tRNA (O-tRNA), an unnatural amino acid, and a nucleic acid that comprises a polynucleotide that encodes a polypeptide of interest, which polynucleotide comprises a selector codon that is recognized by the O-tRNA. The O-RS preferentially aminoacylates the orthogonal tRNA (O-tRNA) with the unnatural amino acid in the vertebrate cell, and the cell produces the polypeptide of interest in the absence of the unnatural amino acid with a yield that is, e.g., less than 30%, less than 20%, less than 15%, less than 10%, less than 5%, less than 2.5%, etc., of the yield of the polypeptide in the presence of the unnatural amino acid.

[95] Methods for producing an O-RS, which are a feature of the invention, optionally include generating a pool of mutant synthetases from the framework of a wild-type synthetase, and then selecting for mutated RSs based on their specificity for an unnatural amino acid relative to the common twenty amino acids. To isolate such a synthetase, the selection methods of the are: (i) sensitive, as the activity of desired synthetases from the initial rounds can be low and the population small; (ii) "tunable", since it is desirable to vary the selection stringency at different selection rounds; and, (iii) general, so that the methods can be used for different unnatural amino acids.

[96] Methods of producing an orthogonal aminoacyl-tRNA synthetase (O-RS) that preferentially aminoacylates an orthogonal tRNA with an unnatural amino acid in a vertebrate cell typically include applying a combination of a positive selection followed by a negative selection. In the positive selection, suppression of the selector codon introduced at nonessential position(s) of a positive marker allows the vertebrate cells to survive under positive selection pressure. In the presence of unnatural amino acids, survivors thus encode active synthetases charging the orthogonal suppressor tRNA with an unnatural amino acid. In the negative selection, suppression of a selector codon introduced at nonessential 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 (or at least preferentially).

[97] For example, the method includes: (a) subjecting to positive selection, in the presence of an unnatural amino acid, a population of vertebrate cells of a first species, where the vertebrate cells each comprise: i) a member of a library of aminoacyl-tRNA synthetases (RSs), ii) an orthogonal tRNA (O-tRNA), iii) a polynucleotide that encodes a positive selection marker, and iv) a polynucleotide that encodes a negative selection marker; wherein cells that survive the positive selection comprise an active RS that aminoacylates the orthogonal tRNA (O-tRNA) in the presence of an unnatural amino acid; and, (b) subjecting the cells that survive the positive selection to negative selection in the absence of the unnatural amino acid to eliminate active RSs that aminoacylate the O-tRNA with a natural amino acid, thereby providing the O-RS that preferentially aminoacylates the O-tRNA with the unnatural amino acid.

[98] The positive selection marker can be any of a variety of molecules. In one embodiment, the positive selection marker is 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-β-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.

[99] 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.

[100] 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.[comment]*, 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.[comment]*, 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. In one embodiment, 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.

[101] 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.

[102] 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 vertebrate regulatory protein*, Science 231:699-704; and, M. Ptashne, (1988), *How vertebrate 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.

[103] 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.[comment]*, 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.[comment]*, Proc. Natl. Acad. Sci. U. S. A. 93:10315-10320.

[104] As with the positive selection marker, the negative selection marker can also be any of a variety of molecules. In one embodiment, the positive selection marker and/or the negative selection marker is 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. In one aspect of the invention, the positive selection marker and/or the negative selection marker is detected by fluorescence-activated cell sorting (FACS) or by luminescence. In another example, the positive selection marker and/or negative selection marker comprise an affinity based screening marker. The same polynucleotide can encode both the positive selection marker and the negative selection marker.

[105] Additional levels of selection/screening stringency can also be used in the methods of the invention. 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 response elements in a polynucleotide that encodes the positive and/or negative selection marker, adding a varying amount of an inactive synthetase to one or both of the steps, varying the amount of selection/screening agent that is used, etc. Additional rounds of positive and/or negative selections can also be performed.

[106] 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 polynucleotides that comprise or encode components of an orthogonal tRNA-tRNA synthetase pair that are used to produce protein.

[107] Model enrichment studies can also be used to rapidly select an active synthetase from an excess of inactive synthetases. Positive and/or negative model selection studies can be done. For example, vertebrate cells that comprise potential active aminoacyl-

tRNA synthetases are mixed with a varying fold excess of inactive aminoacyl-tRNA synthetases. A ratio comparison is made between cells grown in a nonselective media and assayed by, e.g., X-GAL overlay, and those grown and able to survive in a selective media (e.g., in the absence of histidine and/or uracil) and assayed by, e.g., an X-GAL assay. For a negative model selection, potential active aminoacyl-tRNA synthetases are mixed with a varying fold excess of inactive aminoacyl-tRNA synthetases and selection is performed with a negative selection substance, e.g., 5-FOA.

[108] Typically, the library of RSs (e.g., a library of mutant RSs) comprises RSs derived from at least one aminoacyl-tRNA synthetase (RS), e.g., from a non-vertebrate organism. In one embodiment, the library of RSs is derived from an inactive RS, e.g., where the inactive RS is generated by mutating an active RS, 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. For example, residues in the active site of the RS are mutated to, e.g., alanine residues. The polynucleotide that encodes the alanine mutated RS is used as a template to mutagenize the alanine residues to all 20 amino acids. The library of mutant RSs is selected/screened to produce the O-RS. In another embodiment, the inactive RS comprises an amino acid binding pocket and one or more amino acids that comprise the binding pocket are substituted with one or more different amino acids. In one example, the substituted amino acids are substituted with alanines. Optionally, the polynucleotide that encodes the alanine mutated RS is used as a template to mutagenize the alanine residues to all 20 amino acids and screened/selected.

[109] The method of producing an O-RS can further include producing the library of RSs by 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." Once the synthetases are subjected 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 invention, the steps are performed at least two times.

[110] Additional details for producing O-RS can be found in WO 2002/086075 entitled "Methods and compositions for the production of orthogonal tRNA-aminoacyltRNA synthetase pairs." *See also*, 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 vertebrate 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.

[111] Orthogonal tRNA's

[112] Eukaryotic cells that include an orthogonal tRNA (O-tRNA) are provided by the invention. The orthogonal tRNA mediates incorporation of an unnatural amino acid into a protein that is encoded by a polynucleotide that comprises a selector codon that is recognized by the O-tRNA, in vivo. In certain embodiments, an O-tRNA of the invention mediates the incorporation of an unnatural amino acid into a protein with, e.g., at least 40%, at least 45%, at least 50%, at least 60%, at least 75%, at least 80%, or even 90% or more as efficiently as tRNA that comprises or is processed in a cell from a polynucleotide sequence as set forth in SEQ ID NO.: 65. *See*, Table 5, herein.

[113] An example of an O-tRNA of the invention is SEQ ID NO.: 65. (*See* Example 6 and Table 5, herein). SEQ ID NO.: 65 is a pre-splicing/processing transcript that is optionally processed in the cell, e.g., using the standard endogenous cellular splicing and processing machinery, and modified to form an active O-tRNA. Typically, a population of such pre-splicing transcripts forms a population of active tRNA's in the cell. The invention also includes conservative variations of the O-tRNA and its processed cellular products. For example, conservative variations of O-tRNA include those molecules that function like the O-tRNA of SEQ ID NO.:65 and maintain the tRNA L-shaped structure in processed form, but do not have the same sequence (and are other than wild type tRNA molecules). Typically, an O-tRNA of the invention is a recyclable O-tRNA, because the O-tRNA can be

reaminoacylated in vivo to again mediate the incorporation of the unnatural amino acid into a protein that is encoded by a polynucleotide in response to a selector codon.

[114] The transcription of the tRNA in eukaryotes, but not in prokaryotes, is carried out by RNA Polymerase III, which places restrictions on the primary sequence of the tRNA structural genes that can be transcribed in vertebrate cells. In addition, in vertebrate cells, tRNA's need to be exported from the nucleus, where they are transcribed, to the cytoplasm, to function in translation. Nucleic acids that encode an O-tRNA of the invention or a complementary polynucleotide thereof are also a feature of the invention. In one aspect of the invention, a nucleic acid that encodes an O-tRNA of the invention includes an internal promoter sequence, e.g., an A box (e.g., TRGCNNAGY) and a B box (e.g., GGTTCGANTCC, SEQ ID NO: 87). Additional examples of A box and B box sequences can be found in Geiduschek, (1988), *Transcription By RNA Polymerase III,* Ann. Rev. Biochem. 57:873-914. The O-tRNA of the invention can also be post-transcriptionally modified. For example, post-transcriptional modification of tRNA genes in eukaryotes includes removal of the 5'- and 3'- flanking sequences by Rnase P and a 3'- endonuclease, respectively. The addition of a 3'- CCA sequence is also a post-transcriptional modification of a tRNA gene in eukaryotes.

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

[116] Alternatively, or in combination with others methods described above to incorporate an unnatural amino acid into a polypeptide, a trans-translation system can be used. This system involves a molecule called tmRNA present in *Escherichia coli.* This RNA molecule is structurally related to an alanyl tRNA and is aminoacylated by the alanyl synthetase. The difference between tmRNA and tRNA is that the anticodon loop is replaced with a special large sequence. This sequence allows the ribosome to resume translation on sequences that have stalled using an open reading frame encoded within the tmRNA as template. In the invention, an orthogonal tmRNA can be generated that is preferentially aminoacylated with an orthogonal synthetase and loaded with an unnatural amino acid. By

transcribing a gene by the system, the ribosome stalls at a specific site; the unnatural amino acid is introduced at that site, and translation resumes using the sequence encoded within the orthogonal tmRNA.

[117] Additional methods for producing a recombinant orthogonal tRNA's can be found, e.g., in International patent applications WO 2002/086075, entitled "Methods and compositions for the production of orthogonal tRNA-aminoacyltRNA synthetase pairs." *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.

[118] Orthogonal TRNA and Orthogonal aminoacyl-TRNA synthetase pairs

[119] An orthogonal pair is composed of an O-tRNA, e.g., a suppressor tRNA, a frameshift tRNA, or the like, and an O-RS. The O-tRNA is not acylated by endogenous synthetases and is capable of mediating incorporation of an unnatural amino acid into a protein that is encoded by a polynucleotide that comprises a selector codon that is recognized by the O-tRNA in vivo. The O-RS recognizes the O-tRNA and preferentially aminoacylates the O-tRNA with an unnatural amino acid in a vertebrate cell. Methods for producing orthogonal pairs along with orthogonal pairs produced by such methods and compositions of orthogonal pairs for use in vertebrate cells are included in the invention. The development of multiple orthogonal tRNA/synthetase pairs can allow the simultaneous incorporation of multiple unnatural amino acids using different codons in a vertebrate cell.

[120] An orthogonal O-tRNA/O-RS pair in a vertebrate cell can be produced by importing a pair, e.g., a nonsense suppressor pair, from a different organism with inefficient cross species aminoacylation. The O-tRNA and O-RS are efficiently expressed and processed in the vertebrate cell and the O-tRNA is efficiently exported from the nucleus to the cytoplasm. For example, one such pair is the tyrosyl-tRNA synthetase/$tRNA_{CUA}$ pair from *E. coli (see, e.g.,* H. M. Goodman, et al., (1968), Nature 217:1019-24; and, D. G. Barker, et al., (1982), FEBS Letters 150:419-23). *E. coli* tyrosyl-tRNA synthetase efficiently aminoacylates its cognate *E. coli* $tRNA_{CUA}$ when both are expressed in the cytoplasm of *S. cerevisiae*, but does not aminoacylate *S. cerevisiae* tRNA's. *See, e.g.,* H. Edwards, & P. Schimmel, (1990), Molecular & Cellular Biology 10:1633-41; and, H. Edwards, et al., (1991), PNAS United States of America 88:1153-6. In addition, *E. coli*

tyrosyl $tRNA_{CUA}$ is a poor substrate for *S. cerevisiae* aminoacyl-tRNA synthetases (*see, e.g.*, V. Trezeguet, et al., (1991), Molecular & Cellular Biology 11:2744-51), but functions efficiently in protein translation in *S. cerevisiae*. *See, e.g.*, H. Edwards, & P. Schimmel, (1990) Molecular & Cellular Biology 10:1633-41; H. Edwards, et al., (1991), PNAS United States of America 88:1153-6; and, V. Trezeguet, et al., (1991), Molecular & Cellular Biology 11:2744-51. Moreover, *E. coli* TyrRS does not have an editing mechanism to proofread an unnatural amino acid ligated to the tRNA.

[121] The O-tRNA and O-RS can be naturally occurring or can be derived by mutation of a naturally occurring tRNA and/or RS, which generates libraries of tRNA's and/or libraries of RSs, from a variety of organism. See the section entitled "Sources and Hosts" herein. 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 non-vertebrate organisms are the same. Alternatively, the first and second non-vertebrate organisms can be different.

[122] See sections herein entitled "Orthogonal aminoacyl-tRNA synthetases" and "O-tRNA" for methods of producing O-RSs and O-tRNA's. *See also*, International patent application WO 2002/086075, entitled "Methods and compositions for the production of orthogonal tRNA-aminoacyltRNA synthetase pairs."

[123] Fidelity, Efficiency, and Yield

[124] Fidelity refers to the accuracy with which a desired molecule, e.g., an unnatural amino acid or amino acid, is incorporated into a growing polypeptide at a desired position. The translational components of the invention incorporate unnatural amino acids, with high fidelity, into proteins in response to a selector codon. For example, using the components of the invention, the efficiency of incorporation of a desired unnatural amino acid into a growing polypeptide chain at a desired position (e.g., in response to a selector codon) is, e.g., greater than 75%, greater than 85%, greater than 95%, or even greater than 99% or more as efficient as compared to unwanted incorporation a specific natural amino acid being incorporated into the growing polypeptide chain the desired position.

[125] Efficiency can also refer to the degree with which the O-RS aminoacylates the O-tRNA with the unnatural amino acid as compared to a relevant control. O-RSs of the

invention can be defined by their efficiency. In certain embodiments of the invention, an O-RS is compared to another O-RS. For example, a O-RS of the invention aminoacylates a O-tRNA with an unnatural amino acid, e.g., at least 40%, at least 50%, at least 60%, at least 75%, at least 80%, at least 90%, at least 95%, or even 99% or more as efficiently as an O-RS having an amino acid sequence, e.g., as set forth in SEQ ID NO.: 86 or 45)or another specific RS in Table 5) aminoacylates an O-tRNA. In another embodiment, an O-RS of the invention aminoacylates the O-tRNA with the unnatural amino acid at least 10-fold, at least 20-fold, at least 30-fold, etc., more efficiently than the O-RS aminoacylates the O-tRNA with a natural amino acid.

[126] Using the translational components of the invention, the yield of the polypeptide of interest comprising the unnatural amino acid is, e.g., at least 5%, at least 10%, at least 20%, at least 30%, at least 40%, 50% or more, of that obtained for the naturally occurring polypeptide of interest from a cell in which the polynucleotide lacks the selector codon. In another aspect, the cell produces the polypeptide of interest in the absence of the unnatural amino acid with a yield that is, e.g., less than 30%, less than 20%, less than 15%, less than 10%, less than 5%, less than 2.5%, etc., of the yield of the polypeptide in the presence of the unnatural amino acid.

[127] Source and Host Organisms

[128] The orthogonal translational components of the invention are typically derived from non-vertebrate organisms for use in vertebrate cells or translation systems. For example, the orthogonal O-tRNA can be derived from a non-vertebrate organism, e.g., a eubacterium, such as *Escherichia coli, Thermus thermophilus, Bacillus stearothermphilus,* or the like, or an archaebacterium, 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, while the orthogonal O-RS can be derived from a non-vertebrate organism, e.g., a eubacterium, such as *Escherichia coli, Thermus thermophilus, Bacillus stearothermphilus,* or the like, or an archaebacterium, 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. Alternately, vertebrate sources can also be used, e.g., plants, algae, protists, fungi, yeasts, animals (e.g., mammals, insects,

arthropods, etc.), or the like, e.g., where the components are orthogonal to a cell or translation system of interest, or where they are modified (e.g., mutated) to be orthogonal to the cell or translation system.

[129] 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. For example, the O-tRNA/O-RS pair can be derived from a tyrosyl-tRNA synthetase/$tRNA_{CUA}$ pair from *E. coli.* Alternatively, the O-tRNA and the O-RS of the O-tRNA/O-RS pair are optionally from different organisms.

[130] The orthogonal O-tRNA, O-RS or O-tRNA/O-RS pair can be selected or screened and/or used in a vertebrate cell to produce a polypeptide with an unnatural amino acid. A vertebrate cell can be from a variety of sources, e.g., any vertebrate animal (e.g., a mammal, an amphibian, birds, reptiles, fish, etc.), or the like. Compositions of vertebrate cells with translational components of the invention are also a feature of the invention.

[131] The invention also provides for the efficient screening in one species for optional use in that species and/or a second species (optionally, without additional selection/screening). For example, the components of the O-tRNA/O-RS are selected or screened in one species, e.g., an easily manipulated species (such as a yeast cell, etc.) and introduced into a second vertebrate species, e.g., a plant (e.g., complex plant such as monocots, or dicots), an algae, a protist, a fungus, a yeast, an animal (e.g., a mammal, an insect, an arthropod, etc.), or the like, for use in the in vivo incorporation of an unnatural amino acid in the second species.

[132] For example, *Saccharomyces cerevisiae* (*S. cerevisiae*) can be chosen as the vertebrate first species, as it is unicellular, has a rapid generation time, and relatively well-characterized genetics. *See, e.g.,* D. Burke, et al., (2000) Methods in Yeast Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. Moreover, since the translational machinery of eukaryotes is highly conserved (*see, e.g.,* (1996) Translational Control. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; Y. Kwok, & J.T. Wong, (1980), *Evolutionary relationship between Halobacterium cutirubrum and eukaryotes determined by use of aminoacyl-tRNA synthetases as phylogenetic probes*, Canadian Journal of Biochemistry 58:213-218; and, (2001) The Ribosome. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY), aaRSs genes for the incorporation of unnatural amino acids discovered in *S. cerevisiae* can be introduced into higher vertebrate organisms and

used, in partnership with cognate tRNA's (*see, e.g.*, K. Sakamoto, et al., (2002) *Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells,* Nucleic Acids Res. 30:4692-4699; and, C. Kohrer, et al., (2001), *Import of amber and ochre suppressor tRNA's into mammalian cells: a general approach to site-specific insertion of amino acid analogues into proteins*, Proc. Natl. Acad. Sci. U. S. A. 98:14310-14315) to incorporate unnatural amino acids.

[133] In one example, the method of producing O-tRNA/O-RS in a first species as described herein further includes introducing a nucleic acid that encodes the O-tRNA and a nucleic acid that encodes the O-RS into a vertebrate cell of a second species (e.g., a mammal, an insect, a fungus, an algae, a plant and the like). In another example, a method of producing an orthogonal aminoacyl-tRNA synthetase (O-RS) that preferentially aminoacylates an orthogonal tRNA with an unnatural amino acid in a vertebrate cell includes: (a) subjecting to positive selection, in the presence of an unnatural amino acid, a population of vertebrate cells of a first species (e.g., yeast and the like). Each of the vertebrate cells comprise: i) a member of a library of aminoacyl-tRNA synthetases (RSs), ii) an orthogonal tRNA (O-tRNA), iii) a polynucleotide that encodes a positive selection marker, and iv) a polynucleotide that encodes a negative selection marker. The cells that survive the positive selection comprise an active RS that aminoacylates the orthogonal tRNA (O-tRNA) in the presence of an unnatural amino acid. The cells that survive the positive selection are subjected to negative selection in the absence of the unnatural amino acid to eliminate active RSs that aminoacylate the O-tRNA with a natural amino acid. This provides an O-RS that preferentially aminoacylates the O-tRNA with the unnatural amino acid. A nucleic acid that encodes the O-tRNA and a nucleic acid that encodes the O-RS (or the components O-tRNA and/or O-RS) are introduced into a vertebrate cell of a second species e.g., a mammal, an insect, a fungus, an algae, a plant and/or the like. Typically, the O-tRNA is obtained by subjecting to negative selection a population of vertebrate cells of a first species, where the vertebrate cells comprise a member of a library of tRNA's. The negative selection eliminates cells that comprise a member of the library of tRNA's that is aminoacylated by an aminoacyl-tRNA synthetase (RS) that is endogenous to the vertebrate cells, which provides a pool of tRNA's that are orthogonal to the vertebrate cell of the first species and the second species.

[134] Selector Codons

[135] Selector codons of the invention expand the genetic codon framework of the protein biosynthetic machinery. For example, a selector codon includes, e.g., a unique three base codon, a nonsense codon, such as a stop codon, e.g., an amber codon (UAG), an opal codon (UGA), an unnatural codon, at least a four base codon, a rare codon, or the like. A number of selector codons can be introduced into a desired gene, e.g., one or more, two or more, more than three, etc. Once gene can include multiple copies of a given selector codon, or can include multiple different selector codons, or any combination thereof.

[136] In one embodiment, the methods involve the use of a selector codon that is a stop codon for the incorporation of unnatural amino acids in vivo in a vertebrate cell. For example, an O-tRNA is produced that recognizes the stop codon, e.g., UAG, and is aminoacylated by an O-RS with a desired unnatural 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, e.g., TAG, at the site of interest in a polypeptide of interest. *See, e.g.,* Sayers, J.R., et al. (1988), *5',3' Exonuclease in phosphorothioate-based oligonucleotide-directed mutagenesis.* Nucleic Acids Res, 791-802. When the O-RS, O-tRNA and the nucleic acid that encodes the polypeptide of interest are combined in vivo, the unnatural amino acid is incorporated in response to the UAG codon to give a polypeptide containing the unnatural amino acid at the specified position.

[137] The incorporation of unnatural amino acids in vivo can be done without significant perturbation of the vertebrate host cell. For example, because the suppression efficiency for the UAG codon depends upon the competition between the O-tRNA, e.g., the amber suppressor tRNA, and a vertebrate 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.

[138] 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, e.g., AGGA, CUAG, UAGA, CCCU and the like. Examples of five base codons include, e.g., AGGAC, CCCCU, CCCUC, CUAGA, CUACU, UAGGC and the like. A feature of the invention includes using extended codons based on frameshift suppression. Four or more base codons can insert, e.g., one or multiple 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 at least 8-10 nt anticodon loops, the four or more base codon is read as single amino acid. 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.

[139] 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 tRNALeu 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 tRNALeu 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.

[140] 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.

[141] 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. Other relevant publications are listed below.

[142] 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., 121:11586; 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 invention can take advantage of this property to generate orthogonal tRNA's for them.

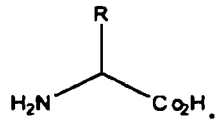
[143] A translational bypassing system can also be used to incorporate an unnatural amino acid in a desired polypeptide. In a translational bypassing system, a large sequence is inserted into a gene but is not translated into protein. The sequence contains a structure that

serves as a cue to induce the ribosome to hop over the sequence and resume translation downstream of the insertion.

[144] Unnatural amino acids

[145] 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:

I



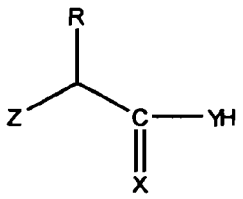
[146] An unnatural amino acid is typically any structure having Formula I wherein the R group is any substituent other than one used in the twenty natural amino acids. See, e.g., *Biochemistry* by L. Stryer, 3rd ed. 1988, Freeman and Company, New York, for structures of the twenty natural amino acids.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

[147] Because the unnatural amino acids of the invention typically differ from the natural amino acids in side chain, the unnatural amino acids form amide bonds with other amino acids, e.g., 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 optionally comprises 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 unnatural amino acids of interest 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,

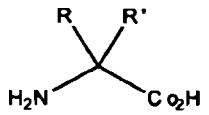
photocaged and/or photoisomerizable amino acids, biotin or biotin-analogue containing amino acids, keto containing amino acids, amino acids comprising polyethylene glycol or polyether, heavy atom substituted amino acids, chemically cleavable or photocleavable amino acids, amino acids with an elongated side chain as compared to natural amino acids (e.g., polyethers or long chain hydrocarbons, e.g., greater than about 5, greater than about 10 carbons, etc.), carbon-linked sugar-containing amino acids, redox-active amino acids, amino thioacid containing amino acids, and amino acids containing one or more toxic moiety. In some embodiments, the unnatural amino acids have a photoactivatable cross-linker that is used, e.g., to link a protein to a solid support. In one embodiment, the unnatural amino acids have a saccharide moiety attached to the amino acid side chain (e.g., glycosylated amino acids) and/or other carbohydrate modification.

[148] 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



[149] wherein Z typically comprises OH, NH_2, 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 of the invention optionally 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.For example, 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, where the substituted tyrosine comprises, e.g., a keto group (e.g., 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_6 - C_{20} straight chain or branched hydrocarbon, a saturated or unsaturated hydrocarbon, an O-methyl group, a polyether group, a nitro group, an alkynyl group or the like. In addition, multiply substituted aryl rings are also contemplated. Glutamine analogs of the invention include, but are not limited to, α-hydroxy derivatives, γ-substituted derivatives, cyclic derivatives, and amide substituted glutamine derivatives. Example phenylalanine analogs include, but are not limited to, para-substituted phenylalanines, ortho-substituted phenyalanines, and meta-substituted phenylalanines, where the substituent comprises, e.g., a hydroxy group, a methoxy group, a methyl group, an allyl group, an aldehyde, an azido, an iodo, a bromo, a keto group (e.g., an acetyl group), a benzoyl, an alkynyl group, or the like. Specific examples of unnatural amino acids include, but are not limited to, a *p*-acetyl-L- phenylalanine, a *p*-propargyloxyphenylalanine, 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-GlcNAcβ-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 phosphonoserine, a phosphonotyrosine, a *p*-iodo-phenylalanine, a *p*-bromophenylalanine, a *p*-amino-L-phenylalanine, and an isopropyl-L-phenylalanine, and the like. Additional 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." See also, figure 1 structures 2-5 of Kiick et al., (2002) *Incorporation of azides into recombinant proteins for chemoselective*

modification by the Staudinger ligtation, PNAS 99:19-24, for additional methionine analogs.

[150] In one embodiment, compositions that include an unnatural amino acid (such as *p*-(propargyloxy)-phenyalanine) are provided. Various compositions comprising *p*-(propargyloxy)-phenyalanine and, e.g., proteins and/or cells, are also provided. In one aspect, a composition that includes the *p*-(propargyloxy)-phenyalanine unnatural amino acid further includes an orthogonal tRNA. The unnatural amino acid can be bonded (e.g., covalently) to the orthogonal tRNA, e.g., covalently bonded to the orthogonal tRNA though an amino-acyl bond, covalently bonded to a 3'OH or a 2'OH of a terminal ribose sugar of the orthogonal tRNA, etc.

[151] The chemical moieties via unnatural amino acids that can be incorporated into proteins offer a variety of advantages and manipulations of the protein. For example, the unique reactivity of a keto functional group allows selective modification of proteins with any of a number of hydrazine- or hydroxylamine-containing reagents in vitro and in vivo. A heavy atom unnatural amino acid, for example, can be useful for phasing x-ray structure data. The site-specific introduction of heavy atoms using unnatural amino acids also provides selectivity and flexibility in choosing positions for heavy atoms. Photoreactive unnatural amino acids (e.g., amino acids with benzophenone and arylazides (e.g., phenylazide) side chains), for example, allow for efficient in vivo and in vitro photocrosslinking of proteins. Examples of photoreactive unnatural amino acids include, but are not limited to, e.g., p-azido-phenylalanine and p-benzoyl-phenylalanine. The protein with the photoreactive unnatural amino acids can then be crosslinked at will by excitation of the photoreactive group-providing temporal (and/or spatial) control. In one example, the methyl group of an unnatural amino can be substituted with an isotopically labeled, e.g., methyl group, as a probe of local structure and dynamics, e.g., with the use of nuclear magnetic resonance and vibrational spectroscopy. Alkynyl or azido functional groups, for example, allow the selective modification of proteins with molecules through a [3+2] cycloaddition reaction.

[152] Chemical Synthesis of Unnatural Amino Acids

[153] Many of the unnatural amino acids provided above are commercially available, e.g., from Sigma (USA) or Aldrich (Milwaukee, WI, USA). Those that are not commercially available are optionally synthesized as provided herein or as provided in

various publications or using standard methods known to those of skill in the art. For organic synthesis techniques, see, e.g., Organic Chemistry by Fessendon and Fessendon, (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 γ-Dipeptides of Glutamic Acid from Phthylated Intermediates.* J. Chem. Soc., 3315-3319; Friedman, O.M. & Chatterrji, 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 Pipecolates from L-Asparagine. Application to the Total Synthesis of (+)-Apovincamine through Amino Acid Decarbonylation and Iminium Ion Cyclization.* J. Org. Chem. 1989:1859-1866; Barton et al., (1987) *Synthesis of Novel a-Amino-Acids and Derivatives Using Radical Chemistry: Synthesis of L- and D-a-Amino-Adipic Acids, L-a-aminopimelic Acid and Appropriate Unsaturated Derivatives.* Tetrahedron Lett. 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.

[154] Cellular uptake of unnatural amino acids

[155] Unnatural amino acid uptake by a vertebrate 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 α-amino acids suggests that these compounds are unlikely to be cell permeable. Natural amino acids are taken up into the vertebrate 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., the application entitled "Protein Arrays," attorney docket number

P1001US00 filed on December 22, 2002; 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*.

[156] Biosynthesis of Unnatural Amino Acids

[157] 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, e.g., in a vertebrate 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 vertebrate 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, e.g., 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.

[158] 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 www.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, e.g., 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, e.g., to create new pathways.

[159] Typically, the unnatural amino acid produced with an engineered biosynthetic pathway of the 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.

[160] Polypeptides with Unnatural Amino Acids

[161] Proteins or polypeptides of interest with at least one unnatural amino acid are a feature of the invention. The invention also includes polypeptides or proteins with at least one unnatural amino acid produced using the compositions and methods of the invention. An excipient (e.g., a pharmaceutically acceptable excipient) can also be present with the protein.

[162] By producing proteins or polypeptides of interest with at least one unnatural amino acid in vertebrate cells, proteins or polypeptides will typically include vertebrate posttranslational 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 vertebrate cell, where the post-translational modification is not made by a prokaryotic cell. For example, the post-translation modification includes, e.g., acetylation, acylation, lipid-modification, palmitoylation, palmitate addition, phosphorylation, glycolipid-linkage modification, glycosylation, and the like. In one aspect, the post-translational modification includes attachment of an oligosaccharide (e.g., $(GlcNAc\text{-}Man)_2\text{-}Man\text{-}GlcNAc\text{-}GlcNAc$) to an asparagine by a GlcNAc-asparagine linkage. See also, Table 7, which lists some examples of N-linked oligosaccharides of vertebrate proteins (additional residues can also be present, which are not shown). In another aspect, the post-translational modification includes attachment of an oligosaccharide (e.g., Gal-GalNAc, Gal-GlcNAc, etc.) to a serine

or threonine by a GalNAc-serine or GalNAc-threonine linkage, or a GlcNAc-serine or a GlcNAc-threonine linkage.

TABLE 7: EXAMPLES OF OLIGOSACCHARIDES THROUGH GlcNAc-LINKAGE

Type	Base Structure
High-mannose	Manα1-6 / Manα1-3 > Manα1-6 / Manα1-3 > Manβ1-4GlcNAcβ1-4GlcNAcβ1-Asn
Hybrid	Manα1-6 / GlcNAcβ1-2 — Manα1-3 > Manβ1-4GlcNAcβ1-4GlcNAcβ1-Asn
Complex	GlcNAcβ1-2 — Manα1-6 / GlcNAcβ1-2 — Manα1-3 > Manβ1-4GlcNAcβ1-4GlcNAcβ1-Asn
Xylose	Manα1-6 / Xylβ1-2 > Manβ1-4GlcNAcβ1-4GlcNAcβ1-Asn

[163] In yet another aspect, the post-translation modification includes proteolytic processing of precursors (e.g., 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 (e.g., 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.

[164] 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 vertebrate cell, or in vitro. Thus, in certain embodiments, the post-

translational modification is through the unnatural amino acid. 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, e.g. the reaction of α-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) 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) 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, in press. 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,* patent application USSN 10/686,944 entitled "Glycoprotein synthesis" filed October 15, 2003. Post-translational modifications, e.g., through an azido amino acid, can also made through the Staudinger ligation (e.g., with triarylphosphine reagents). *See, e.g.,* Kiick et al., (2002) *Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligtation,* PNAS 99:19-24.

[165] This invention provides another highly efficient method for the selective modification of proteins, which involves the genetic incorporation of unnatural amino acids, e.g., containing an azide or alkynyl moiety into proteins in response to a selector codon. These amino acid side chains can then be modified by, e.g., a Huisgen [3+2] cycloaddition reaction (*see, e.g.,* Padwa, A. in Comprehensive Organic Synthesis, Vol. 4, (1991) Ed. Trost, B. M., Pergamon, Oxford, p. 1069-1109; and, Huisgen, R. in 1,3-Dipolar Cycloaddition Chemistry, (1984) Ed. Padwa, A., Wiley, New York, p. 1-176) with, e.g., alkynyl or azide derivatives, respectively. *See, e.g.,* **Figure 16**. Because this method involves a cycloaddition rather than a nucleophilic substitution, proteins can be modified with extremely high selectivity. This reaction can be carried out at room temperature in aqueous conditions with excellent regioselectivity (1,4 > 1,5) by the addition of catalytic amounts of Cu(I) salts to the reaction mixture. *See, e.g.,* Tornoe, et al., (2002) Org. Chem. 67:3057-3064; and, Rostovtsev, et al., (2002) Angew. Chem. Int. Ed. 41:2596-2599.

Another method that can be used is the ligand exchange on a bisarsenic compound with a tetracysteine motif, *see, e.g.,* Griffin, et al., (1998) Science 281:269-272.

[166] A molecule that can be added to a protein of the invention through a functional group of a non-naturally encoded amino acid includes virtually any molecule with complementary functional group. Such molecules include, but are not limited to, dyes, fluorophores, crosslinking agents, saccharide derivatives, polymers (e.g., derivatives of polyethylene glycol), photocrosslinkers, cytotoxic compounds, affinity labels, derivatives of biotin, resins, beads, a second protein or polypeptide (or more), polynucleotide(s) (e.g., DNA, RNA, etc.), metal chelators, cofactors, fatty acids, carbohydrates, and the like.

In another aspect, the invention provides compositions including such molecules and methods of producing these molecules, e.g., polyethylene glycol derivatives, where n is an integer between, e.g., 50 and 10,000, 75 and 5,000, 100 and 2,000, 100 and 1,000, etc. In embodiment of the invention, the polyethylene glycol has a molecular weight of, e.g., about 5,000 to about 100,000 Da, about 20,000 to about 30,000, about 40,000, or about 50, 000 Da, about 20,000 to about 10,000 Da, etc.

[167] Various compositions comprising these compounds, e.g., with proteins and cells, are also provided. In one aspect of the invention, a protein comprising an azido dye (e.g., of chemical structure 4 or chemical structure 6), further includes at least one unnatural amino acid (e.g., an alkynyl amino acid), where the azido dye is attached to the unnatural amino acid through a [3+2] cycloaddition.

[168] A vertebrate cell of the invention provides the ability to synthesize proteins that comprise unnatural amino acids in large useful quantities. In one aspect, the composition optionally includes, e.g., at least 10 micrograms, at least 50 micrograms, at least 75 micrograms, at least 100 micrograms, at least 200 micrograms, at least 250 micrograms, at least 500 micrograms, at least 1 milligram, at least 10 milligrams or more of the protein that comprises an unnatural amino acid, or an amount that can be achieved with in vivo protein production methods (details on recombinant protein production and purification are provided herein). In another aspect, the protein is optionally present in the composition at a concentration of, e.g., at least 10 micrograms of protein per liter, at least 50 micrograms of protein per liter, at least 75 micrograms of protein per liter, at least 100 micrograms of protein per liter, at least 200 micrograms of protein per liter, at least 250 micrograms of protein per liter, at least 500 micrograms of protein per liter, at least 1

milligram of protein per liter, or at least 10 milligrams of protein per liter or more, in, e.g., a cell lysate, a buffer, a pharmaceutical buffer, or other liquid suspension (e.g., in a volume of, e.g., anywhere from about 1 nl to about 100 L). The production of large quantities (e.g., greater that that typically possible with other methods, e.g., in vitro translation) of a protein in a vertebrate cell including at least one unnatural amino acid is a feature of the invention.

[169] The incorporation of an unnatural amino acid can be done to, e.g., tailor changes in protein structure and/or function, e.g., to change size, acidity, nucleophilicity, hydrogen bonding, hydrophobicity, accessibility of protease target sites, target to a moiety (e.g., for a protein array), etc. Proteins that include an unnatural amino acid can have enhanced or even entirely new catalytic or physical 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 (e.g., serum half-life), ability to react with other molecules, e.g., covalently or noncovalently, and the like. The compositions including proteins that include at least one unnatural amino acid are useful for, e.g., novel therapeutics, diagnostics, catalytic enzymes, industrial enzymes, binding proteins (e.g., antibodies), and e.g., 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.

[170] In one aspect of the invention, a composition includes at least one protein with at least one, e.g., 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, e.g., 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. In another aspect, a composition includes a protein with 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 acid, the unnatural amino acids can be identical or different (e.g., 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.

[171] Essentially any protein (or portion thereof) that includes 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. 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.

[172] 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 unnatural amino acid. Examples of therapeutic, diagnostic, and other proteins that can be modified to comprise one or more unnatural amino acids include, but are not limited to, e.g., Alpha-1 antitrypsin, Angiostatin, Antihemolytic factor, antibodies (further details on antibodies are found below), Apolipoprotein, Apoprotein, Atrial natriuretic factor, Atrial natriuretic polypeptide, Atrial peptides, C-X-C chemokines (e.g., T39765, NAP-2, ENA-78, Gro-a, Gro-b, Gro-c, IP-10, GCP-2, NAP-4, SDF-1, PF4, MIG), Calcitonin, CC chemokines (e.g., Monocyte chemoattractant protein-1, Monocyte chemoattractant protein-2, Monocyte chemoattractant protein-3, Monocyte inflammatory protein-1 alpha, Monocyte inflammatory protein-1 beta, RANTES, I309, R83915, R91733, HCC1, T58847, D31065, T64262), CD40 ligand, C-kit Ligand, Collagen, Colony stimulating factor (CSF), Complement factor 5a, Complement inhibitor, Complement receptor 1, cytokines, (e.g., epithelial Neutrophil Activating Peptide-78, GROα/MGSA, GROβ, GROγ, MIP-1α, MIP-1δ, MCP-1), Epidermal Growth Factor (EGF), Erythropoietin ("EPO", representing a preferred target for modification by the incorporation of one or more unnatural amino acid), Exfoliating toxins A and B, Factor IX, Factor VII, Factor VIII, Factor X, Fibroblast Growth Factor (FGF), Fibrinogen, Fibronectin, G-CSF, GM-CSF, Glucocerebrosidase, Gonadotropin, growth factors, Hedgehog proteins (e.g., Sonic, Indian, Desert), Hemoglobin, Hepatocyte Growth Factor (HGF), Hirudin, Human serum albumin, Insulin, Insulin-like Growth Factor (IGF), interferons (e.g., IFN-α, IFN-β, IFN-γ), interleukins (e.g., IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, etc.), Keratinocyte Growth Factor (KGF), Lactoferrin, leukemia inhibitory factor, Luciferase, Neurturin, Neutrophil inhibitory factor (NIF), oncostatin M, Osteogenic

protein, Parathyroid hormone, PD-ECSF, PDGF, peptide hormones (e.g., Human Growth Hormone), Pleiotropin, Protein A, Protein G, Pyrogenic exotoxins A, B, and C, Relaxin, Renin, SCF, Soluble complement receptor I, Soluble I-CAM 1, Soluble interleukin receptors (IL-1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15), Soluble TNF receptor, Somatomedin, Somatostatin, Somatotropin, Streptokinase, Superantigens, i.e., Staphylococcal enterotoxins (SEA, SEB, SEC1, SEC2, SEC3, SED, SEE), Superoxide dismutase (SOD), Toxic shock syndrome toxin (TSST-1), Thymosin alpha 1, Tissue plasminogen activator, Tumor necrosis factor beta (TNF beta), Tumor necrosis factor receptor (TNFR), Tumor necrosis factor-alpha (TNF alpha), Vascular Endothelial Growth Factor (VEGEF), Urokinase, and many others.

[173] One class of proteins that can be made using the compositions and methods for in vivo incorporation of unnatural amino acids described herein includes transcriptional modulators or portions thereof. Example transcriptional modulators include genes and transcriptional modulator proteins that modulate cell growth, differentiation, regulation, or the like. Transcriptional modulators are found in prokaryotes, viruses, and eukaryotes, including fungi, plants, yeasts, insects, and animals, including mammals, providing a wide range of therapeutic targets. It will be appreciated that expression and transcriptional activators regulate transcription by many mechanisms, e.g., by binding to receptors, stimulating a signal transduction cascade, regulating expression of transcription factors, binding to promoters and enhancers, binding to proteins that bind to promoters and enhancers, unwinding DNA, splicing pre-mRNA, polyadenylating RNA, and degrading RNA. For example, compositions of GAL4 protein or portion thereof in a vertebrate cell are also a feature of the invention. Typically, the GAL4 protein or portion thereof comprises at least one unnatural amino acid. See also the section herein entitled "Orthogonal aminoacyl-tRNA synthetases."

[174] One class of proteins of the invention (e.g., proteins with one or more unnatural amino acids) include expression activators such as cytokines, inflammatory molecules, growth factors, their receptors, and oncogene products, e.g., interleukins (e.g., IL-1, IL-2, IL-8, etc.), interferons, FGF, IGF-I, IGF-II, FGF, PDGF, TNF, TGF-α, TGF-β, EGF, KGF, SCF/c-Kit, CD40L/CD40, VLA-4/VCAM-1, ICAM-1/LFA-1, and hyalurin/CD44; signal transduction molecules and corresponding oncogene products, e.g., Mos, Ras, Raf, and Met; and transcriptional activators and suppressors, e.g., p53, Tat, Fos,

Myc, Jun, Myb, Rel, and steroid hormone receptors such as those for estrogen, progesterone, testosterone, aldosterone, the LDL receptor ligand and corticosterone.

[175] Enzymes (e.g., industrial enzymes), or portions thereof with at least one unnatural amino acid, are also provided by the invention. Examples of enzymes include, but are not limited to, e.g., amidases, amino acid racemases, acylases, dehalogenases, dioxygenases, diarylpropane peroxidases, epimerases, epoxide hydrolases, esterases, isomerases, kinases, glucose isomerases, glycosidases, glycosyl transferases, haloperoxidases, monooxygenases (e.g., p450s), lipases, lignin peroxidases, nitrile hydratases, nitrilases, proteases, phosphatases, subtilisins, transaminase, and nucleases.

[176] Many of these proteins are commercially available (*See,* e.g., the Sigma BioSciences 2002 catalogue and price list), and the corresponding protein sequences and genes and, typically, many variants thereof, are well-known (*see*, e.g., Genbank). Any of them can be modified by the insertion of one or more unnatural amino acid according to the invention, e.g., to alter the protein with respect to one or more therapeutic, diagnostic or enzymatic properties of interest. Examples of therapeutically relevant properties include serum half-life, shelf half-life, stability, immunogenicity, therapeutic activity, detectability (e.g., by the inclusion of reporter groups (e.g., labels or label binding sites) in the unnatural amino acids), reduction of LD_{50} or other side effects, ability to enter the body through the gastric tract (e.g., oral availability), or the like. Examples of diagnostic properties include shelf half-life, stability, diagnostic activity, detectability, or the like. Examples of relevant enzymatic properties include shelf half-life, stability, enzymatic activity, production capability, or the like.

[177] A variety of other proteins can also be modified to include one or more unnatural amino acid of the invention. For example, the invention can include substituting one or more natural amino acids in one or more vaccine proteins with an unnatural amino acid, e.g., in proteins from infectious fungi, e.g., *Aspergillus*, *Candida* species; bacteria, particularly *E. coli*, which serves a model for pathogenic bacteria, as well as medically important bacteria such as *Staphylococci* (e.g., *aureus*), or *Streptococci* (e.g., *pneumoniae*); protozoa such as sporozoa (e.g., *Plasmodia*), *rhizopods* (e.g., *Entamoeba*) and flagellates (*Trypanosoma, Leishmania, Trichomonas, Giardia*, etc.); viruses such as (+) RNA viruses (examples include Poxviruses e.g., *vaccinia*; Picornaviruses, e.g. *polio*; Togaviruses, e.g., *rubella*; Flaviviruses, e.g., HCV; and Coronaviruses), (-) RNA viruses (e.g.,

Rhabdoviruses, e.g., VSV; Paramyxovimses, e.g., RSV; Orthomyxovimses, e.g., influenza; Bunyaviruses; and Arenaviruses), dsDNA viruses (Reoviruses, for example), RNA to DNA viruses, i.e., Retroviruses, e.g., HIV and HTLV, and certain DNA to RNA viruses such as Hepatitis B.

[178] Agriculturally related proteins such as insect resistance proteins (e.g., the Cry proteins), starch and lipid production enzymes, plant and insect toxins, toxin-resistance proteins, Mycotoxin detoxification proteins, plant growth enzymes (e.g., Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase, "RUBISCO"), lipoxygenase (LOX), and Phosphoenolpyruvate (PEP) carboxylase are also suitable targets for unnatural amino acid modification.

[179] The invention also provides methods for producing in a vertebrate cell at least one protein comprising at least one unnatural amino acid (and proteins produced by such methods). For example, a method includes: growing, in an appropriate medium, a vertebrate cell that comprises a nucleic acid that comprises at least one selector codon and encodes the protein. The vertebrate cell also 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 unnatural amino acid, and the medium comprises an unnatural amino acid.

[180] In one embodiment, the method further includes incorporating into the protein the unnatural amino acid, where the unnatural amino acid comprises a first reactive group; and contacting the protein with a molecule (e.g., a dye, a polymer, e.g., a derivative of polyethylene glycol, a photocrosslinker, a cytotoxic compound, an affinity label, a derivative of biotin, a resin, a second protein or polypeptide, a metal chelator, a cofactor, a fatty acid, a carbohydrate, a polynucleotide (e.g., DNA, RNA, etc.), and the like) that comprises a second reactive group. The first reactive group reacts with the second reactive group to attach the molecule to the unnatural amino acid through a [3+2] cycloaddition. In one embodiment, the first reactive group is an alkynyl or azido moiety and the second reactive group is an azido or alkynyl moiety. For example, the first reactive group is the alkynyl moiety (e.g., in unnatural amino acid p-propargyloxyphenylalanine) and the second reactive group is the azido moiety. In another example, the first reactive group is the azido moiety (e.g., in the unnatural amino acid p-azido-L-phenylalanine) and the second reactive group is the alkynyl moiety.

[181] In one embodiment, the O-RS aminoacylates the O-tRNA with the unnatural amino acid at least 50% as efficiently as does an O-RS having an amino acid sequence, e.g., as set forth in SEQ ID NO.: 86 or 45. In another embodiment, the O-tRNA comprises, is processed from, or is encoded by SEQ ID NO.: 65 or 64, or a complementary polynucleotide sequence thereof. In yet another embodiment, the O-RS comprises an amino acid set forth in any one of SEQ ID NO.: 36-63 and/or 86.

[182] The encoded protein can comprise, e.g., a therapeutic protein, a diagnostic protein, an industrial enzyme, or portion thereof. Optionally, the protein that is produced by the method is further modified through the unnatural amino acid. For example, the protein produced by the method is optionally modified by at least one post-translational modification in vivo.

[183] Methods of producing a screening or selecting transcriptional modulator protein are also provided (and screening or selecting transcriptional modulator proteins produced by such methods). For example, a method includes: selecting a first polynucleotide sequence, where the polynucleotide sequence encodes a nucleic acid binding domain; and mutating the first polynucleotide sequence to include at least one selector codon. This provides a screening or selecting polynucleotide sequence. The method also includes: selecting a second polynucleotide sequence, where the second polynucleotide sequence encodes a transcriptional activation domain; providing a construct that comprises the screening or selecting polynucleotide sequence operably linked to the second polynucleotide sequence; and, introducing the construct, an unnatural amino acid, an orthogonal tRNA synthetase (O-RS) and an orthogonal tRNA (O-tRNA) into a cell. With these components, the O-RS preferentially aminoacylates the O-tRNA with the unnatural amino acid and the O-tRNA recognizes the selector codon and incorporates the unnatural amino acid into the nucleic acid binding domain, in response to the selector codon in the screening or selecting polynucleotide sequence, thereby providing the screening or selecting transcriptional modulator protein.

[184] In certain embodiments, the protein or polypeptide of interest (or portion thereof) in the methods and/or compositions of the 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.

[185] 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 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 unnatural amino acids. The invention includes any such variant, e.g., mutant, versions of any protein, e.g., including at least one unnatural 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 unnatural amino acid.

[186] Purifying recombinant proteins comprising unnatural amino acids

[187] Proteins of the invention, e.g., proteins comprising unnatural amino acids, antibodies to proteins comprising unnatural amino acids, etc., can be purified, either partially or substantially to homogeneity, according to standard procedures known to and used by those of skill in the art. Accordingly, polypeptides of the invention can be recovered and purified by any of a number of methods well known in the art, including, e.g., ammonium sulfate or ethanol precipitation, acid or base extraction, column chromatography, affinity column chromatography, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, hydroxylapatite chromatography, lectin chromatography, gel electrophoresis and the like. Protein refolding steps can be used, as desired, in making correctly folded mature proteins. High performance liquid chromatography (HPLC), affinity chromatography or other suitable methods can be employed in final purification steps where high purity is desired. In one embodiment, antibodies made against unnatural amino acids (or proteins comprising unnatural amino acids) are used as purification reagents, e.g., for affinity-based purification of proteins comprising one or more unnatural amino acid(s). Once purified, partially or to homogeneity, as desired, the polypeptides are optionally used e.g., as assay components, therapeutic reagents or as immunogens for antibody production.

[188] In addition to other references noted herein, a variety of purification/protein folding methods are well known in the art, including, e.g., those set forth in R. Scopes, Protein Purification, Springer-Verlag, N.Y. (1982); Deutscher, Methods in Enzymology Vol. 182: Guide to Protein Purification, Academic Press, Inc. N.Y. (1990); Sandana (1997) Bioseparation of Proteins, Academic Press, Inc.; Bollag et al. (1996) Protein Methods, 2nd Edition Wiley-Liss, NY; Walker (1996) The Protein Protocols Handbook Humana Press, NJ, Harris and Angal (1990) Protein Purification Applications: A Practical Approach IRL Press at Oxford, Oxford, England; Harris and Angal Protein Purification Methods: A Practical Approach IRL Press at Oxford, Oxford, England; Scopes (1993) Protein Purification: Principles and Practice 3rd Edition Springer Verlag, NY; Janson and Ryden (1998) Protein Purification: Principles, High Resolution Methods and Applications, Second Edition Wiley-VCH, NY; and Walker (1998) Protein Protocols on CD-ROM Humana Press, NJ; and the references cited therein.

[189] One advantage of producing a protein or polypeptide of interest with an unnatural amino acid in a vertebrate cell is that typically the proteins or polypeptides will be folded in their native conformations. However, in certain embodiments of the invention, those of skill in the art will recognize that, after synthesis, expression and/or purification, proteins can possess a conformation different from the desired conformations of the relevant polypeptides. In one aspect of the invention, the expressed protein is optionally denatured and then renatured. This is accomplished, e.g., by adding a chaperonin to the protein or polypeptide of interest, and/or by solubilizing the proteins in a chaotropic agent such as guanidine HCl, etc.

[190] In general, it is occasionally desirable to denature and reduce expressed polypeptides and then to cause the polypeptides to re-fold into the preferred conformation. For example, guanidine, urea, DTT, DTE, and/or a chaperonin can be added to a translation product of interest. Methods of reducing, denaturing and renaturing proteins are well known to those of skill in the art (see, the references above, and Debinski, et al. (1993) J. Biol. Chem., 268: 14065-14070; Kreitman and Pastan (1993) Bioconjug. Chem.,4: 581-585; and Buchner, et al., (1992) Anal. Biochem., 205: 263-270). Debinski, et al., for example, describe the denaturation and reduction of inclusion body proteins in guanidine-DTE. The proteins can be refolded in a redox buffer containing, e.g., oxidized glutathione and L-arginine. Refolding reagents can be flowed or otherwise moved into contact with the one or more polypeptide or other expression product, or vice-versa.

[191] Antibodies

[192] In one aspect, the invention provides antibodies to molecules of the invention, e.g., synthetases, tRNA, and proteins comprising unnatural amino acids. Antibodies to molecules of the invention are useful as purification reagents, e.g., for purifying the molecules of the invention. In addition, the antibodies can be used as indicator reagents to indicate the presence of a synthetase, a tRNA, or protein comprising an unnatural amino acid, e.g., to track the presence or location (e.g., in vivo or in situ) of the molecule.

[193] An antibody of the invention can be a protein comprising one or more polypeptides substantially or partially encoded by immunoglobulin genes or fragments of immunoglobulin genes. The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon and mu constant region genes, as well as myriad immunoglobulin variable region genes. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively. A typical immunoglobulin (e.g., antibody) structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kD) and one "heavy" chain (about 50-70 kD). The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain (VL) and variable heavy chain (V_H) refer to these light and heavy chains, respectively.

[194] Antibodies exist as intact immunoglobulins or as a number of well-characterized fragments produced by digestion with various peptidases. Thus, for example, pepsin digests an antibody below the disulfide linkages in the hinge region to produce $F(ab')_2$, a dimer of Fab which itself is a light chain joined to V_H-C_H1 by a disulfide bond. The $F(ab')_2$ may be reduced under mild conditions to break the disulfide linkage in the hinge region thereby converting the $F(ab')_2$dimer into an Fab' monomer. The Fab' monomer is essentially an Fab with part of the hinge region (*see*, Fundamental Immunology, 4th addition, W.E. Paul, ed., Raven Press, N.Y. (1999), for a more detailed description of other antibody fragments). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such Fab' fragments, etc. may be synthesized de novo either chemically or by utilizing recombinant DNA

methodology. Thus, the term antibody, as used herein, also optionally includes antibody fragments either produced by the modification of whole antibodies or synthesized *de novo* using recombinant DNA methodologies. Antibodies include single chain antibodies, including single chain Fv (sFv or scFv) antibodies in which a variable heavy and a variable light chain are joined together (directly or through a peptide linker) to form a continuous polypeptide. Antibodies of the invention can be, *e.g.*, polyclonal, monoclonal, chimeric, humanized, single chain, Fab fragments, fragments produced by an Fab expression library, or the like.

[195] In general, antibodies of the invention are valuable, both as general reagents and as therapeutic reagents in a variety of molecular biological or pharmaceutical processes. Methods of producing polyclonal and monoclonal antibodies are available, and can be applied to making the antibodies of the invention. A number of basic texts describe standard antibody production processes, including, e.g., Borrebaeck (ed) (1995) Antibody Engineering, 2nd Edition Freeman and Company, NY (Borrebaeck); McCafferty et al. (1996) Antibody Engineering, A Practical Approach IRL at Oxford Press, Oxford, England (McCafferty), and Paul (1995) Antibody Engineering Protocols Humana Press, Towata, NJ (Paul); Paul (ed.), (1999) Fundamental Immunology, Fifth edition Raven Press, N.Y.; Coligan (1991) Current Protocols in Immunology Wiley/Greene, NY; Harlow and Lane (1989) Antibodies: A Laboratory Manual Cold Spring Harbor Press, NY; Stites *et al.* (eds.) Basic and Clinical Immunology (4th ed.) Lange Medical Publications, Los Altos, CA, and references cited therein; Goding (1986) Monoclonal Antibodies: Principles and Practice (2d ed.) Academic Press, New York, NY; and Kohler and Milstein (1975) Nature 256: 495-497.

[196] A variety of recombinant techniques for antibody preparation which do not rely on, e.g., injection of an antigen into an animal have been developed and can be used in the context of the present invention. For example, it is possible to generate and select libraries of recombinant antibodies in phage or similar vectors. *See*, e.g., Winter et al. (1994) *Making Antibodies by Phage Display Technology* Annu. Rev. Immunol. 12:433-55 and the references cited therein for a review. *See also*, Griffiths and Duncan (1998) *Strategies for selection of antibodies by phage display* Curr Opin Biotechnol 9: 102-8; Hoogenboom et al. (1998) *Antibody phage display technology and its applications* Immunotechnology 4: 1-20; Gram et al. (1992) *in vitro selection and affinity maturation of antibodies from a naïve combinatorial immunoglobulin library* PNAS 89:3576-3580; Huse et al. (1989) Science 246: 1275-1281; and Ward, et al. (1989) Nature 341: 544-546.

[197] In one embodiment, antibody libraries can include repertoires of V genes (e.g., harvested from populations of lymphocytes or assembled in vitro) which are cloned for display of associated heavy and light chain variable domains on the surface of filamentous bacteriophage. Phage are selected by binding to an antigen. Soluble antibodies are expressed from phage infected bacteria and the antibody can be improved, e.g., via mutagenesis. *See* e.g., Balint and Larrick (1993) *Antibody Engineering by Parsimonious Mutagenesis* Gene 137:109-118; Stemmer et al. (1993) *Selection of an Active Single Chain Fv Antibody From a Protein Linker Library Prepared by Enzymatic Inverse PCR* Biotechniques 14(2):256-65; Crameri et al. (1996) *Construction and evolution of antibody-phage libraries by DNA shuffling* Nature Medicine 2:100-103; and Crameri and Stemmer (1995) *Combinatorial multiple cassette mutagenesis creates all the permutations of mutant and wildtype cassettes* BioTechniques 18:194-195.

[198] Kits for cloning and expression of recombinant antibody phage systems are also known and available, e.g., the "recombinant phage antibody system, mouse ScFv module," from Amersham-Pharmacia Biotechnology (Uppsala, Sweden). Bacteriophage antibody libraries have also been produced for making high affinity human antibodies by chain shuffling (*See*, e.g., Marks et al. (1992) *By- Passing Immunization: Building High Affinity Human Antibodies by Chain Shuffling* Biotechniques 10:779-782. It will also be recognized that antibodies can be prepared by any of a number of commercial services (e.g., Bethyl Laboratories (Montgomery, TX), Anawa (Switzerland), Eurogentec (Belgium and in the US in Philadelphia, PA, etc.) and many others.

[199] In certain embodiments, it is useful to "humanize" antibodies of the invention, e.g., where the antibodies are to be administered therapeutically. The use of humanized antibodies tends to reduce the incidence of unwanted immune responses against the therapeutic antibodies (e.g., when the patient is a human). The antibody references above describe humanization strategies. In addition to humanized antibodies, human antibodies are also a feature of the invention. Human antibodies consist of characteristically human immunoglobulin sequences. Human antibodies can be produced in using a wide variety of methods (*see, e.g.*, Larrick *et al.*, U.S. Pat. No. 5,001,065, for a review). A general approach for producing human antibodies by trioma technology is described by Ostberg *et al.* (1983), *Hybridoma* 2: 361-367, Ostberg, U.S. Pat. No. 4,634,664, and Engelman *et al.*, U.S. Pat. No. 4,634,666.

[200] A variety of methods of using antibodies in the purification and detection of proteins are known and can be applied to detecting and purifying proteins comprising unnatural amino acids as noted herein. In general, antibodies are useful reagents for ELISA, western blotting, immunochemistry, affinity chromatograpy methods, SPR, and many other methods. The references noted above provide details on how to perform ELISA assays, western blots, surface plasmon resonance (SPR) and the like.

[201] In one aspect of the invention, antibodies of the invention themselves include unnatural amino acids, providing the antibodies with properties of interest (e.g., improved half-life, stability, toxicity, or the like). See also, the section herein entitled "Polypeptides with unnatural amino acids." Antibodies account for nearly 50% of all compounds currently in clinical trials (Wittrup, (1999) *Phage on display* Tibtech 17: 423-424 and antibodies are used ubiquitously as diagnostic reagents. Accordingly, the ability to modify antibodies with unnatural amino acids provides an important tool for modifying these valuable reagents.

[202] For example, there are many applications of MAbs to the field of diagnostics. Assays range from simple spot tests to more involved methods such as the radio-labeled NR-LU-10 MAb from DuPont Merck Co. used for tumor imaging (Rusch et al. (1993) *NR-LU-10 monoclonal antibody scanning. A helpful new adjunct to computed tomography in evaluating non-small-cell lung cancer.* J Thorac Cardiovasc Surg 106: 200-4). As noted, MAbs are central reagents for ELISA, western blotting, immunochemistry, affinity chromatograpy methods and the like. Any such diagnostic antibody can be modified to include one or more unnatural amino acid, altering, e.g., the specificity or avidity of the Ab for a target, or altering one or more detectable property, e.g., by including a detectable label (e.g., spectrographic, fluorescent, luminescent, etc.) in the unnatural amino acid.

[203] One class of valuable antibody reagents are therapeutic Abs. For example, antibodies can be tumor-specific MAbs that arrest tumor growth by targeting tumor cells for destruction by antibody-dependent cell-mediated cytotoxicity (ADCC) or complement-mediated lysis (CML) (these general types of Abs are sometimes referred to as "magic bullets"). One example is Rituxan, an anti-CD20 MAb for the treatment of Non-Hodgkins lymphoma (Scott (1998) *Rituximab: a new therapeutic monoclonal antibody for non-Hodgkin's lymphoma* Cancer Pract 6: 195-7). A second example relates to antibodies which

interfere with a critical component of tumor growth. Herceptin is an anti-HER-2 monoclonal antibody for treatment of metastatic breast cancer, and provides an example of an antibody with this mechanism of action (Baselga et al. (1998) *Recombinant humanized anti-HER2 antibody (Herceptin) enhances the antitumor activity of paclitaxel and doxorubicin against HER2/neu overexpressing human breast cancer xenografts* [published erratum appears in Cancer Res (1999) 59(8):2020], Cancer Res 58: 2825-31). A third example relates to antibodies for delivery of cytotoxic compounds (toxins, radionuclides, etc.) directly to a tumor or other site of interest. For example, one application Mab is CYT-356, a 90Y-linked antibody that targets radiation directly to prostate tumor cells (Deb et al. (1996) *Treatment of hormone-refractory prostate cancer with 90Y-CYT-356 monoclonal antibody* Clin Cancer Res 2: 1289-97. A fourth application is antibody-directed enzyme prodrug therapy, where an enzyme co-localized to a tumor activates a systemically-administered pro-drug in the tumor vicinity. For example, an anti-Ep-CAM1 antibody linked to carboxypeptidase A is being developed for treatment of colorectal cancer (Wolfe et al. (1999) *Antibody-directed enzyme prodrug therapy with the T268G mutant of human carboxypeptidase A1: in vitro and in vivo studies with prodrugs of methotrexate and the thymidylate synthase inhibitors GW1031 and GW1843* Bioconjug Chem 10: 38-48). Other Abs (e.g., antagonists) are designed to specifically inhibit normal cellular functions for therapeutic benefit. An example is Orthoclone OKT3, an anti-CD3 MAb offered by Johnson and Johnson for reducing acute organ transplant rejection (Strate et al. (1990) *Orthoclone OKT3 as first-line therapy in acute renal allograft rejection* Transplant Proc 22: 219-20. Another class of antibody products are agonists. These Mabs are designed to specifically enhance normal cellular functions for therapeutic benefit. For example, Mab-based agonists of acetylcholine receptors for neurotherapy are under development (Xie et al. (1997) *Direct demonstration of MuSK involvement in acetylcholine receptor clustering through identification of agonist ScFv* Nat. Biotechnol. 15: 768-71. Any of these antibodies can be modified to include one or more unnatural amino acid to enhance one or more therapeutic property (specificity, avidity, serum-half-life, etc.).

[204] Another class of antibody products provide novel functions. The main antibodies in this group are catalytic antibodies such as Ig sequences that have been engineered to mimic the catalytic abilities of enzymes (Wentworth and Janda (1998) *Catalytic antibodies* Curr Opin Chem Biol 2: 138-44. For example, an interesting application involves using the catalytic antibody mAb-15A10 to hydrolyze cocaine *in vivo*

for addiction therapy (Mets et al. (1998) *A catalytic antibody against cocaine prevents cocaine's reinforcing and toxic effects in rats* Proc Natl Acad Sci U S A 95: 10176-81). Catalytic antibodies can also be modified to include one or more unnatural amino acid to improve one or more property of interest.

[205] Defining Polypeptides by Immunoreactivity

[206] Because the polypeptides of the invention provide a variety of new polypeptide sequences (e.g., comprising unnatural amino acids in the case of proteins synthesized in the translation systems herein, or, e.g., in the case of the novel synthetases herein, 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 antibodies or antibodies which specifically bind the polypeptides of the invention, as well as the polypeptides which are bound by such antibodies or antisera, are a feature of the invention.

[207] For example, the invention includes synthetase proteins that specifically bind to or that are specifically immunoreactive with an antibody or antisera generated against an immunogen comprising an amino acid sequence selected from one or more of (SEQ ID NO: 36-63, and/or 86). To eliminate cross-reactivity with other homologues, the antibody or antisera is subtracted with available control synthetase homologues, such as the wild-type *E. coli* tyrosyl synthetase (TyrRS) (e.g., SEQ ID NO.:2).

[208] In one typical format, the immunoassay uses a polyclonal antiserum which was raised against one or more polypeptide comprising one or more of the sequences corresponding to one or more of SEQ ID NO: 36-63, and/or 86, 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 SEQ ID NO: 36-63 and 86 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 control synthetase homologues, prior to use of the polyclonal antiserum in the immunoassay.

[209] 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. Additional references and discussion of antibodies is also found herein and can be applied here to make antibodies that define/detect 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.

[210] 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 titered polyclonal antisera.

[211] The subtracted pooled titered 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 titered polyclonal antisera which result in at least about a 5-10 fold higher signal to noise ratio for binding of the titered polyclonal antisera to the immunogenic synthetase as compared to binding to a control synthetase homologue. That is, the stringency of the binding/washing reaction(s) is/are 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/washing 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. In particular, test polypeptides which show at least a 2-5x higher signal to noise ratio than the control synthetase homologue under discriminatory binding conditions, and at least about a ½ signal to noise ratio as compared to the immunogenic polypeptide(s), shares substantial structural similarity with the immunogenic polypeptide as compared to known synthetases, and is, therefore a polypeptide of the invention.

[212] 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.

[213] 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.

[214] 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.

[215] 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 immunosorbtion 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.

[216] Pharmaceutical Compositions

[217] The polypeptides or proteins of the invention (e.g., synthetases, proteins comprising one or more unnatural amino acid, etc.) are optionally employed for therapeutic uses, e.g., in combination with a suitable pharmaceutical carrier. Such compositions, e.g., comprise a therapeutically effective amount of the compound, and a pharmaceutically acceptable carrier or excipient. Such a carrier or excipient includes, but is not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol, and/or combinations thereof. The formulation is made to suit the mode of administration. In general, methods of administering proteins are well known in the art and can be applied to administration of the polypeptides of the invention.

[218] Therapeutic compositions comprising one or more polypeptide of the invention are optionally tested in one or more appropriate in vitro and/or in vivo animal models of disease, to confirm efficacy, tissue metabolism, and to estimate dosages, according to methods well known in the art. In particular, dosages can be initially determined by activity, stability or other suitable measures of unnatural herein to natural amino acid homologues (e.g., comparison of an EPO modified to include one or more unnatural amino acids to a natural amino acid EPO), i.e., in a relevant assay.

[219] Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells. The unnatural amino acid polypeptides of the invention are administered in any suitable manner, optionally with one or more pharmaceutically acceptable carriers. Suitable methods of administering such polypeptides in the context of the present invention to a patient are available, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective action or reaction than another route.

[220] Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions of the present invention.

[221] Polypeptide compositions can be administered by a number of routes including, but not limited to: oral, intravenous, intraperitoneal, intramuscular, transdermal, subcutaneous, topical, sublingual, or rectal means. Unnatural amino acid polypeptide compositions can also be administered via liposomes. Such administration routes and appropriate formulations are generally known to those of skill in the art.

[222] The unnatural amino acid polypeptide, alone or in combination with other suitable components, can also be made into aerosol formulations (i.e., they can be "nebulized") to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like.

[223] Formulations suitable for parenteral administration, such as, for example, by intraarticular (in the joints), intravenous, intramuscular, intradermal, intraperitoneal, and subcutaneous routes, include aqueous and non-aqueous, isotonic sterile injection solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood of the intended recipient, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. The formulations of packaged nucleic acid can be presented in unit-dose or multi-dose sealed containers, such as ampules and vials.

[224] Parenteral administration and intravenous administration are preferred methods of administration. In particular, the routes of administration already in use for natural amino acid homologue therapeutics (e.g., those typically used for EPO, GCSF, GMCSF, IFNs, interleukins, antibodies, and/or any other pharmaceutically delivered protein), along with formulations in current use, provide preferred routes of administration and formulation for the proteins that include unnatural amino acids of the invention (e.g., pegylated variants of current thereputic proteins, etc.).

[225] The dose administered to a patient, in the context of the present invention, is sufficient to effect a beneficial therapeutic response in the patient over time, or, e.g., to inhibit infection by a pathogen, or other appropriate activity, depending on the application.

The dose is determined by the efficacy of a particular composition/formulation, and the activity, stability or serum half-life of the unnatural amino acid polypeptide employed and the condition of the patient, as well as the body weight or surface area of the patient to be treated. The size of the dose is also determined by the existence, nature, and extent of any adverse side-effects that accompany the administration of a particular composition/ formulation, or the like in a particular patient.

[226] In determining the effective amount of the composition/formulation to be administered in the treatment or prophylaxis of disease (e.g., cancers, inherited diseases, diabetes, AIDS, or the like), the physician evaluates circulating plasma levels, formulation toxicities, progression of the disease, and/or where relevant, the production of anti-unnatural amino acid polypeptide antibodies.

[227] The dose administered, e.g., to a 70 kilogram patient, is typically in the range equivalent to dosages of currently-used therapeutic proteins, adjusted for the altered activity or serum half-life of the relevant composition. The compositions/formulations of this invention can supplement treatment conditions by any known conventional therapy, including antibody administration, vaccine administration, administration of cytotoxic agents, natural amino acid polypeptides, nucleic acids, nucleotide analogues, biologic response modifiers, and the like.

[228] For administration, formulations of the present invention are administered at a rate determined by the LD-50 of the relevant formulation, and/or observation of any side-effects of the unnatural amino acids at various concentrations, e.g., as applied to the mass and overall health of the patient. Administration can be accomplished via single or divided doses.

[229] If a patient undergoing infusion of a formulation develops fevers, chills, or muscle aches, he/she receives the appropriate dose of aspirin, ibuprofen, acetaminophen or other pain/fever controlling drug. Patients who experience reactions to the infusion such as fever, muscle aches, and chills are premedicated 30 minutes prior to the future infusions with either aspirin, acetaminophen, or, e.g., diphenhydramine. Meperidine is used for more severe chills and muscle aches that do not quickly respond to antipyretics and antihistamines. Treatment is slowed or discontinued depending upon the severity of the reaction.

[230] Nucleic acid and polypeptide sequence and variants

[231] As described above and below, the invention provides for nucleic acid polynucleotide sequences and polypeptide amino acid sequences, e.g., O-tRNA's and O-RSs, and, e.g., compositions and methods comprising said sequences. Examples of said sequences, e.g., O-tRNA's and O-RSs are disclosed herein (see, Table 5, e.g., SEQ ID NO. 3-65, 86, and other than SEQ ID NO.: 1 and 2). However, one of skill in the art will appreciate that the invention is not limited to those sequences disclosed herein, e.g., the Examples and Table 5. One of skill will appreciate that the invention also provides many related and even unrelated sequences with the functions described herein, e.g., encoding an O-tRNA or an O-RS.

[232] The invention also provides polypeptides (O-RSs) and polynucleotides, e.g., O-tRNA, polynucleotides that encode O-RSs or portions thereof (e.g., the active site of the synthetase), oligonucleotides used to construct aminoacyl-tRNA synthetase mutants, etc. For example, a polypeptide of the invention includes a polypeptide that comprises an amino acid sequence as shown in any one of SEQ ID NO.: 36-63, and/or 86, a polypeptide that comprises an amino acid sequence encoded by a polynucleotide sequence as shown in any one of SEQ ID NO.: 3-35, and a polypeptide that is specifically immunoreactive with an antibody specific for a polypeptide that comprises an amino acid sequence as shown in any one of SEQ ID NO.: 36-63, and/or 86, or a polypeptide that comprises an amino acid sequence encoded by a polynucleotide sequence as shown in any one of SEQ ID NO.: 3-35.

[233] Also included among the polypeptides of the invention are polypeptides that comprise an amino acid sequence that is at least 90% identical to that of a naturally occurring tyrosyl aminoacyl-tRNA synthetase (TyrRS) (e.g., SEQ ID NO.:2) and comprises two or more amino acids of groups A-E. For example, group A includes valine, isoleucine, leucine, glycine, serine, alanine, or threonine at a position corresponding to Tyr37 of *E. coli* TyrRS; group B includes aspartate at a position corresponding to Asn126 of *E. coli* TyrRS; group C includes threonine, serine, arginine, asparagine or glycine at a position corresponding to Asp182 of *E. coli* TyrRS; group D includes methionine, alanine, valine, or tyrosine at a position corresponding to Phe183 of *E. coli* TyrRS; and, group E includes serine, methionine, valine, cysteine, threonine, or alanine at a position corresponding to Leu186 of *E. coli* TyrRS. Similarly, polypeptides of the invention also include a polypeptide that comprises at least 20 contiguous amino acids of SEQ ID NO.: 36-63, and/or 86, and two or more amino acid substitutions as indicated above in groups A-E. *See also*, Table 4, Table 6, and/or Table 8, herein. An amino acid sequence comprising a

conservative variation of any of the above polypeptides is also included as a polypeptide of the invention.

[234] In one embodiment, a composition includes a polypeptide of the invention and an excipient (e.g., buffer, water, pharmaceutically acceptable excipient, etc.). The invention also provides an antibody or antisera specifically immunoreactive with a polypeptide of the invention.

[235] Polynucleotides are also provided in the invention. Polynucleotides of the invention include those that encode proteins or polypeptides of interest of the invention, or that include one or more selector codon, or both. For example, polynucleotides of the invention include, e.g., a polynucleotide comprising a nucleotide sequence as set forth in any one of SEQ ID NO.: 3-35, 64-85; a polynucleotide that is complementary to or that encodes a polynucleotide sequence thereof; and/or a polynucleotide encoding a polypeptide that comprises an amino acid sequence as set forth in any one of SEQ ID NO.: 36-63, and/or 86, or a conservative variation thereof. A polynucleotide of the invention also includes a polynucleotide that encodes a polypeptide of the invention. 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 invention.

[236] A polynucleotide of the invention also includes a polynucleotide that encodes a polypeptide that comprises an amino acid sequence that is at least 90% identical to that of a naturally occurring tyrosyl aminoacyl-tRNA synthetase (TyrRS) (e.g., SEQ ID NO.: 2) and comprises two or more mutations as indicated above in groups A-E in paragraph 11. A polynucleotide that is that is at least 70%, (or at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or least 99% or more) identical to a polynucleotide indicated above and/or a polynucleotide comprising a conservative variation of any of the polynucleotides indicated above are also included among the polynucleotides of the invention.

[237] In certain embodiments, a vector (e.g., a plasmid, a cosmid, a phage, a virus, etc.) comprises a polynucleotide of the 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 invention. In another embodiment, a cell comprises a vector that includes a polynucleotide of the invention.

[238] 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.

[239] Conservative variations

[240] 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.

[241] "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 skill will recognize that individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than 5%, more typically less than 4%, 2% or 1%) in an encoded sequence are "conservatively modified variations" 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 2% or 1%, of the amino acids of the polypeptide sequence, with a conservatively selected amino acid of the same conservative substitution group. Finally, the addition of sequences that 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.

[242] Conservative substitution tables providing functionally similar amino acids are well known in the art. The following sets forth example groups which contain natural amino acids that include "conservative substitutions" for one another.

Conservative Substitution Groups

1	Alanine (A)	Serine (S)	Threonine (T)	
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)	

[243] Nucleic Acid Hybridization

[244] Comparative hybridization can be used to identify nucleic acids of the invention, including conservative variations of nucleic acids of the invention, and this comparative hybridization method is a preferred method of distinguishing nucleic acids of the invention. In addition, target nucleic acids which hybridize to the nucleic acids represented by SEQ ID NO: 3-35, 64-85 under high, ultra-high and ultra-ultra high stringency conditions are a feature of the 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.

[245] A test nucleic acid is said to specifically hybridize to a probe nucleic acid when it hybridizes at least ½ as well to the probe as to the perfectly matched complementary target, i.e., with a signal to noise ratio at lest ½ 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.

[246] 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. 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, *supra*. 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.

[247] 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, *supra* 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.

[248] "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. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993), *supra.* and in Hames and Higgins, 1 and 2. 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.

[249] "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.

[250] "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 ½ that of the perfectly matched complementary target nucleic acid is said to bind to the probe under ultra-high stringency conditions.

[251] 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.

[252] Nucleic acids that 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.

[253] Unique subsequences

[254] 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 invention.

[255] 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 known polypeptide sequence.

[256] 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 invention were derived, e.g., by mutation). Unique sequences are determined as noted above.

[257] Sequence comparison, identity, and homology

[258] 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, as measured using one of the sequence comparison algorithms described below (or other algorithms available to persons of skill) or by visual inspection.

[259] 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%, preferably 80%, most preferably 90-95% nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using a sequence comparison algorithm or by visual inspection. Such "substantially identical" sequences are typically considered to be "homologous," without reference to actual ancestry. Preferably, the "substantial identity" exists over a region of the sequences that is at least about 50 residues in length, more preferably over a region of at least about 100 residues, and most preferably, the sequences are substantially identical over at least about 150 residues, or over the full length of the two sequences to be compared.

[260] 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.

[261] Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, Adv. Appl. Math. 2:482 (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 visual inspection (*see generally*, Ausubel *et al.*, *infra*).

[262] One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in 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 (www.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 (1989) Proc. Natl. Acad. Sci. USA 89:10915).

[263] 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 is 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 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

[264] Mutagenesis and Other Molecular Biology Techniques

[265] 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 that include selector codons for production of proteins that include unnatural amino acids, orthogonal tRNA's, orthogonal synthetases, and pairs thereof.

[266] Various types of mutagenesis are used in the invention, e.g., to produce libraries of tRNA's, to produce libraries of synthetases, to insert selector codons that encode unnatural amino acids 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, e.g., involving chimeric constructs, are 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, e.g., sequence, sequence comparisons, physical properties, crystal structure or the like.

[267] The above 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); Methods in Enzymol. 100: 468-500 (1983); Methods in Enzymol. 154: 329-350 (1987); 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: 8765-8787 (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., *Y-T 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., *Point Mismatch Repair*, 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); Sakamar and Khorana, *Total synthesis and expression of a gene for the a-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.

[268] The invention also relates to vertebrate host cells and organisms for the in vivo incorporation of an unnatural amino acid via orthogonal tRNA/RS pairs. Host cells are genetically engineered (e.g., transformed, transduced or transfected) with the

polynucleotides of the invention or constructs which include a polynucleotide of the invention, e.g., a vector of the invention, which can be, for example, a cloning vector or an expression vector. The vector can be, for example, in the form of a plasmid, 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 (From 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)).

[269] 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.

[270] The invention also relates to vertebrate cell lines with the ability to incorporate an unnatural amino acid or acids via orthogonal tRNA/RS pairs. These cell lines can be established using cell culture techniques known in the art on host cells which have been transformed, transduced, or transfected with the polynucleotides of the invention or constructs which include a polynucleotide of the invention. The methods of introducing exogenous nucleic acids into host cells are well known in the art, and will vary with the host cell used. Techniques include, but are not limited to, dextran-mediated transfection, calcium phosphate precipitation, calcium chloride treatment, polybrene mediated transfection, protoplast fusion, electroporation, viral or phage infection, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection.

[271] Cells may be transformed or transfected in a manner to allow either transient or stable incorporation of DNA. For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express the

antibody molecule may be engineered. Rather than using expression vectors which contain viral origins of replication, host cells can be transformed with DNA controlled by appropriate expression control elements (e.g., promoter, enhancer, sequences, transcription terminators, polyadenylation sites, etc.), and a selectable marker. Following the introduction of the foreign DNA, engineered cells may be allowed to grow for 1-2 days in an enriched media, and then are switched to a selective media. The selectable marker in the recombinant plasmid confers resistance to the selection and allows cells to stably integrate the plasmid into their chromosomes and grow to form foci which in turn can be cloned and expanded into cell lines. This method may advantageously be used to engineer cell lines which express the antibody molecule. Such engineered cell lines may be particularly useful in screening and evaluation of compounds that interact directly or indirectly with the antibody molecule. Alternatively, other techniques, such as some viral-mediated vector transfection techniques, well known to those in the art, can permit transient transfection of cells.

[272] 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, a plethora of 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, (e.g., shuttle vectors) and selection markers for both prokaryotic and vertebrate systems. Vectors are suitable for replication and integration in prokaryotes, eukaryotes, or preferably both. *See,* Giliman & Smith, Gene 8:81 (1979); Roberts, *et al.*, Nature, 328:731 (1987); Schneider, B., *et al.*, Protein Expr. Purif. 6435:10 (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 mcrc.com), The Great American Gene Company (Ramona, CA available on the World Wide Web at genco.com), ExpressGen Inc. (Chicago, IL available on the World Wide Web at expressgen.com), Operon Technologies Inc. (Alameda, CA) and many others.

[273] Kits

[274] Kits are also a feature of the invention. For example, a kit for producing a protein that comprises at least one 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 one unnatural amino acid. In another embodiment, the kit further comprises instructional materials for producing the protein.

[275] EXAMPLES

[276] 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.

[277] Example 1: Methods of producing and compositions of Aminoacyl-tRNA synthetases that Incorporate unnatural amino acids in vertebrate cells

[278] The expansion of the vertebrate genetic code to include unnatural amino acids with novel physical, chemical or biological properties would provide powerful tools for analyzing and controlling protein function in these cells. Towards this goal, a general approach for the isolation of aminoacyl-tRNA synthetases that incorporate unnatural amino acids with high fidelity into proteins in response to an amber codon in Saccharomyces cerevisiae (S. cerevisiae) is described. The method is based on the activation of GAL4

responsive reporter genes, HIS3, URA3 or LacZ, by suppression of amber codons between the DNA binding domain and transcriptional activation domain of GAL4. The optimization of a GAL4 reporter for positive selection of active *Escherichia coli* tyrosyl-tRNA synthetase (EcTyrRS) variants is described. A negative selection of inactive EcTyrRS variants has also been developed with the URA3 reporter by use of a small molecule (5-fluroorotic acid (5-FOA)) added to the growth media as a 'toxic allele.' Importantly both positive and negative selections can be performed in a single yeast strain and with a range of stringencies. This can facilitate the isolation of a range of aminoacyl-tRNA synthetase (aaRS) activities from large libraries of mutant synthetases. The power of the method for isolating desired aaRS phenotypes is demonstrated by model selections.

[279] Example 2

[280] Site specific incorporation of pAF in mammalian cells

[281] **Plasmid constructions:**

[282] Wild-type human growth hormone (hGH) and hGH amber mutant expression vectors were constructed by ligating a DNA insert encoding hGH that has an N-terminal native secretion signal with pM1-MT vector (Roche) at Sal I and EcoR V restriction sites.

[283] Single copy *B. stearothermophilus* tRNA expression insert which includes 5' restriction sites EcoR I and Bgl II , 5' flanking sequence of human $tRNA^{Tyr}$ (GGATTACGCATGCTCAGTGCAATCTTCGGTTGCCTGGACTAGCGCTCCGGTTTT TCTGTGCTGAACCTCAGGGGACGCCGACACACGTACACGTC (SEQ ID NO: 88)), *B. stearothermophilus tRNA* amber suppression mutant lacking 3'-CCA, 3' flanking sequence of human $tRNA^{Tyr}$ (GACAAGTGCGGTTTTTTTCTCCAGCTCCCGATGACTTATGGC (SEQ ID NO: 89)) and 3' restriction sites BamH I and Hind III, was constructed by overlap PCR using primers:

FTam 73: forward primer with EcoR I and Bgl II site
GTACGAATTCCCGAGATCTGGATTACGCATGCTCAGTGCAATCTTCGGTTGCCT GGACTAGCGCTCCGGTTTTTCTGTGC (SEQ ID NO: 90)
FTam 74: Reverse primer overlap with FTam73
AGTCCGCCGCGTTTAGCCACTTCGCTACCCCTCCGACGTGTACGTGTGTCGGCG TCCCCTGAGGTTCAGCACAGAAAAACCGGAGCGC (SEQ ID NO: 91)

FTam 75: Forward primer overlap with FTam74 and FTam 76
GAAGTGGCTAAACGCGGCGGACTCTAAATCCGCTCCCTTTGGGTTCGGCGGTTC
GAATCCGTCCCCCTCCAGACAAGTG (SEQ ID NO: 92)
FTam 76: Reverse primer with BamH I and Hind III sites, overlap with FTam 75
GATGCAAGCTTGATGGATCCGCCATAAGTCATCGGGAGCTGGAGAAAAAAACC
GCACTTGTCTGGAGGGGGACGG (SEQ ID NO: 93)

[284] To construct a single copy tRNA expression vector, the insert described above was digested with EcoRI/HindIII and ligated to pUC 19 vector cut with the same restriction enzymes. To construct a two-copy tRNA expression vector, the single copy insert was digested with EcoR I and BamH I and ligated to the single copy expression vector cut with EcoR I and Bgl II. The ligated product regenerates a 5' EcoR I and 3' Bgl II sites. A similar strategy can be used in an iterative fashion to construct expression vectors containing tandem copies of tRNA sequence.

[285] The FLAG tag (DYKDDDDK) was added to the C-termini of wild-type *E. coli* Tyr tRNA synthetase and its mutants that charge non-natural amino acids. The RS gene was amplified by PCR and ligated to pcDNA3.1/Zeo(+) (Invitrogen).

[286] Cell culture

[287] One day prior to transfection, approximately 3.5 x 10^5 CHO K1 cells were plated in each well of a 6-well tissue culture plate (BD bioscience) in F-12 + Glutmax medium (Gibco) supplemented with 10% Fetal Bovine Serum (FBS) (Hyclone) and 100 U/ml penicillin G sodium and 100 ug/ml streptomycin sulfate (Gibco). The plates were incubated at 37°C, 5% CO_2. At 95% confluence, transfections were carried out according to the standard transfection protocol for lipofectamine 2000 (Invitrogen). In order to observe transient suppression, 1 μg of each plasmid (ie. tRNA, RS, GOI plasmids) was added to each well. After 4 hours of incubation, the transfection solution was replaced with 2 ml of growth medium (F-12 + Glutmax medium supplemented with 10% FBS serum and 100 U/ml penicillin G sodium and 100 ug/ml streptomycin sulfate). For those wells transfected with a non-natural amino acid tRNA/RS pair, 1 mM of the corresponding non-natural amino acid was added as a supplement. In most cases, experiments were performed in triplicate. The expression of hGH was assayed after 40 hours using the Active Human Growth Hormone ELISA kit (Diagnostic Systems Laboratories Inc.).

[288] Results:

[289] **Site specific amber suppression with non-natural amino acids in CHO-K1. (Figure 1).** The non-natural amino acid dependence of amber suppression was evaluated with the hGH G131 amber mutant. CHO K1 cells were co-transfected with: a plasmid carrying 6 tandem copies of an amber-suppressing *B. stearothermophilus* tRNA mutant; a gene of interest plasmid encoding the hGH-G131 amber mutant; and a plasmid encoding the *E. coli* Tyr tRNA synthetase mutant that aminoacylates its cognate tRNA with tyrosine (Tyr), *para*-acetylphenylalanine (pAF), *para*-azidophenylalanine (pAz) and *para*-benzoylphenylalanine (pBz).. After a 4 hour incubation with transfection solution, the transfection solutions were replaced with growth medium +/- corresponding 1 mM non-natural amino acids. The expression of hGH was assayed after 24 hours. For pAF and pBz, full length hGH expression was only observed in the presence of the corresponding non-natural amino acid. In the absence of non-natural amino acid, no full length hGH expression were detected. In the case of amber suppression with pAZ, no hGH expression was observed in the presence and absence of pAZ. This is likely due to the inability of expression of pAZ tRNA synthetase, which was observed by anti-FLAG western blot.

[290] The effect of pAF concentration on amber suppression of Tyr and pAF was evaluated using the hGH E88 amber mutant. Plasmids encoding hGH E88 amber mutant, 6-copy *B* (Figure 2). *Stearothermophilus* tRNA, and the corresponding *E. coli* tRNA synthetases that charge Tyr and pAF were co-transfected into CHO K1 cell. After a 4 hour incubation with transfection solution, the transfection solution was replaced with growth media with or without the addition of 1 M HCl (1:1000 relative to the growth media volume). The resulting media was next supplemented with 1, 2, 4, 6, 8 and 10 mM pAF using 1M pAF stock solution in 1 M HCl, and neutralized with an equal volume of 1 M NaOH. Expression of hGH was assayed after 42 hours. In the case of Tyr suppression, the suppression efficiency decreased slightly with the HCl/NaOH-treatment. In the case of pAF suppression, no hGH expression was detected in the absence of pAF. For both Tyr and pAF-based suppressions, the efficiency was optimum at 1 and 2 mM pAF. Suppression efficiency decreased with the increase of pAF concentration from 4 to 10 mM.

[291] Example 3

Amber suppression of human Fc in suspension cells with pAF using hybrid tRNA

Plasmids containing human Fc I21 amber mutant, human tRNA and the *E. coli* tRNA synthetase mutant charging pAF were co-transfected into CHO-S FreeStyle suspension cells. Four copies of htRNA were used in the experiment. The expression medium contained 1mM pAF. The expression of human Fc was assayed 72 hours after transfection. Higher suppression of human Fc was detected by deceasing amount of transfected pAF specific synthetase (pAFRS).

hIgG1-Fc2 DNA sequence:

CTGAGATCACCGGCGAAGGAGGGCCACCATGTACAGGATGCAACTCCTGTCTT
GCATTGCACTAAGTCTTGCACTTGTCACGAATTCGATATCGGCCATGGTTAGAT
CTGACAAAACTCACACATGCCCACCGTGCCCAGCACCTGAACTCCTGGGGGGA
CCGTCAGTCTTCCTCTTCCCCCCAAAACCCAAGGACACCCTCATGATCTCCCGG
ACCCCTGAGGTCACATGCGTGGTGGTGGACGTGAGCCACGAAGACCCTGAGGT
CAAGTTCAACTGGTACGTGGACGGCGTGGAGGTGCATAATGCCAAGACAAAGC
CGCGGGAGGAGCAGTACAACAGCACGTACCGTGTGGTCAGCGTCCTCACCGTC
CTGCACCAGGACTGGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAA
AGCCCTCCCAGCCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCC
GAGAACCACAGGTGTACACCCTGCCCCCATCCCGGGAGGAGATGACCAAGAAC
CAGGTCAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCCAGCGACATCGCCGTG
GAGTGGGAGAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCCCGT
GCTGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAG
CAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGGTCTGCA
CAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAA (SEQ ID NO: 94)

5' IL2 signal sequence:

ATGTACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCT (SEQ ID NO: 95)

hIgG1-Fc2 protein sequence

MYRMQLLSCIALSLALVTNSISAMVRSDKTHTCPPCPAPELLGGPSVFLFPPKPKDT
LMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSV
LTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKN
QVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRW
QQGNVFSCSVMHEGLHNHYTQKSLSLSPGK (SEQ ID NO: 96)

IL2 signal sequence :

MYRMQLLSCIALSLALVTNS (SEQ ID NO: 97)

In one aspect, the invention provides methods and related compositions of proteins comprising unnatural amino acids coupled to additional substituent molecules.

[292] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims.

[293] 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 invention. For example, all the techniques and apparatus described herein 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 5:

SEQ ID NO.:	Label	SEQUENCE
SEQ ID NO.: 1	E. coli wild-type TyrRS (synthetase) polynucleotide	ATGGCAAGCAGTAACTTGATTAAACAATTGCAAGAGCGGGGGCTGGTA GCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGACTGGCGCAAGG CCCGATCGCGCTCTATTGCGGCTTCGATCCTACCGCTGACAGCTTGCAT TTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTCCAGCAGGCGG GCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGTCTGATTGGCG ACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACCGAAGAAACTG TTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCCCCGTTCCTCG ATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCGAACAACTATGACT GGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATATTGGCAAACA CTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAAGCAGCGTCT CAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCCTACAACCTG TTGCAGGGTTATGACTTCGCCTGTCTGAACAAACAGTACGGTGTGGTGC TGCAAATTGGTGGTTCTGACCAGTGGGGTAACATCACTTCTGGTATCGA CCTGACCCGTCGTCTGCATCAGAATCAGGTGTTTGGCCTGACCGTTCCG CTGATCACTAAAGCAGATGGCACCAAATTTGGTAAAACTGAAGGCGGC GCAGTCTGGTTGGATCCGAAGAAAACCAGCCCGTACAAATTCTACCAG TTCTGGATCAACACTGCGGATGCCGACGTTTACCGCTTCCTGAAGTTCT TCACCTTTATGAGCATTGAAGAGATCAACGCCCTGGAAGAAGAAGATA AAAACAGCGGTAAAGCACCGCGCGCCCAGTATGTACTGGCGGAGCAG GTGACTCGTCTGGTTCACGGTGAAGAAGGTTTACAGGCGGCAAAACGT ATTACCGAATGCCTGTTCAGCGGTTCTTTGAGTGCGCTGAGTGAAGCGG ACTTCGAACAGCTGGCGCAGGACGGCGTACCGATGGTTGAGATGGAAA AGGGCGCAGACCTGATGCAGGCACTGGTCGATTCTGAACTGCAACCTT CCCGTGGTCAGGCACGTAAAACTATCGCCTCCAATGCCATCACCATTAA CGGTGAAAAACAGTCCGATCCTGAATACTTCTTTAAAGAAGAAGATCG TCTGTTTGGTCGTTTTACCTTACTGCGTCGCGGTAAAAAGAATTACTGT CTGATTTGCTGGAAATAA
SEQ ID NO.: 2	E. coli wild-type TyrRS (synthetase) Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALYCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYDFACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 3	pOMe-1 Synthetase polynucleotide	ATGGCAAGCAGTAACTTGATTAAACAATTGCAAGAGCGGGGGCTGGTA GCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGACTGGCGCAAGGC CCGATCGCACTCGTGTGTGGCTTCGATCCTACCGCTGACAGCTTGCATTT GGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTCCAGCAGGCGGGC CACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGTCTGATTGGCGAC CCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACCGAAGAAACTGTT CAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCCCCGTTCCTCGATT TCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATAATTATGACTGGTT CGGCAATATGAATGTGCTGACCTTCCTGCGCGATATTGGCAAACACTTC TCCGTTAACCAGATGATCAACAAAGAAGCGGTTAAGCAGCGTCTCAAC CGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCCTACAACCTGCTGC AGGGTTATAGTATGGCCTGTTTGAACAAACAGTACGGTGTGGTGCTGCA

		AATTGGTGGTTCTGACCAGTGGGGTAACATCACTTCTGGTATCGACCTG ACCCGTCGTCTGCATCAGAATCAGGTGTTTGGCCTGACCGTTCCGCTGA TCACTAAAGCAGATGGCACCAAATTTGGTAAAACTGAAGGCGGCGCAG TCTGGTTGGATCCGAAGAAAACCAGCCCGTACAAATTCTACCAGTTCTG GATCAACACTGCGGATGCCGACGTTTACCGCTTCCTGAAGTTCTTCACC TTTATGAGCATTGAAGAGATCAACGCCCTGGAAGAAGAAGATAAAAAC AGCGGTAAAGCACCGCGCGCCCAGTATGTACTGGCGGAGCAGGTGACT CGTCTGGTTCACGGTGAAGAAGGTTTACAGGCGGCAAAACGTATTACC GAATGCCTGTTCAGCGGTTCTTTGAGTGCGCTGAGTGAAGCGGACTTCG AACAGCTGGCGCAGGACGGCGTACCGATGGTTGAGATGGAAAAGGGCG CAGACCTGATGCAGGCACTGGTCGATTCTGAACTGCAACCTTCCcGTGG TCAGGCACGTAAAACTATCGCCTCCAATGCCATCACCATTAACGGTGAA AAACAGTCCGATCCTGAATACTTCTTTAAAGAAGAAGATCGTCTGTTTG GTCGTTTTACCTTACTGCGTCGCGGTAAAAAGAATTACTGTCTGATTTGC TGGAAATAA
SEQ ID NO.: 4	pOMe-2 Synthetase polynucleotide	ATGGCAAGCAGTAACTTGATTAAACAATTGCAAGAGCGGGGGCTGGTA gCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGACTGGCGCAAGGC CCGATCGCACTCACTTGTGGCTTCGATCCTACCGCTGACAGCTTGCATTT GGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTCCAGCAGGCGGGC CACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGTCTGATTGGCGAC CCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACCGAAGAAACTGTT CAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCCCCGTTCCTCGATT TCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATAATTATGACTGGTT CAGCAATATGAATGTGCTGACCTTCCTGCGCGATATTGGCAAACACTTC TCCGTTAACCAGATGATCAACAAAGAAGCGGTTAAGCAGCGTCTCAAC CGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCCTACAACCTGCTGC AGGGTTATACGTATGCCTGTCTGAACAAACAGTACGGTGTGGTGCTGCA AATTGGTGGTTCTGACCAGTGGGGTAACATCACTTCTGGTATCGACCTG ACCCGTCGTCTGCATCAGAATCAGGTGTTTGGCCTGACCGTTCCGCTGA TCACTAAAGCAGATGGCACCAAATTTGGTAAAACTGAAGGCGGCGCAG TCTGGTTGGATCCGAAGAAAACCAGCCCGTACAAATTCTACCAGTTCTG GATCAACACTGCGGATGCCGACGTTTACCGCTTCCTGAAGTTCTTCACC TTTATGAGCATTGAAGAGATCAACGCCCTGGAAGAAGAAGATAAAAAC AGCGGTAAAGCACCGCGCGCCCAGTATGTACTGGCGGAGCAGGTGACT CGTCTGGTTCACGGTGAAGAAGGTTTACAGGCGGCAAAACGTATTACC GAATGCCTGTTCAGCGGTTCTTTGAGTGCGCTGAGTGAAGCGGACTTCG AACAGCTGGCGCAGGACGGCGTACCGATGGTTGAGATGGAAAAGGGCG CAGACCTGATGCAGGCACTGGTCGATTCTGAACTGCAACCTTCCCGTGG TCAGGCACGTAAAACTATCGCCTCCAATGCCATCACCATTAACGGTGAA AAACAGTCCGATCCTGAATACTTCTTTAAAGAAGAAGATCGTCTGTTTG GTCGTTTTACCTTACTGCGTCGCGGTAAAAAGAATTACTGTCTGATTTGC TGGAAATAA
SEQ ID NO.: 5	pOMe-3 Synthetase polynucleotide	ATGGCAAGCAGTAACTTGATTAAACAATTGCAAGAGCGGGGGCTGGTA GCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGACTGGCGCAAGGC CCGATCGCACTCGTGTGTGGCTTCGATCCTACCGCTGACAGCTTGCATTT GGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTCCAGCAGGCGGGC CACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGTCTGATTGGCGAC CCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACCGAAGAAACTGTT CAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCCCCGTTCCTCGATT TCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATAATTATGACTGGTT CGGCAATATGAATGTGCTGACCTTCCTGCGCGATATTGGCAAACACTTC TCCGTTAACCAGATGATCAACAAAGAAGCGGTTAAGCAGCGTCTCAAC CGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCCTACAACCTGCTGC AGGGTTATAGTATGGCCTGTTTGAACAAACAGTACGGTGTGGTGCTGCA AATTGGTGGTTCTGACCAGTGGGGTAACATCACTTCTGGTATCGACCTG ACCCGTCGTCTGCATCAGAATCAGGTGTTTGGCCTGACCGTTCCGCTGA TCACTAAAGCAGATGGCACCAAATTTGGTAAAACTGAAGGCGGCGCAG TCTGGTTGGATCCGAAGAAAACCAGCCCGTACAAATTCTACCAGTTCTG GATCAACACTGCGGATGCCGACGTTTACCGCTTCCTGAAGTTCTTCACC

		TTTATGAGCATTGAAGAGATCAACGCCCTGGAAGAAGAAGATAAAAAC AGCGGTAAAGCACCGCGCGCCCAGTATGTACTGGCGGAGCAGGTGACT CGTCTGGTTCACGGTGAAGAAGGTTTACAGGCGGCAAAACGTATTACC GAATGCCTGTTCAGCGGTTCTTTGAGTGCGCTGAGTGAAGCGGACTTCG AACAGCTGGCGCAGGACGGCGTACCGATGGTTGAGATGGAAAAGGGCG CAGACCTGATGCAGGCACTGGTCGATTCTGAACTGCAACCTTCCCGTGG TCAGGCACGTAAAACTATCGCCTCCAATGCCATCACCATTAACGGTGAA AAACAGTCCGATCCTGAATACTTCTTTAAAGAAGAAGATCGTCTGTTTG GTCGTTTTACCTTACTGCGTCGCGGTAAAAAGAATTACTGTCTGATTTGC TGGAAATAA
SEQ ID NO.: 6	pOMe-4 Synthetase polynucleotide	ATGGCAAGCAGTAACTTGATTAAACAATTGCAAGAgCGGGGGCTGGTA GCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGACTGGCGCAAGGC CCGATCGCACTCGTGTGTGGCTTCGATCCTACCGCTGACAGCTTGCATTT GGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTCCAGCAGGCGGGC CACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGTCTGATTGGCGAC CCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACCGAAGAAACTGTT CAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCCCCGTTCCTCGATT TCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATAATTATGACTGGTT CGGCAATATGAATGTGCTGACCTTCCTGCGCGATATTGGCAAACACTTC TCCGTTAACCAGATGATCAACAAAGAAGCGGTTAAGCAGCGTCTCAAC CGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCCTACAACCTGCTGC AGGGTTATAGTATGGCCTGTTTGAACAAACAGTACGGTGTGGTGCTGCA AATTGGTGGTTCTGACCAGTGGGGTAACATCACTTCTGGTATCGACCTG ACCCGTCGTCTGCATCAGAATCAGGTGTTTGGCCTGACCGTTCCGCTGA TCACTAAAGCAGATGGCACCAAATTTGGTAAAACTGAAGGCGGCGCAG TCTGGTTGGATCCGAAGAAAACCAGCCCGTACAAATTCTACCAGTTCTG GATCAACACTGCGGATGCCGACGTTTACCGCTTCCTGAAGTTCTTCACC TTTATGAGCATTGAAGAGATCAACGCCCTGGAAGAAGAAGATAAAAAC AGCGGTAAAGCACCGCGCGCCCAGTATGTACTGGCGGAGCAGGTGACT CGTCTGGTTCACGGTGAAGAAGGTTTACAGGCGGCAAAACGTATTACC GAATGCCTGTTCAGCGGTTCTTTGAGTGCGCTGAGTGAAgCGGACTTCG AACAGCTGGCGCAGGACGGCGTACCGATGGTTGAGATGGAAAAGGGCG CAGACCTGATGCAGGCACTGGTCGATTCTGAACTGCAACCTTCCCGTGG TCAGGCACGTAAAACTATCGCCTCCAATGCCATCACCATTAACGGTGAA AAACAGTCCGATCCTGAATACTTCTTTAAAGAAGAAGATCGTCTGTTTG GTCGTTTTACCTTACTGCGTCGCGGTAAAAAGAATTACTGTCTGATTTGC TGGAAATAA
SEQ ID NO.: 7	pOMe-5 Synthetase polynucleotide	ATGGCAAGCAGTAACTTGATTAAACAATTGCAAGAGCGGGGGCTGGTA gCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGACTGGCGCAAGGC CCGATCGCACTCACGTGTGGCTTCGATCCTACCGCTGACAGCTTGCATT TGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTCCAGCAGGCGGG CCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGTCTGATTGGCGA CCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACCGAAGAAACTGT TCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCCCCGTTCCTCGAT TTCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATAATTATGACTGGT TCGGCAATATGAATGTGCTGACCTTCCTGCGCGATATTGGCAAACACTT CTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAAGCAGCGTCTCAA CCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCCTACAGCCTGCTG CAGGGTTATACGATGGCCTGTCTGAACAAACAGTACGGTGTGGTGCTGC AAATTGGTGGTTCTGACCAGTGGGGTAACATCACTTCTGGTATCGACCT GACCCGTCGTCTGCATCAGAATCAGGTGTTTGGCCTGACCGTTCCGCTG ATCACTAAAGCAGATGGCACCAAATTTGGTAAAACTGAAGGCGGCGCA GTCTGGTTGGATCCGAAGAAAACCAGCCCGTACAAATTCTACCAGTTCT GGATCAACACTGCGGATGCCGACGTTTACCGCTTCCTGAAGTTCTTCAC CTTTATGAGCATTGAAGAGATCAACGCCCTGGAAGAAGAAGATAAAAA CAGCGGTAAAGCACCGCGCGCCCAGTATGTACTGGCGGAGCAGGTGAC TCGTCTGGTTCACGGTGAAGAAGGTTTACAGGCGGCAAAACGTATTACC GAATGCCTGTTCAGCGGTTCTTTGAGTGCGCTGAGTGAAGCGGACTTCG

		AACAGCTGGCGCAGGACGGCGTACCGATGGTTGAGATGGAAAAGGGCG CAGACCTGATGCAGGCACTGGTCGATTCTGAACTGCAACCTTCCCGTGG TCAGGCACGTAAAACTATCGCCTCCAATGCCATCACCATTAACGGTGAA AAACAGTCCGATCCTGAATACTTCTTTAAAGAAGAAGATCGTCTGTTTG GTCGTTTTACCTTACTGCGTCGCGGTAAAAAGAATTACTGTCTGATTTGC TGGAAATAA
SEQ ID NO.: 8	pOMe-6 (active site) Synthetase polynucleotide	CGGGGGCTGGTAGCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCACTTGTGGCTTCGATCCTACCGCTG ACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTC CAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATA ATTATGACTGGTTCAGCAATATGAATGTGCTGACCTTCCTGCGCGATAT TGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCC TACAACCTGCTGCAGGGTTATACGTATGCCTGTCTGAACAAACAGTACG GTGTG
SEQ ID NO.: 9	pOMe-7 (active site) Synthetase polynucleotide	CGGGGGCTGGTACCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCACTTGTGGCTTCGATCCTACCGCTG ACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTC CAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATA ATTATGACTGGTTCAGCAATATGAATGTGCTGACCTTCCTGCGCGATAT TGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCC TACAACCTGCTGCAGGGTTATACGTATGCCTGTCTGAACAAACAGTACG GTGTG
SEQ ID NO.: 10	pOMe-8 (active site) Synthetase polynucleotide	CGGGGGCTGGTAGCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCACTTGTGGCTTCGATCCTACCGCTG ACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTC CAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATA ATTATGACTGGTTCAGCAATATGAATGTGCTGACCTTCCTGCGCGATAT TGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCC TACAACCTGCTGCAGGGTTATACGTATGCCTGTCTGAACAAACAGTACG GTGTG
SEQ ID NO.: 11	pOMe-9 (active site) Synthetase polynucleotide	CGGGGGCTGGTAGCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCACTTGTGGCTTCGATCCTACCGCTG ACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTC CAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATA ATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATAT TGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCC TACAACCTGCTGCAGGGTTATTCGTATGCCTGTGCGAACAAACAGTACG GTGTG
SEQ ID NO.: 12	pOMe-10 (active site) Synthetase polynucleotide	CGGGGGCTGGTAGCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCACTTGTGGCTTCGATCCTACCGCTG ACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTC CAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGT

		CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATA ATTATGACTGGTTCAGCAATATGAATGTGCTGACCTTCCTGCGCGATAT TGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCC TACAACCTGCTGCAGGGTTATACGTATGCCTGTCTGAACAAACAGTACG GTGTG
SEQ ID NO.: 13	pOMe-11 (active site) Synthetase polynucleotide	CGGGGGCTGGTACCcCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCCTTTGTGGCTTCGATCCTACCGCTG ACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTC CAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATA ATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATAT TGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCC TACAACCTGCTGCAGGGTTATTCTATTGCCTGTTCGAACAAACAGTACG GTGTG
SEQ ID NO.: 14	pOMe-12 (active site) Synthetase polynucleotide	CGGGGGCTGGTAGCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCGTGTGTGGCTTCGATCCTACCGCTG ACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTC CAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATA ATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATAT TGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCC TACAACCTGCTGCAGGGTTATAGTATTGCCTGTTTGAACAAACAGTACG GTGTG
SEQ ID NO.: 15	pOMe-13 (active site) Synthetase polynucleotide	CGGGGGCTGGTACCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCGTGTGTGGCTTCGATCCTACCGCTG ACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTC CAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATA ATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATAT TGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCC TACAACCTGCTGCAGGGTTATAGTATTGCCTGTTTGAACAAACAGTACG GTGTG
SEQ ID NO.: 16	pOMe-14 (active site) Synthetase polynucleotide	CGGGGGCTGGTAGCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCTGGTGTGGCTTCGATCCTACCGCTG ACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTC CAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGT CTGATTGGCGACCCGAGCTTCAAGGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATT GTTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATAT TGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCC TACAACCTGCTGCAGGGTTATATGCGTGCCTGTGAGAACAAACAGTACG GTGTG
SEQ ID NO.: 17	p-acetylPhe-1 (active site) Synthetase	CGGGGGCTGGTAGCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCATTTGTGGCTTCGATCCTACCGCTG ACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTC

	polynucleotide	CAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATA ATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATAT TGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAA GCAGCGTCTCAACCGTGAAGGTCAGGGGATTTCGTTCACTGAGTTTTCC TACAACCTGCTGCAGGGTTATGGTATGGCCTGTGCTAACAAACAGTACG GTGTGGTGCTGCAAATTGGTGGTTCTGACCAATGGGGTAACATCACTTC TGGTATCGACCTGACCCGTCGTCTGCATCAGAATCAGGTG
SEQ ID NO.: 18	pBenzophenon -1 (active site) Synthetase polynucleotide	CAGGTGACGGACGAGGAAGCGTTAGCAGAGCGACTGGCGCAAGGCCCG ATCGCACTCGGTTGTGGCTTCGATCCTACCGCTGACAGCTTGCATTTGG GGCATCTTGTTCCATTGTTATGCCTGAAACGCTTCCAGCAGGCGGGCCA CAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGTCTGATTGGCGACCC GAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACCGAAGAAACTGTTCA GGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCCCCGTTCCTCGATTTC GACTGTGGAGAAAACTCTGCTATCGCGGCCAATAATTATGACTGGTTCG GCAATATGAATGTGCTGACCTTCCTGCGCGATATTGGCAAACACTTCTC CGTTAACCAGATGATCAACAAAGAAGCGGTTAAGCAGCGTCTCAACCG TGAAGATCAGGGGATTTCGTTCACTGAGTTTTCCTACAACCTGCTGCAG GGTTATGGTTTTGCCTGTTTGAACAAACAGTACGGTGTGGTGCTGCAAA TTGGTGGTTCTGACCAGTGGGGTAACATCACTTCTGGTATCGACCTGAC CCGTCGTCTGCATCAGAATCAGGTG
SEQ ID NO.: 19	pBenzophenon e-2 (active site) Synthetase polynucleotide	GCGTTAGCAGAGCGACTGGCGCAAGGCCCGATCGCACTCGGGTGTGGC TTCGATCCTACCGCTGACAGCTTGCATTTGGGGCATCTTGTTCCATTGTT ATGCCTGAAACGCTTCCAGCAGGCGGGCCACAAGCCGGTTGCGCTGGT AGGCGGCGCGACGGGTCTGATTGGCGACCCGAGCTTCAAAGCTGCCGA GCGTAAGCTGAACACCGAAGAAACTGTTCAGGAGTGGGTGGACAAAAT CCGTAAGCAGGTTGCCCCGTTCCTCGATTTCGACTGTGGAGAAAACTCT GCTATCGCGGCCAATAATTATGACTGGTTCGGCAATATGAATGTGCTGA CCTTCCTGCGCGATATTGGCAAACACTTCTCCGTTAACCAGATGATCAA CAAAGAAGCGGTTAAGCAGCGTCTCAACCGTGAAGATCAGGGGATTTC GTTCACTGAGTTTTCCTACAACCTGCTGCAGGGTTATGGTTATGCCTGTA TGAACAAACAGTACGGTGTGGTGCTGCAAATTGGTGGTTCTGACCAGTG GGGTAACATCACTTCTGGTATCGACCTGACCCGTCGTCTGCATCAGAAT CAGGTG
SEQ ID NO.: 20	pAzidoPhe-1 (active site) Synthetase polynucleotide	GGGCTGGTAGCCCAGGTGACGGACGNAGAAGCGTTAGCAGAGCGACTG GCGCAAGGCCCGATCGCACTCCTTTGTGGCTTCGATCCTACCGCTGACA GCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTCCAG CAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGTCTG ATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACCGAA GAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCCCCG TTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATAATT ATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATATTGG CAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAAGCA GCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCCTAC AACCTGCTGCAGGGTTATTCTATGGCCTGTGCGAACAAACAGTACGGTG TGGTGCTGCAAATTGGTGGTTCTGACCAGTGGGGTAACATCACTTCTGG TATCGACCTGACCCGTCGTCTGCATCANAATCANGTG
SEQ ID NO.: 21	pAzidoPhe-2 (active site) Synthetase polynucleotide	TTAGCAGAGCGACTGGCGCAAGGCCCGATCGCACTCGTTTGTGGCTTCG ATCCTACCGCTGACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGC CTGAAACGCTTCCAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGC GGCGCGACGGGTCTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGT AAGCTGAACACCGAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGT AAGCAGGTTGCCCCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTA TCGCGGCCAATAATTATGACTGGTTCGGCAATATGAATGTGCTGACCTT CCTGCGCGATATTGGCAAACACTTCTCCGTTAACCAGATGATCAACAAA GAAGCGGTTAAGCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTC

		ACTGAGTTTTCCTACAACCTGCTGCAGGGTTATTCTGCGGCCTGTGCGA ACAAACAGTACGGTGTGGTGCTGCAAATTGGTGGTTCTGACCAGTGGG GTAACATCACTTCTGGTATCGACCTGACCCGTCGTCTGCATCAGAATCA GGTG
SEQ ID NO.: 22	pAzidoPhe-3 (active site) Synthetase polynucleotide	GACGAGGAAGCGTTAGCAGAGCGACTGGCGCAAGGCCCGATCGCACTC CTGTGTGGCTTCGATCCTACCGCTGACAGCTTGCATTTGGGGCATCTTGT TCCATTGTTATGCCTGAAACGCTTCCAGCAGGCGGGCCACAAGCCGGTT GCGCTGGTAGGCGGCGCGACGGGTCTGATTGGCGACCCGAGCTTCAAA GCTGCCGAGCGTAAGCTGAACACCGAAGAAACTGTTCAGGAGTGGGTG GACAAAATCCGTAAGCAGGTTGCCCCGTTCCTCGATTTCGACTGTGGAG AAAACTCTGCTATCGCGGCCAATAATTATGACTGGTTCGGCAATATGAA TGTGCTGACCTTCCTGCGCGATATTGGCAAACACTTCTCCGTTAACCAG ATGATCAACAAANAAGCGGTTAAGCAGCGTCTCAACCGTGAAGATCAG GGGATTTCGTTCACTGAGTTTTCCTACAACCTGCTGCAGGGTTATTCGGC TGCCTGTGCGAACAAACAGTACGGNGNGGNGCTGCAAATTGGNGGTTC TGACCAGGGGGGTAACATCACTTCTGGTATCGACCTGACCCGTCGTCTG CATCAAAATCAGGTG
SEQ ID NO.: 23	pAzidoPhe-4 (active site) Synthetase polynucleotide	GCGTTAGCAGAGCGACTGGCGCAAGGCCCGATCGCACTCGTTTGTGGCT TCGATCCTACCGCTGACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTG TGCCTGAAACGCTTCCAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTA GGCGGCGCGACGGGTCTGATTGGCGACCCGAGCTTCAAAGCTGCCGAG CGTAAGCTGAACACCGAAGAAACTGTTCAGGAGTGGGTGGACAAAATC CGTAAGCAGGTTGCCCCGTTCCTCGATTTCGACTGTGGAGAAAACTCTG CTATCGCGGCCAATAATTATGACTGGTTCGGCAATATGAATGTGCTGAC CTTCCTGCGCGATATTGGCAAACACTTCTCCGTTAACCAGATGATCAAC AAAGAAGCGGTTAAGCAGCGTCTCAACCGTGAAGATCAGGGGATTTCG TTCACTGAGTTTTCCTACAACCTGCTGCAGGGTTATAGTGCGGCCTGTGT TAACAAACAGTACGGTGTGGTGCTGCAAATTGGTGGTTCTGACCAGTGG GGTAACATCACTTCTGGTATCGACCTGACCCGTCGTCTGCATCAGAATC ANGTG
SEQ ID NO.: 24	pAzidoPhe-5 (active site) Synthetase polynucleotide	GACGAGGAAGCGTTAGCAGAGCGACTGGCGCAAGGCCCGATCGCACTC ATTTGTGGCTTCGATCCTACCGCTGACAGCTTGCATTTGGGGCATCTTGT TCCATTGTTATGCCTGAAACGCTTCCAGCAGGCGGGCCACAAGCCGGTT GCGCTGGTAGGCGGCGCGACGGGTCTGATTGGCGACCCGAGCTTCAAA GCTGCCGAGCGTAAGCTGAACACCGAAGAAACTGTTCAGGAGTGGGTG GACAAAATCCGTAAGCAGGTTGCCCCGTTCCTCGATTTCGACTGTGGAG AAAACTCTGCTATCGCGGCCAATGATTATGACTGGTTCGGCAATATGAA TGTGCTGACCTTCCTGCGCGATATTGGCAAACACTTCTCCGTTAACCAG ATGATCAACAAAGAAGCGGTTAAGCAGCGTCTCAACCGTGAAGATCAG GGGATTTCGTTCACTGAGTTTTCCTACAACCTGCTGCAGGGTTATAATTT TGCCTGTGTGAACAAACAGTACGGTGTGGTGCTGCAAATTGGTGGTTCT GACCAGTGGGGTAACATCACTTCTGGTATCGACCTGACCCGTCGTCTGC ATCAGAATCAGGTG
SEQ ID NO.: 25	pAzidoPhe-6 (active site) Synthetase polynucleotide	CGACTGGCGCAAGGCCCGATCGCACTCACGTGTGGCTTCGATCCTACCG CTGACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGC TTCCAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACG GGTCTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAAC ACCGAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTT GCCCCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCCA ATAATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGA TATTGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTT AAGCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTT CCTACAATCTGCTGCAGGGTTATTCGGCTGCCTGTCTTAACAAACAGTA CGGTGTGGTGCTGCAAATTGGTGGTTCTGACCAGTGGGGTAACATCACT TCTGGTATCGACCTGACCCGTCGTCTGCATCAGAATCAGGTG
SEQ ID NO.: 26	pPR-EcRS-1 (propargyloxy phenylalanine synthetase)	CGGGGGCTGGTANCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCGGGTGTGGCTTCGATCCTACCGCTG ACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTC CAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGT

	(active site) Synthetase polynucleotide	CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATA ATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATAT TGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCC TACAACCTGCTGCAGGGTTATTCTATGGCCTGTTTGAACAAACAGTACG GTGTGGTGCTGCAAATTGGTGGTTCTGACCAGTGGGGTAACATCACTTC TGGTATCGACCTGANCCGTCGTCTGCATCAGAATCAGGTG
SEQ ID NO.: 27	pPR-EcRS -2 (active site) Synthetase polynucleotide	CGGGGGCTGGTAGCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCACGTGTGGCTTCGATCCTACCGCTG ACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTC CAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATA ATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATAT TGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCC TACAATCTGCTGCAGGGTTATTCGGCTGCCTGTCTTAACAAACAGTACG GTGTGGTGCTGCAAATTGGTGGTTCTGACCAGTGGGGTAACATCACTTC TGGTATCGAACCTGANCCGTCGTCTGCATCAAAATCAAGTG
SEQ ID NO.: 28	pPR-EcRS -3 (active site) Synthetase polynucleotide	CGGGGGCTGGTACCCCAAGTGACGGACGAGGAAACGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCTCTTGTGGCTTCGATCCTACCGCTG ACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTC CAGCAGGCAGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATA ATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATAT TGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCC TACAACCTGCTGCAGGGTTATACGATGGCCTGTGTGAACAAACAGTACG GTGTGGTGCTGCAAATTGGTGGTTCTGACCAGTGGGGTAACATCACTTC TGGTATCGACCTGACCCGTCGTCTGCATCAGAATCAGGTG
SEQ ID NO.: 29	pPR-EcRS -4 (active site) Synthetase polynucleotide	CGGGGGCTGGTAGCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCGCGTGCGGCTTCGATCCTACCGCTG ACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTC CAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGT CTGATTGGCGACCCGAGCTTCAAGGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATA ATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATAT TGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCC TACAACCTGCTGCAGGGTTATTCTTATGCCTGTCTTAACAAACAGTACG GTGTGGTGCTGCAAATTGGTGGTTCTGACCAGTGGGGTAACATCACTTC TGGTATCGACCTGACCCGTCGTCTGCATCAGAATCAGGTG
SEQ ID NO.: 30	pPR-EcRS-5 (active site) Synthetase polynucleotide	CGGGGGCTGGTAGCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCGCGTGTGGCTTCGATCCTACCGCTG ACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTC CAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATA ATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATAT TGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCC TACAACCTGCTGCAGGGTTATACGATGGCCTGTTGTAACAAACAGTACG

		GTGTGGTGCTGCAAATTGGTGGTTCTGACCAGTGGGGTAACATCACTTC TGGTATCGACCTGACCCGTCGTCTGCATCAGAATCAGGTG
SEQ ID NO.: 31	pPR-EcRS-6 (active site) Synthetase polynucleotide	CGGGGGCTGGTACCCCAAGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCACGTGTGGCTTCGATCCTACCGCTG ACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTC CAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATA ATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATAT TGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCGCTGAGTTTTCC TACAACCTGCTGCAGGGTTATACGTTTGCCTGTATGAACAAACAGTACG GTGTGGTGCTGCAAATTGGTGGTTCTGACCAGTGGGGTAACATCACTTC TGGTATCGACCTGACCCGTCGTCTGCATCAGAATCAGGTG
SEQ ID NO.:32	pPR-EcRS-7 (active site) Synthetase polynucleotide	GTGACGGACGAGGAAGCGTTAGCAGAGCGACTGGCGCAAGGCCCGATC GCACTCACGTGTGGCTTCGATCCTACCGCTGACAGCTTGCATTTGGGGC ATCTTGTTCCATTGTTATGCCTGAAACGCTTCCAGCAGGCGGGCCACAA GCCGGTTGCGCTGGTAGGCGGCGCGACGGGTCTGATTGGCGACCCGAG CTTCAAAGCTGCCGAGCGTAAGCTGAACACCGAAGAAACTGTTCAGGA GTGGGTGGACAAAATCCGTAAGCAGGTTGCCCCGTTCCTCGATTTCGAC TGTGGAGAAAACTCTGCTATCGCGGCCAATAATTATGACTGGTTCGGCA ATATGAATGTGCTGACCTTCCTGCGCGATATTGGCAAACACTTCTCCGT TAACCAGATGATCAACAAAGAAGCGGTTAAGCAGCGTCTCAACCGTGA AGATCAGGGGATTTCGTTCACTGAGTTTTCCTACAATCTGCTGCAGGGT TATTCGGCTGCCTGTCTTAACAAACAGTACGGTGTGGTGCTGCAAATTG GTGGTTCTGACCAGTGGGGTAACATCACTTCTGGTATCGACCTGACCCG TCGTCTGCATCAGAATCAGGTG
SEQ ID NO.: 33	pPR-EcRS-8 (active site) Synthetase polynucleotide	CGGGGGCTGGTAGCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCGTTTGTGGCTTCGATCCTACCGCTG ACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTC CAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATA ATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATAT TGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCC TACAACCTGCTGCAGGGTTATTCGATGGCCTGTACGAACAAACAGTACG GTGTGGTGCTGCAAATTGGTGGTTCTGACCAGTGGGGTAACATCACTTC TGGTATCGACCTGACCCGTCGTCTGCATCAGAATCAGGTG
SEQ ID NO.: 34	pPR-EcRS-9 (active site) Synthetase polynucleotide	CGGGGGCTGGTANCCCAAGTGACGGACGGGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCAGTTGTGGCTTCGATCCTACCGCTG ACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTC CAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCCTCGATCTCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATA ATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATAT TGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCC TACAACCTGCTGCAGGGTTATAGTTTTGCCTGTCTGAACAAACAGTACG GTGTGGTGCTGCAAATTGGTGGTTCTGACCAGTGGGGTAACATCACTTC TGGTATCGACCTGACCCGTCGTCTGCATCAGAATCAGGTG
SEQ ID NO.: 35	pPR-EcRS-10 (active site) Synthetase polynucleotide	CGGGGGCTGGTAGCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCACGTGTGGCTTCGATCCTACCGCTG ACAGCTTGCATTTGGGGCATCTTGTTCCATTGTTATGCCTGAAACGCTTC CAGCAGGCGGGCCACAAGCCGGTTGCGCTGGTAGGCGGCGCGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC

		GAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCCTCGATTTCGACTGTGGAGAAAACTCTGCTATCGCGGCCAATA ATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATAT TGGCAAACACTTCTCCGTTAACCAGATGATCAACAAAGAAGCGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTTCC TACAACCTGCTGCAGGGTTATACGTTTGCCTGTACTAACAAACAGTACG GTGTGGTGCTGCAAATTGGTGGTTCTGACCAGTGGGGTAACATCACTTC TGGTATCGACCTGACCCGTCGTCTGCATCAGAATCAGGTG
SEQ ID NO.: 36	p-iodoPheRS-1 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALVCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSYACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 37	p-iodoPheRS-2 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALICGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSMACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 38	p-iodoPheRS-3 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALVCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSMACANKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 39	OMeTyrRS-1 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALVCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSMACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 40	OMeTyrRS-2 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALTCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYTMACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 41	OMeTyrRS-3 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALTCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYTYACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL

		AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 42	OMeTyrRS-4 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALLCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSMACSNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 43	OMeTyrRS-5 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALLCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSMACANKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 44	OMeTyrRS-6 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALTCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYRMACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 45	p-acetylPheRS-1 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALICGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYGMACANKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 46	p-benzoylPheRS-1 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALGCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYGFACANKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 47	p-benzoylPheRS-2 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALGCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYGYACMNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 48	p-azidoPheRS-1	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALLCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV

	Synthetase Amino acid (aa)	DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSMACANKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 49	p-azidoPheRS-2 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALVCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSAACANKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 50	p-azidoPheRS-3 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALLCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSAACANKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 51	p-azidoPheRS-4 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALVCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSAACVNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 52	p-azidoPheRS-5 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALICGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANDYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYNFACVNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 53	p-azidoPheRS-6 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALTCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSAACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 54	pPR-EcRS-1 Synthetase Amino acid (aa) *p-propargyloxyphenylalanine*	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALGCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSMACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM

	synthetase	EKGADLMQALVDSELQPSRGQARKTIASNAITNGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 55	pPR-EcRS-2 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALTCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSAACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITNGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 56	pPR-EcRS-3 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALSCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYTMACVNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITNGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 57	pPR-EcRS-4 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALACGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSYACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITNGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 58	pPR-EcRS-5 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALACGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYTMACCNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITNGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 59	pPR-EcRS-6 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALTCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYTFACMNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITNGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 60	pPR-EcRS-7 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALTCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSVACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITNGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 61	pPR-EcRS-8 Synthetase Amino acid	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALVCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM

	(aa)	INKEAVKQRLNREDQGISFTEFSYNLLQGYSMACTNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 62	pPR-EcRS-9 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALSCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSFACLNKQYGVVLQIGGSDQW GNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTSP YKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVLA EQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEME KGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 63	pPR-EcRS-10 Synthetase Amino acid (aa)	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALTCGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYTFACTNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 64	tRNA/Tyr polynucleotide	AGCTTCCCGATAAGGGAGCAGGCCAGTAAAAAGCATTACCCCGTGGTG GGGTTCCCGAGCGGCCAAAGGGAGCAGACTCTAAATCTGCCGTCATCG ACCTCGAAGGTTCGAATCCTTCCCCCACCACCA
SEQ ID NO.: 65	tRNA/Tyr	AGCUUCCCGAUAAGGGAGCAGGCCAGUAAAAAGCAUUACCCCGUGGU GGGGUUCCCGAGCGGCCAAAGGGAGCAGACUCUAAAUCUGCCGUCAU CGACCUCGAAGGUUCGAAUCCUUCCCCCACCACCA
SEQ ID NO.: 66	Amber Mutants L3TAG	5'-ATGAAGTAGCTGTCTTCTATCGAACAAGCATGCG-3'
SEQ ID NO.: 67	Amber Mutants I13TAG	5'-CGAACAAGCATGCGATTAGTGCCGACTTAAAAAG-3'
SEQ ID NO.: 68	Amber Mutants T44TAG	5'-CGCTACTCTCCCAAATAGAAAAGGTCTCCGCTG-3'
SEQ ID NO.: 69	Amber Mutants F68TAG	5'-CTGGAACAGCTATAGCTACTGATTTTTCCTCG-3'
SEQ ID NO.: 70	Amber Mutants R110TAG	5'-GCCGTCACAGATTAGTTGGCTTCAGTGGAGACTG-3'
SEQ ID NO.: 71	Amber Mutants V114TAG	5'-GATTGGCTTCATAGGAGACTGATATGCTCTAAC-3'
SEQ ID NO.:72	Amber Mutants T121TAG	5'-GCCTCTATAGTTGAGACAGCATAGAATAATGCG-3'
SEQ ID NO.: 73	Amber Mutants I127TAG	5'-GAGACAGCATAGATAGAGTGCGACATCATCATCGG-3'
SEQ ID NO.: 74	Amber Mutants S131TAG	5'-GAATAAGTGCGACATAGTCATCGGAAGAGAGTAGTAG-3'

SEQ ID NO.: 75	Amber Mutants T145TAG	5'-GGTCAAAGACAGTTGTAGGTATCGATTGACTCGGC-3'
SEQ ID NO.: 76	Permissive Site Mutants T44F	5'-CGCTACTCTCCCCAAATTTAAAAGGTCTCCGCTG-3'
SEQ ID NO.: 77	Permissive Site Mutants T44Y	5'-CGCTACTCTCCCCAAATATAAAAGGTCTCCGCTG-3'
SEQ ID NO.: 78	Permissive Site Mutants T44W	5'-CGCTACTCTCCCCAAATGGAAAAGGTCTCCGCTG-3'
SEQ ID NO.:79	Permissive Site Mutants T44D	5'-CGCTACTCTCCCCAAAGATAAAAGGTCTCCGCTG-3'
SEQ ID NO.:80	Permissive Site Mutants T44K	5'-CGCTACTCTCCCCAAAAAAAAAAGGTCTCCGCTG-3'
SEQ ID NO.: 81	Permissive Site Mutants R110F	5'-GCCGTCACAGATTTTTTGGCTTCAGTGGAGACTG-3'
SEQ ID NO.: 82	Permissive Site Mutants R110Y	5'-GCCGTCACAGATTATTTGGCTTCAGTGGAGACTG-3'
SEQ ID NO.: 83	Permissive Site Mutants R110W	5'-GCCGTCACAGATTGGTTGGCTTCAGTGGAGACTG-3'
SEQ ID NO.: 84	Permissive Site Mutants R110D	5'-GCCGTCACAGATGATTTGGCTTCAGTGGAGACTG-3'
SEQ ID NO.: 85	Permissive Site Mutants R110K	5'-GCCGTCACAGATAAATTGGCTTCAGTGGAGACTG-3'
SEQ ID NO.: 86	p-acetylPheRS-1 Synthetase Amino acid (aa)[a]	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALICGFDPTADSLHLGH LVPLLCLKRFQQAGHKPVALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREGQGISFTEFSYNLLQGYGMACANKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTKFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITECLFSGSLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK
	A box	TRGCNNAGY
SEQ ID NO: 87	B box	GGTTCGANTCC
SEQ ID NO: 88	Single copy *B. stearothermophilus* tRNA expression insert	GGATTACGCATGCTCAGTGCAATCTTCGGTTGCCTGGACTAGCGCTCCG GTTTTTCTGTGCTGAACCTCAGGGGACGCCGACACACGTACACGTC
SEQ ID NO: 89	*B. stearothermophilus tRNA* amber suppression mutant	GACAAGTGCGGTTTTTTTCTCCAGCTCCCGATGACTTATGGC
SEQ ID NO: 90	FTam 73: forward primer	GTACGAATTCCCGAGATCTGGATTACGCATGCTCAGTGCAATCTTCGGT TGCCTGGACTAGCGCTCCGGTTTTTCTGTGC
SEQ ID	FTam 74:	AGTCCGCCGCGTTTAGCCACTTCGCTACCCCTCCGACGTGTACGTGTGT

NO: 91	Reverse primer	CGGCGTCCCCTGAGGTTCAGCACAGAAAAACCGGAGCGC
SEQ ID NO: 92	FTam 75: Forward primer	GAAGTGGCTAAACGCGGCGGACTCTAAATCCGCTCCCTTTGGGTTCGGC GGTTCGAATCCGTCCCCCTCCAGACAAGTG
SEQ ID NO: 93	FTam 76: Reverse primer	GATGCAAGCTTGATGGATCCGCCATAAGTCATCGGGAGCTGGAGAAAA AAACCGCACTTGTCTGGAGGGGGACGG
SEQ ID NO: 94	hIgG1-Fc2 DNA	CTGAGATCACCGGCGAAGGAGGGCCACCATGTACAGGATGCAACTCCT GTCTTGCATTGCACTAAGTCTTGCACTTGTCACGAATTCGATATCGGCC ATGGTTAGATCTGACAAAACTCACACATGCCCACCGTGCCCAGCACCTG AACTCCTGGGGGGACCGTCAGTCTTCCTCTTCCCCCCAAAACCCAAGGA CACCCTCATGATCTCCCGGACCCCTGAGGTCACATGCGTGGTGGTGGAC GTGAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACGGC GTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAAC AGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGC TGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAG CCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAAC CACAGGTGTACACCCTGCCCCCATCCCGGGAGGAGATGACCAAGAACC AGGTCAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCCAGCGACATCGC CGTGGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACTACAAGACCA CGCCTCCCGTGCTGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTC ACCGTGGACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCC GTGATGCATGAGGGTCTGCACAACCACTACACGCAGAAGAGCCTCTCC CTGTCTCCGGGTAAA
SEQ ID NO: 95	5' IL2 signal sequence	ATGTACAGGATGCAACTCCTGTCTTGCATTGCACTAAGTCT
SEQ ID NO: 96	hIgG1-Fc2 protein	MYRMQLLSCIALSLALVTNSISAMVRSDKTHTCPPCPAPELLGGPSVFLFPP KPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQ YNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPRE PQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPP VLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEGLHNHYTQKSLSLSPG K
SEQ ID NO: 97	IL 2 signal sequence	MYRMQLLSCIALSLALVTNS

[a] These clones also contain a Asp165Gly mutation

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A vertebrate cell or cell line comprising (i) an orthogonal aminoacyl-tRNA synthetase (O-RS) comprising an amino acid sequence set forth in SEQ ID NO: 45 and (ii) an orthogonal tRNA (O-tRNA) comprising a transcript of the nucleotide sequence set forth in SEQ ID NO: 88, wherein said O-RS is capable of preferentially aminoacylating the O-tRNA with para-acetyl-phenylalanine (*p*AF) or *p*-amino-phenylalanine.

2. The vertebrate cell or cell line according to claim 1, wherein the O-RS is capable of preferentially aminoacylating the O-tRNA with *p*AF.

3. The vertebrate cell or cell line according to claim 1 or claim 2, when used to preferentially aminoacylate and O-tRNA with *p*AF, or when used in the production of a polypeptide comprising *p*AF.

4. The vertebrate cell or cell line according to any one of claims 1 to 3, wherein the O-RS aminoacylates the O-tRNA with the *p*AF at least 10-fold more efficiently than the O-RS aminoacylates the O-tRNA with a natural amino acid.

5. The vertebrate cell or cell line according to claim 1, wherein the O-RS is capable of preferentially aminoacylating the O-tRNA with *p*-amino-phenylalanine.

6. The vertebrate cell or cell line according to claim 1 or 5, when used to preferentially aminoacylate an O-tRNA with *p*-amino-phenylalanine, or when used in the production of a polypeptide comprising *p*-amino-phenylalanine.

7. The vertebrate cell or cell line according to claim 5 or 6, wherein the O-RS aminoacylates the O-tRNA with the *p*-amino-phenylalanine at least 10-fold more efficiently than the O-RS aminoacylates the O-tRNA with a natural amino acid.

8. The vertebrate cell or cell line according to any one of claims 1 to 7, wherein the O-RS has one or more improved or enhanced enzymatic properties for *p*AF or *p*-amino-phenylalanine compared to a natural amino acid, which properties are selected from the group consisting of: higher K_m, lower K_m, higher k_{cat}, lower k_{cat}, lower k_{cat}/k_m, and higher k_{cat}/k_m.

9. The vertebrate cell or cell line according to any one of claims 1 to 8, further comprising a nucleic acid that comprises a nucleotide sequence that encodes a polypeptide, wherein the nucleotide sequence comprises a selector codon that is recognisable by an aminoacylated O-tRNA.

10. The vertebrate cell or cell line according to claim 9, wherein the selector codon is selected from the group consisting of: amber codon, ochre codon, opal codon, or four or more base codons.

11. The vertebrate cell or cell line according to claim 9, wherein the selector codon is amber codon.

12. The vertebrate cell or cell line according to claim 9, wherein the selector codon is ochre codon.

13. The vertebrate cell or cell line according to claim 9, wherein the selector codon is opal codon.

14. The vertebrate cell or cell line according to claim 9, wherein the selector codon is a a four-base codon or a codon comprising more than four bases.

15. The vertebrate cell or cell line according to any one of claims 1 to 11, wherein the O-tRNA is capable of mediating incorporation of *p*-acetyl-phenylalanine (*p*AF) or *p*-amino-phenylalanine into a polypeptide encoded by a polynucleotide that comprises a selector codon recognisable by an aminoacylated form of the O-tRNA.

16. The vertebrate cell or cell line according to claim 15, wherein the O-tRNA is capable of mediating incorporation of *p*AF into a polypeptide.

17. The vertebrate cell or cell line according to claim 15 or claim 16, when used to mediate incorporation of *p*AF into a polypeptide.

18. The vertebrate cell or cell line according to claim 15, wherein the O-tRNA is capable of mediating incorporation of *p*-amino-phenylalanine into a polypeptide.

19. The vertebrate cell or cell line according to claim 15 or claim 18, when used to mediate incorporation of *p*-amino-phenylalanine into a polypeptide.

20. The vertebrate cell or cell line according to any one of claims 1 to 19, wherein the O-RS and O-tRNA are derived from a non-vertebrate organism.

21. The vertebrate cell or cell line according to claim 20, wherein the non-vertebrate organism is *Escherichia coli* or *Bacillus stearothermophilus*.

22. The vertebrate cell or cell line according to claim 17 or 19, wherein said cell or cell line yields a polypeptide comprising *p*AF or *p*-amino-phenylalanine at a level which is at least 5% of the yield obtainable for a corresponding polypeptide comprising natural amino acids instead of *p*AF or *p*-amino-phenylalanine.

23. The vertebrate cell or cell line according to claim 17 or 19, wherein said cell or cell line yields a polypeptide in the absence of *p*AF or *p*-amino-phenylalanine at a level that is less than 30% of the yield of the polypeptide obtained in the presence of *p*AF or *p*-amino-phenylalanine.

24. The vertebrate cell or cell line according to claim 17 or 19, when used to produce a polypeptide comprising *p*AF or *p*-amino-phenylalanine, and wherein said polypeptide is a therapeutic protein, a diagnostic protein, an industrial enzyme, or a functional portion thereof.

25. The vertebrate cell or cell line according to any one of claims 17, 19 or 24, wherein the polypeptide is selected from the group consisting of: a cytokine, a growth factor, a growth factor receptor, an interferon, an interleukin, an inflammatory molecule, an oncogene product, a peptide hormone, a signal transduction molecule, a steroid hormone receptor, erythropoietin (EPO), insulin, human growth hormone, an Alpha-1 antitrypsin, an Angiostatin, an Antihemolytic factor, an antibody, an Apolipoprotein, an Apoprotein, an Atrial natriuretic factor, an Atrial natriuretic polypeptide, an Atrial peptide, a C-X-C chemokine, T39765, NAP-2, ENA-78, a Gro-a, a Gro-b, a Gro-c, an IP-10, a GCP-2, an NAP-4, an SDF-1, a PF4, a MIG, a Calcitonin, a c-kit ligand, a cytokine, a CC chemokine, a Monocyte chemoattractant protein-1, a Monocyte chemoattractant protein-2, a Monocyte chemoattractant protein-3, a Monocyte inflammatory protein-1 alpha, a Monocyte inflammatory protein-1 beta, RANTES, I309, R83915, R91733, HCC1, T58847, D31065, T64262, a CD40, a CD40 ligand, a C-kit Ligand, a Collagen, a Colony stimulating factor (CSF), a Complement factor 5a, a Complement inhibitor, a Complement receptor 1, a cytokine, DHFR, an epithelial Neutrophil Activating Peptide-78, a GROα/MGSA, a GROβ, a GROγ a MIP-1α, a MIP-1δ, a MCP-1, an Epidermal Growth Factor (EGF), an epithelial Neutrophil Activating Peptide, an Erythropoietin (EPO), an Exfoliating toxin, a Factor IX, a Factor VII, a Factor VIII, a Factor X, a Fibroblast Growth Factor (FGF), a Fibrinogen, a Fibronectin, a G-CSF, a GM-CSF, a Glucocerebrosidase, a Gonadotropin, a growth factor, a growth factor receptor, a Hedgehog protein, a Hemoglobin, a Hepatocyte Growth Factor (HGF), a Hirudin, a Human serum albumin, an ICAM-1, an ICAM-1 receptor, an LFA-1, an LFA-1 receptor, an Insulin, an Insulin-like Growth Factor (IGF), an IGF-I, an IGF-II, an interferon, an IFN-α, an IFN-β, an IFN-γ, an interleukin, an IL-1, an IL-2, an IL-3, an IL-4, an IL-5, an IL-6, an IL-7, an IL-8, an IL-9, an IL-10, an IL-11, an IL-12, a Keratinocyte Growth Factor (KGF), a Lactoferrin, a leukemia inhibitory factor, a Luciferase, a Neurturin, a Neutrophil inhibitory factor (NIF), an oncostatin M, an Osteogenic protein, an oncogene product, a Parathyroid hormone, a PD-ECSF, a PDGF, a peptide hormone, a Human Growth Hormone, a Pleiotropin, a Protein A, a Protein G, a Pyrogenic exotoxins A, B, or C, a Relaxin, a Renin, an SCF, a Soluble complement receptor I, a Soluble I-CAM 1, a Soluble interleukin receptor, a Soluble TNF receptor, a Somatomedin, a Somatostatin, a Somatotropin, a Streptokinase, a Superantigen, a Staphylococcal enterotoxins, an SEA, an SEB, an SEC1, an SEC2, an SEC3, an SED, an SEE, a steroid hormone

receptor, a Superoxide dismutase (SOD), a Toxic shock syndrome toxin, a Thymosin alpha 1, a Tissue plasminogen activator, a tumor growth factor (TGF), a TGF-α, a TGF-β, a Tumor Necrosis Factor, a Tumor Necrosis Factor alpha, a Tumor necrosis factor beta, a Tumor necrosis factor receptor (TNFR), a VLA-4 protein, a VCAM-1 protein, a Vascular Endothelial Growth Factor (VEGEF), a Urokinase, a Mos, a Ras, a Raf, a Met; a p53, a Tat, a Fos, a Myc, a Jun, a Myb, a Rel, an estrogen receptor, a progesterone receptor, a testosterone receptor, an aldosterone receptor, an LDL receptor, a SCF/c-Kit, a CD40L/CD40, a VLA-4/VCAM-1, an ICAM-1/LFA-1, a hyalurin/CD44, and a corticosterone.

26. The vertebrate cell or cell line according to any one of claims 17, 19, 24 or 25, wherein the polypeptide comprises *p*AF or *p*-amino-phenylalanine.

27. The vertebrate cell or cell line according to any one of claims 17, 19 or 24 to 26, wherein the polypeptide further comprises at least one unnatural amino acid other than *p*AF or *p*-amino-phenylalanine.

28. The vertebrate cell or cell line according to any one of claims 1 to 27, wherein the cell or cell line has been transiently transfected.

29. The vertebrate cell or cell line according to any one of claims 1 to 28, wherein the cell or cell line has been stably transfected.

30. The vertebrate cell or cell line according to any one of claims 1 to 29, wherein the cell or cell line is mammalian.

31. The vertebrate cell line according to any one of claims 1 to 30, wherein the cell or cell line is human.

32. A kit for producing the vertebrate cell or cell line according to any one of claims 1 to 31, wherein said kit comprises:(i) an isolated nucleic acid comprising the sequence set forth in SEQ ID NO: 88 or an orthogonal tRNA (O-tRNA) comprising a transcript of the sequence set forth in SEQ ID NO: 88; and (ii) an isolated nucleic acid comprising a nucleotide sequence encoding an orthogonal aminoacyl-tRNA synthetase (O-RS) or an O-RS, wherein said O-RS comprises an amino acid sequence set forth in SEQ ID NO: 45, wherein the O-RS is capable of preferentially aminoacylating the O-tRNA with at least one unnatural amino acid selected from *p*-acetyl-phenylalanine (pAF) or *p*-amino-phenylalanine.

33. The kit according to claim 32, further comprising pAF and/or *p*-amino-phenylalanine.

34. The kit according to claim 33, comprising *p*AF.

35. The kit according to claim 33, comprising *p*-amino-phenylalanine.

36. The kit according to any one of claim 32 to 35 when used to produce the vertebrate cell or cell line according to any one of claims 1 to 32.