(19) **DANMARK**

(10) **DK/EP 1828224 T3**

Patent- og
Varemærkestyrelsen

(12) Oversættelse af
europæisk patentskrift

---

(51) Int.Cl.: *C 07 C 22/00 (2006.01)* *A 61 K 38/00 (2006.01)* *C 07 C 2/00 (2006.01)*
C 07 C 227/00 (2006.01)

(45) Oversættelsen bekendtgjort den: **2016-07-18**

(80) Dato for Den Europæiske Patentmyndigheds
bekendtgørelse om meddelelse af patentet: **2016-04-06**

(86) Europæisk ansøgning nr.: **05855215.9**

(86) Europæisk indleveringsdag: **2005-12-21**

(87) Den europæiske ansøgnings publiceringsdag: **2007-09-05**

(86) International ansøgning nr.: **US2005046618**

(87) Internationalt publikationsnr.: **WO2006069246**

(30) Prioritet:
2004-12-22 US 638418 P
2004-12-22 US 639195 P
2005-07-01 US 696302 P
2004-12-22 US 638527 P
2005-07-01 US 696210 P
2005-07-01 US 696068 P

(84) Designerede stater: **AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU LV MC NL PL PT RO SE SI SK TR**

(73) Patenthaver: **Ambrx, Inc., 10975 North Torrey Pines Road, Suite 100, La Jolla CA 92037, USA**

(72) Opfinder: **MIAO, Zhenwei, 1588 Sunnyfield Place, San Diego, CA 92127, USA**
LIU, Junjie, 5004 Field Street, San Diego, CA 92110, USA
NORMAN, Thea, 1729 Bancroft Street, San Diego, CA 92102, USA
DRIVER, Russell, 6709 Jacobs Way No. 4, Madison, Wisconsin 53711, USA

(74) Fuldmægtig i Danmark: **Plougmann Vingtoft A/S, Rued Langgaards Vej 8, 2300 København S, Danmark**

(54) Benævnelse: **SAMMENSÆTNINGER INDEHOLDENDE, FREMGANGSMÅDER INDBEFATTENDE, OG ANVENDELSER AF IKKE-NATURLIGE AMINOSYRER OG POLYPEPTIDER**

(56) Fremdragne publikationer:
CA-A1- 2 110 543
US-A- 4 272 432
US-A- 4 680 311
US-A- 5 364 840
US-A- 5 484 771
US-A1- 2002 065 418
US-A1- 2003 072 746
US-B1- 6 344 483
US-B1- 6 449 281
M. KINOSHITA ET AL.: "Studies on Antibiotics an Related Substances. XXIX. Syntheses of 2-Acetamido-5-oxo-6-heptenoic Acid and 2-Acetamido-4-oxo-5-hexenoic Acid" BULLETIN OF THE CHEMICAL SOCIETY OF JAPAN, vol. 40, 1967, pages 926-931, XP002509698

Fortsættes ...

V. JULLIAN, V. MONJARDET-BAS, C. FOSSE, S. LAVIELLE,G. CHASSAING: "Synthesis of New Photoactivatable Phenylalanine Analogues and Their Incorporation into a Model Peptide - Phenylseleno Derivatives as Precursors of alpha,beta-Unsaturated Ketones in Peptide Synthesis" EUROPEAN JOURNAL OF ORGANIC CHEMISTRY, 2002, pages 1677-1684, XP002509699

GAO Y.: 'Inhibition of Grb2 SH2 Domain Binding by Non-Phosphate-Containing Ligands. 2. 4-(2-Malonyl)phenylalanine as Potent Phosphotyrosyl Mimetic' J. MED. CHEM. vol. 43, 2000, pages 911 - 920, XP002143953

JIANG G ET AL: "BETIDAMINO ACID-SCAN OF THE GNRH ANTAGONIST ACYLINE", JOURNAL OF MEDICINAL CHEMISTRY, AMERICAN CHEMICAL SOCIETY, WASHINGTON, US LNKD- DOI:10.1021/JM970024P, vol. 40, 1 January 1997 (1997-01-01), pages 3739-3748, XP002071836, ISSN: 0022-2623

RIVIER J E ET AL: "GONADOTROPIN-RELEASING HORMONE ANTAGONISTS: NOVEL MEMBERS OF THE AZALINE B FAMILY", 1 January 1995 (1995-01-01), JOURNAL OF MEDICINAL CHEMISTRY, AMERICAN CHEMICAL SOCIETY, WASHINGTON, US LNKD- DOI:10.1021/JM00014A017, PAGE(S) 2649 - 2662, XP002071775, ISSN: 0022-2623

K. SHEN, P.A. COLE: "Conversion of a Tyrosine Kinase Protein Substrate to a High Affinity Ligand by ATP Linkage", JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, vol. 125, 2003, pages 16172-16173, ISSN: 0002-7863

E. FISCHER, M. BERNIER, P. PARENT: "128. Angiotensin II Analogues. Part II. Synthesis and Incorporation of the Sulfur-Containing Aromatic Amino Acids: L-(4'-SH)Phe,L-(4'-SO2NH2)Phe and L-(4'-S-CH3)Phe", HELVETICA CHIMICA ACTA, vol. 66, 1983, pages 1355-1365, ISSN: 0018-019X

S. SELTZER, M. LIN: "Maleylacetone cis-trans-Isomerase. Mechanism of the Interaction of Coenzyme Glutathione and Substrate Maleylacetone in the Presence and Absence of Enzyme", JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, 1979, pages 3091-3097, ISSN: 0002-7863

M.M. MEHROTRA ET AL.: "alpha-Dicarbonyls as "non-charged" Arginine-Directed Affinity Labels. Novel Synthetic Routes to alpha-Dicarbonyl Analogs of the PP60c-src SH2 Domain-Targeted Phosphopeptide Ac-Tyr(OPO3H2)-Glu-Glu-Ile-Glu", BIOORGANIC & MEDICINAL CHEMISTRY LETTERS, vol. 6, 1996, pages 1941-1946, ISSN: 0960-894X

T. BARSBY ET AL.: "Bogorol A Produced in Culture by a Marine Bacillus sp. Reveals a Novel Template for Cationic Peptide Antibiotics", ORGANIC LETTERS, vol. 3, no. 3, 2001, pages 437-440, ISSN: 1523-7060

HENDRICKSON TAMARA L ET AL: "Incorporation of nonnatural amino acids into proteins", ANNUAL REVIEW OF BIOCHEMISTRY, PALTO ALTO, CA, US LNKD- DOI:10.1146/ANNUREV.BIOCHEM.73.012803.0924 29, vol. 73, 26 March 2004 (2004-03-26), pages 147-176, XP009093569, ISSN: 0066-4154

JIANG GUANGCHENG ET AL: "GnRH antagonists: A new generation of long acting analogues incorporating p-ureido-phenylalanines at positions 5 and 6", JOURNAL OF MEDICINAL CHEMISTRY, AMERICAN CHEMICAL SOCIETY, WASHINGTON, US LNKD- DOI:10.1021/JM0003900, [Online] vol. 44, no. 3, 1 February 2001 (2001-02-01), pages 453-467, XP009119476, ISSN: 0022-2623 Retrieved from the Internet: URL:http://pubs.acs.org/journals/jmcmar/in dex.html ORD - 2001-02-00>

T. HOHSAKA; M. SISIDO: "Incorporation of non-natural amino acids into proteins", BIOPOLYMERS, 2002, pages 809-815, ISSN: 0006-3525

HOHSAKA T ET AL: "Incorporation of non-natural amino acids into proteins", CURRENT OPINION IN CHEMICAL BIOLOGY, CURRENT BIOLOGY LTD, LONDON, GB LNKD- DOI:10.1016/S1367-5931(02)00376-9, vol. 6, no. 6, 1 December 2002 (2002-12-01), pages 809-815, XP008099492, ISSN: 1367-5931 [retrieved on 2002-10-18]

WANG LEI ET AL: "Addition of the keto functional group to the genetic code of Escherichia coli", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES - PNAS, NATIONAL ACADEMY OF SCIENCES, US, vol. 100, no. 1, 7 January 2003 (2003-01-07), pages 56-61, XP002405595, ISSN: 0027-8424, DOI: 10.1073/PNAS.0234824100

Jeet Kalia ET AL: "Hydrolytic Stability of Hydrazones and Oximes", Angewandte Chemie International Edition, vol. 47, no. 39, 15 September 2008 (2008-09-15), pages 7523-7526, XP055136521, ISSN: 1433-7851, DOI: 10.1002/anie.200802651

DESCRIPTION

BACKGROUND OF THE INVENTION

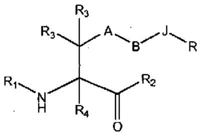
[0001] The ability to incorporate non-genetically encoded amino acids (i.e., "non-natural amino acids") into proteins permits the introduction of chemical functional groups that could provide valuable alternatives to the naturally-occurring functional groups, such as the epsilon -NH$_2$ of lysine, the sulfhydryl -SH of cysteine, the imino group of histidine, etc. Certain chemical functional groups are known to be inert to the functional groups found in the 20 common, genetically-encoded amino acids but react cleanly and efficiently to form stable linkages with functional groups that can be incorporated onto non-natural amino acids.

[0002] Methods are now available to selectively introduce chemical functional groups that are not found in proteins, that are chemically inert to all of the functional groups found in the 20 common, genetically-encoded amino acids and that may be used to react efficiently and selectively with reagents comprising certain functional groups to form stable covalent linkages.

[0003] Wang Lei et al, PNAS, 2003, describes an orthogonal tRNA-synthetase pair that allows for the efficient incorporation of a keto amino acid, *p*-acetyl-L-phenylalanine into proteins in *E. coli.*

SUMMARY OF THE INVENTION

[0004] The invention provides a method for derivatizing a polypeptide comprising an amino acid of Formula (I), the method comprising contacting the polypeptide with a reagent of Formula (XXVII), wherein Formula (I) corresponds to:



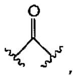
(I)

wherein:

A is optional, and when present is lower alkylene, substituted lower alkylene, lower cycloalkylene, substituted lower cycloalkylene, lower alkenylene, substituted lower alkenylene, alkynylene, lower heteroalkylene, substituted heteroalkylene, lower heterocycloalkylene, substituted lower heterocycloalkylene, arylene, substituted arylene, heteroarylene, substituted heteroarylene, alkarylene, substituted alkarylene, aralkylene, or substituted aralkylene;

B is optional, and when present is a linker selected from the group consisting of lower alkylene, substituted lower alkylene, lower alkenylene, substituted lower alkenylene, lower heteroalkylene, substituted lower heteroalkylene, -0-, -O-(alkylene or substituted alkylene)-, -S-, -S-(alkylene or substituted alkylene)-, -S(O)$_k$- where k is 1, 2, or 3, -S(O)$_k$(alkylene or substituted alkylene)-, -C(O)-, -C(O)-(alkylene or substituted alkylene)-, -C(S)-, -C(S)-(alkylene or substituted alkylene)-,-N(R')-, -NR'-(alkylene or substituted alkylene)-, -C(O)N(R')-, -CON(R')-(alkylene or substituted alkylene)-, -CSN(R')-, -CSN(R')-(alkylene or substituted alkylene)-, -N(R')CO-(alkylene or substituted alkylene)-, -N(R')C(0)0-, -S(O)$_k$N(R')-, -N(R')C(O)N(R')-, -N(R')C(S)N(R')-, -N(R')S(O)$_k$N(R')-, -N(R')-N=, -C(R')=N-, -C(R')=N-N(R')-, -C(R')=N-N=, -C(R')$_2$-N=N-, and -C(R')$_2$-N(R')-N(R')-, where each R' is independently H, alkyl, or substituted alkyl

J is



,

R is H, alkyl, substituted alkyl, cycloalkyl, or substituted cycloalkyl;

each R' is independently H, alkyl, or substituted alkyl;

R$_1$ is H, an amino protecting group, resin, amino acid, polypeptide, or polynucleotide; and

R$_2$ is OH, an ester protecting group, resin, amino acid, polypeptide, or polynucleotide;

each R$_3$ and R$_4$ is independently H, halogen, lower alkyl, or substituted lower alkyl;

wherein Formula (XXVII) corresponds to:

$$\boxed{\text{PEG}}—L—O—NH_2$$

(XXVII)

wherein:

each L is a linker independently selected from the group consisting of alkylene, substituted alkylene, alkenylene, substituted alkenylene, -0-, -O-(alkylene or substituted alkylene)-, -S-, -S-(alkylene or substituted alkylene)-, $-S(O)_k-$ where k is 1, 2, or 3, -$S(O)_k$(alkylene or substituted alkylene)-, -C(O)-, -C(O)-(alkylene or substituted alkylene)-, -C(S)-, -C(S)-(alkylene or substituted alkylene)-, -N(R')-, -NR'-(alkylene or substituted alkylene)-, -C(O)N(R')-, -CON(R')-(alkylene or substituted alkylene)-, -(alkylene or substituted alkylene)NR'C(O)O-(alkylene or substituted alkylene)-, -O-CON(R')-(alkylene or substituted alkylene)-, -CSN(R')-, -CSN(R')-(alkylene or substituted alkylene)-, -N(R')CO-(alkylene or substituted alkylene)-, -N(R')C(O)O-, -N(R')C(O)O-(alkylene or substituted alkylene)-, -$S(O)_k$N(R')-, -N(R')C(O)N(R')-, -N(R')C(O)N(R')-(alkylene or substituted alkylene)-, -N(R')C(S)N(R')-, -N(R')$S(O)_k$N(R')-, -N(R')-N=, -C(R')=N-, -C(R')=N-N(R')-, -C(R')=N-N=, -$C(R')_2$-N=N-, and -$C(R')_2$-N(R')-N(R')-, where each R' is independently H, alkyl, or substituted alkyl;

wherein "lower alkyl" is a chain alkyl having eight or fewer carbon atoms.

[0005] Described herein are methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides. In one aspect of the invention are methods for derivatizing a non-natural amino acid polypeptide. The method involves chemical derivatization.. Such derivatizations may be regioselective. Such derivatizations may be regiospecific. Such derivatizations may be rapid at ambient temperature. In further or additional embodiments, such derivatizations occur in aqueous solutions. In further or additional aspects, such derivatizations occur at a pH between about 4 and about 10. With the addition of an accelerant, such derivations may be stoichiometric, near stoichiometric or stoichiometric -like in both the non-natural amino acid containing reagent and the derivatizing reagent. Also described are methods which, with the addition of an accelerant, allow the stoichiometric, near stoichiometric or stoichiometric-like incorporation of a desired group onto a non-natural amino acid polypeptide. Also disclosed are strategies, reaction mixtures, synthetic conditions which, with the addition of an accelerant, allow the stoichiometric, near stoichiometric or stoichiometric-like incorporation of a desired group onto a non-natural amino acid polypeptide.

[0006] In one aspect of the invention are non-natural amino acids for the chemical derivatization of peptides and proteins based upon an oxime linkage. In further or additional embodiments, the non-natural amino acid is incorporated into a polypeptide, that is, such embodiments are non-natural amino acid polypeptides. In further or additional embodiments, the non-natural amino acids are functionalized on their sidechains such that their reaction with a derivatizing molecule generates an oxime linkage. In further or additional embodiments are non-natural amino acid polypeptides that can react with a derivatizing molecule to generate an oxime-containing non-natural amino acid polypeptide. In further or additional embodiments, the non-natural amino acids are selected from amino acids having carbonyl, dicarbonyl or oxime sidechains. In further or additional embodiments, the non-natural amino acids are selected from amino acids having protected or masked carbonyl, dicarbonyl or oxime sidechains. In further or additional embodiments, the non-natural amino acids comprise an oxime-masked sidechain. In further or additional embodiments, the non-natural amino acids comprise carbonyl or dicarbonyl sidechains where the carbonyl or dicarbonyl is selected from a ketone or an aldehyde. In another embodiment are non-natural amino acids containing a functional group that is capable of forming an oxime upon treatment with an appropriately functionalized co-reactant. In a further or additional embodiment, the non-natural amino acids resemble a natural amino acid in structure but contain one of the aforementioned functional groups. The non-natural amino acids may resemble phenylalanine or tyrosine (aromatic amino acids).The non-natural amino acids may resemble alanine and leucine (hydrophobic amino acids). The non-natural amino acids may have properties that are distinct from those of the natural amino acids. Such distinct properties are the chemical reactivity of the sidechain. This distinct chemical reactivity permits the sidechain of the non-natural amino acid to undergo a reaction while being a unit of a polypeptide. even though the sidechains of the naturally-occurring amino acid units in the same polypeptide do not undergo the aforementioned reaction. In a further embodiment, the sidechain of the non-natural amino acid has a chemistry orthogonal to those of the naturally-occurring amino acids. In a further embodiment, the sidechain of the non-natural amino acid comprises an electrophile-containing moiety; in a further embodiment, the electrophile-containing moiety on the sidechain of the non-natural amino acid can undergo nucleophilic attack to generate an oxime-derivatized protein. In any of the aforementioned embodiments in this paragraph, the non-natural amino acid may exist as a separate molecule or may be incorporated into a polypeptide of any length; if the latter, then the polypeptide may further incorporate naturally-occurring or non-natural amino acids.

[0007] In another aspect are carbonyl- or dicarbonyl-substituted molecules for the production of derivatized non-natural amino acid polypeptides based upon an oxime linkage. In a further embodiment are carbonyl- or dicarbonyl-substituted molecules used to derivatize oxime-containing non-natural amino acid polypeptides via an oxime exchange reaction between the derivatizing molecule and the oxime-containing peptide or protein. Also described are carbonyl- or dicarbonyl-substituted molecules that can undergo oxime exchange with an oxime-containing non-natural amino acid polypeptide in a pH range between about 4 and about 8. Also described are carbonyl- or dicarbonyl-substituted molecules used to derivatize oxime-containing or hydroxylamine-containing non-natural amino acid polypeptides via the formation of an oxime linkage between the derivatizing molecule and the oxime-containing (thus forming a new oxime linkage via an exchange-type reaction) or hydroxylamine-containing non-natural amino acid polypeptides. In a further embodiment the carbonyl- or dicarbonyl-substituted molecules are aldehyde substituted molecules. In further aspects, the carbonyl- or dicarbonyl-substituted molecules comprise a group selected from: a label; a dye; a polymer; a water-soluble polymer; a derivative of polyethylene glycol; a photocrosslinker; a cytotoxic compound; a drug; an affinity label; a photoaffinity label; a reactive compound; a resin; a second protein or polypeptide or polypeptide analog; an antibody or antibody fragment; a metal chelator; a cofactor; a fatty acid; a carbohydrate; a polynucleotide; a DNA; a RNA; an antisense polynucleotide; a saccharide, a water-soluble dendrimer, a cyclodextrin, a biomaterial; a nanoparticle; a spin label; a fluorophore, a metal-containing moiety; a radioactive moiety; a novel functional group; a group that covalently or noncovalently interacts with other molecules; a photocaged moiety; an actinic radiation excitable moiety, a ligand, a photoisomerizable moiety; biotin; a biotin analogue; a moiety incorporating a heavy atom; a chemically cleavable group; a photocleavable group; an elongated side chain; a carbon-linked sugar; a redox-active agent; an amino thioacid; a toxic moiety; an isotopically labeled moiety; a biophysical probe; a phosphorescent group; a chemiluminescent group; an electron dense group; a magnetic group; an intercalating group; a chromophore; an energy transfer agent; a biologically active agent; a detectable label; a small molecule; an inhibitory ribonucleic acid, a radionucleotide; a neutron-capture agent; a derivative of biotin; quantum dot(s); a hanotransmitter; a radiotransmitter; an abzyme, an activated complex activator, a virus, an adjuvant, an aglycan, an allergan, an angiostatin, an antihormone, an antioxidant, an aptamer, a guide RNA, a saponin, a shuttle vector, a macromolecule, a mimotope, a receptor, a reverse micelle, and any combination thereof. In further or additional embodiments, the aldehyde-substituted molecules are aldehyde-substituted polyethylene glycol (PEG) molecules. In a further embodiment, the sidechain of the non-natural amino acid has a chemistry orthogonal to those of the naturally-occurring amino acids that allows the non-natural amino acid to react selectively with the carbonyl- or dicarbonyl-substituted molecules. In a further embodiment, the sidechain of the non-natural amino acid comprises a moiety, by way of example an oxime group, which reacts selectively with the carbonyl- or dicarbonyl-containing molecule; in a further embodiment, the nucleophilic moiety on the sidechain of the non-natural amino acid can undergo electrophilic attack to generate an oxime-derivatized protein. In a further aspect related to the embodiments described in this paragraph are the modified non-natural amino acid polypeptides that result from the reaction of the derivatizing molecule with the non-natural amino acid polypeptides.

[0008] The compounds of the invention may comprise mono-, bi- and multi-functional linkers for the generation of derivatized non-natural amino acid polypeptides based upon an oxime linkage. In one embodiment are molecular linkers (bi- and multi-functional) that can be used to connect the carbonyl- or dicarbonyl-containing non-natural amino acid polypeptides to other molecules. The compounds of the invention may comprise molecular linkers (bi- and multi-functional) that can be used to connect oxime- containing non-natural amino acid polypeptides to other molecules. In another embodiment the carbonyl- or dicarbonyl-containing non-natural amino acid polypeptides comprise a ketone and/or an aldehyde sidechain. In an aspect utilizing an oxime- or hydroxylamine-containing non-natural amino acid polypeptide, the molecular linker contain a carbonyl or dicarbonyl group at one of its termini; in further embodiments, the carbonyl or dicarbonyl group is selected from an aldehyde group or a ketone group. In further or additional aspects, the hydroxylamine-substituted linker molecules are hydroxylamine-substituted polyethylene glycol (PEG) linker molecules. In further or additional embodiments, the carbonyl- or dicarbonyl-substituted linker molecules are carbonyl- or dicarbonyl-substituted polyethylene glycol (PEG) linker molecules. The phrase "other molecules" includes, by way of example only, proteins, other polymers and small molecules. In further or additional aspects, the hydroxylamine-containing molecular linkers comprise the same or equivalent groups on all termini so that upon reaction with a carbonyl- or dicarbonyl-containing non-natural amino acid polypeptide, the resulting product is the homo-multimerization of the carbonyl- or dicarbonyl-containing non-natural amino acid polypeptide. In further aspects, the homo-multimerization is a homo-dimerization. In further or additional embodiments, the carbonyl- or dicarbonyl-containing molecular linkers comprise the same or equivalent groups on all termini so that upon reaction with an oxime- containing non-natural amino acid polypeptide, the resulting product is the homo-multimerization of the oxime- containing non-natural amino acid polypeptide. In further aspects, the homo-multimerization is a homo-dimerization. In a further aspect, the sidechain of the non-natural amino acid has a chemistry orthogonal to those of the naturally-occurring amino acids that allows the non-natural amino acid to react selectively with the hydroxylamine-substituted linker molecules. In a further embodiment, the sidechain of the non-natural amino acid has a chemistry orthogonal to those of the naturally-occurring amino acids that allows the non-natural amino acid to react selectively with the carbonyl- or dicarbonyl-substituted linker molecules. In a further aspect, the sidechain of the non-natural amino acid comprises an electrophile-containing moiety that reacts selectively with the hydroxylamine-containing linker molecule; in a further aspect, the electrophile-containing moiety on the sidechain of the non-natural amino acid can undergo nucleophilic attack by the hydroxylamine-containing linker molecule to generate an oxime-derivatized protein. In a further aspect related to the embodiments described in this paragraph are the linked "modified or unmodified" non-natural amino acid polypeptides that result from the reaction of the linker molecule with the non-natural amino acid polypeptides.

[0009] The methods of the invention derivatize proteins via the condensation of carbonyl or dicarbonyl reactants to generate an oxime-based product. Included within this aspect are methods for the derivatization of proteins based upon the condensation of carbonyl- or dicarbonyl- containing reactants to generate an oxime-derivatized protein adduct. In additional or further aspects are methods to derivatize keto-containing proteins with hydroxylamine-functionalized polyethylene glycol (PEG) molecules. In yet additional or further aspects are methods to derivatize oxime-containing proteins via an oxime exchange reaction between a carbonyl- or dicarbonyl-containing derivatizing molecule and the oxime-containing peptide or protein. In yet additional or further aspects, the hydroxylamine-substituted molecule can include proteins, other polymers, and small molecules.

[0010] Described herein are methods for the chemical synthesis of carbonyl- or dicarbonyl-substituted molecules for the derivatization of oxime-substituted non-natural amino acid polypeptides.The carbonyl- or dicarbonyl-substituted molecule may be a keto-substituted molecule.The carbonyl- or dicarbonyl-substituted molecule may be an aldehyde-substituted molecule.The carbonyl- or dicarbonyl-substituted molecules may include proteins, polymers (non-branched and branched) and small molecules. Such methods may complement technology that enables the site-specific incorporation of non-natural amino acids during the *in vivo* translation of proteins. Also described are general methods for the preparation of carbonyl- or dicarbonyl-substituted molecules suitable for reaction with oxime-containing non-natural amino acid polypeptides to provide site-specifically derivatized non-natural amino acid polypeptides. Also described are methods for synthesizing carbonyl- or dicarbonyl-substituted polyethylene glycol (PEG) molecules.

[0011] Another aspect of the methods of the invention is the chemical derivatization of carbonyl- or dicarbonyl-substituted non-natural amino acid polypeptides using a hydroxylamine-containing bi-functional linker. In one aspect are methods for attaching a hydroxylamine-substituted linker to a carbonyl- or dicarbonyl-substituted protein via a condensation reaction to generate an oxime linkage. The carbonyl- or dicarbonyl-substituted non-natural amino acid may be a keto-substituted non-natural amino acid. In further or additional aspects, the non-natural amino acid polypeptides are derivatized site-specifically and/or with precise control of three-dimensional structure, using a hydroxylamine-containing bi-functional linker. In one aspect, such methods are used to attach molecular linkers (including mono- bi- and multi-functional linkers) to carbonyl- or dicarbonyl-containing (including keto-containing and aldehyde-containing) non-natural amino acid polypeptides, wherein at least one of the linker termini contains a hydroxylamine group which can link to the carbonyl- or dicarbonyl-containing non-natural amino acid polypeptides via an oxime linkage. In a further or additional aspect, these linkers are used to connect the carbonyl- ordicarbonyl-containing non-natural amino acid polypeptides to other molecules, including by way of example, proteins, other polymers (branched and non-branched).

[0012] In some embodiments, the non-natural amino acid polypeptide of the invention is linked to a water soluble polymer. In some embodiments, the water soluble polymer comprises a polyethylene glycol moiety. In some embodiments, the polyethylene glycol molecule is a bifunctional polymer. In some embodiments, the bifunctional polymer is linked to a second polypeptide. In some embodiments, the second polypeptide is identical to the first polypeptide, in other embodiments, the second polypeptide is a different polypeptide. In some embodiments, the non-natural amino acid polypeptide comprises at least two amino acids linked to a water soluble polymer comprising a poly(ethylene glycol) moiety.

[0013] The non-natural amino acid polypeptide may comprise a substitution, addition or deletion that increases affinity of the non-natural amino acid polypeptide for a receptor. In some aspects, the non-natural amino acid polypeptide comprises a substitution, addition, or deletion that increases the stability of the non-natural amino acid polypeptide. In some aspects, the non-natural amino acid polypeptide comprises a substitution, addition, or deletion that increases the aqueous solubility of the non-natural amino acid polypeptide. In some aspects, the non-natural amino acid polypeptide comprises a substitution, addition, or deletion that increases the solubility of the non-natural amino acid polypeptide produced in a host cell. In some aspects, the non-natural amino acid polypeptide comprises a substitution, addition, or deletion that modulates protease resistance, serum half-life, immunogenicity, and/or expression relative to the amino-acid polypeptide without the substitution, addition or deletion.

[0014] The non-natural amino acid polypeptidemay be an agonist, partial agonist, antagonist, partial antagonist, or inverse agonist.The agonist, partial agonist, antagonist, partial antagonist, or inverse agonist may comprise a non-natural amino acid linked to a water soluble polymer. In some embodiments, the water polymer comprises a polyethylene glycol moiety. In some aspects, the polypeptide comprising a non-natural amino acid linked to a water soluble polymer, for example, may prevent dimerization of the corresponding receptor. In some aspects, the polypeptide comprising a non-natural amino acid linked to a water soluble polymer modulates binding of the polypeptide to a binding partner, ligand or receptor. In some aspects, the polypeptide comprising a non-natural amino acid linked to a water soluble polymer modulates one or more properties or activities of the polypeptide.

[0015] In some aspects, the selector codon is selected from the group consisting of an amber codon, ochre codon, opal codon, a unique codon, a rare codon, an unnatural codon, a five-base codon, and a four-base codon.

[0016] The method of the invention are useful for making a non-natural amino acid polypeptide linked to a water soluble polymer. In some embodiments, the method comprises contacting an isolated polypeptide comprising a non-natural amino acid with a water soluble polymer comprising a moiety that reacts with the non-natural amino acid. In some embodiments, the non-natural amino acid incorporated into is reactive toward a water soluble polymer that is otherwise unreactive toward any of the 20 common amino acids. In some embodiments, the water polymer comprises a polyethylene glycol moiety. The molecular weight of the polymer may be of a wide range, including between about 100 Da and about 100,000 Da or more. The molecular weight of the polymer may be between about 100 Da and about 100,000 Da, including 100,000 Da, 95,000 Da, 90,000 Da, 85,000 Da, 80,000 Da, 75,000 Da, 70,000 Da, 65,000 Da, 60,000 Da, 55,000 Da, 50,000 Da, 45,000 Da, 40,000 Da, 35,000 Da, 30,000 Da, 25,000 Da, 20,000 Da, 15,000 Da, 10,000 Da, 9,000 Da, 8,000 Da, 7,000 Da, 6,000 Da, 5,000 Da, 4,000 Da, 3,000 Da, 2,000 Da, 1,000 Da, 900 Da, 800 Da, 700 Da, 600 Da, 500 Da, 400 Da, 300 Da, 200 Da, and 100 Da. In some embodiments, the molecular weight of the polymer is between about 100 Da and 50,000 Da. In some embodiments, the molecular weight of the polymer is between about 100 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 1,000 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 5,000 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 10,000 Da and 40,000 Da. In some embodiments, the polyethylene glycol molecule is a branched polymer. The molecular weight of the branched chain PEG may be between about 1,000 Da and about 100,000 Da, including 100,000 Da, 95,000 Da, 90,000 Da, 85,000 Da, 80,000 Da, 75,000 Da, 70,000 Da, 65,000 Da, 60,000 Da, 55,000 Da, 50,000 Da, 45,000 Da, 40,000 Da, 35,000 Da, 30,000 Da, 25,000 Da, 20,000 Da, 15,000 Da, 10,000 Da, 9,000 Da, 8,000 Da, 7,000 Da, 6,000 Da, 5,000 Da, 4,000 Da, 3,000 Da, 2,000 Da, and 1,000 Da. In some embodiments, the molecular weight of the branched chain PEG is between about 1,000 Da and 50,000 Da. In some embodiments, the molecular weight of the branched chain PEG is between about 1,000 Da and 40,000 Da. In some embodiments, the molecular weight of the branched chain PEG is between about 5,000 Da and 40,000 Da. In some embodiments, the molecular weight of the branched chain PEG is between about 5,000 Da and 20,000 Da.

[0017] Also described herein are compositions comprising a polypeptide comprising at least one of the non-natural amino acids described herein and a pharmaceutically acceptable carrier. In some embodiments, the non-natural amino acid is linked to a water soluble polymer. Also described herein are pharmaceutical compositions comprising a pharmaceutically acceptable carrier and a polypeptide, wherein at least one amino acid is substituted by a non-natural amino acid.The non-natural amino acid may comprise a saccharide moiety. The water soluble polymer may be linked to the polypeptide via a saccharide moiety. Also described herein are prodrugs of the non-natural amino acids, non-natural amino acid polypeptides, and modified non-natural amino acid polypeptides; further described herein are compositions comprising such prodrugs and a pharmaceutically acceptable carrier. Also described herein are metabolites of the non-natural amino acids, non-natural amino acid polypeptides, and modified non-natural amino acid polypeptides; such metabolites may have a desired activity that complements or synergizes with the activity of the non-natural amino acids, non-natural amino acid polypeptides, and modified non-natural amino acid polypeptides. Also described herein are the use of the non-natural amino acids, non-natural amino acid polypeptides, and modified non-natural amino acid polypeptides described herein to provide a desired metabolite to an organism, including a patient in need of such metabolite.

[0018] Also described herein are cells comprising a polynucleotide encoding the polypeptide comprising a selector codon. The cells may comprise an orthogonal RNA synthetase and/or an orthogonal tRNA for substituting a non-natural amino acid into the polypeptide. The cells may be in a cell culture, or part of a multicellular organism, including amphibians, reptiles, birds, and mammals. The non-natural amino acid polypeptide may be produced by expression of the polypeptide. Described herein are organisms that can utilize the non-natural amino acids to produce a non-natural amino acid polypeptide, including a modified non-natural amino acid polypeptide. Also described herein are organisms containing the non-natural amino acids, the non-natural amino acid polypeptides, and/or the modified non-natural amino acid polypeptides. Such organisms include unicellular and multicellular organisms, including amphibians, reptiles, birds, and mammals.The non-natural amino acid polypeptide may be produced *in vitro.The* non-natural amino acid polypeptide may be produced in cell lysate. The non-natural amino acid polypeptide may be produced by ribosomal translation.

[0019] Also described herein are methods of making a polypeptide comprising a non-natural amino acid. The methods comprise culturing cells comprising a polynucleotide or polynucleotides encoding a polypeptide, an orthogonal RNA synthetase and/or an orthogonal tRNA under conditions to permit expression of the polypeptide; and purifying the polypeptide from the cells and/or culture medium.

[0020] Also described herein are libraries of the non-natural amino acids described herein or libraries of the non-natural amino acid polypeptides described herein, or libraries of the modified non-natural amino acid polypeptides described herein, or combination libraries thereof. Also described herein are arrays containing at least one non-natural amino acid, at least one non-natural amino acid polypeptide, and/or at least one modified non-natural amino acid. Also described herein are arrays containing at least one polynucleotide encoding a polypeptide comprising a selector codon. The arrays described herein may be used to screen for the production of the non-natural amino acid polypeptides in an organism (either by detecting transcription of the polynucleotide encoding the polypeptide or by detecting the translation of the polypeptide).

[0021] Also described herein are methods for screening libraries described herein for a desired activity, or for using the arrays described herein to screen the libraries described herein, or for other libraries of compounds and/or polypeptides and/or polynucleotides for a desired activity. Also described herein is the use of such activity data from library screening to develop and discover new therapeutic agents, as well as the therapeutic agents themselves.

[0022] Also described herein are methods of increasing therapeutic half-life, serum half-life or circulation time of a polypeptide. In some aspects, the methods comprise substituting at least one non-natural amino acid for any one or more amino acids in a naturally occurring polypeptide and/or coupling the polypeptide to a water soluble polymer.

[0023] Also described herein are methods of treating a patient in need of such treatment with an effective amount of a pharmaceutical composition which comprises a polypeptide comprising a non-natural amino acid and a pharmaceutically acceptable carrier. In some embodiments, the non-natural amino acid is coupled to a water soluble polymer.

[0024] Also described are methods for treating a disorder, condition or disease, the method comprising administering a therapeutically effective amount of a non-natural amino acid polypeptide comprising at least one non-natural amino acid selected from the group consisting of an oxime-containing non-natural amino acid, a carbonyl-containing non-natural amino acid, a dicarbonyl-containing non-natural amino acid, and a hydroxylamine-containing non-natural amino acid. Such non-natural amino acids may have been biosynthetically incorporated into the polypeptide as described herein. Such non-natural amino acid polypeptide may comprise at least one non-natural amino acid selected from amino acids of Formula I-XVIII, XXX-XXXIV(A&B), or XXXX-XXXXIII.

[0025] Also described herein are methods for treating a disorder, condition or disease, the method comprising administering a therapeutically effective amount of a non-natural amino acid polypeptide comprising at least one oxime-containing non-natural amino acid and the resulting biosynthetic oxime-containing non-natural amino acid polypeptide increases the bioavailability of the polypeptide relative to the homologous naturally-occurring amino acid polypeptide.

[0026] Also described herein are methods for treating a disorder, condition or disease, the method comprising administering a therapeutically effective amount of a non-natural amino acid polypeptide comprising at least one oxime-containing non-natural amino acid and the resulting biosynthetic oxime-containing non-natural amino acid polypeptide increases the safety profile of the polypeptide relative to the homologous naturally-occurring amino acid polypeptide.

[0027] Also described herein are methods for treating a disorder, condition or disease, the method comprising administering a therapeutically effective amount of a non-natural amino acid polypeptide comprising at least one oxime-containing non-natural amino acid and the resulting biosynthetic oxime-containing non-natural amino acid polypeptide increases the water solubility of the polypeptide relative to the homologous naturally-occurring amino acid polypeptide.

[0028] Also described herein are methods for treating a disorder, condition or disease, the method comprising administering a therapeutically effective amount of a non-natural amino acid polypeptide comprising at least one oxime-containing non-natural amino acid and the resulting biosynthetic oxime-containing non-natural amino acid polypeptide increases the therapeutic half-life of the polypeptide relative to the homologous naturally-occurring amino acid polypeptide.

[0029] Also described herein are methods for treating a disorder, condition or disease, the method comprising administering a therapeutically effective amount of a non-natural amino acid polypeptide comprising at least one oxime-containing non-natural amino acid and the resulting biosynthetic oxime-containing non-natural amino acid polypeptide increases the serum half-life of the polypeptide relative to the homologous naturally-occurring amino acid polypeptide.

[0030] Also described herein are methods for treating a disorder, condition or disease, the method comprising administering a therapeutically effective amount of a non-natural amino acid polypeptide comprising at least one oxime-containing non-natural amino acid and the resulting biosynthetic oxime-containing non-natural amino acid polypeptide extends the circulation time of the polypeptide relative to the homologous naturally-occurring amino acid polypeptide.

[0031] Also described herein are methods for treating a disorder, condition or disease, the method comprising administering a therapeutically effective amount of a non-natural amino acid polypeptide comprising at least one oxime-containing non-natural amino acid and the resulting biosynthetic oxime-containing non-natural amino acid polypeptide modulates the biological activity of the polypeptide relative to the homologous naturally-occurring amino acid polypeptide.

[0032] Also described herein are methods for treating a disorder, condition or disease, the method comprising administering a therapeutically effective amount of a non-natural amino acid polypeptide comprising at least one oxime-containing non-natural amino acid and the resulting biosynthetic oxime-containing non-natural amino acid polypeptide modulates the immunogenicity of the polypeptide relative to the homologous naturally-occurring amino acid polypeptide.

[0033] It is to be understood that the methods and compositions described herein are not limited to the particular methodology, protocols, cell lines, constructs, and reagents described herein and as such may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the methods and compositions described herein, which will be limited only by the appended claims.

[0034] As used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly indicates otherwise.

[0035] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which the inventions described herein belong. Although any methods, devices, and materials similar or equivalent to those described herein can be used in the practice or testing of the inventions described herein, the preferred methods, devices and materials are now described.

[0036] The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors described herein are not entitled to antedate such disclosure by virtue of prior invention or for any other reason.

[0037] The term "affinity label," as used herein, refers to a label which reversibly or irreversibly binds another molecule, either to modify it, destroy it, or form a compound with it. By way of example, affinity labels include, enzymes and their substrates, or antibodies and their antigens.

[0038] The terms "alkoxy," "alkylamino" and "alkylthio" are used in their conventional sense, and refer to alkyl groups linked to molecules via an oxygen atom, an amino group, a sulfur atom, respectively.

[0039] The term "alkyl," by itself or as part of another molecule, means, unless otherwise stated, a straight or branched chain, or cyclic hydrocarbon radical, or combination thereof, which may be fully saturated, mono- or polyunsaturated and can include di- and multivalent radicals, having the number of carbon atoms designated (i.e. C1-C10 means one to ten carbons). Examples of saturated hydrocarbon radicals include groups such as methyl, ethyl, n-propyl, isopropyl, n-butyl, t-butyl, isobutyl, sec-butyl, cyclohexyl, (cyclohexyl)methyl, cyclopropylmethyl, homologs and isomers of, for example, n-pentyl, n-hexyl, n-heptyl, n-octyl, and the like. An unsaturated alkyl group is one having one or more double bonds or triple bonds. Examples of unsaturated alkyl groups include vinyl, 2-propenyl, crotyl, 2-isopentenyl, 2-(butadienyl), 2,4-pentadienyl, 3-(1,4-pentadienyl), ethynyl, 1- and 3-propynyl, 3-butynyl, and the higher homologs and isomers. The term "alkyl," unless otherwise noted, is also meant to include those derivatives of alkyl defined in more detail herein, such as "heteroalkyl", "haloalkyl" and "homoalkyl".

[0040] The term "alkylene" by itself or as part of another molecule means a divalent radical derived from an alkane, as exemplified by (-CH2-)n, wherein n may be 1 to about 24. By way of example only, such groups include groups having 10 or fewer carbon atoms such as the structures -CH2CH2- and -CH2CH2CH2CH2-. A "lower alkyl" or "lower alkylene" is a shorter chain alkyl or alkylene group, generally having eight or fewer carbon atoms. The term "alkylene," unless otherwise noted, is also meant to include those groups described herein as "heteroalkylene."

[0041] The term "amino acid" refers to naturally occurring and non-natural amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally encoded amino acids are the 20 common amino acids (alanine, arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, and valine) and pyrolysine and selenocysteine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, by way of example only, an
⍰-

-carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group. Such analogs may have modified R groups (by way of example, norleucine) or may have modified peptide backbones, while still retaining the same basic chemical structure as a naturally occurring amino acid. Examples of amino acid analogs include homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium.

[0042] Amino acids may be referred to herein by either their name, their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Additionally, nucleotides, may be referred to by their commonly accepted single-letter codes.

[0043] An "amino terminus modification group" refers to any molecule that can be attached to a terminal amine group. By way of example, such terminal amine groups may be at the end of polymeric molecules, wherein such polymeric molecules include polypeptides, polynucleotides, and polysaccharides. Terminus modification groups include various water soluble polymers, peptides or proteins. By way of example only, terminus modification groups include polyethylene glycol or serum albumin. Terminus modification groups may be used to modify therapeutic characteristics of the polymeric molecule, including increasing the serum half-life of peptides.

[0044] By "antibody fragment" is meant any form of an antibody other than the full-length form. Antibody fragments herein include antibodies that are smaller components that exist within full-length antibodies, and antibodies that have been engineered. Antibody fragments include Fv, Fc, Fab, and (Fab')2, single chain Fv (scFv), diabodies, triabodies, tetrabodies, bifunctional hybrid antibodies, CDR1, CDR2, CDR3, combinations of CDR's, variable regions, framework regions, constant regions, heavy chains, light chains, and variable regions, and alternative scaffold non-antibody molecules, bispecific antibodies, and the like (Maynard & Georgiou, 2000, Annu. Rev. Biomed. Eng. 2:339-76; Hudson, 1998, Curr. Opin. Biotechnol. 9:395-402). Another functional substructure is a single chain Fv (scFv), comprised of the variable regions of the immunoglobulin heavy and light chain, covalently connected by a peptide linker (S-z Hu et al., 1996, Cancer Research, 56, 3055-3061). These small (Mr 25,000) proteins generally retain specificity and affinity for antigen in a single polypeptide and can provide a convenient building block for larger, antigen-specific molecules. Unless specifically noted otherwise, statements and claims that use the term "antibody" or "antibodies" specifically includes "antibody fragment" and "antibody fragments."

[0045] The term "aromatic" or "aryl", as used herein, refers to a closed ring structure which has at least one ring having a conjugated pi electron system and includes both carbocyclic aryl and heterocyclic aryl (or "heteroaryl" or "heteroaromatic") groups. The carbocyclic or heterocyclic aromatic group may contain from 5 to 20 ring atoms. The term includes monocyclic rings linked covalently or fused-ring polycyclic (i.e., rings which share adjacent pairs of carbon atoms) groups. An aromatic group can be unsubstituted or substituted. Examples of "aromatic" or "aryl", groups include phenyl, 1-naphthyl, 2-naphthyl, 4-biphenyl, anthracenyl, and phenanthracenyl. Substituents for each of the above noted aryl and heteroaryl ring systems are selected from the group of acceptable substituents described herein.

[0046] For brevity, the term "aromatic" or "aryl" when used in combination with other terms (including aryloxy, arylthioxy, aralkyl) includes both aryl and heteroaryl rings as defined above. Thus, the term "aralkyl" or "alkaryl" is meant to include those radicals in which an aryl group is attached to an alkyl group (including benzyl, phenethyl, pyridylmethyl and the like) including those alkyl groups in which a carbon atom (including, a methylene group) has been replaced by a heteroatom, by way of example only, by an oxygen atom. Examples of such aryl groups include phenoxymethyl, 2-pyridyloxymethyl, 3-(1-naphthyloxy)propyl, and the like.

[0047] The term "arylene", as used herein, refers to a divalent aryl radical. Examples of "arylene" include phenylene, pyridinylene, pyrimidinylene and thiophenylene. Substituents for arylene groups are selected from the group of acceptable substituents described herein.

[0048] A "bifunctional polymer", also referred to as a "bifunctional linker", refers to a polymer comprising two functional groups that are capable of reacting specifically with other moieties to form covalent or non-covalent linkages. Such moieties may include the side groups on natural or non-natural amino acids or peptides which contain such natural or non-natural amino acids. By way of example only, a bifunctional linker may have a functional group reactive with a group on a first peptide, and another functional group which is reactive with a group on a second peptide, whereby forming a conjugate that includes the first peptide, the bifunctional linker and the second peptide. Many procedures and linker molecules for attachment of various compounds to peptides are known. See, e.g., European Patent Application No. 188,256; U.S. Patent Nos. 4,671,958, 4,659,839, 4,414,148, 4,699,784; 4,680,338; and 4,569,789. A "multi-functional polymer" also referred to as a "multi-functional linker", refers to a polymer comprising two or more functional groups that are capable of reacting with other moieties. Such moieties may include the side groups on natural or non-natural amino acids or peptides which contain such natural or non-natural amino acids. (including amino acid side groups) to form covalent or non-covalent linkages. A bi-functional polymer or multi-functional polymer may be any desired length or molecular weight, and may be selected to provide a particular desired spacing or conformation between one or more molecules linked to a compound and molecules it binds to or the compound.

[0049] The term "bioavailability," as used herein, refers to the rate and extent to which a substance or its active moiety is delivered from a pharmaceutical dosage form and becomes available at the site of action or in the general circulation. Increases in bioavailability refers to increasing the rate and extent a substance or its active moiety is delivered from a pharmaceutical dosage form and becomes available at the site of action or in the general circulation. By way of example, an increase in bioavailability may be indicated as an increase in concentration of the substance or its active moiety in the blood when compared to other substances or active moieties. An example of a method to evaluate increases in bioavailability is given in examples 88-92. This method may be used for evaluating the bioavailability of any polypeptide.

[0050] The term "biologically active molecule", "biologically active moiety" or "biologically active agent" when used herein means any substance which can affect any physical or biochemical properties of a biological system, pathway, molecule, or interaction relating to an organism, including viruses, bacteria, bacteriophage, transposon, prion, insects, fungi, plants, animals, and humans. In particular, as used herein, biologically active molecules include any substance intended for diagnosis, cure, mitigation, treatment, or prevention of disease in humans or other animals, or to otherwise enhance physical or mental well-being of humans or animals. Examples of biologically active molecules include, peptides, proteins, enzymes, small molecule drugs, hard drugs, soft drugs, carbohydrates, inorganic atoms or molecules, dyes, lipids, nucleosides, radionuclides, oligonucleotides, toxins, cells, viruses, liposomes, microparticles and micelles. Classes of biologically active agents that are suitable for use with the methods and compositions described herein include drugs, prodrugs, radionuclides, imaging agents, polymers, antibiotics, fungicides, anti-viral agents, anti-inflammatory agents, antitumor agents, cardiovascular agents, anti-anxiety agents, hormones, growth factors, steroidal agents, microbially derived toxins, and the like.

[0051] By "modulating biological activity" is meant increasing or decreasing the reactivity of a polypeptide, altering the selectivity of the polypeptide, enhancing or decreasing the substrate selectivity of the polypeptide. Analysis of modified biological activity can be performed by comparing the biological activity of the non-natural polypeptide to that of the natural polypeptide.

[0052] The term "biomaterial," as used herein, refers to a biologically-derived material material obtained from bioreactors and/or from recombinant methods and techniques.

[0053] The term "biophysical probe," as used herein, refers to probes which can detect or monitor structural changes in molecules. Such molecules include to, proteins and the "biophysical probe" may be used to detect or monitor interaction of proteins with other macromolecules. Examples of biophysical probes include spin-labels, a fluorophores, and photoactivatible groups.

[0054] The term "biosynthetically," as used herein, refers to any method utilizing a translation system (cellular or non-cellular), including use of at least one of the following components: a polynucleotide, a codon, a tRNA, and a ribosome. By way of example, non-natural amino acids may be "biosynthetically incorporated" into non-natural amino acid polypeptides using the methods and techniques described in section VIII *"In vivo generation of polypeptides comprising non-natural amino acids"*, and in example 14. Additionally, the methods for the selection of useful non-natural amino acids which may be "biosynthetically incorporated" into non-natural amino acid polypeptides are described in examples 15-16.

[0055] The term "biotin analogue," or also referred to as "biotin mimic", as used herein, is any molecule, other than biotin, which bind with high affinity to avidin and/or streptavidin.

[0056] The term "carbonyl" as used herein refers to a group containing at a moiety selecting from the group consisting of -C(O)-, -S(O)-, -S(O)$_2$-, and -C(S)-, including groups containing a least one ketone group, and/or at least one aldehyde groups, and/or at least one ester group, and/or at least one carboxylic acid group, and/or at least one thioester group. Such carbonyl groups include ketones, aldehydes, carboxylic acids, esters, and thioesters. In addition, such groups may be part of linear, branched, or cyclic molecules.

[0057] The term "carboxy terminus modification group" refers to any molecule that can be attached to a terminal carboxy group. By way of example, such terminal carboxy groups may be at the end of polymeric molecules, wherein such polymeric molecules include polypeptides, polynucleotides, and polysaccharides. Terminus modification groups include various water soluble polymers, peptides or proteins. By way of example only, terminus modification groups include polyethylene glycol or serum albumin. Terminus modification groups may be used to modify therapeutic characteristics of the polymeric molecule, including increasing the serum half-life of peptides.

[0058] The term "chemically cleavable group," also referred to as "chemically labile", as used herein, refers to a group which breaks or cleaves upon exposure to acid, base, oxidizing agents, reducing agents, chemical inititiators, or radical initiators.

[0059] The term "chemiluminescent group," as used herein, refers to a group which emits light as a result of a chemical reaction without the addition of heat. By way of example only, luminol (5-amino-2,3-dihydro-1,4-phthalazinedione) reacts with oxidants like hydrogen peroxide (H202) in the presence of a base and a metal catalyst to produce an excited state product (3-aminophthalate, 3-APA).

[0060] The term "chromophore," as used herein, refers to a molecule which absorbs light of visible wavelengths, UV wavelengths or IR wavelengths.

[0061] The term "cofactor," as used herein, refers to an atom or molecule essential for the action of a large molecule. Cofactors include inorganic ions, coenzymes, proteins, or some other factor necessary for the activity of enzymes. Examples include, heme in hemoglobin, magnesium in chlorophyll, and metal ions for proteins.

[0062] "Cofolding," as used herein, refers to refolding processes, reactions, or methods which employ at least two molecules which interact with each other and result in the transformation of unfolded or improperly folded molecules to properly folded molecules. By way of example only, "cofolding," employ at least two polypeptides which interact with each other and result in the transformation of unfolded or improperly folded polypeptides to native, properly folded polypeptides. Such polypeptides may contain natural amino acids and/or at least one non-natural amino acid.

[0063] A "comparison window," as used herein, refers a segment of any one of contiguous positions used to compare a sequence to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Such contiguous positions include a group consisting of from about 20 to about 600 sequential units, including about 50 to about 200 sequential units, and about 100 to about 150 sequential units. By way of example only, such sequences include polypeptides and polypeptides containing non-natural amino acids, with the sequential units include natural and non-natural amino acids. In addition, by way of example only, such sequences include polynucleotides with nucleotides being the corresponding sequential units. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, including by the local homology algorithm of Smith and Waterman (1970) Adv. Appl. Math. 2:482c , by the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48:443, by the search for similarity method of Pearson and Lipman (1988) Proc. Nat'l. Acad. Sci. USA 85:2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (see, e.g., Ausubel et al., Current Protocols in Molecular Biology (1995 supplement)).

[0064] By way of example, an algorithm which may be used to determine percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al. (1997) Nuc. Acids Res. 25:3389-3402, and Altschul et al. (1990) J. Mol. Biol. 215:403-410, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. 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 word length (W) of 11, an expectation (E) or 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1992) Proc. Natl. Acad. Sci. USA 89:10915) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands. The BLAST algorithm is typically performed with the "low complexity" filter turned off.

[0065] The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin and Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-5787). 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 0.2, or less than about 0.01, or less than about 0.001.

[0066] The term "conservatively modified variants" applies to both natural and non-natural amino acid and natural and non-natural nucleic acid sequences, and combinations thereof. With respect to particular nucleic acid sequences, "conservatively modified variants" refers to those natural and non-natural nucleic acids which encode identical or essentially identical natural and non-natural amino acid sequences, or where the natural and non-natural nucleic acid does not encode a natural and non-natural amino acid sequence, to essentially identical sequences. By way of example, because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and

GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Thus by way of example every natural or non-natural nucleic acid sequence herein which encodes a natural or non-natural polypeptide also describes every possible silent variation of the natural or non-natural nucleic acid. One of skill will recognize that each codon in a natural or non-natural nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a natural and non-natural nucleic acid which encodes a natural and non-natural polypeptide is implicit in each described sequence.

[0067] As to amino acid sequences, individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single natural and non-natural amino acid or a small percentage of natural and non-natural amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the deletion of an amino acid, addition of an amino acid, or substitution of a natural and non-natural amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar natural amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the methods and compositions described herein.

[0068] The following eight groups each contain amino acids that are conservative substitutions for one another:

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

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

[0069] The terms "cycloalkyl" and "heterocycloalkyl", by themselves or in combination with other terms, represent, unless otherwise stated, cyclic versions of "alkyl" and "heteroalkyl", respectively. Thus, a cycloalkyl or heterocycloalkyl include saturated, partially unsaturated and fully unsaturated ring linkages. Additionally, for heterocycloalkyl, a heteroatom can occupy the position at which the heterocycle is attached to the remainder of the molecule. The heteroatom may include oxygen, nitrogen or sulfur. Examples of cycloalkyl include cyclopentyl, cyclohexyl, 1-cyclohexenyl, 3-cyclohexenyl, cycloheptyl, and the like. Examples of heterocycloalkyl include 1-(1,2,5,6-tetrahydropyridyl), 1-piperidinyl, 2-piperidinyl, 3-piperidinyl, 4-morpholinyl, 3-morpholinyl, tetrahydrofuran-2-yl, tetrahydrofuran-3-yl, tetrahydrothien-2-yl, tetrahydrothien-3-yl, 1-piperazinyl, 2-piperazinyl, and the like. Additionally, the term encompasses multicyclic structures bicyclic and tricyclic ring structures. Similarly, the term "heterocycloalkylene" by itself or as part of another molecule means a divalent radical derived from heterocycloalkyl, and the term "cycloalkylene" by itself or as part of another molecule means a divalent radical derived from cycloalkyl.

[0070] The term "cyclodextrin," as used herein, refers to cyclic carbohydrates consisting of at least six to eight glucose molecules in a ring formation. The outer part of the ring contains water soluble groups; at the center of the ring is a relatively nonpolar cavity able to accommodate small molecules.

[0071] The term "cytotoxic," as used herein, refers to a compound which harms cells.

[0072] "Denaturing agent" or "denaturant," as used herein, refers to any compound or material which will cause a reversible unfolding of a polymer. By way of example only, "denaturing agent" or "denaturants," may cause a reversible unfolding of a protein. The strength of a denaturing agent or denaturant will be determined both by the properties and the concentration of the particular denaturing agent or denaturant. By way of example, denaturing agents or denaturants include chaotropes, detergents, organic, water miscible solvents, phospholipids, or a combination thereof. Examples of chaotropes include urea, guanidine, and sodium thiocyanate. Examples of detergents may include strong detergents such as sodium dodecyl sulfate, or polyoxyethylene ethers (e.g. Tween or Triton detergents), Sarkosyl, mild non-ionic detergents (e.g., digitonin), mild cationic detergents such as N->2,3-(Dioleyoxy)-propyl-N,N,N-trimethylammonium, mild ionic detergents (e.g. sodium cholate or sodium deoxycholate) or zwitterionic detergents including sulfobetaines (Zwittergent), 3-(3-chlolamidopropyl)dimethylammonio-1-propane sulfate (CHAPS), and 3-(3-chlolamidopropyl)dimethylammonio-2-hydroxy-1-propane sulfonate (CHAPSO). Examples of organic, water miscible solvents include acetonitrile, lower alkanols (especially C_2 - C_4 alkanols such as ethanol or isopropanol), or lower alkandiols (C_2 -

C_4 alkandiols such as ethylene-glycol) may be used as denaturants. Examples of phospholipids include naturally occurring phospholipids such as phosphatidylethanolamine, phosphatidylcholine, phosphatidylserine, and phosphatidylinositol or synthetic phospholipid derivatives or variants such as dihexanoylphosphatidylcholine or diheptanoylphosphatidylcholine.

[0073] The term "detectable label," as used herein, refers to a label which may be observable using analytical techniques including fluorescence, chemiluminescence, electron-spin resonance, ultraviolet/visible absorbance spectroscopy, mass spectrometry, nuclear magnetic resonance, magnetic resonance, and electrochemical methods.

[0074] The term "dicarbonyl" as used herein refers to a group containing at least two moieties selected from the group consisting of -C(O)-, -S(O)-, $-S(O)_2-$, and -C(S)-, including 1,2-dicarbonyl groups, a 1,3-dicarbonyl groups, and 1,4-dicarbonyl groups, and groups containing a least one ketone group, and/or at least one aldehyde groups, and/or at least one ester group, and/or at least one carboxylic acid group, and/or at least one thioester group. Such dicarbonyl groups include diketones, ketoaldehydes, ketoacids, ketoesters, and ketothioesters. In addition, such groups may be part of linear, branched, or cyclic molecules. The two moieties in the dicarbonyl group may be the same or different, and may include substituents that would produce, by way of example only, an ester, a ketone, an aldehyde, a thioester, or an amide, at either of the two moieties.

[0075] The term "drug," as used herein, refers to any substance used in the prevention, diagnosis, alleviation, treatment, or cure of a disease or condition.

[0076] The term "dye," as used herein, refers to a soluble, coloring substance which contains a chromophore.

[0077] The term "effective amount," as used herein, refers to a sufficient amount of an agent or a compound being administered which will relieve to some extent one or more of the symptoms of the disease or condition being treated. The result can be reduction and/or alleviation of the signs, symptoms, or causes of a disease, or any other desired alteration of a biological system. By way of example, an agent or a compound being administered includes a natural amino acid polypeptide, non-natural amino acid polypeptide, modified natural amino acid polypeptide, or modified non-amino acid polypeptide. Compositions containing such natural amino acid polypeptides, non-natural amino acid polypeptides, modified natural amino acid polypeptides, or modified non-natural amino acid polypeptides can be administered for prophylactic, enhancing, and/or therapeutic treatments. An appropriate "effective" amount in any individual case may be determined using techniques, such as a dose escalation study.

[0078] The term "electron dense group," as used herein, refers to a group which scatters electrons when irradiated with an electron beam. Such groups include ammonium molybdate, bismuth subnitrate cadmium iodide, 99%, carbohydrazide, ferric chloride hexahydrate, hexamethylene tetramine, 98.5%, indium trichloride anhydrous, lanthanum nitrate, lead acetate trihydrate, lead citrate trihydrate, lead nitrate, periodic acid, phosphomolybdic acid, phosphotungstic acid, potassium ferricyanide, potassium ferrocyanide, ruthenium red, silver nitrate, silver proteinate (Ag Assay: 8.0-8.5%) "Strong", silver tetraphenylporphin (S-TPPS), sodium chloroaurate, sodium tungstate, thallium nitrate, thiosemicarbazide (TSC), uranyl acetate, uranyl nitrate, and vanadyl sulfate.

[0079] The term "energy transfer agent," as used herein, refers to a molecule which can either donate or accept energy from another molecule. By way of example only, fluorescence resonance energy transfer (FRET) is a dipole-dipole coupling process by which the excited-state energy of a fluorescence donor molecule is non-radiatively transferred to an unexcited acceptor molecule which then fluorescently emits the donated energy at a longer wavelength.

[0080] The terms "enhance" or "enhancing" means to increase or prolong either in potency or duration a desired effect. By way of example, "enhancing" the effect of therapeutic agents refers to the ability to increase or prolong, either in potency or duration, the effect of therapeutic agents on during treatment of a disease, disorder or condition. An "enhancing-effective amount," as used herein, refers to an amount adequate to enhance the effect of a therapeutic agent in the treatment of a disease, disorder or condition. When used in a patient, amounts effective for this use will depend on the severity and course of the disease, disorder or condition, previous therapy, the patient's health status and response to the drugs, and the judgment of the treating physician.

[0081] As used herein, the term "eukaryote" refers to organisms belonging to the phylogenetic domain Eucarya, including animals (including mammals, insects, reptiles, birds, etc.), ciliates, plants (including monocots, dicots, and algae), fungi, yeasts, flagellates, microsporidia, and protists.

[0082] The term "fatty acid," as used herein, refers to carboxylic acids with about C6 or longer hydrocarbon side chain.

[0083] The term "fluorophore," as used herein, refers to a molecule which upon excitation emits photons and is thereby fluorescent.

[0084] The terms "functional group", "active moiety", "activating group", "leaving group", "reactive site", "chemically reactive group" and "chemically reactive moiety," as used herein, refer to portions or units of a molecule at which chemical reactions occur. The terms are somewhat synonymous in the chemical arts and are used herein to indicate the portions of molecules that perform some function or activity and are reactive with other molecules.

[0085] The term "halogen" includes fluorine, chlorine, iodine, and bromine.

[0086] The term "haloacyl," as used herein, refers to acyl groups which contain halogen moieties, including $-C(O)CH_3$, $-C(O)CF_3$, $-C(O)CH_2OCH_3$, and the like.

[0087] The term "haloalkyl," as used herein, refers to alkyl groups which contain halogen moieties, including $-CF_3$ and $-CH_2CF_3$ and the like.

[0088] The term "heteroalkyl," as used herein, refers to straight or branched chain, or cyclic hydrocarbon radicals, or combinations thereof, consisting of an alkyl group and at least one heteroatom selected from the group consisting of O, N, Si and S, and wherein the nitrogen and sulfur atoms may optionally be oxidized and the nitrogen heteroatom may optionally be quaternized. The heteroatom(s) O, N and S and Si may be placed at any interior position of the heteroalkyl group or at the position at which the alkyl group is attached to the remainder of the molecule. Examples include $-CH_2-CH_2-O-CH_3$, $-CH2-CH_2-NH-CH_3$, $-CH_2-CH_2-N(CH_3)-CH_3$, $-CH_2-S-CH_2-CH_3$, $-CH_2-CH_2,-S(O)-CH_3$, $-CH_2-CH_2-S(O)_2-CH_3$, $-CH=CH-O-CH_3$,$-Si(CH_3)_3$, $-CH_2-CH=N-OCH_3$, and $-CH=CH-N(CH_3)-CH_3$. In addition, up to two heteroatoms may be consecutive, such as, by way of example, $-CH_2-NH-OCH_3$ and $-CH_2-O-Si(CH_3)_3$.

[0089] The term "heteroalkylene," as used herein, refers to a divalent radical derived from heteroalkyl, as exemplified $-CH_2-CH_2-S-CH_2CH_2-$ and $-CH_2-S-CH_2-CH_2-NH-CH_2-$. For heteroalkylene groups, the same or different heteroatoms can also occupy either or both of the chain termini (including alkyleneoxy, alkylenedioxy, alkyleneamino, alkylenediamino, aminooxyalkylene, and the like). Still further, for alkylene and heteroalkylene linking groups, no orientation of the linking group is implied by the direction in which the formula of the linking group is written. By way of example, the formula $-C(O)_2R'-$ represents both $-C(O)_2R'-$ and $-R'C(O)_2-$.

[0090] The term "heteroaryl" or "heteroaromatic," as used herein, refers to aryl groups which contain at least one heteroatom selected from N, O, and S; wherein the nitrogen and sulfur atoms may be optionally oxidized, and the nitrogen atom(s) may be optionally quaternized. Heteroaryl groups may be substituted or unsubstituted. A heteroaryl group may be attached to the remainder of the molecule through a heteroatom. Examples of heteroaryl groups include 1-pyrrolyl, 2-pyrrolyl, 3-pyrrolyl, 3-pyrazolyl, 2-imidazolyl, 4-imidazolyl, pyrazinyl, 2-oxazolyl, 4-oxazolyl, 2-phenyl-4-oxazolyl, 5-oxazolyl, 3-isoxazolyl, 4-isoxazolyl, 5-isoxazolyl, 2-thiazolyl, 4-thiazolyl, 5-thiazolyl, 2-furyl, 3-furyl, 2-thienyl, 3-thienyl, 2-pyridyl, 3-pyridyl, 4-pyridyl, 2-pyrimidyl, 4-pyrimidyl, 5-benzothiazolyl, purinyl, 2-benzimidazolyl, 5-indolyl, 1-isoquinolyl, 5-isoquinolyl, 2-quinoxalinyl, 5-quinoxalinyl, 3-quinolyl, and 6-quinolyl.

[0091] The term "homoalkyl," as used herein refers to alkyl groups which are hydrocarbon groups.

[0092] The term "identical," as used herein, refers to two or more sequences or subsequences which are the same. In addition, the term "substantially identical," as used herein, refers to two or more sequences which have a percentage of sequential units which are the same when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using comparison algorithms or by manual alignment and visual inspection. By way of example only, two or more sequences may be "substantially identical" if the sequential units are about 60% identical, about 65% identical, about 70% identical, about 75% identical, about 80% identical, about 85% identical, about 90% identical, or about 95% identical over a specified region. Such percentages to describe the "percent identity" of two or more sequences. The identity of a sequence can exist over a region that is at least about 75-100 sequential units in length, over a region that is about 50 sequential units in length, or, where not specified, across the entire sequence. This definition also refers to the complement of a test sequence. By way of example only, two or more polypeptide sequences are identical when the amino acid residues are the same, while two or more polypeptide sequences are "substantially identical" if the amino acid residues are about 60% identical, about 65% identical, about 70% identical, about 75% identical, about 80% identical, about 85% identical, about 90% identical, or about 95% identical over a specified region. The identity can exist over a region that is at least about 75-100 amino acids in length, over a region that is about 50 amino acids in length, or, where not specified, across the entire sequence of a polypeptide sequence. In addition, by way of example only, two or more polynucleotide sequences are identical when the nucleic acid residues are the same, while two or more polynucleotide sequences are "substantially identical" if the nucleic acid residues are about 60% identical, about 65% identical, about 70% identical, about 75% identical, about 80% identical, about 85% identical, about 90% identical, or about 95% identical over a specified region. The identity can exist over a region that is at least about 75-100 nucleic acids in length, over a region that is about 50 nucleic acids in length, or, where not specified, across the entire sequence of a polynucleotide sequence.

[0093] For sequence comparison, 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 entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

[0094] The term "immunogenicity," as used herein, refers to an antibody response to administration of a therapeutic drug. The immunogenicity toward therapeutic non-natural amino acid polypeptides can be obtained using quantitative and qualitative assays for detection of anti-non-natural amino acid polypeptides antibodies in biological fluids. Such assays include Radioimmunoassay (RIA), Enzyme-linked immunosorbent assay (ELISA), luminescent immunoassay (LIA), and fluorescent immunoassay (FIA). Analysis of immunogenicity toward therapeutic non-natural amino acid polypeptides involves comparing the antibody response upon administration of therapeutic non-natural amino acid polypeptides to the antibody response upon administration of therapeutic natural amino acid polypeptides.

[0095] The term "intercalating agent," also referred to as "intercalating group," as used herein, refers to a chemical that can insert into the intramolecular space of a molecule or the intermolecular space between molecules. By way of example only an intercalating agent or group may be a molecule which inserts into the stacked bases of the DNA double helix.

[0096] The term "isolated," as used herein, refers to separating and removing a component of interest from components not of interest. Isolated substances can be in either a dry or semi-dry state, or in solution, including an aqueous solution. The isolated component can be in a homogeneous state or the isolated component can be a part of a pharmaceutical composition that comprises additional pharmaceutically acceptable carriers and/or excipients. Purity and homogeneity may be determined using analytical chemistry techniques including polyacrylamide gel electrophoresis or high performance liquid chromatography. In addition, when a component of interest is isolated and is the predominant species present in a preparation, the component is described herein as substantially purified. The term "purified," as used herein, refers to a component of interest which is at least 85% pure, at least 90% pure, at least 95% pure, at least 99% or greater pure. By way of example only, nucleic acids or proteins are "isolated" when such nucleic acids or proteins are free of at least some of the cellular components with which it is associated in the natural state, or that the nucleic acid or protein has been concentrated to a level greater than the concentration of its *in vivo* or *in vitro* production. Also, by way of example, a gene is isolated when separated from open reading frames which flank the gene and encode a protein other than the gene of interest.

[0097] The term "label," as used herein, refers to a substance which is incorporated into a compound and is readily detected, whereby its physical distribution may be detected and/or monitored.

[0098] The term "linkage," as used herein to refer to bonds or chemical moiety formed from a chemical reaction between the functional group of a linker and another molecule. Such bonds may include covalent linkages and non-covalent bonds, while such chemical moieties may include esters, carbonates, imines phosphate esters, hydrazones, acetals, orthoesters, peptide linkages, and oligonucleotide linkages. Hydrolytically stable linkages means that the linkages are substantially stable in water and do not react with water at useful pH values, including under physiological conditions for an extended period of time, perhaps even indefinitely. Hydrolytically unstable or degradable linkages means that the linkages are degradable in water or in aqueous solutions, including for example, blood. Enzymatically unstable or degradable linkages means that the linkage can be degraded by one or more enzymes. By way of example only, PEG and related polymers may include degradable linkages in the polymer backbone or in the linker group between the polymer backbone and one or more of the terminal functional groups of the polymer molecule. Such degradable linkages include ester linkages formed by the reaction of PEG carboxylic acids or activated PEG carboxylic acids with alcohol groups on a biologically active agent, wherein such ester groups generally hydrolyze under physiological conditions to release the biologically active agent. Other hydrolytically degradable linkages include carbonate linkages; imine linkages resulted from reaction of an amine and an aldehyde; phosphate ester linkages formed by reacting an alcohol with a phosphate group; hydrazone linkages which are reaction product of a hydrazide and an aldehyde; acetal linkages that are the reaction product of an aldehyde and an alcohol; orthoester linkages that are the reaction product of a formate and an alcohol; peptide linkages formed by an amine group, including at an end of a polymer such as PEG, and a carboxyl group of a peptide; and oligonucleotide linkages formed by a phosphoramidite group, including at the end of a polymer, and a 5' hydroxyl group of an oligonucleotide.

[0099] The terms "medium" or "media," as used herein, refer to any culture medium used to grow and harvest cells and/or products expressed and/or secreted by such cells. Such "medium" or "media" include solution, solid, semi-solid, or rigid supports that may support or contain any host cell, including, by way of example, bacterial host cells, yeast host cells, insect host cells, plant host cells, eukaryotic host cells, mammalian host cells, CHO cells, prokaryotic host cells, E. coli, or Pseudomonas host cells, and cell contents. Such "medium" or "media" includes medium or media in which the host cell has been grown into which a polypeptide has been secreted, including medium either before or after a proliferation step. Such "medium" or "media" also includes buffers or reagents that contain host cell lysates, by way of example a polypeptide produced intracellularly and the host cells are lysed or disrupted to release the polypeptide.

[0100] The term "metabolite," as used herein, refers to a derivative of a compound, by way of example natural amino acid polypeptide, a non-natural amino acid polypeptide, a modified natural amino acid polypeptide, or a modified non-natural amino acid polypeptide, that is formed when the compound, by way of example natural amino acid polypeptide, non-natural amino acid polypeptide, modified natural amino acid polypeptide, or modified non-natural amino acid polypeptide, is metabolized. The term "pharmaceutically active metabolite" or "active metabolite" refers to a biologically active derivative of a compound, by way of example natural amino acid polypeptide, a non-natural amino acid polypeptide, a modified natural amino acid polypeptide, or a modified non-natural amino acid polypeptide, that is formed when such a compound, by way of example a natural amino acid polypeptide, non-natural amino acid polypeptide, modified natural amino acid polypeptide, or modified non-natural amino acid polypeptide, is metabolized.

[0101] The term "metabolized," as used herein, refers to the sum of the processes by which a particular substance is changed by an organism. Such processes include hydrolysis reactions and reactions catalyzed by enzymes. Further information on metabolism may be obtained from The Pharmacological Basis of Therapeutics, 9th Edition, McGraw-Hill (1996). By way of example only, metabolites of natural amino acid polypeptides, non-natural amino acid polypeptides, modified natural amino acid polypeptides, or modified non-natural amino acid polypeptides may be identified either by administration of the natural amino acid polypeptides, non-natural amino acid polypeptides, modified natural amino acid polypeptides, or modified non-natural amino acid polypeptides to a host and analysis of tissue samples from the host, or by incubation of natural amino acid polypeptides, non-natural amino acid polypeptides, modified natural amino acid polypeptides, or modified non-natural amino acid polypeptides with hepatic cells *in vitro* and analysis of the resulting compounds.

[0102] The term "metal chelator," as used herein, refers to a molecule which forms a metal complex with metal ions. By way of example, such molecules may form two or more coordination bonds with a central metal ion and may form ring structures.

[0103] The term "metal-containing moiety," as used herein, refers to a group which contains a metal ion, atom or particle. Such moieties include cisplatin, chelated metals ions (such as nickel, iron, and platinum), and metal nanoparticles (such as nickel, iron, and platinum).

[0104] The term "moiety incorporating a heavy atom," as used herein, refers to a group which incorporates an ion of atom which is usually heavier than carbon. Such ions or atoms include silicon, tungsten, gold, lead, and uranium.

[0105] The term "modified," as used herein refers to the presence of a change to a natural amino acid, a non-natural amino acid, a natural amino acid polypeptide or a non-natural amino acid polypeptide. Such changes, or modifications, may be obtained by post synthesis modifications of natural amino acids, non-natural amino acids, natural amino acid polypeptides or non-natural amino acid polypeptides, or by co-translational, or by post-translational modification of natural amino acids, non-natural amino acids, natural amino acid polypeptides or non-natural amino acid polypeptides. The form "modified or unmodified" means that the natural amino acid, non-natural amino acid, natural amino acid polypeptide or non-natural amino acid polypeptide being discussed are optionally modified, that is, he natural amino acid, non-natural amino acid, natural amino acid polypeptide or non-natural amino acid polypeptide under discussion can be modified or unmodified.

[0106] As used herein, the term "modulated serum half-life" refers to positive or negative changes in the circulating half-life of a modified biologically active molecule relative to its non-modified form. By way of example, the modified biologically active molecules include natural amino acid, non-natural amino acid, natural amino acid polypeptide or non-natural amino acid polypeptide. By way of example, serum half-life is measured by taking blood samples at various time points after administration of the biologically active molecule or modified biologically active molecule, and determining the concentration of that molecule in each sample. Correlation of the serum concentration with time allows calculation of the serum half-life. By way of example, modulated serum half-life may be an increased in serum half-life, which may enable an improved dosing regimens or avoid toxic effects. Such increases in serum may be at least about two fold, at least about three-fold, at least about five-fold, or at least about ten-fold. An example of a method to evaluate increases in serum half-life is given in examples 88-92. This method may be used for evaluating the serum half-life of any polypeptide.

[0107] The term "modulated therapeutic half-life," as used herein, refers to positive or negative change in the half-life of the therapeutically effective amount of a modified biologically active molecule, relative to its non-modified form. By way of example, the modified biologically active molecules include natural amino acid, non-natural amino acid, natural amino acid polypeptide or non-natural amino acid polypeptide. By way of example, therapeutic half-life is measured by measuring pharmacokinetic and/or pharmacodynamic properties of the molecule at various time points after administration. Increased therapeutic half-life may enable a particular beneficial dosing regimen, a particular beneficial total dose, or avoids an undesired effect. By way of example, the increased therapeutic half-life may result from increased potency, increased or decreased binding of the modified molecule to its target, an increase or decrease in another parameter or mechanism of action of the non-modified molecule, or an increased or decreased breakdown of the molecules by enzymes such as, by way of example only, proteases. An example of a method to evaluate increases in therapeutic half-life is given in examples 88-92. This method may be used for evaluating the therapeutic half-life of any polypeptide.

[0108] The term "nanoparticle," as used herein, refers to a particle which has a particle size between about 500 nm to about 1 nm.

[0109] The term "near-stoichiometric," as used herein, refers to the ratio of the moles of compounds participating in a chemical reaction being about 0.75 to about 1.5.

[0110] As used herein, the term "non-eukaryote" refers to non-eukaryotic organisms. By way of example, a non-eukaryotic organism may belong to the Eubacteria, (which includes Escherichia coli, Thermus thermophilus, or Bacillus stearothermophilus, Pseudomonas fluorescens, Pseudomonas aeruginosa, Pseudomonas putida), phylogenetic domain, or the Archaea, which includes Methanococcus jannaschii, Methanobacterium thermoautotrophicum, Archaeoglobus fulgidus, Pyrococcus furiosus, Pyrococcus horikoshii, Aeuropyrum pernix, or Halobacterium such as Haloferax volcanii and Halobacterium species NRC-1, or phylogenetic domain.

[0111] A "non-natural amino acid" refers to an amino acid that is not one of the 20 common amino acids or pyrolysine or selenocysteine. Other terms that may be used synonymously with the term "non-natural amino acid" is "non-naturally encoded amino acid," "unnatural amino acid," "non-naturally-occurring amino acid," and variously hyphenated and non-hyphenated versions thereof. The term "non-natural amino acid" includes amino acids which occur naturally by modification of a naturally encoded amino acid (including the 20 common amino acids or pyrrolysine and selenocysteine) but are not themselves incorporated into a growing polypeptide chain by the translation complex. Examples of naturally-occurring amino acids that are not naturally-encoded include N-acetylglucosaminyl-L-serine, N-acetylglucosaminyl-L-threonine, and O-phosphotyrosine. Additionally, the term "non-natural amino acid" includes amino acids which do not occur naturally and may be obtained synthetically or may be obtained by modification of non-natural amino acids.

[0112] The term "nucleic acid," as used herein, refers to deoxyribonucleotides, deoxyribonucleosides, ribonucleosides or ribonucleotides and polymers thereof in either single- or double-stranded form. By way of example only, such nucleic acids and nucleic acid polymers include (i) analogues of natural nucleotides which have similar binding properties as a reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides; (ii) oligonucleotide analogs including PNA (peptidonucleic acid), analogs of DNA used in antisense technology (phosphorothioates, phosphoroamidates, and the like); (iii) conservatively modified variants thereof (including degenerate codon substitutions) and complementary sequences and sequence explicitly indicated. By way of example, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer et al., Nucleic Acid Res. 19:5081 (1991); Ohtsuka et al., J. Biol. Chem. 260:2605-2608 (1985); and Rossolini et al., Mol. Cell. Probes 8:91-98 (1994)).

[0113] The term "oxidizing agent," as used herein, refers to a compound or material which is capable of removing an electron from a compound being oxidized. By way of example oxidizing agents include oxidized glutathione, cystine, cystamine, oxidized dithiothreitol, oxidized erythreitol, and oxygen. A wide variety of oxidizing agents are suitable for use in the methods and compositions described herein.

[0114] The term "pharmaceutically acceptable", as used herein, refers to a material, including a salt, carrier or diluent, which does not abrogate the biological activity or properties of the compound, and is relatively nontoxic, i.e., the material may be administered to an individual without causing undesirable biological effects or interacting in a deleterious manner with any of the components of the composition in which it is contained.

[0115] The term "photoaffinity label," as used herein, refers to a label with a group, which, upon exposure to light, forms a linkage with a molecule for which the label has an affinity. By way of example only, such a linkage may be covalent or non-covalent.

[0116] The term "photocaged moiety," as used herein, refers to a group which, upon illumination at certain wavelengths, covalently or non-covalently binds other ions or molecules.

[0117] The term "photocleavable group," as used herein, refers to a group which breaks upon exposure to light.

[0118] The term "photocrosslinker," as used herein, refers to a compound comprising two or more functional groups which, upon exposure to light, are reactive and form a covalent or non-covalent linkage with two or more monomeric or polymeric molecules.

[0119] The term "photoisomerizable moiety," as used herein, refers to a group wherein upon illumination with light changes from one isomeric form to another.

[0120] The term "polyalkylene glycol," as used herein, refers to linear or branched polymeric polyether polyols. Such polyalkylene glycols, including polyethylene glycol, polypropylene glycol, polybutylene glycol, and derivatives thereof. Other exemplary embodiments are listed, for example, in commercial supplier catalogs, such as Shearwater Corporation's catalog "Polyethylene Glycol and Derivatives for Biomedical Applications" (2001). By way of example only, such polymeric polyether polyols have average molecular weights between about 0.1 kDa to about 100 kDa. By way of example, such polymeric polyether polyols include between about 100 Da and about 100,000 Da or more. The molecular weight of the polymer may be between about 100 Da and about 100,000 Da, including 100,000 Da, 95,000 Da, 90,000 Da, 85,000 Da, 80,000 Da, 75,000 Da, 70,000 Da, 65,000 Da, 60,000 Da, 55,000 Da, 50,000 Da, 45,000 Da, 40,000 Da, 35,000 Da, 30,000 Da, 25,000 Da, 20,000 Da, 15,000 Da, 10,000 Da, 9,000 Da, 8,000 Da, 7,000 Da, 6,000 Da, 5,000 Da, 4,000 Da, 3,000 Da, 2,000 Da, 1,000 Da, 900 Da, 800 Da, 700 Da, 600 Da, 500 Da, 400 Da, 300 Da, 200 Da, and 100 Da. In some embodiments, the molecular weight of the polymer is between about 100 Da and 50,000 Da. In some embodiments, the molecular weight of the polymer is between about 100 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 1,000 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 5,000 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 10,000 Da and 40,000 Da. In some embodiments, the poly(ethylene glycol) molecule is a branched polymer. The molecular weight of the branched chain PEG may be between about 1,000 Da and about 100,000 Da, including 100,000 Da, 95,000 Da, 90,000 Da, 85,000 Da, 80,000 Da, 75,000 Da, 70,000 Da, 65,000 Da, 60,000 Da, 55,000 Da, 50,000 Da, 45,000 Da, 40,000 Da, 35,000 Da, 30,000 Da, 25,000 Da, 20,000 Da, 15,000 Da, 10,000 Da, 9,000 Da, 8,000 Da, 7,000 Da, 6,000 Da, 5,000 Da, 4,000 Da, 3,000 Da, 2,000 Da, and 1,000 Da. In some embodiments, the molecular weight of the branched chain PEG is between about 1,000 Da and 50,000 Da. In some embodiments, the molecular weight of the branched chain PEG is between about 1,000 Da and 40,000 Da. In some embodiments, the molecular weight of the branched chain PEG is between about 5,000 Da and 40,000 Da. In some embodiments, the molecular weight of the branched chain PEG is between about 5,000 Da and 20,000 Da.

[0121] The term "polymer," as used herein, refers to a molecule composed of repeated subunits. Such molecules include polypeptides, polynucleotides, or polysaccharides or polyalkylene glycols.

[0122] The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. That is, a description directed to a polypeptide applies equally to a description of a peptide and a description of a protein, and vice versa. The terms apply to naturally occurring amino acid polymers as well as amino acid polymers in which one or more amino acid residues is a non-natural amino acid. Additionally, such "polypeptides," "peptides" and "proteins" include amino acid chains of any length, including full length proteins, wherein the amino acid residues are linked by covalent peptide bonds.

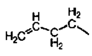
[0123] The term "post-translationally modified" refers to any modification of a natural or non-natural amino acid which occurs after such an amino acid has been translationally incorporated into a polypeptide chain. Such modifications include co-translational *in vivo* modifications, co-translational *in vitro* modifications (such as in a cell-free translation system), post-translational *in vivo* modifications, and post-translational *in vitro* modifications.

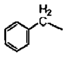
[0124] The terms "prodrug" or "pharmaceutically acceptable prodrug," as used herein, refers to an agent that is converted into the parent drug *in vivo* or *in vitro,* wherein which does not abrogate the biological activity or properties of the drug, and is relatively nontoxic, i.e., the material may be administered to an individual without causing undesirable biological effects or interacting in a deleterious manner with any of the components of the composition in which it is contained. Prodrugs are generally drug precursors that, following administration to a subject and subsequent absorption, are converted to an active, or a more active species via some process, such as conversion by a metabolic pathway. Some prodrugs have a chemical group present on the prodrug that renders it less active and/or confers solubility or some other property to the drug. Once the chemical group has been cleaved and/or modified from the prodrug the active drug is generated. Prodrugs are converted into active drug within the body through enzymatic or non-enzymatic reactions. Prodrugs may provide improved physiochemical properties such as better solubility, enhanced delivery characteristics, such as specifically,targeting a particular cell, tissue, organ or ligand, and improved therapeutic value of the drug. The benefits of such prodrugs include (i) ease of administration compared with the parent drug; (ii) the prodrug may be bioavailable by oral administration whereas the parent is not; and (iii) the prodrug may also have improved solubility in pharmaceutical compositions compared with the parent drug. A pro-drug includes a pharmacologically inactive, or reduced-activity, derivative of an active drug. Prodrugs may be designed to modulate the amount of a drug or biologically active molecule that reaches a desired site of action through the manipulation of the properties of a drug, such as physiochemical, biopharmaceutical, or pharmacokinetic properties. An example of a prodrug would be a non-natural amino acid polypeptides which is administered as an ester (the "prodrug") to facilitate transmittal across a cell membrane where water solubility is detrimental to mobility but which then is metabolically hydrolyzed to the carboxylic acid, the active entity, once inside the cell where water-solubility is beneficial. Prodrugs may be designed as reversible drug derivatives, for use as modifiers to enhance drug transport to site-specific tissues.

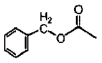
[0125] The term "prophylactically effective amount," as used herein, refers that amount of a composition containing at least one non-natural amino acid polypeptide or at least one modified non-natural amino acid polypeptide prophylactically applied to a patient which will relieve to some extent one or more of the symptoms of a disease, condition or disorder being treated. In such prophylactic applications, such amounts may depend on the patient's state of health, weight, and the like. It is considered well within the skill of the art for one to determine such prophylactically effective amounts by routine experimentation, including a dose escalation clinical trial.

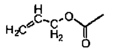
[0126] The term "protected," as used herein, refers to the presence of a "protecting group" or moiety that prevents reaction of the chemically reactive functional group under certain reaction conditions. The protecting group will vary depending on the type of chemically reactive group being protected. By way of example only, (i) if the chemically reactive group is an amine or a hydrazide, the protecting group may be selected from tert-butyloxycarbonyl (t-Boc) and 9-fluorenylmethoxycarbonyl (Fmoc); (ii) if the chemically reactive group is a thiol, the protecting group may be orthopyridyldisulfide; and (iii) if the chemically reactive group is a carboxylic acid, such as butanoic or propionic acid, or a hydroxyl group, the protecting group may be benzyl or an alkyl group such as methyl, ethyl, or tert-butyl.

[0127] By way of example only, blocking/protecting groups may also be selected from:

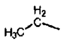

allyl

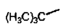

Bn

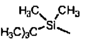

Cbz

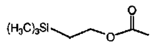

alloc

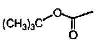

Me

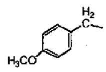

Et

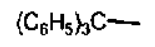

t-butyl

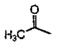

TBDMS

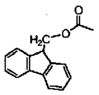

Teoc


Boc


pMBn


trityl


acetyl


Fmoc

[0128] Additionally, protecting groups include including photolabile groups such as Nvoc and MeNvoc and other protecting groups known in the art. Other protecting groups are described in Greene and Wuts, Protective Groups in Organic Synthesis, 3rd Ed., John Wiley & Sons, New York, NY, 1999.

[0129] The term "radioactive moiety," as used herein, refers to a group whose nuclei spontaneously give off nuclear radiation, such as alpha, beta, or gamma particles; wherein, alpha particles are helium nuclei, beta particles are electrons, and gamma particles are high energy photons.

[0130] The term "reactive compound," as used herein, refers to a compound which under appropriate conditions is reactive toward another atom, molecule or compound.

[0131] The term "recombinant host cell," also referred to as "host cell," refers to a cell which includes an exogenous polynucleotide, wherein the methods used to insert the exogenous polynucleotide into a cell include direct uptake, transduction, f-mating, or other methods known in the art to create recombinant host cells. By way of example only, such exogenous polynucleotide may be a nonintegrated vector, including a plasmid, or may be integrated into the host genome.

[0132] The term "redox-active agent," as used herein, refers to a molecule which oxidizes or reduces another molecule, whereby the redox active agent becomes reduced or oxidized. Examples of redox active agent include ferrocene, quinones, Ru2+/3+complexes, Co2+/3+complexes, and Os2+/3+complexes.

[0133] The term "reducing agent," as used herein, refers to a compound or material which is capable of adding an electron to a compound being reduced. By way of example reducing agents include dithiothreitol (DTT), 2-mercaptoethanol, dithioerythritol, cysteine, cysteamine (2-aminoethanethiol), and reduced glutathione. Such reducing agents may be used, by way of example only, to maintain sulfhydryl groups in the reduced state and to reduce intra- or intermolecular disulfide bonds.

[0134] "Refolding," as used herein describes any process, reaction or method which transforms an improperly folded or unfolded state to a native or properly folded conformation. By way of example only, refolding transforms disulfide bond containing polypeptides from an improperly folded or unfolded state to a native or properly folded conformation with respect to disulfide bonds. Such disulfide bond containing polypeptides may be natural amino acid polypeptides or non-natural amino acid polypeptides.

[0135] The term "resin," as used herein, refers to high molecular weight, insoluble polymer beads. By way of example only, such beads may be used as supports for solid phase peptide synthesis, or sites for attachment of molecules prior to purification.

[0136] The term "saccharide," as used herein, refers to a series of carbohydrates including sugars, monosaccharides, oligosaccharides, and polysaccharides.

[0137] The term "safety" or "safety profile," as used herein, refers to side effects that might be related to administration of a drug relative to the number of times the drug has been administered. By way of example, a drug which has been administered many times and produced only mild or no side effects is said to have an excellent safety profile. An example of a method to evaluate the safety profile is given in example 92. This method may be used for evaluating the safety profile of any polypeptide.

[0138] The phrase "selectively hybridizes to" or "specifically hybridizes to," as used herein, refers to the binding, duplexing, or hybridizing of a molecule to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture including total cellular or library DNA or RNA.

[0139] The term "spin label," as used herein, refers to molecules which contain an atom or a group of atoms exhibiting an unpaired electron spin (i.e. a stable paramagnetic group) that can be detected by electron spin resonance spectroscopy and can be attached to another molecule. Such spin-label molecules include nitryl radicals and nitroxides, and may be single spin-labels or double spin-labels.

[0140] The term "stoichiometric," as used herein, refers to the ratio of the moles of compounds participating in a chemical reaction being about 0.9 to about 1.1.

[0141] The term "stoichiometric-like," as used herein, refers to a chemical reaction which becomes stoichiometric or near-stoichiometric upon changes in reaction conditions or in the presence of additives. Such changes in reaction conditions include an increase in temperature or change in pH. Such additives include accelerants.

[0142] The phrase "stringent hybridization conditions" refers to hybridization of sequences of DNA, RNA, PNA or other nucleic acid mimics, or combinations thereof, under conditions of low ionic strength and high temperature. By way of example, under stringent conditions a probe will hybridize to its target subsequence in a complex mixture of nucleic acid (including total cellular or library DNA or RNA) but does not hybridize to other sequences in the complex mixture. Stringent conditions are sequence-dependent and will be different in different circumstances. By way of example, longer sequences hybridize specifically at higher temperatures. Stringent hybridization conditions include (i) about 5-10o C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH; (ii) the salt concentration is about 0.01 M to about 1.0 M at about pH 7.0 to about pH 8.3 and the temperature is at least about 30oC for short probes (including 10 to 50 nucleotides) and at least about 60o C for long probes (including greater than 50 nucleotides); (iii) the addition of destabilizing agents including formamide, (iv) 50% formamide, 5X SSC, and 1% SDS, incubating at 42oC, or 5X SSC, 1% SDS, incubating at 65oC, with wash in 0.2X SSC, and 0.1% SDS at 65oC for between about 5 minutes to about 120 minutes. By way of example only, detection of selective or specific hybridization, includes a positive signal at least two times background. An extensive guide to the hybridization of nucleic acids is found in Tijssen, Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993).

[0143] The term "subject" as used herein, refers to an animal which is the object of treatment, observation or experiment. By way of example only, a subject may be a mammal including a human.

[0144] The term "substantially purified," as used herein, refers to a component of interest that may be substantially or essentially free of other components which normally accompany or interact with the component of interest prior to purification. By way of example only, a component of interest may be "substantially purified" when the preparation of the component of interest contains less than about 30%, less than about 25%, less than about 20%, less than about 15%, less than about 10%, less than about 5%, less than about 4%, less than about 3%, less than about 2%, or less than about 1% (by dry weight) of contaminating components. Thus, a "substantially purified" component of interest may have a purity level of about 70%, about 75%, about 80%, about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, about 99% or greater. By way of example only, a natural amino acid polypeptide or a non-natural amino acid polypeptide may be purified from a native cell, or host cell in the case of recombinantly produced natural amino acid polypeptides or non-natural amino acid polypeptides. By way of example a preparation of a natural amino acid polypeptide or a non-natural amino acid polypeptide may be "substantially purified" when the preparation contains less than about 30%, less than about 25%, less than about 20%, less than about 15%, less than about 10%, less than about 5%, less than about 4%, less than about 3%, less than about 2%, or less than about 1% (by dry weight) of contaminating material. By way of example when a natural amino acid polypeptide or a non-natural amino acid polypeptide is recombinantly produced by host cells, the natural amino acid polypeptide or non-natural amino acid polypeptide may be present at about 30%, about 25%, about 20%, about 15%, about 10%, about 5%, about 4%, about 3%, about 2%, or about 1% or less of the dry weight of the cells. By way of example when a natural amino acid polypeptide or a non-natural amino acid polypeptide is recombinantly produced by host cells, the natural amino acid polypeptide or non-natural amino acid polypeptide may be present in the culture medium at about 5g/L, about 4g/L, about 3g/L, about 2g/L, about 1g/L, about 750mg/L, about 500mg/L, about 250mg/L, about 100mg/L, about 50mg/L, about 10mg/L, or about 1mg/L or less of the dry weight of the cells. By way of example, "substantially purified" natural amino acid polypeptides or non-natural amino acid polypeptides may have a purity level of about 30%, about 35%, about 40%, about 45%, about 50%, about 55%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 95%, about 99% or greater as determined by appropriate methods, including SDS/PAGE analysis, RP-HPLC, SEC, and capillary electrophoresis.

[0145] The term "substituents" also referred to as "non-interfering substituents" "refers to groups which may be used to replace another group on a molecule. Such groups include halo, C1-C10 alkyl, C2-C10 alkenyl, C_2-C_{10} alkynyl, C_1-C_{10} alkoxy, C_5-C_{12} aralkyl, C_3-C_{12} cycloalkyl, C_4-C_{12} cycloalkenyl, phenyl, substituted phenyl, toluolyl, xylenyl, biphenyl, C_2-C_{12}. alkoxyalkyl, C_5-C_{12} alkoxyaryl, C_5-C_{12} aryloxyalkyl, C_7-C_{12} oxyaryl, C_1-C_6 alkylsulfinyl, C1-C10 alkylsulfonyl, -$(CH_2)_m$-O-(C_1-C_{10} alkyl) wherein m is from 1 to 8, aryl, substituted aryl, substituted alkoxy, fluoroalkyl, heterocyclic radical, substituted heterocyclic radical, nitroalkyl, -NO_2, -CN, -NRC(O)-(C_1-C_{10} alkyl), -C(O)-(C_1-C_{10} alkyl), C_2-C_{10} alkthioalkyl, -C(O)O-(C_1-C_{10} alkyl), -OH, -SO_2, =S, -COOH, -NR_2, carbonyl, -C(O)-(C_1-C_{10} alkyl)-CF_3, -C(O)-CF_3, -C(O)NR_2, -(C_1-C_{10} aryl)-S-(C_6-C_{10} aryl), -C(O)-(C_6-C_{10} aryl), -(CH_2)m-O-(CH_2)m-O-(C_1-C_{10} alkyl) wherein each m is from 1 to 8, -C(O)NR_2, -C(S)NR_2, -SO_2NR_2, -NRC(O)NR_2,-NRC(S)NR_2, salts thereof, and the like. Each R group in the preceding list includes H, alkyl or substituted alkyl, aryl or substituted aryl, or alkaryl. Where substituent groups are specified by their conventional chemical formulas, written from left to right, they equally encompass the chemically identical substituents that would result from writing the structure from right to left, for example, -CH_2O- is equivalent to -OCH_2-.

[0146] By way of example only, substituents for alkyl and heteroalkyl radicals (including those groups referred to as alkylene, alkenyl, heteroalkylene, heteroalkenyl, alkynyl, cycloalkyl, heterocycloalkyl, cycloalkenyl, and heterocycloalkenyl) includes: -OR, =0, =NR, =N-OR, -NR_2, -SR, -halogen, -SiR_3, -OC(O)R, -C(O)R, -CO_2R, -$CONR_2$, -OC(O)NR_2, -NRC(O)R, - NRC(O)NR_2, -NR$(O)_2$R, -NR-C(NR_2)=NR, -S(O)R, -S$(O)_2$R, -S$(O)_2NR_2$, -$NRSO_2R$, -CN and -NO_2. Each R group in the preceding list includes hydrogen, substituted or unsubstituted heteroalkyl, substituted or unsubstituted aryl, including aryl substituted with 1-3 halogens, substituted or unsubstituted alkyl, alkoxy or thioalkoxy groups, or aralkyl groups. When two R groups are attached to the same nitrogen atom, they can be combined with the nitrogen atom to form a 5-, 6-, or 7-membered ring. For example, -NR2 is meant to include 1-pyrrolidinyl and 4-morpholinyl.

[0147] By way of example, substituents for aryl and heteroaryl groups include -OR, =O, =NR, =N-OR, -NR_2, -SR,-halogen, -SiR_3, -OC(O)R, -C(O)R, -CO_2R, -$CONR_2$, -OC(O)NR_2, -NRC(O)R, -NRC(O)NR_2, -NR$(O)_2$R, -NR-C(NR_2)=NR, -S(O)R,-S$(O)_2$R, -S$(O)_2NR_2$, -$NRSO_2R$, -CN, -NO_2, -R, -N_3, -CH$(Ph)_2$, fluoro(C_1-C_4)alkoxy, and fluoro(C_1-C_4)alkyl, in a number ranging from zero to the total number of open valences on the aromatic ring system; and where each R group in the preceding list includes hydrogen, alkyl, heteroalkyl, aryl and heteroaryl.

[0148] The term "therapeutically effective amount," as used herein, refers to the amount of a composition containing at least one non-natural amino acid polypeptide and/or at least one modified non-natural amino acid polypeptide administered to a patient already suffering from a disease, condition or disorder, sufficient to cure or at least partially arrest, or relieve to some extent one or more of the symptoms of the disease, disorder or condition being treated. The effectiveness of such compositions depend conditions including the severity and course of the disease, disorder or condition, previous therapy, the patient's health status and response to the drugs, and the judgment of the treating physician. By way of example only, therapeutically effective amounts may be determined by routine experimentation, including a dose escalation clinical trial.

[0149] The term "thioalkoxy," as used herein, refers to sulfur containing alkyl groups linked to molecules via an oxygen atom.

[0150] The term "thermal melting point" or Tm is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of probes complementary to a target hybridize to the target sequence at equilibrium.

[0151] The term "toxic moiety," as used herein, refers to a compound which can cause harm or death.

[0152] The terms "treat," "treating" or "treatment", as used herein, include alleviating, abating or ameliorating a disease or condition symptoms, preventing additional symptoms, ameliorating or preventing the underlying metabolic causes of symptoms, inhibiting the disease or condition, e.g., arresting the development of the disease or condition, relieving the disease or condition, causing regression of the disease or condition, relieving a condition caused by the disease or condition, or stopping the symptoms of the disease or condition. The terms "treat," "treating" or "treatment", include prophylactic and/or therapeutic treatments.

[0153] As used herein, the term "water soluble polymer" refers to any polymer that is soluble in aqueous solvents. Such water soluble polymers include polyethylene glycol, polyethylene glycol propionaldehyde, mono C_1-C_{10} alkoxy or aryloxy derivatives thereof (described in U.S. Patent No. 5,252,714), monomethoxy-polyethylene glycol, polyvinyl pyrrolidone, polyvinyl alcohol, polyamino acids, divinylether maleic anhydride, N-(2-Hydroxypropyl)-methacrylamide, dextran, dextran derivatives including dextran sulfate, polypropylene glycol, polypropylene oxide/ethylene oxide copolymer, polyoxyethylated polyol, heparin, heparin fragments, polysaccharides, oligosaccharides, glycans, cellulose and cellulose derivatives, including methylcellulose and carboxymethyl cellulose, serum albumin, starch and starch derivatives, polypeptides, polyalkylene glycol and derivatives thereof, copolymers of polyalkylene glycols and derivatives thereof, polyvinyl ethyl ethers, and alpha-beta-poly[(2-hydroxyethyl)-DL-aspartamide, and the like, or mixtures thereof. By way of example only, coupling of such water soluble polymers to natural amino acid polypeptides or non-natural polypeptides may result in changes including increased water solubility, increased or modulated serum half-life, increased or modulated therapeutic half-life relative to the unmodified form, increased bioavailability, modulated biological activity, extended circulation time, modulated immunogenicity, modulated physical association characteristics including aggregation and multimer formation, altered receptor binding, altered binding to one or more binding partners, and altered receptor dimerization or multimerization. In addition, such water soluble polymers may or may not have their own biological activity.

[0154] Unless otherwise indicated, conventional methods of mass spectroscopy, NMR, HPLC, protein chemistry, biochemistry, recombinant DNA techniques and pharmacology, within the skill of the art are employed.

[0155] Compounds, (includingnon-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides, and reagents for producing the aforementioned compounds) presented herein include isotopically-labeled compounds, which are identical to those recited in the various formulas and structures presented herein, but for the fact that one or more atoms are replaced by an atom having an atomic mass or mass number different from the atomic mass or mass number usually found in nature. Examples of isotopes that can be incorporated into the present compounds include isotopes of hydrogen, carbon, nitrogen, oxygen, fluorine and chlorine, such as ^{2}H, ^{3}H, ^{13}C, ^{14}C, ^{15}N, ^{18}O, ^{17}O, ^{35}S, ^{18}F, ^{36}Cl, respectively. Certain isotopically-labeled compounds described herein, for example those into which radioactive isotopes such as ^{3}H and ^{14}C are incorporated, are useful in drug and/or substrate tissue distribution assays. Further, substitution with isotopes such as deuterium, i.e., ^{2}H, can afford certain therapeutic advantages resulting from greater metabolic stability, for example increased *in vivo* half-life or reduced dosage requirements.

[0156] Some of the compounds herein (including non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides, and reagents for producing the aforementioned compounds) have asymmetric carbon atoms and can therefore exist as enantiomers or diastereomers. Diasteromeric mixtures can be separated into their individual diastereomers on the basis of their physical chemical differences by methods known, for example, by chromatography and/or fractional crystallization. Enantiomers can be separated by converting the enantiomeric mixture into a diastereomeric mixture by reaction with an appropriate optically active compound (e.g., alcohol), separating the diastereomers and converting (e.g., hydrolyzing) the individual diastereomers to the corresponding pure enantiomers. All such isomers, including diastereomers, enantiomers, and mixtures thereof are considered as part of the compositions described herein.

[0157] In additional or further aspects, the compounds described herein (including, non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides, and reagents for producing the aforementioned compounds) are used in the form of pro-drugs. In additional or further embodiments, the compounds described herein ((including non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides, and reagents for producing the aforementioned compounds) are metabolized upon administration to an organism in need to produce a metabolite that is then used to produce a desired effect, including a desired therapeutic effect. In further or additional embodiments are active metabolites of non-natural amino acids and "modified or unmodified" non-natural amino acid polypeptides.

[0158] The methods and formulations described herein include the use of N-oxides, crystalline forms (also known as polymorphs), or pharmaceutically acceptable salts of non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides. In certain embodiments, non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides may exist as tautomers. All tautomers are included within the scope of the non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides presented herein. In addition, the non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein can exist in unsolvated as well as solvated forms with pharmaceutically acceptable solvents such as water, ethanol, and the like. The solvated forms of the non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides presented herein are also considered to be disclosed herein.

[0159] Some of the compounds herein (including non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides and reagents for producing the aforementioned compounds) may exist in several tautomeric forms. All such tautomeric forms are considered as part of the compositions described herein. Also, for example all enol-keto forms of any compounds (including non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides and reagents for producing the aforementioned compounds) herein are considered as part of the compositions described herein.

[0160] Some of the compounds herein (including to non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides and reagents for producing either of the aforementioned compounds) are acidic and may form a salt with a pharmaceutically acceptable cation. Some of the compounds herein (including non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides and reagents for producing the aforementioned compounds) can be basic and accordingly, may form a salt with a pharmaceutically acceptable anion. All such salts, including di-salts are within the scope of the compositions described herein and they can be prepared by conventional methods. For example, salts can be prepared by contacting the acidic and basic entities, in either an aqueous, non-aqueous or partially aqueous medium. The salts are recovered by using at least one of the following techniques: filtration, precipitation with a non-solvent followed by filtration, evaporation of the solvent, or, in the case of aqueous solutions, lyophilization.

[0161] Pharmaceutically acceptable salts of the non-natural amino acid polypeptides disclosed herein may be formed when an acidic proton present in the parent non-natural amino acid polypeptides either is replaced by a metal ion, by way of example an alkali metal ion, an alkaline earth ion, or an aluminum ion; or coordinates with an organic base. In addition, the salt forms of the disclosed non-natural amino acid polypeptides can be prepared using salts of the starting materials or intermediates. The non-natural amino acid polypeptides described herein may be prepared as a pharmaceutically acceptable acid addition salt (which is a type of a pharmaceutically acceptable salt) by reacting the free base form of non-natural amino acid polypeptides described herein with a pharmaceutically acceptable inorganic or organic acid. Alternatively, the non-natural amino acid polypeptides described herein may be prepared as pharmaceutically acceptable base addition salts (which is a type of a pharmaceutically acceptable salt) by reacting the free acid form of non-natural amino acid polypeptides described herein with a pharmaceutically acceptable inorganic or organic base.

[0162] The type of pharmaceutical acceptable salts, include: (1) acid addition salts, formed with inorganic acids such as hydrochloric acid, hydrobromic acid, sulfuric acid, nitric acid, phosphoric acid, and the like; or formed with organic acids such as acetic acid, propionic acid, hexanoic acid, cyclopentanepropionic acid, glycolic acid, pyruvic acid, lactic acid, malonic acid, succinic acid, malic acid, maleic acid, fumaric acid, tartaric acid, citric acid, benzoic acid, 3-(4-hydroxybenzoyl)benzoic acid, cinnamic acid, mandelic acid, methanesulfonic acid, ethanesulfonic acid, 1,2-ethanedisulfonic acid, 2-hydroxyethanesulfonic acid, benzenesulfonic acid, 2-naphthalenesulfonic acid, 4-methylbicyclo-[2.2.2]oct-2-ene-1-carboxylic acid, glucoheptonic acid, 4,4'-methylenebis-(3-hydroxy-2-ene-1 -carboxylic acid), 3-phenylpropionic acid, trimethylacetic acid, tertiary butylacetic acid, lauryl sulfuric acid, gluconic acid, glutamic acid, hydroxynaphthoic acid, salicylic acid, stearic acid, muconic acid, and the like; (2) salts formed when an acidic proton present in the parent compound either is replaced by a metal ion, e.g., an alkali metal ion, an alkaline earth ion, or an aluminum ion; or coordinates with an organic base. Acceptable organic bases include ethanolamine, diethanolamine, triethanolamine, tromethamine, N-methylglucamine, and the like. Acceptable inorganic bases include aluminum hydroxide, calcium hydroxide, potassium hydroxide, sodium carbonate, sodium hydroxide, and the like.

[0163] The corresponding counterions of the non-natural amino acid polypeptide pharmaceutical acceptable salts may be analyzed and identified using various methods including ion exchange chromatography, ion chromatography, capillary electrophoresis, inductively coupled plasma, atomic absorption spectroscopy, mass spectrometry, or any combination thereof. In addition, the therapeutic activity of such non-natural amino acid polypeptide pharmaceutical acceptable salts may be tested using the techniques and methods described in examples 87-91.

[0164] It should be understood that a reference to a salt includes the solvent addition forms or crystal forms thereof, particularly solvates or polymorphs. Solvates contain either stoichiometric or non-stoichiometric amounts of a solvent, and are often formed during the process of crystallization with pharmaceutically acceptable solvents such as water, ethanol, and the like. Hydrates are formed when the solvent is water, or alcoholates are formed when the solvent is alcohol. Polymorphs include the different crystal packing arrangements of the same elemental composition of a compound. Polymorphs usually have different X-ray diffraction patterns, infrared spectra, melting points, density, hardness, crystal shape, optical and electrical properties, stability, and solubility. Various factors such as the recrystallization solvent, rate of crystallization, and storage temperature may cause a single crystal form to dominate.

[0165] The screening and characterization of non-natural amino acid polypeptide pharmaceutical acceptable salts polymorphs and/or solvates may be accomplished using a variety of techniques including thermal analysis, x-ray diffraction, spectroscopy, vapor sorption, and microscopy. Thermal analysis methods address thermo chemical degradation or thermo physical processes including polymorphic transitions, and such methods are used to analyze the relationships between polymorphic forms, determine weight loss, to find the glass transition temperature, or for excipient compatibility studies. Such methods include Differential scanning calorimetry (DSC), Modulated Differential Scanning Calorimetry (MDCS), Thermogravimetric analysis (TGA), and Thermogravi-metric and Infrared analysis (TG/IR). X-ray diffraction methods include single crystal and powder diffractometers and synchrotron sources. The various spectroscopic techniques used include Raman, FTIR, UVIS, and NMR (liquid and solid state). The various microscopy techniques include polarized light microscopy, Scanning Electron Microscopy (SEM) with Energy Dispersive X-Ray Analysis (EDX), Environmental Scanning Electron Microscopy with EDX (in gas or water vapor atmosphere), IR microscopy, and Raman microscopy.

BRIEF DESCRIPTION OF THE FIGURES

[0166] A better understanding of the features and advantages of the present methods and compositions may be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of our methods, compositions, devices and apparatuses are utilized, and the accompanying drawings which illustrate embodiments of the invention, as well as illustrative aspects to aid working of the invention:

FIG. 1 presents a schematic representation of the relationship of certain aspects of the methods, compositions, strategies and techniques described herein.

FIG. 2 presents illustrative examples of the types of non-natural amino acids described herein. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 3 presents illustrative examples of the types of non-natural amino acids described herein. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 4 presents an illustrative example of the synthetic methodology used to make the non-natural amino acids of the invention and as described herein. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 5 presents illustrative examples of the synthetic methodology used to make the non-natural amino acids of the invention and as described herein. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 6 presents illustrative examples of the synthetic methodology used to make the non-natural amino acids of the invention and as described herein. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 7 presents illustrative examples of the post-translational modification of carbonyl-containing non-natural amino acid polypeptides with hydroxylamine-containing reagents to form modified oxime-containing non-natural amino acid polypeptides. Such non-natural amino acid polypeptides may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 8 presents illustrative examples of additives that can be used to enhance the reaction of carbonyl-containing non-natural amino acid polypeptides with hydroxylamine-containing reagents to form modified oxime-containing non-natural amino acid polypeptides.

FIG. 9 presents illustrative examples of the post-translational modification of oxime-containing non-natural amino acid polypeptides with carbonyl-containing reagents to form modified oxime-containing non-natural amino acid polypeptides. Such non-natural amino acid polypeptides may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 10 presents an illustrative example of the post-translational modification of hydroxylamine-containing non-natural amino acid polypeptides with carbonyl-containing reagents to form modified oxime-containing non-natural amino acid polypeptides. Such non-natural amino acid polypeptides may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 11 presents illustrative examples of PEG-containing reagents that can be used to modify non-natural amino acid polypeptides to form PEG-containing, oxime-linked non-natural amino acid polypeptides.

FIG. 12 presents illustrative, non-limiting examples of the synthesis of PEG-containing reagents that can be used to modify non-natural amino acid polypeptides to form PEG-containing, oxime-linked non-natural amino acid polypeptides.

FIG. 13 presents an illustrative example of the synthesis of an amide-based PEG-containing reagent that can be used to modify non-natural amino acid polypeptides to form PEG-containing, oxime-linked non-natural amino acid polypeptides.

FIG. 14 presents an illustrative example of the synthesis of a carbamate-based PEG-containing reagent that can be used to modify non-natural amino acid polypeptides to form PEG-containing, oxime-linked non-natural amino acid polypeptides.

FIG. 15 presents an illustrative example of the synthesis of a carbamate-based PEG-containing reagent that can be used to modify non-natural amino acid polypeptides to form PEG-containing, oxime-linked non-natural amino acid polypeptides.

FIG. 16 presents illustrative examples of the synthesis of simple PEG-containing reagents that can be used to modify non-natural amino acid polypeptides to form PEG-containing, oxime-linked non-natural amino acid polypeptides.

FIG. 17 presents illustrative examples of branched PEG-containing reagents that can be used to modify non-natural amino acid polypeptides to form PEG-containing, oxime-linked non-natural amino acid polypeptides, and the use of one such reagent to modify a carbonyl-based non-natural amino acid polypeptide.

FIG. 18 presents an illustrative example of the synthesis of a bifunctional linker group that can be used to modify and link non-natural amino acid polypeptides.

FIG. 19 presents illustrative examples of multifunctional linker groups that can be used to modify and link non-natural amino acid polypeptides.

FIG. 20 presents an illustrative representation of the use of a bifunctional linker group to modify and link a non-natural amino acid polypeptide to a PEG group.

FIG. 21 presents illustrative examples of the use of bifunctional linker groups to modify and link non-natural amino acid polypeptides to a PEG group.

FIG. 22 presents an illustrative representation of the use of a bifunctional linker group to link together two non-natural amino acid polypeptides to form a homodimer.

FIG. 23 presents an illustrative representation of the use of a bifunctional linker group to link together two different non-natural amino acid polypeptides to form a heterodimer.

FIG. 24 presents an illustrative non representation of the synthesis of a carbonyl-containing non-natural amino acid. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 25 presents an illustrative representation of the synthesis of a dicarbonyl-containing non-natural amino acid. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 26 presents an illustrative representation of the synthesis of a dicarboriyl-containing non-natural amino acid.

FIG. 27 presents an illustrative representation of the synthesis of a dicarbonyl-containing non-natural amino acid. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 28 presents an illustrative representation of the synthesis of a dicarbonyl-containing non-natural amino acid. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 29 presents an illustrative representation of the synthesis of a dicarbonyl-containing non-natural amino acid. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 30 presents an illustrative representation of the synthesis of a dicarbonyl-containing non-natural amino acid. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 31 presents an illustrative representation of the synthesis of a dicarbonyl-containing non-natural amino acid. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 32 presents an illustrative representation of the synthesis of a dicarbonyl-containing non-natural amino acid. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 33 presents an illustrative representation of the synthesis of a dicarbonyl-containing non-natural amino acid. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 34 presents illustrative representations of the syntheses of dicarbonyl-containing non-natural amino acids. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 35 presents illustrative representations of carbonyl- and dicarbonyl-containing non-natural amino acids. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 36 presents illustrative representations of the syntheses of non-natural amino acids. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 37 presents illustrative representations of the syntheses of carbonyl- and dicarbonyl-containing non-natural amino acids. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 38 presents illustrative representations of the syntheses of carbonyl- and dicarbonyl-containing non-natural amino acids. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 39 presents illustrative representations of the syntheses of carbonyl- and dicarbonyl-containing non-natural amino acids. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 40 presents illustrative representations of dicarbonyl-containing non-natural amino acids. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 41 presents illustrative representations of dicarbonyl-containing non-natural amino acids. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 42 presents illustrative representations of dicarbonyl-containing non-natural amino acids. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 43 presents illustrative representations of (a) protected or unprotected 1,3-ketoaldehyde-containing non-natural amino acids, and (b) 1-3-ketocarboxylyl (thio)ester-containing non-natural amino acids. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 44 presents illustrative representations of hydrazide-containing non-natural amino acids. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 45 presents illustrative representations of hydrazide-containing non-natural amino acids. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 46A and 46B presents illustrative representations of oxime-containing non-natural amino acids, and FIG. 46C presents illustrative representations of hydrazine-containing non-natural amino acids. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 47 presents illustrative representations of one-step conjugation to non-natural amino acid polypeptides and two-step conjugation to non-natural amino acid polypeptides. By way of example, such conjugations include PEGylation of to non-natural amino acid polypeptides.

FIG. 48 presents illustrative representations of the synthesis of mPEG-hydroxylamine compounds. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 49 presents illustrative representations of the synthesis of mPEG-hydroxylamine compounds. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 50 presents illustrative representations of the synthesis of mPEG-hydroxylamine compounds. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 51 presents illustrative representations of the synthesis of mPEG-hydroxylamine compounds. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 52 presents illustrative representations of the synthesis of mPEG-hydroxylamine compounds. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 53 presents illustrative representations of the synthesis of mPEG-hydroxylamine compounds. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 54 presents illustrative representations of the synthesis of mPEG-hydroxylamine compounds. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 55 presents illustrative representations of the synthesis of mPEG-hydroxylamine compounds. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 56 presents illustrative representations of the synthesis of mPEG-hydroxylamine compounds. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 57 presents illustrative representations of the synthesis of mPEG-hydroxylamine compounds. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 58 presents illustrative representations of the synthesis of mPEG-hydroxylamine compounds. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 59 presents illustrative representations of the synthesis of mPEG-hydroxylamine compounds. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 60 presents illustrative representations of the synthesis of hydroxylamine compounds. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 61 presents illustrative representations of the synthesis of mPEG-hydroxylamine compounds. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 62A presents illustrative representations of the synthesis of hydroxylamine compounds; FIG. 62B presents illustrative, representations of the synthesis of mPEG compounds. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 63 presents illustrative examples of (A) the modification of non-natural amino acid polypeptides by chemical conversion into carbonyl-containing (including dicarbonyl-containing) non-natural amino acid polypeptides and (B) the modification of non-natural amino acid polypeptides by chemical conversion into hydroxylamine-containing non-natural amino acid polypeptides. Such non-natural amino acid polypeptides may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 64 presents illustrative representations of the synthesis of PEG-hydroxylamine compounds. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

FIG. 65 presents illustrative representations of the synthesis of PEG-hydroxylamine compounds. Such non-natural amino acids may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein.

DETAILED DESCRIPTION OF THE INVENTION

I. Introduction

[0167] Recently, an entirely new technology in the protein sciences has been reported, which promises to overcome many of the limitations associated with site-specific modifications of proteins. Specifically, new components have been added to the protein biosynthetic machinery of the prokaryote *Escherichia coli (E. coli)* (e.g., L. Wang, et al., (2001), Science 292:498-500) and the eukaryote *Sacchromyces cerevisiae (S. cerevisiae)* (e.g., J. Chin et al., Science 301:964-7 (2003)), which has enabled the incorporation of non-natural amino acids to proteins *in vivo.* A number of new amino acids with novel chemical, physical or biological properties, including photoaffinity labels and photoisomerizable amino acids, keto amino acids, and glycosylated amino acids have been incorporated efficiently and with high fidelity into proteins in *E. coli* and in yeast in response to 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 3(11):1135-1137 ; J. W. Chin, et al., (2002), PNAS United States of America 99(17):11020-11024; and, L. Wang, & P. G. Schultz, (2002), Chem. Comm., 1-11 . These studies have demonstrated that it is possible to selectively and routinely introduce chemical functional groups that are not found in proteins, that are chemically inert to all of the functional groups found in the 20 common, genetically-encoded amino acids and that may be used to react efficiently and selectively to form stable covalent linkages.

II. Overview

[0168] Figure 1 presents an overview of the compositions, methods and techniques that are described herein. At one level, described herein are the tools (methods, compositions, techniques) for creating and using a polypeptide comprising at least one non-natural amino acid or modified non-natural amino acid with a carbonyl, dicarbonyl, oxime or hydroxylamine group. Such non-natural amino acids may contain further functionality, including a label; a dye; a polymer; a water-soluble polymer; a derivative of polyethylene glycol; a photocrosslinker; a cytotoxic compound; a drug; an affinity label; a photoaffinity label; a reactive compound; a resin; a second protein or polypeptide or polypeptide analog; an antibody or antibody fragment; a metal chelator; a cofactor; a fatty acid; a carbohydrate; a polynucleotide; a DNA; a RNA; an antisense polynucleotide; a saccharide, a water-soluble dendrimer, a cyclodextrin, a biomaterial; a nanoparticle; a spin label; a fluorophore; a metal-containing moiety; a radioactive moiety; a novel functional group; a group that covalently or noncovalently interacts with other molecules; a photocaged moiety; an actinic radiation excitable moiety; a ligand; a photoisomerizable moiety; biotin; a biotin analogue; a moiety incorporating a heavy atom; a chemically cleavable group; a photocleavable group; an elongated side chain; a carbon-linked sugar; a redox-active agent; an amino thioacid; a toxic moiety; an isotopically labeled moiety; a biophysical probe; a phosphorescent group; a chemiluminescent group; an electron dense group; a magnetic group; an intercalating group; a chromophore; an energy transfer agent; a biologically active agent (in which case, the biologically active agent can include an agent with therapeutic activity and the non-natural amino acid polypeptide or modified non-natural amino acid can serve either as a co-therapeutic agent with the attached therapeutic agent or as a means for delivery the therapeutic agent to a desired site within an organism); a detectable label; a small molecule; an inhibitory ribonucleic acid; a radionucleotide; a neutron-capture agent; a derivative of biotin; quantum dot(s); a nanotransmitter; a radiotransmitter; an abzyme, an activated complex activator, a virus, an adjuvant, an aglycan, an allergan, an angiostatin, an antihormone, an antioxidant, an aptamer, a guide RNA, a saponin, a shuttle vector, a macromolecule, a mimotope, a receptor, a reverse micelle, and any combination thereof. Note that the various aforementioned functionalities are not meant to imply that the members of one functionality can not be classified as members of another functionality. Indeed, there will be overlap depending upon the particular circumstances. By way of example only, a water-soluble polymer overlaps in scope with a derivative of polyethylene glycol, however the overlap is not complete and thus both functionalities are cited above.

[0169] As shown in Figure 1, in one aspect are methods for selecting and designing a polypeptide to be modified using the methods, compositions and techniques described herein. The new polypeptide may be designed de novo, including by way of example only, as part of high-throughput screening process (in which case numerous polypeptides may be designed, synthesized, characterized and/or tested) or based on the interests of the researcher. The new polypeptide may also be designed based on the structure of a known or partially characterized polypeptide. By way of example only, the Growth Hormone Gene Superfamily (see infra) has been the subject of intense study by the scientific community; a new polypeptide may be designed based on the structure of a member or members of this gene superfamily. The principles for selecting which amino acid(s) to substitute and/or modify are described separately herein. The choice of which modification to employ is also described herein, and can be used to meet the need of the experimenter or end user. Such needs may include manipulating the therapeutic effectiveness of the polypeptide, improving the safety profile of the polypeptide, adjusting the pharmacokinetics, pharmacologics and/or pharmacodynamics of the polypeptide, such as, by way of example only, increasing water solubility, bioavailability, increasing serum half-life, increasing therapeutic half-life, modulating immunogenicity, modulating biological activity, or extending the circulation time. In addition, such modifications include, by way of example only, providing additional functionality to the polypeptide, incorporating a tag, label or detectable signal into the polypeptide, easing the isolation properties of the polypeptide, and any combination of the aforementioned modifications.

[0170] Also described herein are non-natural amino acids that have or can be modified to contain an oxime, carbonyl, dicarbonyl, or hydroxylamine group. Included with this aspect are methods for producing, purifying, characterizing and using such non-natural amino acids. In another aspect described herein are methods, strategies and techniques for incorporating at least one such non-natural amino acid into a polypeptide. Also included with this aspect are methods for producing, purifying, characterizing and using such polypeptides containing at least one such non-natural amino acid. Also included with this aspect are compositions of and methods for producing, purifying, characterizing and using oligonucleotides (including DNA and RNA) that can be used to produce, at least in part, a polypeptide containing at least one non-natural amino acid. Also included with this aspect are compositions of and methods for producing, purifying, characterizing and using cells that can express such oligonucleotides that can be used to produce, at least in part, a polypeptide containing at least one non-natural amino acid.

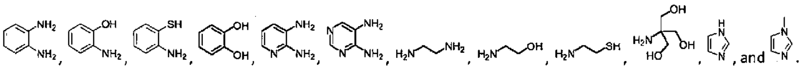
[0171] Thus, polypeptides comprising at least one non-natural amino acid or modified non-natural amino acid with a carbonyl, dicarbonyl, oxime or hydroxylamine group are provided and described herein. In certain aspects, polypeptides with at least one non-natural amino acid or modified non-natural amino acid with a carbonyl, dicarbonyl, oxime or hydroxylamine group include at least one post-translational modification at some position on the polypeptide. In some aspects the co-translational or post-translational modification occurs via the cellular machinery (e.g., glycosylation, acetylation, acylation, lipid-modification, palmitoylation, palmitate addition, phosphorylation, glycolipid-linkage modification, and the like), in many instances, such cellular-machinery-based co-translational or post-translational modifications occur at the naturally occurring amino acid sites on the polypeptide, however, in certain aspects, the cellular-machinery-based co-translational or post-translational modifications occur on the non-natural amino acid site(s) on the polypeptide.

[0172] In other aspects the post-translational modification does not utilize the cellular machinery, but the functionality is instead provided by attachment of a molecule (including a label; a dye; a polymer; a water-soluble polymer; a derivative of polyethylene glycol; a photocrosslinker; a cytotoxic compound; a drug; an affinity label; a photoaffinity label; a reactive compound; a resin; a second protein or polypeptide or polypeptide analog; an antibody or antibody fragment; a metal chelator; a cofactor; a fatty acid; a carbohydrate; a polynucleotide; a DNA; a RNA; an antisense polynucleotide; a saccharide, a water-soluble dendrimer, a cyclodextrin, a biomaterial; a nanoparticle; a spin label; a fluorophore, a metal-containing moiety; a radioactive moiety; a novel functional group; a group that covalently or noncovalently interacts with other molecules; a photocaged moiety; an actinic radiation excitable moiety; a ligand; a photoisomerizable moiety; biotin; a biotin analogue; a moiety incorporating a heavy atom; a chemically cleavable group; a photocleavable group; an elongated side chain; a carbon-linked sugar; a redox-active agent; an amino thioacid; a toxic moiety; an isotopically labeled moiety; a biophysical probe; a phosphorescent group; a chemiluminescent group; an electron dense group; a magnetic group; an intercalating group; a chromophore; an energy transfer agent; a biologically active agent; a detectable label; a small molecule; an inhibitory ribonucleic acid, a radionucleotide; a neutron-capture agent; a derivative of biotin; quantum dot(s); a nanotransmitter; a radiotransmitter; an abzyme, an activated complex activator, a virus, an adjuvant, an aglycan, an allergan, an angiostatin, an antihormone, an antioxidant, an aptamer, a guide RNA, a saponin, a shuttle vector, a macromolecule, a mimotope, a receptor, a reverse micelle, and any combination thereof) comprising a second reactive group to the at least one non-natural amino acid comprising a first reactive group (including non-natural amino acid containing a ketone, aldehyde, acetal, hemiacetal, oxime, or hydroxylamine functional group) utilizing chemistry methodology described herein, or others suitable for the particular reactive groups. In certain aspects, the co-translational or post-translational modification is made *in vivo* in a eukaryotic cell or in a non-eukaryotic cell. In certain aspects, the post-translational modification is made *in vitro* not utilizing the cellular machinery. Also included with this aspect are methods for producing, purifying, characterizing and using such polypeptides containing at least one such co-translationally or post-translationally modified non-natural amino acids.

[0173] Also included within the scope of the methods, compositions, strategies and techniques described herein are reagents capable of reacting with a non-natural amino acid (containing a carbonyl or dicarbonyl group, oxime group, hydroxylamine group, or masked or protected forms thereof) that is part of a polypeptide so as to produce any of the aforementioned post-translational modifications. In general, the resulting post-translationally modified non-natural amino acid will contain at least one oxime group; the resulting modified oxime-containing non-natural amino acid may undergo subsequent modification reactions. Also included with this aspect are methods for producing, purifying, characterizing and using such reagents that are capable of any such post-translational modifications of such non-natural amino acid(s).

[0174] In certain aspects, the polypeptide includes at least one co-translational or post-translational modification that is made *in vivo* by one host cell, where the post-translational modification is not normally made by another host cell type. In certain aspects, the polypeptide includes at least one co-translational or post-translational modification that is made *in vivo* by a eukaryotic cell, where the co-translational or post-translational modification is not normally made by a non-eukaryotic cell. Examples of such co-translational or post-translational modifications include glycosylation, acetylation, acylation, lipid-modification, palmitoylation, palmitate addition, phosphorylation, glycolipid-linkage modification, and the like. In one aspect, the co-translational or post-translational modification comprises attachment of an oligosaccharide to an asparagine by a GlcNAc-asparagine linkage (including where the oligosaccharide comprises $(\text{GlcNAc-Man})_2$-Man-GlcNAc-GlcNAc, and the like). In another aspect, the co-translational or post-translational modification comprises attachment of an oligosaccharide (including 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 aspects, a protein or polypeptide can comprise a secretion or localization sequence, an epitope tag, a FLAG tag, a polyhistidine tag, a GST fusion, and/or the like. Also included with this aspect are methods for producing, purifying, characterizing and using such polypeptides containing at least one such co-translational or post-translational modification. In other aspects, the glycosylated non-natural amino acid polypeptide is produced in a non-glycosylated form. Such a non-glycosylated form of a glycosylated non-natural amino acid may be produced by methods that include chemical or enzymatic removal of oligosaccharide groups from an isolated or substantially purified or unpurified glycosylated non-natural amino acid polypeptide; production of the non-natural amino acid in a host that does not glycosylate such a non-natural amino acid polypeptide (such a host including, prokaryotes or eukaryotes engineered or mutated to not glycosylate such a polypeptide), the introduction of a glycosylation inhibitor into the cell culture medium in which such a non-natural amino acid polypeptide is being produced by a eukaryote that normally would glycosylate such a polypeptide, or a combination of any such methods. Also described herein are such non-glycosylated forms of normally-glycosylated non-natural amino acid polypeptides (by normally-glycosylated is meant a polypeptide that would be glycosylated when produced under conditions in which naturally-occurring polypeptides are glycosylated). Of course, such non-glycosylated forms of normally-glycosylated non-natural amino acid polypeptides (or indeed any polypeptide described herein) may be in an unpurified form, a substantially purified form, or in an isolated form.

[0175] In certain embodiment, the non-natural amino acid polypeptide includes at least one post-translational modification that is made in the presence of an accelerant, wherein the post-translational modification is stoichiometric, stoichiometric-like, or near-stoichiometric. In other embodiment the polypeptide is contacted with a reagent of Formula (XIX) in the presence of an accelerant. In other embodiment of the invention the accelerant is selected from the group consisting of:



[0176] The non-natural amino acid containing polypeptide may 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 non-natural amino acids containing either a carbonyl or dicarbonyl group, oxime group, hydroxylamine group, or protected forms thereof. The non-natural amino acids can be the same or different, for example, there can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more different sites in the protein that comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more different non-natural 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 a non-natural amino acid.

[0177] The methods and compositions provided and described herein include polypeptides comprising at least one non-natural amino acid containing a carbonyl or dicarbonyl group, oxime group, hydroxylamine group, or protected or masked forms thereof. Introduction of at least one non-natural amino acid into a polypeptide can allow for the application of conjugation chemistries that involve specific chemical reactions, including with one or more non-natural amino acids while not reacting with the commonly occurring 20 amino acids. Once incorporated, the non-naturally occurring amino acid side chains can also be modified by utilizing chemistry methodologies described herein or suitable for the particular functional groups or substituents present in the naturally encoded amino acid.

[0178] The non-natural amino acid methods and compositions described herein provide conjugates of substances having a wide variety of functional groups, substituents or moieties, with other substances including a label; a dye; a polymer; a water-soluble polymer; a derivative of polyethylene glycol; a photocrosslinker; a cytotoxic compound; a drug; an affinity label; a photoaffinity label; a reactive compound; a resin; a second protein or polypeptide or polypeptide analog; an antibody or antibody fragment; a metal chelator; a cofactor; a fatty acid; a carbohydrate; a polynucleotide; a DNA; a RNA; an antisense polynucleotide; a saccharide, a water-soluble dendrimer, a cyclodextrin, a biomaterial; a nanoparticle; a spin label; a fluorophore, a metal-containing moiety; a radioactive moiety; a novel functional group; a group that covalently or noncovalently interacts with other molecules; a photocaged moiety; an actinic radiation excitable moiety; a ligand; a photoisomerizable moiety; biotin; a biotin analogue; a moiety incorporating a heavy atom; a chemically cleavable group; a photocleavable group; an elongated side chain; a carbon-linked sugar; a redox-active agent; an amino thioacid; a toxic moiety; an isotopically labeled moiety; a biophysical probe; a phosphorescent group; a chemiluminescent group; an electron dense group; a magnetic group; an intercalating group; a chromophore; an energy transfer agent; a biologically active agent; a detectable label; a small molecule; an inhibitory ribonucleic acid, a radionucleotide; a neutron-capture agent; a derivative of biotin; quantum dot(s); a nanotransmitter; a radiotransmitter; an abzyme, an activated complex activator, a virus, an adjuvant, an aglycan, an allergan, an angiostatin, an antihormone, an antioxidant, an aptamer, a guide RNA, a saponin, a shuttle vector, a macromolecule, a mimotope, a receptor, a reverse micelle, and any combination thereof.

[0179] In certain embodiments of the invention the non-natural amino acids, non-natural amino acid polypeptides, linkers and reagents described herein, including compounds of Formulas (I)-(XXXIII) are stable in aqueous solution under mildly acidic conditions (including pH 2 to 8). In other aspects, such compounds are stable for at least one month under mildly acidic conditions. In other aspects, such compounds are stable for at least 2 weeks under mildly acidic conditions. In other aspects, such compounds are stable for at least 5 days under mildly acidic conditions.

[0180] In another aspect of the compositions, methods, techniques and strategies described herein are methods for studying or using any of the aforementioned "modified or unmodified" non-natural amino acid polypeptides. Included within this aspect, by way of example only, are therapeutic, diagnostic, assay-based, industrial, cosmetic, plant biology, environmental, energy-production, consumer-products, and/or military uses which would benefit from a polypeptide comprising a "modified or unmodified" non-natural amino acid polypeptide or protein.

III. Location of non-natural amino acids in polypeptides

[0181] The methods and compositions described herein include incorporation of one or more non-natural amino acids into a polypeptide. One or more non-natural amino acids may be incorporated at one or more particular positions which does not disrupt activity of the polypeptide. This can be achieved by making "conservative" substitutions, including substituting hydrophobic amino acids with non-natural or natural hydrophobic amino acids, bulky amino acids with non-natural or natural bulky amino acids, hydrophilic amino acids with non-natural or natural hydrophilic amino acids) and/or inserting the non-natural amino acid in a location that is not required for activity.

[0182] A variety of biochemical and structural approaches can be employed to select the desired sites for substitution with a non-natural amino acid within the polypeptide. Any position of the polypeptide chain is suitable for selection to incorporate a non-natural amino acid, and selection may be based on rational design or by random selection for any or no particular desired purpose. Selection of desired sites may be based on producing a non-natural amino acid polypeptide (which may be further modified or remain unmodified) having any desired property or activity, including agonists, super-agonists, partial agonists, inverse agonists, antagonists, receptor binding modulators, receptor activity modulators, modulators of binding to binder partners, binding partner activity modulators, binding partner conformation modulators, dimer or multimer formation, no change to activity or property compared to the native molecule, or manipulating any physical or chemical property of the polypeptide such as solubility, aggregation, or stability. For example, locations in the polypeptide required for biological activity of a polypeptide can be identified using methods including point mutation analysis, alanine scanning or homolog scanning methods. Residues other than those identified as critical to biological activity by methods including alanine or homolog scanning mutagenesis may be good candidates for substitution with a non-natural amino acid depending on the desired activity sought for the polypeptide. Alternatively, the sites identified as critical to biological activity may also be good candidates for substitution with a non-natural amino acid, again depending on the desired activity sought for the polypeptide. Another alternative would be to simply make serial substitutions in each position on the polypeptide chain with a non-natural amino acid and observe the effect on the activities of the polypeptide. Any means, technique, or method for selecting a position for substitution with a non-natural amino acid into any polypeptide is suitable for use in the methods, techniques and compositions described herein.

[0183] The structure and activity of naturally-occurring mutants of a polypeptide that contain deletions can also be examined to determine regions of the protein that are likely to be tolerant of substitution with a non-natural amino acid. Once residues that are likely to be intolerant to substitution with non-natural amino acids have been eliminated, the impact of proposed substitutions at each of the remaining positions can be examined using methods including the three-dimensional structure of the relevant polypeptide, and any associated ligands or binding proteins. X-ray crystallographic and NMR structures of many polypeptides are available in the Protein Data Bank (PDB, www.rcsb.org), a centralized database containing three-dimensional structural data of large molecules of proteins and nucleic acids, one can be used to identify amino acid positions that can be substituted with non-natural amino acids.. In addition, models may be made investigating the secondary and tertiary structure of polypeptides, if three-dimensional structural data is not available. Thus, the identity of amino acid positions that can be substituted with non-natural amino acids can be readily obtained.

[0184] Exemplary sites of incorporation of a non-natural amino acid include those that are excluded from potential receptor binding regions, or regions for binding to binding proteins or ligands may be fully or partially solvent exposed, have minimal or no hydrogen-bonding interactions with nearby residues, may be minimally exposed to nearby reactive residues, and/or may be in regions that are highly flexible as predicted by the three-dimensional crystal structure of a particular polypeptide with its associated receptor, ligand or binding proteins.

[0185] A wide variety of non-natural amino acids can be substituted for, or incorporated into, a given position in a polypeptide. By way of example, a particular non-natural amino acid may be selected for incorporation based on an examination of the three dimensional crystal structure of a polypeptide with its associated ligand, receptor and/or binding proteins, a preference for conservative substitutions

[0186] In one embodiment, the methods described herein include incorporating into the polypeptide the non-natural amino acid, where the non-natural amino acid comprises a first reactive group; and contacting the polypeptide with a molecule (including a label; a dye; a polymer; a water-soluble polymer; a derivative of polyethylene glycol; a photocrosslinker; a cytotoxic compound; a drug; an affinity label; a photoaffinity label; a reactive compound; a resin; a second protein or polypeptide or polypeptide analog; an antibody or antibody fragment; a metal chelator; a cofactor; a fatty acid; a carbohydrate; a polynucleotide; a DNA; a RNA; an antisense polynucleotide; a saccharide, a water-soluble dendrimer, a cyclodextrin, a biomaterial; a nanoparticle; a spin label; a fluorophore, a metal-containing moiety; a radioactive moiety; a novel functional group; a group that covalently or noncovalently interacts with other molecules; a photocaged moiety; an actinic radiation excitable moiety; a ligand; a photoisomerizable moiety; biotin; a biotin analogue; a moiety incorporating a heavy atom; a chemically cleavable group; a photocleavable group; an elongated side chain; a carbon-linked sugar; a redox-active agent; an amino thioacid; a toxic moiety; an isotopically labeled moiety; a biophysical probe; a phosphorescent group; a chemiluminescent group; an electron dense group; a magnetic group; an intercalating group; a chromophore; an energy transfer agent; a biologically active agent; a detectable label; a small molecule; an inhibitory ribonucleic acid, a radionucleotide; a neutron-capture agent; a derivative of biotin; quantum dot(s); a nanotransmitter; a radiotransmitter; an abzyme, an activated complex activator, a virus, an adjuvant, an aglycan, an allergan, an angiostatin, an antihormone, an antioxidant, an aptamer, a guide RNA, a saponin, a shuttle vector, a macromolecule, a mimotope, a receptor, a reverse micelle, and any combination thereof) that comprises a second reactive group. In certain aspects, the first reactive group is a carbonyl or dicarbonyl moiety and the second reactive group is a hydroxylamine moiety, whereby an oxime linkage is formed. In certain aspects, the first reactive group is a hydroxylamine moiety and the second reactive group is carbonyl or dicarbonyl moiety, whereby an oxime linkage is formed. In certain embodiments, the first reactive group is a carbonyl or dicarbonyl moiety and the second reactive group is an oxime moiety, whereby an oxime exchange reaction occurs. In certain embodiments, the first reactive group is an oxime moiety and the second reactive group is carbonyl or dicarbonyl moiety, whereby an oxime exchange reaction occurs.

[0187] In some cases, the non-natural amino acid substitution(s) or incorporation(s) will be combined with other additions, substitutions, or deletions within the polypeptide to affect other chemical, physical, pharmacologic and/or biological traits. In some cases, the other additions, substitutions or deletions may increase the stability (including resistance to proteolytic degradation) of the polypeptide or increase affinity of the polypeptide for its appropriate receptor, ligand and/or binding proteins. In some cases, the other additions, substitutions or deletions may increase the solubility (including when expressed in E. coli or other host cells) of the polypeptide. In some embodiments sites are selected for substitution with a naturally encoded or non-natural amino acid in addition to another site for incorporation of a non-natural amino acid for the purpose of increasing the polypeptide solubility following expression in E. coli, or other recombinant host cells. In some embodiments, the polypeptides comprise another addition, substitution, or deletion that modulates affinity for the associated ligand, binding proteins, and/or receptor, modulates (including increases or decreases) receptor dimerization, stabilizes receptor dimers, modulates circulating half-life, modulates release or bio-availability, facilitates purification, or improves or alters a particular route of administration. Similarly, the non-natural amino acid polypeptide can comprise chemical or enzyme cleavage sequences, protease cleavage sequences, reactive groups, antibody-binding domains (including FLAG or poly-His) or other affinity based sequences (including FLAG, poly-His, GST, etc.) or linked molecules (including biotin) that improve detection (including GFP), purification, transport thru tissues or cell membranes, prodrug release or activation, size reduction, or other traits of the polypeptide.

IV. Growth Hormone Supergene Family as Exemplar

[0188] The methods, compositions, strategies and techniques described herein are not limited to a particular type, class or family of polypeptides or proteins. Indeed, virtually any polypeptides may be designed or modified to include at least one "modified or unmodified" non-natural amino acids described herein. By way of example only, the polypeptide can be homologous to a therapeutic protein selected from the group consisting of: alpha-1 antitrypsin, angiostatin, antihemolytic factor, antibody, antibody fragment, apolipoprotein, apoprotein, atrial natriuretic factor, atrial natriuretic polypeptide, atrial peptide, C-X-C chemokine, T39765, NAP-2, ENA-78, gro-a, gro-b, gro-c, IP-10, GCP-2, NAP-4, SDF-1, PF4, MIG, calcitonin, c-kit ligand, cytokine, CC chemokine, monocyte chemoattractant protein-1, monocyte chemoattractant protein-2, monocyte chemoattractant protein-3, monocyte inflammatory protein-1 alpha, monocyte inflammatory protein-i beta, RANTES, 1309, R83915, R91733, HCC1, T58847, D31065, T64262, CD40, CD40 ligand, c-kit ligand, collagen, colony stimulating factor (CSF), complement factor 5a, complement inhibitor, complement receptor 1, cytokine, epithelial neutrophil activating peptide-78, MIP-16, MCP-1, epidermal growth factor (EGF), epithelial neutrophil activating peptide, erythropoietin (EPO), exfoliating toxin, Factor IX, Factor VII, Factor VIII, Factor X, fibroblast growth factor (FGF), fibrinogen, fibronectin, four-helical bundle protein, G-CSF, glp-1, GM-CSF, glucocerebrosidase, gonadotropin, growth factor, growth factor receptor, grf, hedgehog protein, hemoglobin, hepatocyte growth factor (hGF), hirudin, human growth hormone (hGH), human serum albumin, ICAM-1, ICAM-1 receptor, LFA-1, LFA-1 receptor, insulin, insulin-like growth factor (IGF), IGF-I, IGF-II, interferon (IFN), IFN-alpha, IFN-beta, IFN-gamma, interleukin (IL), IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, keratinocyte growth factor (KGF), lactoferrin, leukemia inhibitory factor, luciferase, neurturin, neutrophil inhibitory factor (NIF), oncostatin M, osteogenic protein, oncogene product, paracitonin, parathyroid hormone, PD-ECSF, PDGF, peptide hormone, pleiotropin, protein A, protein G, pth, pyrogenic exotoxin A, pyrogenic exotoxin B, pyrogenic exotoxin C, pyy, relaxin, renin, SCF, small biosynthetic protein, soluble complement receptor I, soluble I-CAM 1, soluble interleukin receptor, soluble TNF receptor, somatomedin, somatostatin, somatotropin, streptokinase, superantigens, staphylococcal enterotoxin, SEA, SEB, SEC1, SEC2, SEC3, SED, SEE, steroid hormone receptor, superoxide dismutase, toxic shock syndrome toxin, thymosin alpha 1, tissue plasminogen activator, tumor growth factor (TGF), tumor necrosis factor, tumor necrosis factor alpha, tumor necrosis factor beta, tumor necrosis factor receptor (TNFR), VLA-4 protein, VCAM-1 protein, vascular endothelial growth factor (VEGF), urokinase, mos, ras, raf, met, p53, tat, fos, myc, jun, myb, rel, estrogen receptor, progesterone receptor, testosterone receptor, aldosterone receptor, LDL receptor, and corticosterone.

[0189] Thus, the following description of the growth hormone (GH) supergene family is provided for illustrative purposes.. Further, reference to GH polypeptides in this application is intended to use the generic term as an example of any member of the GH supergene family. Thus, it is understood that the modifications and chemistries described herein with reference to GH polypeptides or protein can be equally applied to any member of the GH supergene family, including those specifically listed herein.

[0190] The following proteins include those encoded by genes of the growth hormone (GH) supergene family (Bazan, F., Immunology Today 11: 350-354 (1990); Bazan, J. F. Science 257: 410-411 (1992); Mott, H. R. and Campbell, I. D., Current Opinion in Structural Biology 5: 114-121 (1995) ; Silvennoinen, O. and Ihle, J. N., Signalling by the Hematopoietic Cytokine Receptors (1996)): growth hormone, prolactin, placental lactogen, erythropoietin (EPO), thrombopoietin (TPO), interleukin-2 (IL-2), IL-3, IL-4, IL-5, IL-6, IL-7, IL-9, IL-10, IL-11, IL-12 (p35 subunit), IL-13, IL-15, oncostatin M, ciliary neurotrophic factor, leukemia inhibitory factor, alpha interferon, beta interferon, epsilon interferon, gamma interferon, omega interferon, tau interferon, granulocyte-colony stimulating factor (G-CSF), granulocyte-macrophage colony stimulating factor (GM-CSF), macrophage colony stimulating factor (M-CSF) and cardiotrophin-1 (CT-1) ("the GH supergene family"). It is anticipated that additional members of this gene family will be identified in the future through gene cloning and sequencing. Members of the GH supergene family have similar secondary and tertiary structures, despite the fact that they generally have limited amino acid or DNA sequence identity. The shared structural features allow new members of the gene family to be readily identified and the non-natural amino acid methods and compositions described herein similarly applied.

[0191] Structures of a number of cytokines, including G-CSF (Zink et al., FEBS Lett. 314:435 (1992); Zink et al., Biochemistry 33:8453 (1994); Hill et al., Proc. Natl. Acad. Sci.USA 90:5167 (1993)), GM-CSF (Diederichs, K., et al. Science 154: 1779-1782 (1991); Walter et al., J. Mol. Biol. 224:1075-1085 (1992)), IL-2 (Bazan, J: F. and McKay, D. B., Science 257: 410-413 (1992); IL-4 (Redfield et al., Biochemistry 30: 11029-11035 (1991); Powers et al., Science 256:1673-1677 (1992)), and IL-5 (Milburn et al., Nature 363: 172-176 (1993)) have been determined by X-ray diffraction and NMR studies and show striking conservation with the GH structure, despite a lack of significant primary sequence homology. IFN is considered to be a member of this family based upon modeling and other studies (Lee et al., J. Interferon Cytokine Res. 15:341 (1995); Murgolo et al., Proteins 17:62 (1993); Radhakrishnan et al., Structure 4:1453 (1996); Klaus et al., J. Mol. Biol. 274:661 (1997)). A large number of additional cytokines and growth factors including ciliary neurotrophic factor (CNTF), leukemia inhibitory factor (LIF), thrombopoietin (TPO), oncostatin M, macrophage colony stimulating factor (M-CSF), IL-3, IL-6, IL-7, IL-9, IL-12, IL-13, IL-15, and granulocyte-colony stimulating factor (G-CSF), as well as the IFN's such as alpha, beta, omega, tau, epsilon, and gamma interferon belong to this family (reviewed in Mott and Campbell, Current Opinion in Structural Biology 5: 114-121 (1995); Silvennoinen and Ihle (1996) Signalling by the Hematopoietic Cytokine Receptors). All of the above cytokines and growth factors are now considered to comprise one large gene family.

[0192] In addition to sharing similar secondary and tertiary structures, members of this family share the property that they must oligomerize cell surface receptors to activate intracellular signaling pathways. Some GH family members, including; GH and EPO, bind a single type of receptor and cause it to form homodimers. Other family members, including IL-2, IL4. and IL-6, bind more than one type of receptor and cause the receptors to form heterodimers or higher order aggregates (Davis et al., (1993) Science 260: 1805-1808; Paonessa et al., 1995) EMBO J. 14: 1942-1951; Mott and Campbell, Current Opinion in Structural Biology 5: 114-121 (1995)). Mutagenesis studies have shown that, like GH, these other cytokines and growth factors contain multiple receptor binding sites, typically two, and bind their cognate receptors sequentially (Mott and Campbell, Current Opinion in Structural Biology 5: 114-121 (1995); Matthews et al., (1996) Proc. Natl. Acad. Sci. USA 93: 9471-9476). Like GH, the primary receptor binding sites for these other family members occur primarily in the four alpha helices and the A-B loop. The specific amino acids in the helical bundles that participate in receptor binding differ amongst the family members. Most of the cell surface receptors that interact with members of the GH supergene family are structurally related and comprise a second large multi-gene family. See, e.g. U.S. Patent No. 6,608,183.

[0193] A general conclusion reached from mutational studies of various members of the GH supergene family is that the loops joining the alpha helices generally tend to not be involved in receptor binding. In particular the short B-C loop appears to be non-essential for receptor binding in most, if not all, family members. For this reason, the B-C loop may be substituted with non-natural amino acids as described herein in members of the GH supergene family. The A-B loop, the C-D loop (and D-E loop of interferon/ IL-10-like members of the GH superfamily) may also be substituted with a non-natural amino acid. Amino acids proximal to helix A and distal to the final helix also tend not to be involved in receptor binding and also may be sites for introducing non-natural amino acids. In some aspects, a non-natural amino acid is substituted at any position within a loop structure including the first 1, 2, 3,4, 5, 6, 7, or more amino acids of the A-B, B-C, C-D or D-E loop. In some aspects, a non-natural amino acid is substituted within the last 1, 2, 3, 4, 5, 6, 7, or more amino acids of the A-B, B-C, C-D or D-E loop.

[0194] Certain members of the GH family, including EPO, IL-2, IL-3, IL-4, IL-6, IFN, GM-CSF, TPO, IL-10, IL-12 p35, IL-13, IL-15 and beta interferon contain N-linked and/or O-linked sugars. The glycosylation sites in the proteins occur almost exclusively in the loop regions and not in the alpha helical bundles. Because the loop regions generally are not involved in receptor binding and because they are sites for the covalent attachment of sugar groups, they may be useful sites for introducing non-natural amino acid substitutions into the proteins. Amino acids that comprise the N- and O-linked glycosylation sites in the proteins may be sites for non-natural amino acid substitutions because these amino acids are surface-exposed. Therefore, the natural protein can tolerate bulky sugar groups attached to the proteins at these sites and the glycosylation sites tend to be located away from the receptor binding sites.

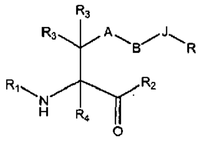
[0195] Additional members of the GH gene family are likely to be discovered in the future. New members of the GH supergene family can be identified through computer-aided secondary and tertiary structure analyses of the predicted protein sequences, and by selection techniques designed to identify molecules that bind to a particular target. Members of the GH supergene family typically possess four or five amphipathic helices joined by non-helical amino acids (the loop regions). The proteins may contain a hydrophobic signal sequence at their N-terminus to promote secretion from the cell. Such later discovered members of the GH supergene family also are included within the methods and compositions described herein.

V. Non-natural Amino Acids

[0196] The chemical moieties incorporated into polypeptides via incorporation of non-natural amino acids into such polypeptides offer a variety of advantages and manipulations of polypeptides. For example, the unique reactivity of a carbonyl or dicarbonyl functional group (including a keto- or aldehyde- functional group) allows selective modification of proteins with any of a number of hydrazine- or hydroxylamine-containing reagents *in vivo* and *in vitro*. A heavy atom non-natural amino acid, for example, can be useful for phasing x-ray structure data. The site-specific introduction of heavy atoms using non-natural amino acids also provides selectivity and flexibility in choosing positions for heavy atoms. Photoreactive non-natural amino acids (including amino acids with benzophenone and arylazides (including phenylazide) side chains), for example, allow for efficient *in vivo* and *in vitro* photocrosslinking of polypeptides. Examples of photoreactive non-natural amino acids include, p-azido-phenylalanine and p-benzoyl-phenylalanine. The polypeptide with the photoreactive non-natural amino acids may then be crosslinked at will by excitation of the photoreactive group-providing temporal control. In an example, the methyl group of a non-natural amino can be substituted with an isotopically labeled, including with a methyl group, as a probe of local structure and dynamics, including with the use of nuclear magnetic resonance and vibrational spectroscopy.

A. Structure and Synthesis of Non-Natural Amino Acids: Carbonyl, Carbonyl like, Masked Carbonyl, and Protected Carbonyl Groups

[0197] Amino acids with an electrophilic reactive group allow for a variety of reactions to link molecules via various chemical reactions, including nucleophilic addition reactions. Such electrophilic reactive groups include a carbonyl- or dicarbonyl-group (including a keto- or aldehyde group), a carbonyl-like- or dicarbonyl-like-group (which has reactivity similar to a carbonyl- or dicarbonyl-group and is structurally similar to a carbonyl- or dicarbonyl-group), a masked carbonyl-or masked dicarbonyl-group (which can be readily converted into a carbonyl- or dicarbonyl-group), or a protected carbonyl-or protected dicarbonyl-group (which has reactivity similar to a carbonyl- or dicarbonyl-group upon deprotection). Such amino acids include amino acids having the structure of Formula (I):



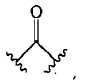
(I),

wherein:

A is optional, and when present is lower alkylene, substituted lower alkylene, lower cycloalkylene, substituted lower cycloalkylene, lower alkenylene, substituted lower alkenylene, alkynylene, lower heteroalkylene, substituted heteroalkylene, lower heterocycloalkylene, substituted lower heterocycloalkylene, arylene, substituted arylene, heteroarylene, substituted heteroarylene, alkarylene, substituted alkarylene, aralkylene, or substituted aralkylene;

B is optional, and when present is a linker selected from the group consisting of lower alkylene, substituted lower alkylene, lower alkenylene, substituted lower alkenylene, lower heteroalkylene, substituted lower heteroalkylene, -O-, -O-(alkylene or substituted alkylene)-, -S-, -S-(alkylene or substituted alkylene)-, -S(O)$_k$- where k is 1, 2, or 3, -S(O)$_k$(alkylene or substituted alkylene)-, -C(O)-, -C(O)-(alkylene or substituted alkylene)-, -C(S)-, -C(S)-(alkylene or substituted alkylene)-, -N(R')-, -NR'-(alkylene or substituted alkylene)-, -C(O)N(R')-, -CON(R')-(alkylene or substituted alkylene)-, -CSN(R')-, -CSN(R')-(alkylene or substituted alkylene)-, -N(R')CO-(alkylene or substituted alkylene)-, -N(R')C(O)O-, -S(O)$_k$N(R')-, -N(R')C(O)N(R')-, -N(R')C(S)N(R')-, -N(R')S(O)$_k$N(R')-, -N(R')-N=, -C(R')=N-, -C(R')=N-N(R')-, -C(R')=N-N=, -C(R')$_2$-N=N-, and -C(R')$_2$-N(R')-N(R')-, where each R' is independently H, alkyl, or substituted alkyl;

J is



R is H, alkyl, substituted alkyl, cycloalkyl, or substituted cycloalkyl;

R$_1$ is H, an amino protecting group, resin, amino acid, polypeptide, or polynucleotide; and

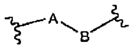
R$_2$ is OH, an ester protecting group, resin, amino acid, polypeptide, or polynucleotide;

each of R$_3$ and R$_4$ is independently H, halogen, lower, alkyl, or substituted lower alkyl;

[0198] Compounds of Formula (I) may be stable in aqueous solution for at least 1 month under mildly acidic conditions.Compounds of Formula (I) may be stable for at least 2 weeks under mildly acidic conditions.Compound of Formula (I) may be stable for at least 5 days under mildly acidic conditions. In certain embodiments, such acidic conditions are pH 2 to 8.

[0199] In certain embodiments of compounds of Formula (I), B is lower alkylene, substituted lower alkylene, -O-(alkylene or substituted alkylene)-, -C(R')=N-N(R')-, -N(R')CO-, -C(O)-, -C(R')=N-, -C(O)-(alkylene or substituted alkylene)-, -CON(R')-(alkylene or substituted alkylene)-, -S(alkylene or substituted alkylene)-, -S(O)(alkylene or substituted alkylene)-, or -S(O)$_2$(alkylene or substituted alkylene)-.

[0200] In certain embodiments of compounds of Formula (I),

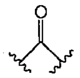


is selected from the group consisting of:

1. (i) A is substituted lower alkylene, substituted arylene, heteroarylene, substituted heteroarylene, alkarylene, substituted alkarylene, aralkylene, or substituted aralkylene;
   B is optional, and when present is a linker selected from the group consisting of lower alkylene, substituted lower alkylene, lower alkenylene, substituted lower alkenylene, -O-, -O-(alkylene or substituted alkylene)-, -S-, -S(O)-, -S(O)$_2$-, -C(O)-, -C(O)-(alkylene or substituted alkylene)-, -C(S)-, -N(R')-, -C(O)N(R')-, -CON(R')-(alkylene or substituted alkylene)-, -CSN(R')-, -N(R')CO-(alkylene or substituted alkylene)-, -N(R')C(O)O-, -S(O)N(R'), -S(O)$_2$N(R'), -N(R')C(O)N(R')-, -N(R')C(S)N(R')-, -N(R')S(O)N(R')-,-N(R')S(O)$_2$N(R')-, -N(R')-N=, -C(R')=N-N(R')-, -C(R')=N-N=, -C(R')$_2$-N=N-, and -C(R')$_2$-N(R')-N(R')-;
2. (ii) A is optional, and when present is substituted lower alkylene, substituted arylene, heteroarylene, substituted heteroarylene, alkarylene, substituted alkarylene, aralkylene, or substituted aralkylene; B is alinker selected from the group consisting of lower alkylene, substituted lower alkylene, lower alkenylene, substituted lower alkenylene, -0-, -O-(alkylene or substituted alkylene)-, -S-, -S(O)-, -S(O)$_2$-, - NS(O)$_2$-, -OS(O)$_2$-, -C(O)-, -C(O)-(alkylene or substituted alkylene)-, -C(S)-, -N(R')-, -C(O)N(R')-, -CON(R')-(alkylene or substituted alkylene)-, -CSN(R')-, -N(R')CO-(alkylene or substituted alkylene)-, -N(R')C(O)O-, - N(R')C(S)-, -S(O)N(R'), -S(O)$_2$N(R'), -N(R')C(O)N(R')-, -N(R')C(S)N(R')-, -N(R')S(O)N(R')-, -N(R')S(O)$_2$N(R')-, -N(R')-N=, -C(R')=N-N(R')-, -C(R')=N-N=, -C(R')$_2$-N=N-, and -C(R')$_2$-N(R')-N(R')-;
3. (iii) A is lower alkylene;
   B is optional, and when present is a linker selected from the group consisting of lower alkylene, substituted lower alkylene, lower alkenylene, substituted lower alkenylene, -0-, -O-(alkylene or substituted alkylene)-, -S-, -S(O)-, -S(O)$_2$-, -C(O)-, -C(O)-(alkylene or substituted alkylene)-, -C(S)-, -N(R')-, -C(O)N(R')-, -CSN(R')-, -CON(R')-(alkylene or substituted alkylene)-, -N(R')C(O)O-, -S(O)N(R'), -S(O)$_2$N(R'), -N(R')C(O)N(R')-, -N(R')C(S)N(R')-, -N(R')S(O)N(R')-, -N(R')S(O)$_2$N(R')-, -N(R')-N=, -C(R')=N-N(R')-, -C(R')=N-N=, -C(R')$_2$-N=N-, and -C(R')$_2$-N(R')-N(R')-; and
4. (iv) A is phenylene;
   B is a linker selected from the group consisting of lower alkylene, substituted lower alkylene, lower alkenylene, substituted

lower alkenylene, -0-, -O-(alkylene or substituted alkylene)-, -S-, -S(O)-, -S(O)$_2$-, -C(O)-, -C(O)-(alkylene or substituted alkylene)-, -C(S)-, -N(R')-, -C(O)N(R')-, -CON(R')-(alkylene or substituted alkylene)-, -CSN(R')-, -N(R')CO-(alkylene or substituted alkylene)-, -N(R')C(O)O-, -S(O)N(R'), -S(O)$_2$N(R'), -N(R')C(O)N(R')-, -N(R')C(S)N(R')-, -N(R')S(O)N(R')-, -N(R')S(O)$_2$N(R')-, -N(R')-N=, -C(R')=N-N(R')-, -C(R')=N-N=, -C(R')$_2$-N=N-, and -C(R')$_2$-N(R')-N(R')-;

J is



R$_1$ is optional, and when present, is H, an amino protecting group, resin, amino acid, polypeptide, or polynucleotide; and

R$_2$ is optional, and when present, is OH, an ester protecting group, resin, amino acid, polypeptide, or polynucleotide; and

each R$_3$ and R$_4$ is independently H, halogen, lower alkyl, or substituted lower alkyl;

R is H, alkyl, substituted alkyl, cycloalkyl, or substituted cycloalkyl;

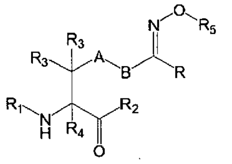
B. Structure and Synthesis of Non-Natural Amino Acids: Hydroxylamine-Containing Amino Acids

[0201] Non-natural amino acids containing a hydroxylamine (also called an aminooxy) group allow for reaction with a variety of electrophilic groups to form conjugates (including with PEG or other water soluble polymers). Like hydrazines, hydrazides and semicarbazides, the enhanced nucleophilicity of the aminooxy group permits it to react efficiently and selectively with a variety of molecules that contain carbonyl- or dicarbonyl-groups, including ketones, aldehydes or other functional groups with similar chemical reactivity. See, e.g., Shao, J. and Tam, J., J. Am. Chem. Soc. 117:3893-3899 (1995); H. Hang and C. Bertozzi, Acc. Chem. Res. 34(9): 727-736 (2001). Whereas the result of reaction with a hydrazine group is the corresponding hydrazone, however, an oxime results generally from the reaction of an aminooxy group with a carbonyl- or dicarbonyl-containing group such as, by way of example, a ketones, aldehydes or other functional groups with similar chemical reactivity.

C. ***Chemical Synthesis of Non-Natural Amino Acids: Oxime-containing amino acids***

[0202] Non-natural amino acids containing an oxime group allow for reaction with a variety of reagents that contain certain reactive carbonyl- or dicarbonyl- groups (including ketones, aldehydes, or other groups with similar reactivity) to form new non-natural amino acids comprising a new oxime group. Such an oxime exchange reaction allow for the further functionalization of non-natural amino acid polypeptides. Further, the original non-natural amino acids containing an oxime group may be useful in their own right as long as the oxime linkage is stable under conditions necessary to incorporate the amino acid into a polypeptide (e.g., the *in vivo, in vitro* and chemical synthetic methods described herein).

[0203] Thus, in certain embodiments described herein are non-natural amino acids with sidechains comprising an oxime group, an oxime-like group (which has reactivity similar to an oxime group and is structurally similar to an oxime group), a masked oxime group (which can be readily converted into an oxime group), or a protected oxime group (which has reactivity similar to an oxime group upon deprotection). Such amino acids include amino acids having the structure of Formula (XI):



(XI)

wherein:

A is optional, and when present is lower alkylene, substituted lower alkylene, lower cycloalkylene, substituted lower cycloalkylene, lower alkenylene, substituted lower alkenylene, alkynylene, lower heteroalkylene, substituted heteroalkylene, lower heterocycloalkylene, substituted lower heterocycloalkylene, arylene, substituted arylene, heteroarylene, substituted heteroarylene, alkarylene, substituted alkarylene, aralkylene, or substituted aralkylene;

B is optional, and when present is a linker selected from the group consisting of lower alkylene, substituted lower alkylene, lower alkenylene, substituted lower alkenylene, lower heteroalkylene, substituted lower heteroalkylene, -0-, -0-(alkylene or substituted alkylene)-, -S-, -S-(alkylene or substituted alkylene)-, -S(O)$_k$- where k is 1, 2, or 3, -S(O)$_k$(alkylene or substituted alkylene)-, -

C(O)- -C(O)-(alkylene or substituted alkylene)-, -C(S)-, -C(S)-(alkylene or substituted alkylene)-, -N(R')-, -NR'-(alkylene or substituted alkylene)-, -C(O)N(R')-, -CON(R')-(alkylene or substituted alkylene)-, - CSN(R')-, -CSN(R')-(alkylene or substituted alkylene)-, -N(R')CO-(alkylene or substituted alkylene)-, -N(R')C(O)O-, -S(O)$_k$N(R')-, -N(R')C(O)N(R')-, -N(R')C(S)N(R')-, -N(R')S(O)$_k$N(R')-, -N(R')-N=, -C(R')=N-, -C(R')=N-N(R')-, -C(R')=N-N=, -C(R')$_2$-N=N-, and -C(R')$_2$-N(R')-N(R')-, where each R' is independently H, alkyl, or substituted alkyl;

R is H, alkyl, substituted alkyl, cycloalkyl, or substituted cycloalkyl;

R_1 is H, an amino protecting group, resin, amino acid, polypeptide, or polynucleotide; and

R_2 is OH, an ester protecting group, resin, amino acid, polypeptide, or polynucleotide;

each of R_3 and R_4 is independently H, halogen, lower alkyl, or substituted lower alkyl;

R_5 is L-X, where X is a derivative of polyethylene glycol.

Such non-natural amino acids may be in the form of a salt, or may be incorporated into a non-natural amino acid polypeptide, polymer, polysaccharide, or a polynucleotide and optionally post translationally modified.

[0204] In certain embodiments of compounds of Formula (XI), B is -O-(alkylene or substituted alkylene)-.

[0205] Compounds of Formula (XI) may be stable in aqueous solution for at least 1 month under mildly acidic conditions. Compounds of Formula (XI) may be stable for at least 2 weeks under mildly acidic conditions.Compound of Formula (XI) may be stable for at least 5 days under mildly acidic conditions. In certain embodiments, such acidic conditions are pH 2 to 8.

D. Cellular uptake of non-natural amino acids

[0206] Non-natural amino acid uptake by a eukaryotic cell is one issue that is typically considered when designing and selecting non-natural amino acids, including 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 eukaryotic cell via a collection of protein-based transport systems. A rapid screen can be done which assesses which non-natural amino acids, if any, are taken up by cells (examples 15 & 16 herein illustrate non-limiting examples of tests which can be done on non-natural amino acids). *See, e.g.,* the toxicity assays in, e.g., the U.S. Patent Publication No. 2004/198637 entitled "Protein Arrays", 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 non-natural amino acids that are amenable to cellular uptake pathways is to provide biosynthetic pathways to create amino acids *in vivo.*

[0207] Typically, the non-natural amino acid produced via cellular uptake as described herein is produced in a concentration sufficient for efficient protein biosynthesis, including 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 this manner are about 10 mM to about 0.05 mM.

E. Biosynthesis of Non-Natural Amino Acids

[0208] Many biosynthetic pathways already exist in cells for the production of amino acids and other compounds. While a biosynthetic method for a particular non-natural amino acid may not exist in nature, including in a cell, the methods and compositions described herein provide such methods. For example, biosynthetic pathways for non-natural amino acids can be generated in host cell by adding new enzymes or modifying existing host cell pathways. Additional new enzymes include 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 eukaryotic 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 herein. Additional enzymes sequences are found, for example, in Genbank. Artificially evolved enzymes can be added into a cell in the same manner. In this manner, the cellular machinery and resources of a cell are manipulated to produce non-natural amino acids.

[0209] 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, including as developed by Maxygen, Inc. (available on the world wide web at www.maxygen.com), can be 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, including to engineer a pathway to create a non-natural amino acid in a cell. This technology reconstructs existing pathways in host organisms using a combination of new genes, including those identified through functional genomics, molecular evolution and design. Diversa Corporation (available on the world wide web at diversa.com) also provides technology for rapidly screening libraries of genes and gene pathways, including to create new pathways for biosynthetically producing non-natural amino acids.

[0210] Typically, the non-natural amino acid produced with an engineered biosynthetic pathway as described herein is produced in a concentration sufficient for efficient protein biosynthesis, including 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 a non-natural amino acid is generated, *in vivo* selections are optionally used to further optimize the production of the non-natural amino acid for both ribosomal protein synthesis and cell growth.

F. Additional Synthetic Methodology

[0211] The non-natural amino acids described herein may be synthesized using methodologies described in the art or using the techniques described herein or by a combination thereof. As an aid, the following table provides various starting electrophiles and nucleophiles which may be combined to create a desired functional group.

Table 1: Examples of Covalent Linkages and Precursors Thereof

Covalent Linkage Product	Electrophile	Nuclophile
Carboxamides	Activated esters	amines/anilines
Carboxamides	acyl azides	amines/anilines
Carboxamides	acyl halides	amines/anilines
Esters	acyl halides	alcohols/phenols
Esters	acyl nitriles	alcohols/phenols
Carboxamides	acyl nitriles	amines/anilines
Imines	Aldehydes	amines/anilines
Hydrazones	aldehydes or ketones	Hydrazines
Oximes	aldehydes or ketones	Hydroxylamines
Alkyl amines	alkyl halides	amines/anilines
Esters	alkyl halides	carboxylic acids
Thioethers	alkyl halides	Thiols
Ethers	alkyl halides	alcohols/phenols
Thioethers	alkyl sulfonates	Thiols
Esters	alkyl sulfonates	carboxylic acids
Ethers	alkyl sulfonates	alcohols/phenols
Esters	Anhydrides	alcohols/phenols
Carboxamides	Anhydrides	amines/anilines
Thiophenols	aryl halides	Thiols
Aryl amines	aryl halides	Amines
Thioethers	Azindines	Thiols
Boronate esters	Boronates	Glycols

Covalent Linkage Product	Electrophile	Nuclophile
Carboxamides	carboxylic acids	amines/anilines
Esters	carboxylic acids	Alcohols
hydrazines	Hydrazides	carboxylic acids
N-acylureas or Anhydrides	carbodiimides	carboxylic acids
Esters	diazoalkanes	carboxylic acids
Thioethers	Epoxides	Thiols
Thioethers	haloacetamides	Thiols
Ammotriazines	halotriazines	amines/anilines
Triazinyl ethers	halotriazines	alcohols/phenols
Amidines	imido esters	amines/anilines
Ureas	Isocyanates	amines/anilines
Urethanes	Isocyanates	alcohols/phenols
Thioureas	isoth iocyanates	amines/anilines
Thioethers	Maleimides	Thiols
Phosphite esters	phosphoramidites	Alcohols
Silyl ethers	silyl halides	Alcohols
Alkyl amines	sulfonate esters	amines/anilines
Thioethers	sulfonate esters	Thiols
Esters	sulfonate esters	carboxylic acids
Ethers	sulfonate esters	Alcohols
Sulfonamides	sulfonyl halides	amines/anilines
Sulfonate esters	sulfonyl halides	phenols/alcohols

[0212] In general, carbon electrophiles are susceptible to attack by complementary nucleophiles, including carbon nucleophiles, wherein an attacking nucleophile brings an electron pair to the carbon electrophile in order to form a new bond between the nucleophile and the carbon electrophile.

[0213] Examples of carbon nucleophiles include alkyl, alkenyl, aryl and alkynyl Grignard, organolithium, organozinc, alkyl-, alkenyl , aryl- and alkynyl-tin reagents (organostannanes), alkyl-, alkenyl-, aryl- and alkynyl-borane reagents (organoboranes and organoboronates); these carbon nucleophiles have the advantage of being kinetically stable in water or polar organic solvents. Other examples of carbon nucleophiles include phosphorus ylids, enol and enolate reagents; these carbon nucleophiles have the advantage of being relatively easy to generate from precursors well known to those skilled in the art of synthetic organic chemistry. Carbon nucleophiles, when used in conjunction with carbon electrophiles, engender new carbon-carbon bonds between the carbon nucleophile and carbon electrophile.

[0214] Examples of non-carbon nucleophiles suitable for coupling to carbon electrophiles include primary and secondary amines, thiols, thiolates, and thioethers, alcohols, alkoxides, azides, semicarbazides, and the like. These non-carbon nucleophiles, when used in conjunction with carbon electrophiles, typically generate heteroatom linkages (C-X-C), wherein X is a hetereoatom, including oxygen, sulfur, or nitrogen.

VI. Polypeptides with Non-natural Amino Acids

[0215] The compositions and methods described herein provide for the incorporation of at least one non-natural amino acid into a polypeptide. The non-natural amino acid may be present at any location on the polypeptide, including any terminal position or any internal position of the polypeptide. Preferably, the non-natural amino acid does not destroy the activity and/or the tertiary structure of the polypeptide relative to the homologous naturally-occurring amino acid polypeptide, unless such destruction of the activity and/or tertiary structure was one of the purposes of incorporating the non-natural amino acid into the polypeptide. Further, the incorporation of the non-natural amino acid into the polypeptide may modify to some extent the activity (e.g., manipulating the therapeutic effectiveness of the polypeptide, improving the safety profile of the polypeptide, adjusting the pharmacokinetics, pharmacologics and/or pharmacodynamics of the polypeptide (e.g., increasing water solubility, bioavailability, increasing serum half-life, increasing therapeutic half-life, modulating immunogenicity, modulating biological activity, or extending the circulation time), providing additional functionality to the polypeptide, incorporating a tag, label or detectable signal into the polypeptide, easing the isolation properties of the polypeptide, and any combination of the aforementioned modifications) and/or tertiary structure of the polypeptide relative to the homologous naturally-occurring amino acid polypeptide without fully causing destruction of the activity and/or tertiary structure. Such modifications of the activity and/or tertiary structure are often one of the goals of effecting such incorporations, although the incorporation of the non-natural amino acid into the polypeptide may also have little effect on the activity and/or tertiary structure of the polypeptide relative to the homologous naturally-occurring amino acid polypeptide. Correspondingly, non-natural amino acid polypeptides, compositions comprising non-natural amino acid polypeptides, methods for making such polypeptides and polypeptide compositions, methods for purifying, isolating, and characterizing such polypeptides and polypeptide compositions, and methods for using such polypeptides and polypeptide compositions are considered within the scope of the present disclosure. Further, the non-natural amino acid polypeptides described herein may also be ligated to another polypeptide (including, by way of example, a non-natural amino acid polypeptide or a naturally-occurring amino acid polypeptide).

[0216] The non-natural amino acid polypeptides described herein may be produced biosynthetically or non-biosynthetically. By biosynthetically is meant any method utilizing a translation system (cellular or non-cellular), including use of at least one of the following components: a polynucleotide, a codon, a tRNA, and a ribosome. By non-biosynthetically is meant any method not utilizing a translation system: this approach can be further divided into methods utilizing solid state peptide synthetic methods, solid phase peptide synthetic methods, methods that utilize at least one enzyme, and methods that do not utilize at least one enzyme; in addition any of this sub-divisions may overlap and many methods may utilize a combination of these sub-divisions.

[0217] The methods, compositions, strategies and techniques described herein are not limited to a particular type, class or family of polypeptides or proteins. Indeed, virtually any polypeptide may include at least one non-natural amino acids described herein. By way of example only, the polypeptide can be homologous to a therapeutic protein selected from the group consisting of: alpha-1 antitrypsin, angiostatin, antihemolytic factor, antibody, apolipoprotein, apoprotein, atrial natriuretic factor, atrial natriuretic polypeptide, atrial peptide, C-X-C chemokine, T39765, NAP-2, ENA-78, gro-a, gro-b, gro-c, IP-10, GCP-2, NAP-4, SDF-1, PF4, MIG, calcitonin, c-kit ligand, cytokine, CC chemokine, monocyte chemoattractant protein-1, monocyte chemoattractant protein-2, monocyte chemoattractant protein-3, monocyte inflammatory protein-1 alpha, monocyte inflammatory protein-i beta, RANTES, 1309, R83915, R91733, HCC1, T58847, D31065, T64262, CD40, CD40 ligand, c-kit ligand, collagen, colony stimulating factor (CSF), complement factor 5a, complement inhibitor, complement receptor 1, cytokine, epithelial neutrophil activating peptide-78, MIP-16, MCP-1, epidermal growth factor (EGF), epithelial neutrophil activating peptide, erythropoietin (EPO), exfoliating toxin, Factor IX, Factor VII, Factor VIII, Factor X, fibroblast growth factor (FGF), fibrinogen, fibronectin, four-helical bundle protein, G-CSF, glp-1, GM-CSF, glucocerebrosidase, gonadotropin, growth factor, growth factor receptor, grf, hedgehog protein, hemoglobin, hepatocyte growth factor (hGF), hirudin, human growth hormone (hGH), human serum albumin, ICAM-1, ICAM-1 receptor, LFA-1, LFA-1 receptor, insulin, insulin-like growth factor (IGF), IGF-I, IGF-II, interferon (IFN), IFN-alpha, IFN-beta, IFN-gamma, interleukin (IL), IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, keratinocyte growth factor (KGF), lactoferrin, leukemia inhibitory factor, luciferase, neurturin, neutrophil inhibitory factor (NIF), oncostatin M, osteogenic protein, oncogene product, paracitonin, parathyroid hormone, PD-ECSF, PDGF, peptide hormone, pleiotropin, protein A, protein G, pth, pyrogenic exotoxin A, pyrogenic exotoxin B, pyrogenic exotoxin C, pyy, relaxin, renin, SCF, small biosynthetic protein, soluble complement receptor I, soluble I-CAM 1, soluble interleukin receptor, soluble TNF receptor, somatomedin, somatostatin, somatotropin, streptokinase, superantigens, staphylococcal enterotoxin, SEA, SEB, SEC1, SEC2, SEC3, SED, SEE, steroid hormone receptor, superoxide dismutase, toxic shock syndrome toxin, thymosin alpha 1, tissue plasminogen activator, tumor growth factor (TGF), tumor necrosis factor, tumor necrosis factor alpha, tumor necrosis factor beta, tumor necrosis factor receptor (TNFR), VLA-4 protein, VCAM-1 protein, vascular endothelial growth factor (VEGF), urokinase, mos, ras, raf, met, p53, tat, fos, myc, jun, myb, rel, estrogen receptor, progesterone receptor, testosterone receptor, aldosterone receptor, LDL receptor, and corticosterone. In a related or further embodiment, the non-natural amino acid polypeptide may also be homologous to any polypeptide member of the growth hormone supergene family.

[0218] The non-natural amino acid polypeptides may be further modified as described elsewhere in this disclosure or the non-natural amino acid polypeptide may be used without further modification. Incorporation of non-natural amino acid into a polypeptides can be done for a variety of purposes, including tailoring changes in protein structure and/or function, changing size, acidity, nucleophilicity, hydrogen bonding, hydrophobicity, accessibility of protease target sites, targeting to a moiety (including for a polypeptide array), etc. Polypeptides that include a non-natural amino acid can have enhanced or even entirely new catalytic or biophysical properties. By way of example only, the following properties can be modified by inclusion of a non-natural amino acid into a polypeptide: toxicity, biodistribution, structural properties, spectroscopic properties, chemical and/or photochemical properties, catalytic ability, half-life (including serum half-life), ability to react with other molecules, including covalently or noncovalently, and the like. Compositions with polypeptides that include at least one non-natural amino acid are useful for, including novel therapeutics, diagnostics, catalytic enzymes, industrial enzymes, binding proteins (including antibodies), and research including, 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.

[0219] Further, the sidechain of the non-natural amino acid component(s) of a polypeptide can provide a wide range of additional functionality to the polypeptide; by way of example only, and not as a limitation, the sidechain of the non-natural amino acid portion of a polypeptide may include any of the following: a label; a dye; a polymer; a water-soluble polymer; a derivative of polyethylene glycol; a photocrosslinker; a cytotoxic compound; a drug; an affinity label; a photoaffinity label; a reactive compound; a resin; a second protein or polypeptide or polypeptide analog; an antibody or antibody fragment; a metal chelator; a cofactor; a fatty acid; a carbohydrate; a polynucleotide; a DNA; a RNA; an antisense polynucleotide; a saccharide, a water-soluble dendrimer, a cyclodextrin, a biomaterial; a nanoparticle; a spin label; a fluorophore, a metal-containing moiety; a radioactive moiety; a novel functional group; a group that covalently or noncovalently interacts with other molecules; a photocaged moiety; an actinic radiation excitable moiety; a ligand; a photoisomerizable moiety; biotin; a biotin analogue; a moiety incorporating a heavy atom; a chemically cleavable group; a photocleavable group; an elongated side chain; a carbon-linked sugar; a redox-active agent; an amino thioacid; a toxic moiety; an isotopically labeled moiety; a biophysical probe; a phosphorescent group; a chemiluminescent group; an electron dense group; a magnetic group; an intercalating group; a chromophore; an energy transfer agent; a biologically active agent; a detectable label; a small molecule; an inhibitory ribonucleic acid, a radionucleotide; a neutron-capture agent; a derivative of biotin; quantum dot(s); a nanotransmitter; a radiotransmitter; an abzyme, an activated complex activator, a virus, an adjuvant, an aglycan, an allergan, an angiostatin, an antihormone, an antioxidant, an aptamer, a guide RNA, a saponin, a shuttle vector, a macromolecule, a mimotope, a receptor, a reverse micelle, and any combination thereof.

[0220] In one aspect, a composition includes at least one polypeptide with at least one, including 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 non-natural amino acids. Such non-natural amino acids may be the same or different. In addition, there may be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more different sites in the polypeptide which comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more different, or the same, non-natural amino acids. In another aspect, a composition includes a polypeptide with at least one, but fewer than all, of a particular amino acid present in the polypeptide is substituted with a non-natural amino acid(s). For a given polypeptide with more than one non-natural amino acids, the non-natural amino acids can be identical or different (such as, by way of example only, the polypeptide can include two or more different types of non-natu ral amino acids, or can include two of the same non-natural amino acid). For a given polypeptide with more than two non-natural amino acids, the non-natural amino acids can be the same, different or a combination of a multiple number of non-natural amino acids of the same kind with at least one different non-natural amino acid.

[0221] Although embodiments of the non-natural amino acid polypeptides described herein may be chemically synthesized via solid phase peptide synthesis methods (such as, by way of example only, on a solid resin), by solution phase peptide synthesis methods, and/or without the aid of enzymes, other embodiments of the non-natural amino acid polypeptides described herein allow synthesis via a cell membrane, cellular extract, or lysate system or via an *in vivo* system, such as, by way of example only, using the cellular machinery of a prokaryotic or eukaryotic cell. In further or additional embodiments, one of the key features of the non-natural amino acid polypeptides described herein is that they may be synthesized utilizing ribosomes. In further or additional embodiments of the non-natural amino acid polypeptides described herein are, the non-natural amino acid polypeptides may be synthesized by a combination of the methods including a combination of solid resins, without the aid of enzymes, via the aid of ribosomes, and/or via an *in vivo* system.

[0222] Synthesis of non-natural amino acid polypeptides via ribosomes and/or an *in vivo* system has distinct advantages and characteristic from a non-natural amino acid polypeptide synthesized on a solid resin or without the aid of enzymes. These advantages or characteristics include different impurity profiles: a system utilizing ribosomes and/or an *in vivo* system will have impurities stemming from the biological system utilized, including host cell proteins, membrane portions, and lipids, whereas the impurity profile from a system utilizing a solid resin and/or without the aid of enzymes may include organic solvents, protecting groups, resin materials, coupling reagents and other chemicals used in the synthetic procedures. In addition, the isotopic pattern of the non-natural amino acid polypeptide synthesized via the use of ribosomes and/or an *in vivo* system may mirror the isotopic pattern of the feedstock utilized for the cells; on the other hand, the isotopic pattern of the non-natural amino acid polypeptide synthesized on a solid resin and/or without the aid of enzymes may mirror the isotopic pattern of the amino acids utilized in the synthesis. Further, the non-natural amino acid synthesized via the use of ribosomes and/or an *in vivo* system may be substantially free of the D-isomers of the amino acids and/or may be able to readily incorporate internal cysteine amino acids into the structure of the polypeptide, and/or may rarely provide internal amino acid deletion polypeptides. On the other hand, a non-natural amino acid polypeptide synthesized via a solid resin and/or without the use of enzymes may have a higher content of D-isomers of the amino acids and/or a lower content of internal cysteine amino acids and/or a higher percentage of internal amino acid deletion polypeptides. Furthermore, one of skill in the art will be able to differentiate a non-natural amino acid polypeptide synthesized by use of a ribosome and/or an *in vivo* system from a non-natural amino acid polypeptide synthesized via a solid resin and/or without the use of enzymes.

VII. Compositions and Methods Comprising Nucleic Acids and Oligonucleotides

A. General Recombinant Nucleic Acid Methods For Use Herein

[0223] In the methods and compositions described herein, nucleic acids encoding a polypeptide of interest (including by way of example a GH polypeptide) will be isolated, cloned and often altered using recombinant methods. Such aspects are used, including for protein expression or during the generation of variants, derivatives, expression cassettes, or other sequences derived from a polypeptide. In some aspects, the sequences encoding the polypeptides are operably linked to a heterologous promoter.

[0224] A nucleotide sequence encoding a polypeptide comprising a non-natural amino acid may be synthesized on the basis of the amino acid sequence of the parent polypeptide, and then changing the nucleotide sequence so as to effect introduction (i.e., incorporation or substitution) or removal (i.e., deletion or substitution) of the relevant amino acid residue(s). The nucleotide sequence may be conveniently modified by site-directed mutagenesis in accordance with conventional methods. Alternatively, the nucleotide sequence may be prepared by chemical synthesis, including by using an oligonucleotide synthesizer, wherein oligonucleotides are designed based on the amino acid sequence of the desired polypeptide, and preferably selecting those codons that are favored in the host cell in which the recombinant polypeptide will be produced. For example, several small oligonucleotides coding for portions of the desired polypeptide may be synthesized and assembled by PCR, ligation or ligation chain reaction. *See, e.g.,* Barany, et al., Proc. Natl. Acad. Sci. 88: 189-193 (1991); U.S. 6,521,427.

[0225] The non-natural amino acid methods and compositions described herein utilize routine techniques in the field of recombinant genetics: Basic texts disclosing the general methods of use for the non-natural amino acid methods and compositions described herein include Sambrook et al., Molecular Cloning, A Laboratory Manual (3rd ed. 2001); Kriegler, Gene Transfer and Expression: A Laboratory Manual (1990); and Current Protocols in Molecular Biology (Ausubel et al., eds., 1994)).

[0226] 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, including the generation of genes or polynucleotides which include selector codons for production of proteins that include non-natural amino acids, orthogonal tRNAs, orthogonal synthetases, and pairs thereof.

[0227] Various types of mutagenesis are used in the non-natural amino acid methods and compositions described herein for a variety of purposes, including to produce novel synthetases or tRNAs, to mutate tRNA molecules, to mutate polynucleotides encoding synthetases, libraries of tRNAs, to produce libraries of synthetases, to produce selector codons, to insert selector codons that encode non-natural amino acids in a protein or polypeptide of interest. They include site-directed mutagenesis, random point mutagenesis, homologous recombination, DNA shuffling or other recursive mutagenesis methods, chimeric construction, mutagenesis using uracil containing templates, oligonucleotide-directed mutagenesis, phosphorothioate-modified DNA mutagenesis, mutagenesis using gapped duplex DNA or the like, or any combination thereof. Additional suitable methods include point mismatch repair, mutagenesis using repair-deficient host strains, restriction-selection and restriction-purification, deletion mutagenesis, mutagenesis by total gene synthesis, double-strand break repair, and the like. Mutagenesis, including involving chimeric constructs, are also included in the non-natural amino acid methods and compositions described herein. In one embodiment, mutagenesis can be guided by known information of the naturally occurring molecule or altered or mutated naturally occurring molecule, including sequence comparisons, physical properties, crystal structure or the like.

[0228] The texts and examples found herein describe these and other relevant 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 selections, 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: 8765-8785 (1985); Nakamaye & Eckstein, Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis, Nucl. Acids Res. 14: 9679-9698 (1986); Sayers et al., 5'-3' Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis, Nucl. Acids Res. 16:791-802 (1988); Sayers et al., Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide, (1988) Nucl. Acids Res. 16: 803-814; Kramer et al., The gapped duplex DNA approach to oligonucleotide-directed mutation construction, Nucl. Acids Res. 12: 9441-9456 (1984); Kramer & Fritz Oligonucleotide-directed construction of mutations via gapped duplex DNA, Methods in Enzymol. 154:350-367 (1987); Kramer et al., Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations, Nucl. Acids Res. 16: 7207 (1988); Fritz et al., Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro, Nucl. Acids Res. 16: 6987-6999 (1988); Kramer et al., 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); Sakmar 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 such methods can be found in Methods in Enzymology Volume 154, which also describes useful controls for trouble-shooting problems with various mutagenesis methods.

[0229] The methods and compositions described herein also include use of eukaryotic host cells, non-eukaryotic host cells, and organisms for the *in vivo* incorporation of a non-natural amino acid via orthogonal tRNA/RS pairs. Host cells are genetically engineered (including transformed, transduced or transfected) with the polynucleotides corresponding to the polypeptides described herein or constructs which include a polynucleotide corresponding to the polypeptides described herein, including a vector corresponding to the polypeptides described herein, which can be, for example, a cloning vector or an expression vector. For example, the coding regions for the orthogonal tRNA, the orthogonal tRNA synthetase, and the protein to be derivatized are operably linked to gene expression control elements that are functional in the desired host cell. The vector can be, for example, in the form of a plasmid, cosmid, a phage, a bacterium, a virus, a naked polynucleotide, or a conjugated polynucleotide. The vectors are introduced into cells and/or microorganisms by standard methods including electroporation (Fromm et al., Proc. Natl. Acad. Sci. USA 82, 5824 (1985)), infection by viral vectors, high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface (Klein et al., Nature 327, 70-73 (1987)), and/or the like.

[0230] 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, including 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.

[0231] Several well-known methods of introducing target nucleic acids into cells are available, any of which can be used in methods and compositions described herein. These include: fusion of the recipient cells with bacterial protoplasts containing the DNA, electroporation, projectile bombardment, and infection with viral vectors (discussed further, herein), etc. Bacterial cells can be used to amplify the number of plasmids containing DNA constructs corresponding to the polypeptides described herein. 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, (including shuttle vectors) and selection markers for both prokaryotic and eukaryotic systems. Vectors are suitable for replication and integration in prokaryotes, eukaryotes, or preferably both. *See*, Gillam & Smith, Gene 8:81 (1979); Roberts, et al., Nature, 328:731 (1987); Schneider, E., et al., Protein Expr. Purif. 6(1):10-14 (1995); Ausubel, Sambrook, Berger (*all supra*). A catalogue of bacteria and bacteriophages useful for cloning is provided, e.g., by the ATCC, e.g., The ATCC Catalogue of bacteria and bacteriophage (1992) Gherna et al. (eds) published by the ATCC. Additional basic procedures for sequencing, cloning and other aspects of molecular biology and underlying theoretical considerations are also found in Watson et al. (1992) Recombinant DNA Second Edition Scientific American Books, NY. In addition, essentially any nucleic acid (and virtually any labeled nucleic acid, whether standard or non-standard) can be custom or standard ordered from any of a variety of commercial sources, such as the Midland Certified Reagent Company (Midland, TX 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.

B. Selector Codons

[0232] Selector codons encompassed within the methods and compositions described herein expand the genetic codon framework of protein biosynthetic machinery. For example, a selector codon includes a unique three base codon, a nonsense codon, such as a stop codon, including an amber codon (UAG), or an opal codon (UGA), a unnatural codon, a four or more base codon, a rare codon, or the like. There is a wide range in the number of selector codons that can be introduced into a desired gene or polynucleotide, including one or more, two or more, more than three, 4, 5, 6, 7, 8, 9, 10 or more in a single polynucleotide encoding at least a portion of a polypeptide of interest.

[0233] In one aspect, the methods involve the use of a selector codon that is a stop codon for the incorporation of one or more non-natural amino acids *in vivo*. For example, an O-tRNA is produced that recognizes the stop codon, including UAG, and is aminoacylated by an O-RS with a desired non-natural 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, including UAG, 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, 16(3):791-802. When the O-RS, O-tRNA and the nucleic acid that encodes the polypeptide of interest are combined *in vivo,* the non-natural amino acid is incorporated in response to the UAG codon to give a polypeptide containing the non-natural amino acid at the specified position.

[0234] The incorporation of non-natural amino acids *in vivo* can be done without significant perturbation of the eukaryotic host cell. For example, because the suppression efficiency for the UAG codon depends upon the competition between the O-tRNA, including the amber suppressor tRNA, and a eukaryotic release factor (including eRF) (which binds to a stop codon and initiates release of the growing peptide from the ribosome), the suppression efficiency can be modulated by, including increasing the expression level of O-tRNA, and/or the suppressor tRNA.

[0235] Selector codons also comprise extended codons, including four or more base codons, such as, four, five, six or more base codons. Examples of four base codons include AGGA, CUAG, UAGA, CCCU and the like. Examples of five base codons include AGGAC, CCCCU, CCCUC, CUAGA, CUACU, UAGGC and the like. A feature of the methods and compositions described herein includes using extended codons based on frameshift suppression. Four or more base codons can insert, including one or multiple non-natural amino acids into the same protein. For example, in the presence of mutated 0-tRNAs, including a special frameshift suppressor tRNAs, with anticodon loops, for example, with at least 8-10 nt anticodon loops, the four or more base codon is read as single amino acid. In other aspects, the anticodon loops can decode, including 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 non-natural 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.

[0236] For example, four-base codons have been used to incorporate non-natural amino acids into proteins using *in vitro* biosynthetic methods. *See, e.g.,* Ma et al., (1993) Biochemistry, 32:7939-7945; and Hohsaka et al., (1999) J. Am. Chem. Soc., 121:34-40. 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 tRNAs. *See, e.g.*, Hohsaka et al., (1999) J. Am. Chem. Soc., 121:12194-12195. 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., 298:195-205. In one aspect, extended codons based on rare codons or nonsense codons can be used in the methods and compositions described herein, which can reduce missense readthrough and frameshift suppression at other unwanted sites.

[0237] 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 tRNAthat recognizes the natural three base codon, and/or a system where the three base codon is a rare codon.

[0238] 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, and see also, Wu, Y., et. al. (2002) J. Am. Chem. Soc. 124:14626-14630. Other relevant publications are listed herein.

[0239] 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., 111:8322-8322 ; and Piccirilli et al., (1990) Nature, 343:33-37; Kool, (2000) Curr. Opin. Chem. Biol., 4:602-608. 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., 4:602-608; and Guckian and Kool, (1998) Angew. Chem. Int. Ed. Engl., 36(24): 2825-2828. In an effort to develop an unnatural base pair satisfying all the above requirements, Schultź, Romesberg and co-workers have systematically synthesized and studied a series of unnatural hydrophobic bases. A PICS:PICS self-pair is found to be more stable than natural base pairs, and can be efficiently incorporated into DNA by Klenow fragment of Escherichia coli DNA polymerase I (KF). *See, e.g.,* McMinn et al., (1999) J. Am. Chem. Soc., 121:11585-11586 ; and Ogawa et al., (2000) J. Am. Chem. Soc., 122:3274-3278. 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., 122:8803-8804. 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., 123:7439-7440 . 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., 122:10714-10715. Because extended codons and unnatural codons are intrinsically orthogonal to natural codons, the non-natural amino acid methods described herein can take advantage of this property to generate orthogonal tRNAs for them.

[0240] A translational bypassing system can also be used to incorporate a non-natural amino acid in a desired polypeptide. In a translational bypassing system, a large sequence is incorporated 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.

[0241] In certain embodiments, the protein or polypeptide of interest (or portion thereof) in the methods and/or compositions described herein 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.

[0242] 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, for example, one or more selector codons for the incorporation of a non-natural amino acid. For example, a nucleic acid for a protein of interest is mutagenized to include one or more selector codons, providing for the incorporation of the one or more non-natural amino acids. The methods and compositions described herein include any such variant, including mutant versions of any protein, for example, including at least one non-natural amino acid. Similarly, the methods and compositions described herein include corresponding nucleic acids, i.e., any nucleic acid with one or more selector codons that encodes or allows for the *in vivo* incorporation of one or more non-natural amino acid.

[0243] Nucleic acid molecules encoding a polypeptide of interest, including by way of example only, GH polypeptide may be readily mutated to introduce a cysteine at any desired position of the polypeptide. Cysteine is widely used to introduce reactive molecules, water soluble polymers, proteins, or a wide variety of other molecules, onto a protein of interest. Methods suitable for the incorporation of cysteine into a desired position of a polypeptide are well known in the art, such as those described in U.S. Patent No. 6,608,183, and standard mutagenesis techniques. The use of such cysteine-introducing and utilizing techniques can be used in conjunction with the non-natural amino acid introducing and utilizing techniques described herein.

VIII. In vivo generation of polypeptides comprising non-natural amino acids

[0244] The polypeptides described herein can be generated *in vivo* using modified tRNA and tRNA synthetases to add to or substitute amino acids that are not encoded in naturally-occurring systems.

[0245] Methods for generating tRNAs and tRNA synthetases which use amino acids that are not encoded in naturally-occurring systems are described in, e.g., U.S. Patent Application Publications 2003/0082575 (Serial No. 10/126,927) and 2003/0108885 (Serial No. 10/126,931). These methods involve generating a translational machinery that functions independently of the synthetases and tRNAs endogenous to the translation system (and are therefore sometimes referred to as "orthogonal"). In one aspect the translation system comprises a polynucleotide encoding the polypeptide; the polynucleotide can be mRNA that was transcribed from the corresponding DNA, or the mRNA may arise from an RNA viral vector; further the polynucleotide comprises a selector codon corresponding to the predesignated site of incorporation for the non-natural amino acid. The translation system further comprises a tRNA for and also when appropriate comprising the non-natural amino acid, where the tRNA is specific to/specifically recognizes the aforementioned selector codon; in further aspects, the non-natural amino acid is aminoacylated. The non-natural amino acids include those having the structure of any one of Formulas I-XVIII, XXX-XXXIV(A&B), and XXXX-XXXXIII described herein. In further or additional aspects, the translation system comprises an aminoacyl synthetase specific for the tRNA, and in other or further aspects, the translation system comprises an orthogonal tRNA and an orthogonal aminoacyl tRNA synthetase. In further or additional aspects, the translation system comprises at least one of the following: a plasmid comprising the aforementioned polynucleotide (such as, by way of example only, in the form of DNA), genomic DNA comprising the aforementioned polynucleotide (such as, by way of example only, in the form of DNA), or genomic DNA into which the aforementioned polynucleotide has been integrated (in further aspects, the integration is stable integration). In further or additional aspects of the translation system, the selector codon is selected from the group consisting of an amber codon, ochre codon, opal codon, a unique codon, a rare codon, an unnatural codon, a five-base codon, and a four-base codon. In further or additional aspects of the translation system, the tRNA is a suppressor tRNA. In further or additional aspects, the non-natural amino acid polypeptide is synthesized by a ribosome.

[0246] In further or additional aspects, the translation system comprises an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl tRNA synthetase (O-RS). Typically, the O-RS preferentially aminoacylates the O-tRNA with at least one non-natural amino acid in the translation system and the O-tRNA recognizes at least one selector codon that is not recognized by other tRNAs in the system. The translation system thus inserts the non-natural amino acid into a polypeptide produced in the system, in response to an encoded selector codon, thereby "substituting" a non-natural amino acid into a position in the encoded polypeptide.

[0247] A wide variety of orthogonal tRNAs and aminoacyl tRNA synthetases have been described in the art for inserting particular synthetic amino acids into polypeptides, and are generally suitable for in the methods described herein to produce the non-natural amino acid polypeptides described herein. For example, keto-specific O-tRNA/aminoacyl-tRNA synthetases are described in Wang, L., et al., Proc. Natl. Acad. Sci. USA 100(1):56-61 (2003) and Zhang, Z. et al., Biochem. 42(22):6735-6746 (2003). Exemplary O-RS, or portions thereof, are encoded by polynucleotide sequences and include amino acid sequences disclosed in U.S. Patent Application Publications 2003/0082575 and 2003/0108885. Corresponding 0-tRNA molecules for use with the O-RSs are also described in U.S. Patent Application Publications 2003/0082575 (Serial No. 10/126,927) and 2003/0108885 (Serial No. 10/126,931). In addition, Mehl et al. in J. Am. Chem. Soc. 2003; 125:935-939 and Santoro et al. Nature Biotechnology 2002 Oct; 20:1044-1048, discuss screening methods and aminoacyl tRNA synthetase and tRNA molecules for the incorporation of p-aminophenylalanine into polypeptides

[0248] Exemplary O-tRNA sequences suitable for use in the methods described herein include nucleotide sequences SEQ ID NOs: 1-3 as disclosed in U.S. Patent Application Publication 2003/0108885 (Serial No. 10/126,931). Other examples of O-tRNA/aminoacyl-tRNA synthetase pairs specific to particular non-natural amino acids are described in U.S. Patent Application Publication 2003/0082575 (Serial No. 10/126,927). O-RS and O-tRNA that incorporate both keto- and azide-containing amino acids in *S. cerevisiae* are described in Chin, J. W., et al., Science 301:964-967 (2003).

[0249] Use of O-tRNA/aminoacyl-tRNA synthetases involves selection of a specific codon which encodes the non-natural amino acid. While any codon can be used, it is generally desirable to select a codon that is rarely or never used in the cell in which the O-tRNA/aminoacyl-tRNA synthetase is expressed. By way of example only, exemplary codons include nonsense codon such as stop codons (amber, ochre, and opal), four or more base codons and other natural three-base codons that are rarely or unused.

[0250] Specific selector codon(s) can be introduced into appropriate positions in the polynucleotide coding sequence using mutagenesis methods known in the art (including site-specific mutagenesis, cassette mutagenesis, restriction selection mutagenesis, etc.).

[0251] Methods for generating components of the protein biosynthetic machinery, such as O-RSs, O-tRNAs, and orthogonal O-tRNA/O-RS pairs that can be used to incorporate a non-natural amino acid are described in Wang, L., et al., Science 292: 498-500 (2001); Chin, J. W., et al., J. Am. Chem. Soc. 124:9026-9027 (2002); Zhang, Z. et al., Biochemistry 42: 6735-6746 (2003). Methods and compositions for the *in vivo* incorporation of non-natural amino acids are described in U.S. Patent Application Publication 2003/0082575 (Serial No. 10/126,927). Methods for selecting an orthogonal tRNA-tRNA synthetase pair for use in *vivo* translation system of an organism are also described in U.S. Patent Application Publications 2003/0082575 (Serial No. 10/126,927) and 2003/0108885 (Serial No. 10/126,931). In addition PCT Publication No. WO 04/035743 entitled "Site Specific Incorporation of Keto Amino Acids into proteins, describes orthogonal RS and tRNA pairs for the incorporation of keto amino acids. PCT Publication No. WO 04/094593 entitled "Expanding the Eukaryotic Genetic Code,", describes orthogonal RS and tRNA pairs for the incorporation of non-naturally encoded amino acids in eukaryotic host cells.

[0252] Methods for producing at least one recombinant orthogonal aminoacyl-tRNA synthetase (O-RS) comprise: (a) generating a library of (optionally mutant) RSs derived from at least one aminoacyl-tRNA synthetase (RS) from a first organism, including a prokaryotic organism, such as, by way of example only, *Methanococcus jannaschii, Methanobacterium thermoautotrophicum, Holobacterium, Escherichia coli, A. fulgidus, P. furiosus, P. horikoshii, A. pernix, T. thermophilus* , or the like, or a eukaryotic organism; (b) selecting (and/or screening) the library of RSs (optionally mutant RSs) for members that aminoacylate an orthogonal tRNA (O-tRNA) in the presence of a non-natural amino acid and a natural amino acid, thereby providing a pool of active (optionally mutant) RSs; and/or, (c) selecting (optionally through negative selection) the pool for active RSs (including mutant RSs) that preferentially aminoacylate the O-tRNA in the absence of the non-natural amino acid, thereby providing the at least one recombinant O-RS; wherein the at least one recombinant O-RS preferentially aminoacylates the O-tRNA with the non-natural amino acid.

[0253] In one aspect, the RS is an inactive RS. The inactive RS can be generated by mutating an active RS. By way of example only, the inactive RS can be generated by mutating at least about 1, at least about 2, at least about 3, at least about 4, at least about 5, at least about 6, or at least about 10 or more amino acids to different amino acids, including alanine.

[0254] Libraries of mutant RSs can be generated using various techniques known in the art, including rational design based on protein three dimensional RS structure, or mutagenesis of RS nucleotides in a random or rational design technique. By way of example only, the mutant RSs can be generated by site-specific mutations, random mutations, diversity generating recombination mutations, chimeric constructs, rational design and by other methods described herein or known in the art.

[0255] In one embodiment, selecting (and/or screening) the library of RSs (optionally mutant RS's) for members that are active, including those which aminoacylate an orthogonal tRNA (O-tRNA) in the presence of a non-natural amino acid and a natural amino acid, includes: introducing a positive selection or screening marker, including an antibiotic resistance gene, or the like, and the library of (optionally mutant) RS's into a plurality of cells, wherein the positive selection and/or screening marker comprises at least one selector codon, including an amber codon, ochre codon, opal codon, a unique codon, a rare codon, an unnatural codon, a five-base codon, and a four-base codon; growing the plurality of cells in the presence of a selection agent; identifying cells that survive (or show a specific response) in the presence of the selection and/or screening agent by suppressing the at least one selector codon in the positive selection or screening marker, thereby providing a subset of positively selected cells that contains the pool of active (optionally mutant) RSs. Optionally, the selection and/or screening agent concentration can be varied.

[0256] In one aspect, the positive selection marker is a chloramphenicol acetyltransferase (CAT) gene and the selector codon is an amber stop codon in the CAT gene. Optionally, the positive selection marker is a
d
-lactamase gene and the selector codon is an amber stop codon in the
d
-lactamase gene. In another aspect the positive screening marker comprises a fluorescent or luminescent screening marker or an affinity based screening marker (including a cell surface marker).

[0257] In one embodiment, negatively selecting or screening the pool for active RS's (optionally mutants), including those which preferentially aminoacylate the O-tRNA in the absence of the non-natural amino acid includes: introducing a negative selection or screening marker with the pool of active (optionally mutant) RS's from the positive selection or screening into a plurality of cells of a second organism, wherein the negative selection or screening marker comprises at least one selector codon (including an antibiotic resistance gene, including a chloramphenicol acetyltransferase (CAT) gene); and, identifying cells that survive or show a specific screening response in a first medium supplemented with the non-natural amino acid and a screening or selection agent, but fail to survive or to show the specific response in a second medium not supplemented with the non-natural amino acid and the selection or screening agent, thereby providing surviving cells or screened cells with the at least one recombinant O-RS. By way of example only, a CAT identification protocol optionally acts as a positive selection and/or a negative screening in determination of appropriate O-RS recombinants. For instance, a pool of clones is optionally replicated on growth plates containing CAT (which comprises at least one selector codon) either with or without one or more non-natural amino acid. Colonies growing exclusively on the plates containing non-natural amino acids are thus regarded as containing recombinant O-RS. In one aspect, the concentration of the selection (and/or screening) agent is varied. In some aspects the first and second organisms are different. Thus, the first and/or second organism optionally comprises: a prokaryote, a eukaryote, a mammal, an *Escherichia coli,* a fungi, a yeast, an archaebecterium, a eubacterium, a plant, an insect, a protist, etc. In other aspects, the screening marker comprises a fluorescent or luminescent screening marker or an affinity based screening marker.

[0258] In another embodiment, screening or selecting (including, negatively selecting) the pool for active (optionally mutant) RS's includes: isolating the pool of active mutant RS's from the positive selection step (b); introducing a negative selection or screening marker, wherein the negative selection or screening marker comprises at least one selector codon (including a toxic marker gene, including a ribonuclease barnase gene, comprising at least one selector codon), and the pool of active (optionally mutant) RS's into a plurality of cells of a second organism; and identifying cells that survive or show a specific screening response in a first medium not supplemented with the non-natural amino acid, but fail to survive or show a specific screening response in a second medium supplemented with the non-natural amino acid, thereby providing surviving or screened cells with the at least one recombinant O-RS, wherein the at least one recombinant O-RS is specific for the non-natural amino acid. In one aspect, the at least one selector codon comprises about two or more selector codons. Such aspects optionally can include wherein the at least one selector codon comprises two or more selector codons, and wherein the first and second organism are different (including each organism is optionally, including a prokaryote, a eukaryote, a mammal, an *Escherichia coli,* a fungi, a yeast, an archaebacteria, a eubacteria, a plant, an insect, a protist, etc.). Also, some aspects include wherein the negative selection marker comprises a ribonuclease barnase gene (which comprises at least one selector codon). Other aspects include wherein the screening marker optionally comprises a fluorescent or luminescent screening marker or an affinity based screening marker. In the aspects herein, the screenings and/or selections optionally include variation of the screening and/or selection stringency.

[0259] In another embodiment, the methods for producing at least one recombinant orthogonal aminoacyl-tRNA synthetase (O-RS) may further comprise: (d) isolating the at least one recombinant O-RS; (e) generating a second set of 0-RS (optionally mutated) derived from the at least one recombinant O-RS; and, (f) repeating steps (b) and (c) until a mutated O-RS is obtained that comprises an ability to preferentially aminoacylate the O-tRNA. Optionally, steps (d)-(f) are repeated, including at least about two times. In one aspect, the second set of mutated O-RS derived from at least one recombinant 0-RS can be generated by mutagenesis, including random mutagenesis, site-specific mutagenesis, recombination or a combination thereof.

[0260] The stringency of the selection/screening steps, including the positive selection/screening step (b), the negative selection/screening step (c) or both the positive and negative selection/screening steps (b) and (c), in the above-described methods, optionally includes varying the selection/screening stringency. In another embodiment, the positive selection/screening step (b), the negative selection/screening step (c) or both the positive and negative selection/screening steps (b) and (c) comprise using a reporter, wherein the reporter is detected by fluorescence-activated cell sorting (FACS) or wherein the reporter is detected by luminescence. Optionally, the reporter is displayed on a cell surface, on a phage display or the like and selected based upon affinity or catalytic activity involving the non-natural amino acid or an analogue. In one embodiment, the mutated synthetase is displayed on a cell surface, on a phage display or the like.

[0261] Methods for producing a recombinant orthogonal tRNA (O-tRNA) include: (a) generating a library of mutant tRNAs derived from at least one tRNA, including a suppressortRNA, from a first organism; (b) selecting (including negatively selecting) or screening the library for (optionally mutant) tRNAs that are aminoacylated by an a min oacyl-tRNA synthetase (RS) from a second organism in the absence of a RS from the first organism, thereby providing a pool of tRNAs (optionally mutant); and, (c) selecting or screening the pool of tRNAs (optionally mutant) for members that are aminoacylated by an introduced orthogonal RS (O-RS), thereby providing at least one recombinant O-tRNA; wherein the at least one recombinant O-tRNA recognizes a selector codon and is not efficiency recognized by the RS from the second organism and is preferentially aminoacylated by the O-RS. In some aspects the at least one tRNA is a suppressor tRNA and/or comprises a unique three base codon of natural and/or unnatural bases, or is a nonsense codon, a rare codon, an unnatural codon, a codon comprising at least 4 bases, an amber codon, an ochre codon, or an opal stop codon. In one embodiment, the recombinant O-tRNA possesses an improvement of orthogonality. It will be appreciated that in some aspects, O-tRNA is optionally imported into a first organism from a second organism without the need for modification. In various aspects, the first and second organisms are either the same or different and are optionally chosen from, including prokaryotes (including *Methonococcus jannoschii, Methanobacterium thermoautotrophicum, Escherichia coli, Halobacterium,* etc.), eukaryotes, mammals, fungi, yeasts, archaebacteria, eubacteria, plants, insects, protists, etc. Additionally, the recombinant tRNA is optionally aminoacylated by a non-natural amino acid, wherein the non-natural amino acid is biosynthesized *in vivo* either naturally or through genetic manipulation. The non-natural amino acid is optionally added to a growth medium for at least the first or second organism, wherein the non-natural amino acid is capable of achieving appropriate intracellular concentrations to allow incorporation into the non-natural amino acid polypeptide.

[0262] In one aspect, selecting (including negatively selecting) or screening the library for (optionally mutant) tRNAs that are aminoacylated by an aminoacyl-tRNA synthetase (step (b)) includes: introducing a toxic marker gene, wherein the toxic marker gene comprises at least one of the selector codons (or a gene that leads to the production of a toxic or static agent or a gene essential to the organism wherein such marker gene comprises at least one selector codon) and the library of (optionally mutant) tRNAs into a plurality of cells from the second organism; and, selecting surviving cells, wherein the surviving cells contain the pool of (optionally mutant) tRNAs comprising at least one orthogonal tRNA or nonfunctional tRNA. For example, surviving cells can be selected by using a comparison ratio cell density assay.

[0263] In another aspect, the toxic marker gene can include two or more selector codons. In another embodiment of the methods described herein, the toxic marker gene is a ribonuclease bernase gene, where the ribonuclease barnase gene comprises at least one amber codon. Optionally, the ribonuclease barnase gene can include two or more amber codons.

[0264] In one embodiment, selecting or screening the pool of (optionally mutant) tRNAs for members that are aminoacylated by an introduced orthogonal RS (O-RS) can include: introducing a positive selection or screening marker gene, wherein the positive marker gene comprises a drug resistance gene (including β-lactamase gene, comprising at least one of the selector codons, such as at least one amber stop codon) or a gene essential to the organism, or a gene that leads to detoxification of a toxic agent, along with the O-RS, and the pool of (optionally mutant) tRNAs into a plurality of cells from the second organism; and, identifying surviving or screened cells grown in the presence of a selection or screening agent, including an antibiotic, thereby providing a pool of cells possessing the at least one recombinant tRNA, where the at least one recombinant tRNA is aminoacylated by the O-RS and inserts an amino acid into a translation product encoded by the positive markergene, in response to the at least one selector codons. In another embodiment, the concentration of the selection and/or screening agent is varied.

[0265] Methods for generating specific O-tRNA/O-RS pairs are provided. Methods include: (a) generating a library of mutant tRNAs derived from at least one tRNA from a first organism; (b) negatively selecting or screening the library for (optionally mutant) tRNAs that are aminoacylated by an aminoacyl-tRNA synthetase (RS) from a second organism in the absence of a RS from the first organism, thereby providing a pool of (optionally mutant) tRNAs; (c) selecting or screening the pool of (optionally mutant) tRNAs for members that are aminoacylated by an introduced orthogonal RS (O-RS), thereby providing at least one recombinant O-tRNA. The at least one recombinant O-tRNA recognizes a selector codon and is not efficiently recognized by the RS from the second organism and is preferentially aminoacylated by the O-RS. The method also includes (d) generating a library of (optionally mutant) RSs derived from at least one aminoacyl-tRNA synthetase (RS) from a third organism; (e) selecting or screening the library of mutant RS's for members that preferentially aminoacylate the at least one recombinant O-tRNA in the presence of a non-natural amino acid and a natural amino acid, thereby providing a pool of active (optionally mutant) RSs; and, (f) negatively selecting or screening the pool for active (optionally mutant) RSs that preferentially aminoacylate the at least one recombinant O-tRNA in the absence of the non-natural amino acid, thereby providing the at least one specific O-tRNA/O-RS pair, wherein the at least one specific O-tRNA/O-RS pair comprises at least one recombinant O-RS that is specific for the non-natural amino acid and the at least one recombinant O-tRNA. Specific O-tRNA/O-RS pairs produced by the methods described herein are included within the scope and methods described herein. For example, the specific O-tRNA/O-RS pair can include, including a mutRNATyr-mutTyrRS pair, such as a mutRNATyr-SS12TyrRS pair, a mutRNALeu-mutLeuRS pair, a mutRNAThr-mutThrRS pair, a mutRNAGlu-mutGluRS pair, or the like. Additionally, such methods include wherein the first and third organism are the same (including *Ethanococcus jannaschii*).

[0266] Methods for selecting an orthogonal tRNA-tRNA synthetase pair for use in an *in vivo* translation system of a second organism are also included in the methods described herein. The methods include: introducing a marker gene, a tRNA and an aminoacyl-tRNA synthetase (RS) isolated or derived from a first organism into a first set of cells from the second organism; introducing the marker gene and the tRNA into a duplicate cell set from a second organism; and, selecting for surviving cells in the first set that fail to survive in the duplicate cell set or screening for cells showing a specific screening response that fail to give such response in the duplicate cell set, wherein the first set and the duplicate cell set are grown in the presence of a selection or screening agent, wherein the surviving or screened cells comprise the orthogonal tRNA-tRNA synthetase pair for use in the in the *in vivo* translation system of the second organism. In one embodiment, comparing and selecting or screening includes an *in vivo* complementation assay. The concentration of the selection or screening agent can be varied.

[0267] The organisms described herein comprise a variety of organism and a variety of combinations. In one embodiment, the organisms are optionally a prokaryotic organism, including *Methanococcus jannaschii, Methanobacterium thermoautotrophicum, Halobacterium, Escherichia coli, A. fulgidus, P. furiosus, P. horikoshii, A. pernix, T. thermophilus,* or the like. Alternatively, the organisms are a eukaryotic organism, including plants (including complex plants such as monocots, or dicots), algae, protists, fungi (including yeast, etc), animals (including mammals, insects, arthropods, etc.), or the like.

A. Expression in Non-Eukaryotes and Eukaryotes

[0268] The techniques disclosed in this section can be applied to the expression in non-eukaryotes and eukaryotes of the non-natural amino acid polypeptides described herein.

[0269] To obtain high level expression of a cloned polynucleotide, one typically subclones polynucleotides encoding a desired polypeptide into an expression vector that contains a strong promoter to direct transcription, a transcription/translation terminator, and if for a nucleic acid encoding a protein, a ribosome binding site for translational initiation. Suitable bacterial promoters are described, e.g., in Sambrook *et al.* and Ausubel *et al.*

[0270] Bacterial expression systems for expressing polypeptides are available in, including *E. coli, Bacillus sp., Pseudomonas fluorescens, Pseudomonas aeruginosa, Pseudomonas putida,* and *Salmonella* (Palva et al., Gene 22:229-235 (1983); Mosbach et al., Nature 302:543-545 (1983). Kits for such expression systems are commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are commercially available. In cases where orthogonal tRNAs and aminoacyl tRNA synthetases (described elsewhere herein) are used to express the polypeptides, host cells for expression are selected based on their ability to use the orthogonal components. Exemplary host cells include Gram-positive bacteria (including *B. brevis* or *B. subtilis,* or *Streptomyces)* and Gram-negative bacteria (E. *coli* or *Pseudomonas fluorescens, Pseudomonas aeruginosa, Pseudomonas putida*), as well as yeast and other eukaryotic cells. Cells comprising O-tRNA/O-RS pairs can be used as described herein.

[0271] A eukaryotic host cell or non-eukaryotic host cell as described herein provides the ability to synthesize polypeptides which comprise non-natural amino acids in large useful quantities. In one aspect, the composition optionally includes, 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, at least 100 milligrams, at least one gram, or more of the polypeptide that comprises a non-natural amino acid, or an amount that can be achieved with *in vivo* polypeptide production methods (details on recombinant protein production and purification are provided herein). In another aspect, the polypeptide is optionally present in the composition at a concentration of, including at least 10 micrograms of polypeptide per liter, at least 50 micrograms of polypeptide per liter, at least 75 micrograms of polypeptide per liter, at least 100 micrograms of polypeptide per liter, at least 200 micrograms of polypeptide per liter, at least 250 micrograms of polypeptide per liter, at least 500 micrograms of polypeptide per liter, at least 1 milligram of polypeptide per liter, or at least 10 milligrams of polypeptide per liter or more, in, including a cell lysate, a buffer, a pharmaceutical buffer, or other liquid suspension (including in a volume of anywhere from about 1 nl to about 100 L). The production of large quantities (including greater that that typically possible with other methods, including in *vitro* translation) of a protein in a eukaryotic cell including at least one non-natural amino acid is a feature of the methods, techniques and compositions described herein.

[0272] A eukaryotic host cell or non-eukaryotic host cell as described herein provides the ability to biosynthesize proteins that comprise non-natural amino acids in large useful quantities. For example, polypeptides comprising a non-natural amino acid can be produced at a concentration of, including 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, at least 1mg/liter, at least 2mg/liter, at least 3 mg/liter, at least 4 mg/liter, at least 5 mg/liter, at least 6 mg/liter, at least 7 mg/liter, at least 8 mg/liter, at least 9 mg/liter, at least 10 mg/liter, at least 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900 mg/liter, 1 g/liter, 5 g/liter, 10 g/liter or more of polypeptide in a cell extract, cell lysate, culture medium, a buffer, and/or the like.

1. Expression Systems, Culture, Isolation

[0273] The techniques disclosed in this section can be applied to the expression systems, culture and isolation of the non-natural amino acid polypeptides described herein. Non-natural amino acid polypeptides may be expressed in any number of suitable expression systems including yeast, insect cells, mammalian cells, and bacteria. A description of exemplary expression systems is provided herein.

[0274] Yeast As used herein, the term "yeast" includes any of the various yeasts capable of expressing a gene encoding the non-natural amino acid polypeptide. Such yeasts include ascosporogenous yeasts (*Endoinycetales*), basidiosporogenous yeasts and yeasts belonging to the Fungi imperfecti (*Blastomycetes*) group. The ascosporogenous yeasts are divided into two families, *Spermophthoraceae* and *Saccharomycetaceae*. The latter is comprised of four subfamilies, *Schizosaccharomycoideae* (e.g., genus *Schizosaccharomyces*), *Nadsonioideae, Lipomycoideae* and *Saccharomycoideae* (e.g., genera *Pichia, Kluyveromyces* and *Sacchoromyces*). The basidiosporogenous yeasts include the genera *Leucosporidium, Rhodosporidium, Sporidiobolus, Filobasidium,* and *Filobasidiella.* Yeasts belonging to the Fungi Imperfecti (*Blastomycetes*) group are divided into two *families, Sporobolomycetaceae* (e.g., genera *Sporobolomyces* and *Bullera*) and *Cryptococcaceae* (e.g., genus *Candida*).

[0275] In certain aspects, the species within the genera *Pichia, Kluyveromyces, Saccharomyces, Schizosaccharomyces, Hansenula, Torulopsis,* and *Candida,* including *P. pastoris, P. guillerimondii, S. cerevisiae, S. carlsbergensis, S. diastaticus, S. douglasii, S. kluyveri, S, norbensis, S. oviformis, K. lactis, K. fragilis, C. albicans, C. maltosa,* and *H. polymorpha* are used in the methods, techniques and compositions described herein.

[0276] The selection of suitable yeast for expression of the non-natural amino acid polypeptide is within the skill of one of ordinary skill in the art. In selecting yeast hosts for expression, suitable hosts may include those shown to have, by way of example, good secretion capacity, low proteolytic activity, and overall robustness. Yeast are generally available from a variety of sources including the Yeast Genetic Stock Center, Department of Biophysics and Medical Physics, University of California (Berkeley, CA), and the American Type Culture Collection ("ATCC") (Manassas, VA).

[0277] The term "yeast host" or "yeast host cell" includes yeast that can be, or has been, used as a recipient for recombinant vectors or other transfer DNA. The term includes the progeny of the original yeast host cell that has received the recombinant vectors or other transfer DNA. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement to the original parent, due to accidental or deliberate mutation. Progeny of the parental cell that are sufficiently similar to the parent to be characterized by the relevant property, such as the presence of a nucleotide sequence encoding a non-natural amino acid polypeptide, are included in the progeny intended by this definition.

[0278] Expression and transformation vectors, including extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeast hosts. For example, expression vectors have been developed for *S. cerevisiae* (Sikorski et al., GENETICS (1998) 112:19; Ito et al., J. BACTERIOL. (1983) 153:163; Hinnen et al., PROC. NATL. ACAD. SCI. USA (1978) 75:1929); C. *albicans* (Kurtz et al., MOL. CELL. BIOL. (1986) 6:142); C. *maltosa* (Kunze et al., J. BASIC MICROBIOL. (1985) 25:141); *H. polymorpha* (Gleeson et al., J. GEN. MICROBIOL. (1986) 132:3459; Roggenkamp et al., MOL. GEN. GENET. (1986) 202:302); *K. fragilis* (Das et al., J. BACTERIOL. (1984) 158:1165); *K. lactis* (De Louvencourt et al., J. BACTERIOL. (1983) 154:737; Van den Berg et al., BIO/TECHNOLOGY (1990) 8:135); *P. guillerimondii* (Kunze et al., J. BASIC MICROBIOL. (1985) 25:141); *P. pastoris* (U.S. Patent Nos. 5,324,639; 4,929,555; and 4,837,148; Cregg et al., MOL. CELL. BIOL. (1985) 5:3376); *Schizosaccharomyces pombe* (Beach and Nurse, NATURE (1981) 300:706); and *Y. lipolytica* (Davidow et al., CURR. GENET. (1985) 10:380 (1985); Gaillardin et al., CURR. GENET. (1985) 10:49); *A. nidulans* (Ballance et al., BIOCHEM. BIOPHYS. RES. COMMUN. (1983) 112:284-89; Tilburn et al., GENE (1983) 26:205-221; and Yelton et al., PROC. NATL. ACAD. SCI. USA (1984) 81:1470-74); A. *niger* (Kelly and Hynes, EMBO J. (1985) 4:475-479); *T. reesia* (EP 0 244 234); and filamentous fungi such as, e.g., *Neurospora, Penicillium, Tolypocladium* (WO 91/00357).

[0279] Control sequences for yeast vectors include promoter regions from genes such as alcohol dehydrogenase (ADH) (EP 0 284 044); enolase; glucokinase; glucose-6-phosphate isomerase; glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH); hexokinase; phosphofructokinase; 3-phosphoglycerate mutase; and pyruvate kinase (PyK) (EP 0 329 203). The yeast PH05 gene, encoding acid phosphatase, also may provide useful promoter sequences (Miyanohara et al., PROC. NATL. ACAD. SCI. USA (1983) 80:1). Other suitable promoter sequences for use with yeast hosts may include the promoters for 3-phosphoglycerate kinase (Hitzeman et al., J. BIOL. CHEM. (1980) 255(4):12073-12080); and other glycolytic enzymes, such as pyruvate decarboxylase, triosephosphate isomerase, and phosphoglucose isomerase (Holland et al., BIOCHEMISTRY (1978) 17(23):4900-4907; Hess et al., J. ADV. ENZYME REG. (1969) 7:149-167). Inducible yeast promoters having the additional advantage of transcription controlled by growth conditions may include the promoter regions for alcohol dehydrogenase 2; isocytochrome C; acid phosphatase; metallothionein; glyceraldehyde-3-phosphate dehydrogenase; degradative enzymes associated with nitrogen metabolism; and enzymes responsible for maltose and galactose utilization. Suitable vectors and promoters for use in yeast expression are further described in EP 0073 657.

[0280] Yeast enhancers also may be used with yeast promoters. In addition, synthetic promoters may also function as yeast promoters. By way of example, the upstream activating sequences (UAS) of a yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region. *See* U.S. Patent Nos. 4,880,734 and 4,876,197. Other examples of hybrid promoters include promoters that consist of the regulatory sequences of the ADH2, GAL4, GAL10, or PH05 genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK. *See* EP 0164 556. Furthermore, a yeast promoter may include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription.

[0281] Other control elements that may comprise part of the yeast expression vectors include terminators, for example, from GAPDH or the enolase genes (Holland et al., J. BIOL. CHEM. (1981) 256:1385). In addition, the origin of replication from the 2µ plasmid origin is suitable for yeast. A suitable selection gene for use in yeast is the *trp1* gene present in the yeast plasmid. See Tschumper et al., GENE (1980) 10:157; Kingsman et al., GENE (1979) 7:141. The *trpl* gene provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan. Similarly, Leu2-deficient yeast strains (ATCC 20,622 or 38,626) are complemented by known plasmids bearing the *Leu2* gene.

[0282] Methods of introducing exogenous DNA into yeast hosts include either the transformation of spheroplasts or of intact yeast host cells treated with alkali cations. By way of example, transformation of yeast can be carried out according to the method described in Hsiao et al., PROC. NATL. ACAD. SCI. USA (1979) 76:3829 and Van Solingen et al., J. BACT. (1977) 130:946. However, other methods for introducing DNA into cells such as by nuclear injection, electroporation, or protoplast fusion may also be used as described generally in SAMBROOK ET AL., MOLECULAR CLONING: A LAB. MANUAL (2001). Yeast host cells may then be cultured using standard techniques known to those of ordinary skill in the art.

[0283] Other methods for expressing heterologous proteins in yeast host cells are described in U.S. Patent Publication No. 20020055169, U.S. Patent Nos. 6,361,969; 6,312;923; 6,183,985; 6,083,723; 6,017,731; 5,674,706; 5,629,203; 5,602,034; and 5,089,398; U.S. Reexamined Patent Nos. RE37,343 and RE35,749; PCT Published Patent Applications WO 99/07862; WO 98/37208; and WO 98/26080; European Patent Applications EP 0 946 736; EP 0 732 403; EP 0 480 480; WO 90/10277; EP 0 460 071; EP 0 340 986; EP 0 329 203; EP 0 324 274; and EP 0 164 556. *See also* Gellissen et al., ANTONIE VAN LEEUWENHOEK (1992) 62(1-2):79-93; Romanos et al., YEAST (1992) 8(6):423-488; Goeddel, METHODS IN ENZYMOLOGY (1990) 185:3-7.

[0284] The yeast host strains may be grown in fermentors during the amplification stage using standard feed batch fermentation methods. The fermentation methods may be adapted to account for differences in a particular yeast host's carbon utilization pathway or mode of expression control. By way of example only, fermentation of a *Saccharomyces* yeast host may require a single glucose feed, complex nitrogen source (e.g., casein hydrolysates), and multiple vitamin supplementation, whereas, the methylotrophic yeast *P. pastoris* may require glycerol, methanol, and trace mineral feeds, but only simple ammonium (nitrogen) salts for optimal growth and expression. *See, e.g.,* U.S. Patent No. 5,324,639; Elliott et al., J. PROTEIN CHEM. (1990) 9:95; and Fieschko et al., BIOTECH. BIOENG. (1987) 29:1113.

[0285] Such fermentation methods, however, may have certain common features independent of the yeast host strain employed. By way of example, a growth limiting nutrient, typically carbon, may be added to the fermentor during the amplification phase to allow maximal growth. In addition, fermentation methods generally employ a fermentation medium designed to contain adequate amounts of carbon, nitrogen, basal salts, phosphorus, and other minor nutrients (vitamins, trace minerals and salts, etc.). Examples of fermentation media suitable for use with *Pichia* are described in U.S. Patent Nos. 5,324,639 and 5,231,178.

[0286] Baculovirus-Infected Insect Cells The term "insect host" or "insect host cell" refers to a insect that can be, or has been, used as a recipient for recombinant vectors or other transfer DNA. The term includes the progeny of the original insect host cell that has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement to the original parent, due to accidental or deliberate mutation. Progeny of the parental cell that are sufficiently similar to the parent to be characterized by the relevant property, such as the presence of a nucleotide sequence encoding a non-natural amino acid polypeptide, are included in the progeny intended by this definition.

[0287] The selection of suitable insect cells for expression of a polypeptide is well known to those of ordinary skill in the art. Several insect species are well described in the art and are commercially available including *Aedes aegypti, Bombyx mori, Drosophila melanogaster, Spodoptera frugiperda*, and *Trichoplusia ni.* In selecting insect hosts for expression, suitable hosts may include those shown to have, *interalia,* good secretion capacity, low proteolytic activity, and overall robustness. Insect are generally available from a variety of sources including the Insect Genetic Stock Center, Department of Biophysics and Medical Physics, University of California (Berkeley, CA); and the American Type Culture Collection ("ATCC") (Manassas, VA).

[0288] Generally, the components of a baculovirus-infected insect expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene to be expressed; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media. The materials, methods and techniques used in constructing vectors, transfecting cells, picking plaques, growing cells in culture, and the like are known in the art and manuals are available describing these techniques.

[0289] After inserting the heterologous gene into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome recombine. The packaged recombinant virus is expressed and recombinant plaques are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, for example, Invitrogen Corp. (Carlsbad, CA). Illustrative techniques are described in SUMMERS AND SMITH, TEXAS AGRICULTURAL EXPERIMENT STATION BULLETIN No. 1555 (1987). See *also*, RICHARDSON, 39 METHODS IN MOLECULAR BIOLOGY: BACULOVIRUS EXPRESSION PROTOCOLS (1995); AUSUBEL ET AL., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY 16.9-16.11 (1994) ; KING AND POSSEE, THE BACULOVIRUS SYSTEM: A LABORATORY GUIDE (1992); and O'REILLY ET AL., BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL (1992).

[0290] The production of various heterologous proteins using baculovirus/insect cell expression systems is described in the following references and such techniques can be adapted to produce the non-natural amino acid polypeptides described herein. *See, e.g.,* U.S. Patent Nos. 6,368,825; 6,342,216; 6,338,846; 6,261,805; 6,245,528, 6,225,060; 6,183,987; 6,168,932; 6,126,944; 6,096,304; 6,013,433; 5,965,393; 5,939,285; 5,891,676; 5,871,986; 5,861,279; 5,858,368; 5,843,733; 5,762,939; 5,753,220; 5,605,827; 5,583,023; 5,571,709; 5,516,657; 5,290,686; WO 02/06305; WO 01/90390; WO 01/27301; WO 01/05956; WO 00/55345; WO 00/20032 WO 99/51721; WO 99/45130; WO 99/31257; WO 99/10515; WO 99/09193; WO 97/26332; WO 96/29400; WO 96/25496; WO 96/06161; WO 95/20672; WO 93/03173; WO 92/16619; WO 92/03628; WO 92/01801; WO 90/14428; WO 90/10078; WO 90/02566; WO 90/02186; WO 90/01556; WO 89/01038; WO 89/01037; WO 88/07082.

[0291] Vectors that are useful in baculovirus/insect cell expression systems include insect expression and transfer vectors derived from the baculovirus *Autographacalifornica* nuclear polyhedrosis virus (AcNPV), which is a helper-independent, viral expression vector. Viral expression vectors derived from this system usually use the strong viral polyhedrin gene promoter to drive expression of heterologous *genes. See generally,* Reilly ET AL., Baculovirus EXPRESSION VECTORS: A LABORATORY MANUAL (1992).

[0292] Prior to inserting the foreign gene into the baculovirus genome, the above-described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are typically assembled into an intermediate transplacement construct (transfer vector). Intermediate transplacement constructs are often maintained in a replicon, such as an extra chromosomal element (e.g., plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification. More specifically, the plasmid may contain the polyhedrin polyadenylation signal (Miller et al., ANN. REV. MICROBIOL. (1988) 42:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli.*

[0293] One commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed including, for example, pVL985, which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 base pairs downstream from the ATT. See Luckow and Summers, VIROLOGY 170:31-39 (1989). Other commercially available vectors include, for example, PBlueBac4.5/V5-His; pBlueBacHis2; pMelBac; pBlueBac4.5 (Invitrogen Corp., Carlsbad, CA).

[0294] After insertion of the heterologous gene, the transfer vector and wild type baculoviral genome are co-transfected into an insect cell host. Illustrative methods for introducing heterologous DNA into the desired site in the baculovirus virus described in SUMMERS AND SMITH, TEXAS AGRICULTURAL EXPERIMENT STATION BULLETIN No. 1555 (1987); Smith et al., MOL. CELL. BIOL. (1983) 3:2156; Luckow and Summers, VIROLOGY (1989) 170:31-39. By way of example, the insertion can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. See Miller et al., BIOESSAYS (1989) 4:91.

[0295] Transfection may be accomplished by electroporation using methods described in TROTTER AND WOOD, 39 METHODS IN MOLECULAR BIOLOGY (1995); Mann and King, J. GEN. VIROL. (1989) 70:3501. Alternatively, liposomes may be used to transfect the insect cells with the recombinant expression vector and the baculovirus. *See, e.g.*, Liebman et al., BIOTECHNIQUES (1999) 26(1):36; Graves et al., BIOCHEMISTRY (1998) 37:6050; Nomura et al., J. BIOL. CHEM. (1998) 273(22):13570; Schmidt et al., PROTEIN EXPRESSION AND PURIFICATION (1998) 12:323; Siffert et al., NATURE GENETICS (1998) 18:45; TILKINS ET AL., CELL BIOLOGY: A LABORATORY HANDBOOK 145-154 (1998); Cai et al., PROTEIN EXPRESSION AND PURIFICATION (1997) 10:263; Dolphin et al., NATURE GENETICS (1997) 17:491; Kost et al., GENE (1997) 190:139; Jakobsson et al., J. BIOL. CHEM. (1996) 271:22203; Rowles et al., J. BIOL. CHEM. (1996) 271(37):22376; Reversey et al., J. BIOL. CHEM. (1996) 271(39):23607-10; Stanley et al., J. BIOL. CHEM. (1995) 270:4121; Sisk et al., J. VIROL. (1994) 68(2):766; and Peng et al., BIOTECHNIQUES (1993) 14.2:274. Commercially available liposomes include, for example, Cellfectin® and Lipofectin® (Invitrogen, Corp., Carlsbad, CA). In addition, calcium phosphate transfection may be used. See TROTTER AND WOOD, 39 METHODS IN MOLECULAR BIOLOGY (1995); Kitts, NAR (1990) 18(19):5667; and Mann and King, J. GEN. VIROL. (1989) 70:3501.

[0296] Baculovirus expression vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (3') transcription of a coding sequence (e.g., structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region typically includes an RNA polymerase binding site and a transcription initiation site. A baculovirus promoter may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Moreover, expression may be either regulated or constitutive.

[0297] Structural genes, abundantly transcribed at late times in the infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein (FRIESEN ET AL., The Regulation of Baculovirus Gene Expression in THE MOLECULAR BIOLOGY OF BACULOVIRUSES (1986); EP 0127 839 and 0155 476) and the gene encoding the p10 protein (Vlak et al., J. GEN. VIROL. (1988) 69:765.

[0298] The newly formed baculovirus expression vector is packaged into an infectious recombinant baculovirus and subsequently grown plaques may be purified by techniques such as those described in Miller et al., BIOESSAYS (1989) 4:91; SUMMERS AND SMITH, TEXAS AGRICULTURAL EXPERIMENT STATION BULLETIN No. 1555 (1987).

[0299] Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia, Aedes aegypti* (ATCC No. CCL-125), *Bombyx mori* (ATCC No. CRL-8910), *Drosophila melanogaster* (ATCC No. 1963), *Spodoptera frugiperda,* and *Trichoplusia ni. See* WO 89/046,699; Wright, NATURE (1986) 321:718; Carbonell et al., J. VIROL. (1985) 56:153; Smith et al., MOL. CELL. BIOL. (1983) 3:2156. *See generally,* Fraser et al., IN VITRO CELL. DEV. BIOL. (1989) 25:225. More specifically, the cell lines used for baculovirus expression vector systems commonly include Sf9 *(Spodoptera frugiperda)* (ATCC No. CRL-1711), Sf21 (*Spodoptera frugiperda*) (Invitrogen Corp., Cat. No. 11497-013 (Carlsbad, CA)), Tri-368 *(Trichopulsia ni),* and High-Five™ BTI-TN-5B1-4 (*Trichopulsia ni*).

[0300] Cells and culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression.

[0301] Bacteria. Bacterial expression techniques are well known in the art. A wide variety of vectors are available for use in bacterial hosts. The vectors may be single copy or low or high multicopy vectors. Vectors may serve for cloning and/or expression. In view of the ample literature concerning vectors, commercial availability of many vectors, and even manuals describing vectors and their restriction maps and characteristics, no extensive discussion is required here. As is well-known, the vectors normally involve markers allowing for selection, which markers may provide for cytotoxic agent resistance, prototrophy or immunity. Frequently, a plurality of markers are present, which provide for different characteristics.

[0302] A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3") transcription of a coding sequence (e.g. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region typically includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in Escherichia coli (E. coli) [Raibaud et al., ANNU. REV. GENET. (1984) 18:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

[0303] Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (lac) [Chang et al., NATURE (1977) 198:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (trp) (Goeddel et al., Nuc. ACIDS RES. (1980) 8:4057; Yelverton et al., NUCL. ACIDS RES. (1981) 9:731 ; U.S. Pat. No. 4,738,921; IFNPub. Nos. 036 776 and 121775). The β-galactosidase (bla) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In Interferon 3 (Ed. I. Gresser)], bacteriophage lambda PL [Shimatake et al., NATURE (1981) 292:128] and T5 [U.S. Pat. No. 4,689,406], promoter systems also provide useful promoter sequences. Preferred methods encompassed herein utilize strong promoters, such as the T7 promoter to induce polypeptide production at high levels. Examples of such vectors include the pET29 series from Novagen, and the pPOP vectors described in WO99/05297. Such expression systems produce high levels of polypeptide in the host without compromising host cell viability or growth parameters.

[0304] In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [U.S. Pat. No. 4,551,433]. For example, the tac promoter is a hybrid trp-lac promoter comprised of both trp promoter and lac operon sequences that is regulated by the lac repressor [Amann et al., GENE (1983) 25:167; de Boer et al., PROC. NATL. ACAD. SCI. (1983) 80:21]. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophase T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier et al., J. MOL. BIOL. (1986) 189:113; Tabor et al., Proc Natl. Acad. Sci. (1985) 82:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an E. coli operator region (IFNPub. No. 267 851).

[0305] In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In E. coli, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine et al., NATURE (1975) 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' and of E. coli 16S rRNA [Steitz et al. "Genetic signals and nucleotide sequences in messenger RNA", In Biological Regulation and Development: Gene Expression (Ed. R. F. Goldberger, 1979)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook et al. "Expression of cloned genes in Escherichia coli", Molecular Cloning: A Laboratory Manual, 1989].

[0306] The term "bacterial host" or "bacterial host cell" refers to a bacterial that can be, or has been, used as a recipient for recombinant vectors or other transfer DNA. The term includes the progeny of the original bacterial host cell that has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement to the original parent, due to accidental or deliberate mutation. Progeny of the parental cell that are sufficiently similar to the parent to be characterized by the relevant property, such as the presence of a nucleotide sequence encoding a desired polypeptide, are included in the progeny intended by this definition.

[0307] The selection of suitable host bacteria for expression of a desired polypeptide is well known to those of ordinary skill in the art. In selecting bacterial hosts for expression, suitable hosts may include those shown to have at least one of the following characteristics, and preferably at least two of the following characteristics, *inter alia,* good inclusion body formation capacity, low proteolytic activity, good secretion capacity, good soluble protein production capability, and overall robustness. Bacterial hosts are generally available from a variety of sources including the Bacterial Genetic Stock Center, Department of Biophysics and

Medical Physics, University of California (Berkeley, CA); and the American Type Culture Collection ("ATCC") (Manassas, VA). Industrial/pharmaceutical fermentation generally use bacterial derived from K strains (e.g. W3110) or from bacteria derived from B strains (e.g. BL21). These strains are particularly useful because their growth parameters are extremely well known and robust. In addition, these strains are non-pathogenic, which is commercially important for safety and environmental reasons. In one embodiment of the methods described and encompassed herein, the *E. coli* host includes strains of BL21, DH10B, or derivatives thereof. In another embodiment of the methods described and encompassed herein, the *E. coli* host is a protease minus strain including OMP- and LON-. In another embodiment, the bacterial host is a species of *Pseudomonas,* such a *P. fluorescens, P. aeruginosa,* and *P. putida.* An example of a *Pseudomonas* expression strain is *P. fluorescens* biovar I, strain MB101 (Dow Chemical).

[0308] Once a recombinant host cell strain has been established (i.e., the expression construct has been introduced into the host cell and host cells with the proper expression construct are isolated), the recombinant host cell strain is cultured under conditions appropriate for production of polypeptides. The method of culture of the recombinant host cell strain will be dependent on the nature of the expression construct utilized and the identity of the host cell. Recombinant host strains are normally cultured using methods that are well known to the art. Recombinant host cells are typically cultured in liquid medium containing assimilatable sources of carbon, nitrogen, and inorganic salts and, optionally, containing vitamins, amino acids, growth factors, and other proteinaceous culture supplements well known to the art. Liquid media for culture of host cells may optionally contain antibiotics or anti-fungals to prevent the growth of undesirable microorganisms and/or compounds including antibiotics to select for host cells containing the expression vector.

[0309] Recombinant host cells may be cultured in batch or continuous formats, with either cell harvesting (in the case where the desired polypeptide accumulates intracellularly) or harvesting of culture supernatant in either batch or continuous formats. For production in prokaryotic host cells, batch culture and cell harvest are preferred.

[0310] In one embodiment, the non-natural amino acid polypeptides described herein are purified after expression in recombinant systems. The polypeptides may be purified from host cells or culture medium by a variety of methods known to the art. Normally, many polypeptides produced in bacterial host cells are poorly soluble or insoluble (in the form of inclusion bodies). In one embodiment, amino acid substitutions may readily be made in the polypeptides that are selected for the purpose of increasing the solubility of the recombinantly produced polypeptide utilizing the methods disclosed herein, as well as those known in the art. In the case of insoluble polypeptides, the polypeptides may be collected from host cell lysates by centrifugation or filtering and may further be followed by homogenization of the cells. In the case of poorly soluble polypeptides, compounds including polyethylene imine (PEI) may be added to induce the precipitation of partially soluble polypeptides. The precipitated polypeptides may then be conveniently collected by centrifugation or filtering. Recombinant host cells may be disrupted or homogenized to release the inclusion bodies from within the cells using a variety of methods well known to those of ordinary skill in the art. Host cell disruption or homogenization may be performed using well known techniques including, enzymatic cell disruption, sonication, dounce homogenization, or high pressure release disruption. In one embodiment of the methods described and encompassed herein, the high pressure release technique is used to disrupt the *E. coli* host cells to release the inclusion bodies of the polypeptides. It has been found that yields of insoluble polypeptides in the form of inclusion bodies may be increased by utilizing only one passage of the *E. coli* host cells through the homogenizer. When handling inclusion bodies of polypeptides, it is advantageous to minimize the homogenization time on repetitions in order to maximize the yield of inclusion bodies without loss due to factors such as solubilization, mechanical shearing or proteolysis.

[0311] Insoluble or precipitated polypeptides may then be solubilized using any of a number of suitable solubilization agents known to the art. By way of example, the polypeptides are solubilized with urea or guanidine hydrochloride. The volume of the solubilized polypeptides should be minimized so that large batches may be produced using conveniently manageable batch sizes. This factor may be significant in a large-scale commercial setting where the recombinant host may be grown in batches that are thousands of liters in volume. In addition, when manufacturing polypeptides in a large-scale commercial setting, in particular for human pharmaceutical uses, the avoidance of harsh chemicals that can damage the machinery and container, or the polypeptide product itself, should be avoided, if possible. It has been shown in the methods described and encompassed herein that the milder denaturing agent urea can be used to solubilize the polypeptide inclusion bodies in place of the harsher denaturing agent guanidine hydrochloride. The use of urea significantly reduces the risk of damage to stainless steel equipment utilized in the manufacturing and purification process of a polypeptide while efficiently solubilizing the polypeptide inclusion bodies.

[0312] In the case of soluble polypeptides, the peptides may be secreted into the periplasmic space or into the culture medium. In addition, soluble peptides may be present in the cytoplasm of the host cells. The soluble peptide may be concentrated prior to performing purification steps. Standard techniques, including those described herein, may be used to concentrate soluble peptide from, by way of example, cell lysates or culture medium. In addition, standard techniques, including those described herein, may be used to disrupt host cells and release soluble peptide from the cytoplasm or periplasmic space of the host cells.

[0313] When the polypeptide is produced as a fusion protein, the fusion sequence is preferably removed. Removal of a fusion sequence may be accomplished by methods including enzymatic or chemical cleavage, wherein enzymatic cleavage is preferred. Enzymatic removal of fusion sequences may be accomplished using methods well known to those in the art. The choice of enzyme for removal of the fusion sequence will be determined by the identity of the fusion, and the reaction conditions will be specified by the choice of enzyme. Chemical cleavage may be accomplished using reagents, including cyanogen bromide, TEV protease, and other reagents. The cleaved polypeptide is optionally purified from the cleaved fusion sequence by well known methods. Such methods will be determined by the identity and properties of the fusion sequence and the polypeptide. Methods for purification may include size-exclusion chromatography, hydrophobic interaction chromatography, ion-exchange chromatography or dialysis or any combination thereof.

[0314] The polypeptide is also optionally purified to remove DNA from the protein solution. DNA may be removed by any suitable method known to the art, including precipitation or ion exchange chromatography. In one embodiment, DNA is removed by precipitation with a nucleic acid precipitating agent, such as protamine sulfate. The polypeptide may be separated from the precipitated DNA using standard well known methods including centrifugation or filtration. Removal of host nucleic acid molecules is an important factor in a setting where the polypeptide is to be used to treat humans and the methods described herein reduce host cell DNA to pharmaceutically acceptable levels.

[0315] Methods for small-scale or large-scale fermentation may also be used in protein expression, including fermentors, shake flasks, fluidized bed bioreactors, hollow fiber bioreactors, roller bottle culture systems, and stirred tank bioreactor systems. Each of these methods can be performed in a batch, fed-batch, or continuous mode process.

[0316] Human forms of the non-natural amino acid polypeptides described herein can generally be recovered using methods standard in the art. For example, culture medium or cell lysate can be centrifuged or filtered to remove cellular debris. The supernatant may be concentrated or diluted to a desired volume or diafiltered into a suitable buffer to condition the preparation for further purification. Further purification of the non-natural amino acid polypeptides described herein include separating deamidated and clipped forms of a polypeptide variant from the corresponding intact form.

[0317] Any of the following exemplary procedures can be employed for purification of a non-natural amino acid polypeptide described herein: affinity chromatography; anion- or cation-exchange chromatography (using, including DEAE SEPHAROSE); chromatography on silica; reverse phase HPLC; gel filtration (using, including SEPHADEX G-75); hydrophobic interaction chromatography; size-exclusion chromatography, metal-chelate chromatography; ultrafiltration/diafiltration; ethanol precipitation; ammonium sulfate precipitation; chromatofocusing; displacement chromatography; electrophoretic procedures (including preparative isoelectric focusing), differential solubility (including ammonium sulfate precipitation), SDS-PAGE, extraction, or any combination thereof.

[0318] Polypeptides encompassed within the methods and compositions described herein, including polypeptides comprising non-natural amino acids, antibodies to polypeptides comprising non-natural amino acids, binding partners for polypeptides comprising non-natural amino acids, may 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 described herein may be recovered and purified by any of a number of methods well known in the art, including 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 any combination thereof. 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 non-natural amino acids (or polypeptides comprising non-natural amino acids) are used as purification reagents, including for affinity-based purification of polypeptides comprising one or more non-natural amino acid(s). Once purified, partially or to homogeneity, as desired, the polypeptides are optionally used for a wide variety of utilities, including as assay components, therapeutics, prophylaxis, diagnostics, research reagents, and/or as immunogens for antibody production.

[0319] In addition to other references noted herein, a variety of purification/protein folding methods are well known in the art, including 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, Engl and; Harris and Angal Protein Purification Methods: A Practical Approach IRL Press at Oxford, Oxford, Engl and; 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.

[0320] One advantage of producing polypeptides comprising at least one non-natural amino acid in a eukaryotic host cell or non-eukaryotic host cell is that typically the polypeptides will be folded in their native conformations. However, in certain aspects of the methods and compositions described herein, after synthesis, expression and/or purification, the polypeptides may possess a conformation different from the desired conformations of the relevant polypeptides. In one aspect of the methods and compositions described herein, the expressed protein is optionally denatured and then renatured. This optional denaturation and renaturation is accomplished utilizing methods known in the art, including by adding a chaperonin to the polypeptide of interest, and by solubilizing the polypeptides in a chaotropic agent including guanidine HCl, and utilizing protein disulfide isomerase.

[0321] 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. By way of example, such re-folding may be accomplished with the addition guanidine, urea, DTT, DTE, and/or a chaperonin 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, including 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.

[0322] In the case of prokaryotic production of a non-natural amino acid polypeptide, the polypeptide thus produced may be misfolded and thus lacks or has reduced biological activity. The bioactivity of the protein may be restored by "refolding". In one embodiment, a misfolded polypeptide is refolded by solubilizing (where the polypeptide is also insoluble), unfolding and reducing the polypeptide chain using, by way of example, one or more chaotropic agents (including urea and/or guanidine) and a reducing agent capable of reducing disulfide bonds (including dithiothreitol, DTT or 2-mercaptoethanol, 2-ME). At a moderate concentration of chaotrope, an oxidizing agent is then added (including oxygen, cystine or cystamine), which allows the reformation of disulfide bonds. An unfolded or misfolded polypeptide may be refolded using standard methods known in the art, such as those described in U.S. Pat. Nos. 4,511,502, 4,511,503, and 4,512,922. The polypeptide may also be cofolded with other proteins to form heterodimers or heteromultimers. After refolding or cofolding, the polypeptide is optionally further purified.

[0323] Purification of non-natural amino acid polypeptides may be accomplished using a variety of techniques, including those described herein, by way of example hydrophobic interaction chromatography, size exclusion chromatography, ion exchange chromatography, reverse-phase high performance liquid chromatography, affinity chromatography, and the like or any combination thereof. Additional purification may also include a step of drying or precipitation of the purified protein.

[0324] After purification, the non-natural amino acid polypeptides may be exchanged into different buffers and/or concentrated by any of a variety of methods known to the art, including diafiltration and dialysis. hGH that is provided as a single purified protein may be subject to aggregation and precipitation. In certain embodiments the purified non-natural amino acid polypeptides may be at least 90% pure (as measured by reverse phase high performance liquid chromatography, RP-HPLC, or sodium dodecyl sulfate-polyacrylamide gel electrophoresis, SDS-PAGE). In certain other embodiments the purified non-natural amino acid polypeptides may be at least 95% pure, or at least 98% pure, or at least 99% or greater purity. Regardless of the exact numerical value of the purity of the non-natural amino acid polypeptides, the non-natural amino acid polypeptides is sufficiently pure for use as a pharmaceutical product or for further processing, including conjugation with a water soluble polymer such as PEG.

[0325] In certain embodiments the non-natural amino acid polypeptides molecules may be used as therapeutic agents in the absence of other active ingredients or proteins (other than excipients, carriers, and stabilizers, serum albumin and the like), and in certain embodiments the non-natural amino acid polypeptides molecules they may be complexed with another polypeptide or a polymer.

2. Purification of Non-Natural Amino Acid Polypeptides

[0326] General Purification Methods The techniques disclosed in this section can be applied to the general purification of the non-natural amino acid polypeptides described herein.

[0327] Any one of a variety of isolation steps may be performed on the cell lysate extract, culture medium, inclusion bodies, periplasmic space of the host cells, cytoplasm of the host cells, or other material comprising the desired polypeptide or on any polypeptide mixtures resulting from any isolation steps including affinity chromatography, ion exchange chromatography, hydrophobic interaction chromatography, gel filtration chromatography, high performance liquid chromatography ("HPLC"), reversed phase-HPLC ("RP-HPLC"), expanded bed adsorption, or any combination and/or repetition thereof and in any appropriate order.

[0328] Equipment and other necessary materials used in performing the techniques described herein are commercially available. Pumps, fraction collectors, monitors, recorders, and entire systems are available from, for example, Applied Biosystems (Foster City, CA), Bio-Rad Laboratories, Inc. (Hercules, CA), and Amersham Biosciences, Inc. (Piscataway, NJ). Chromatographic materials including exchange matrix materials, media, and buffers are also available from such companies.

[0329] Equilibration, and other steps in the column chromatography processes described herein such as washing and elution, may be more rapidly accomplished using specialized equipment such as a pump. Commercially available pumps include HILOAD® Pump P-50, Peristaltic Pump P-1, Pump P-901, and Pump P-903 (Amersham Biosciences, Piscataway, NJ).

[0330] Examples of fraction collectors include RediFrac Fraction Collector, FRAC-100 and FRAC-200 Fraction Collectors, and SUPERFRAC® Fraction Collector (Amersham Biosciences, Piscataway, NJ). Mixers are also available to form pH and linear concentration gradients. Commercially available mixers include Gradient Mixer GM-1 and In-Line Mixers (Amersham Biosciences, Piscataway, NJ).

[0331] The chromatographic process may be monitored using any commercially available monitor. Such monitors may be used to gather information like UV, fluorescence, pH, and conductivity. Examples of detectors include Monitor UV-1, UVICORD® S II, Monitor UV-M II, Monitor UV-900, Monitor UPC-900, Monitor pH/C-900, and Conductivity Monitor (Amersham Biosciences, Piscataway, NJ). Indeed, entire systems are commercially available including the various AKTA® systems from Amersham Biosciences (Piscataway, NJ).

[0332] In one embodiment of the methods and compositions described herein, for example, the polypeptide may be reduced and denatured by first denaturing the resultant purified polypeptide in urea, followed by dilution into TRIS buffer containing a reducing agent (such as DTT) at a suitable pH. In another embodiment, the polypeptide is denatured in urea in a concentration range of between about 2 M to about 9 M, followed by dilution in TRIS buffer at a pH in the range of about 5.0 to about 8.0. The refolding mixture of this embodiment may then be incubated. In one embodiment, the refolding mixture is incubated at room temperature for four to twenty-four hours. The reduced and denatured polypeptide mixture may then be further isolated or purified.

[0333] As stated herein, the pH of the first polypeptide mixture may be adjusted prior to performing any subsequent isolation steps. In addition, the first polypeptide mixture or any subsequent mixture thereof may be concentrated using techniques known in the art. Moreover, the elution buffer comprising the first polypeptide mixture or any subsequent mixture thereof may be exchanged for a buffer suitable for the next isolation step using techniques well known to those of ordinary skill in the art.

[0334] Ion Exchange Chromatography The techniques disclosed in this section can be applied to the ion-chromatography of the non-natural amino acid polypeptides described herein.

[0335] In one embodiment, and as an optional, additional step, ion exchange chromatography may be performed on the first polypeptide mixture. *See generally* ION EXCHANGE CHROMATOGRAPHY: PRINCIPLES AND METHODS (Cat. No. 18-1114-21, Amersham Biosciences (Piscataway, NJ)). Commercially available ion exchange columns include HITRAP®, HIPREP®, and HILOAD® Columns (Amersham Biosciences, Piscataway, NJ). Such columns utilize strong anion exchangers such as Q SEPHAROSE® Fast Flow, Q SEPHAROSE® High Performance, and Q SEPHAROSE® XL; strong cation exchangers such as SP SEPHAROSE® High Performance, SP SEPHAROSE® Fast Flow, and SP SEPHAROSE® XL; weak anion exchangers such as DEAE SEPHAROSE® Fast Flow; and weak cation exchangers such as CM SEPHAROSE® Fast Flow (Amersham Biosciences, Piscataway, NJ). Anion or cation exchange column chromatography may be performed on the polypeptide at any stage of the purification process to isolate substantially purified polypeptide. The cation exchange chromatography step may be performed using any suitable cation exchange matrix. Cation exchange matrices include fibrous, porous, non-porous, microgranular, beaded, or cross-linked cation exchange matrix materials. Such cation exchange matrix materials include cellulose, agarose, dextran, polyacrylate, polyvinyl, polystyrene, silica, polyether, or composites of any of the foregoing. Following adsorption of the polypeptide to the cation exchanger matrix, substantially purified polypeptide may be eluted by contacting the matrix with a buffer having a sufficiently high pH or ionic strength to displace the polypeptide from the matrix. Suitable buffers for use in high pH elution of substantially purified polypeptide include citrate, phosphate, formate, acetate, HEPES, and MES buffers ranging in concentration from at least about 5 mM to at least about 100 mM.

[0336] Reverse-Phase Chromatography The techniques disclosed in this section can be applied to the reverse-phase chromatography of the non-natural amino acid polypeptides described herein.

[0337] RP-HPLC may be performed to purify proteins following suitable protocols that are known to those of ordinary skill in the art. *See, e.g.,* Pearson et al., ANAL BIOCHEM. (1982) 124:217-230 (1982); Rivier et al., J. CHROM. (1983) 268:112-119; Kunitani et al., J. CHROM. (1986) 359:391-402. RP-HPLC may be performed on the polypeptide to isolate substantially purified polypeptide. In this regard, silica derivatized resins with alkyl functionalities with a wide variety of lengths, including at least about C_3 to at least about C_{30}, at least about C_3 to at least about C_{20}, or at least about C_3 to at least about C_{18}, resins may be used. Alternatively, a polymeric resin may be used. For example, TosoHaas Amberchrome CG1000sd resin may be used, which is a styrene polymer resin. Cyano or polymeric resins with a wide variety of alkyl chain lengths may also be used. Furthermore, the RP-HPLC column may be washed with a solvent such as ethanol. A suitable elution buffer containing an ion pairing agent and an organic modifier such as methanol, isopropanol, tetrahydrofuran, acetonitrile or ethanol, may be used to elute the polypeptide from the RP-HPLC column. The most commonly used ion pairing agents include acetic acid, formic acid, perchloric acid, phosphoric acid, trifluoroacetic acid, heptafluorobutyric acid, triethylamine; tetramethylammonium, tetrabutylammonium, triethylammonium acetate. Elution may be performed using one or more gradients or isocratic conditions, with gradient conditions preferred to reduce the separation time and to decrease peak width. Another method involves the use of two gradients with different solvent concentration ranges. Examples of suitable elution buffers for use herein may include ammonium acetate and acetonitrile solutions.

[0338] Hydrophobic Interaction Chromatography Purification Techniques The techniques disclosed in this section can be applied to the hydrophobic interaction chromatography purification of the non-natural amino acid polypeptides described herein.

[0339] Hydrophobic interaction chromatography (HIC) may be performed on the polypeptide. *See generally* HYDROPHOBIC INTERACTION CHROMATOGRAPHY HANDBOOK: PRINCIPLES AND METHODS (Cat. No. 18-1020-90, Amersham Biosciences (Piscataway, NJ). Suitable HIC matrices may include alkyl- or aryl-substituted matrices, such as butyl-, hexyl-, octyl- or phenyl-substituted matrices including agarose, cross-linked agarose, sepharose, cellulose, silica, dextran, polystyrene, poly(methacrylate) matrices, and mixed mode resins, including a polyethyleneamine resin or a butyl- or phenyl-substituted poly(methacrylate) matrix. Commercially available sources for hydrophobic interaction column chromatography include HITRAP®, HIPREP®, and HILOAD® columns (Amersham Biosciences, Piscataway, NJ). Briefly, prior to loading, the HIC column may be equilibrated using standard buffers known to those of ordinary skill in the art, such as an acetic acid/sodium chloride solution or HEPES containing ammonium sulfate. Ammonium sulfate may be used as the buffer for loading the HIC column. After loading the polypeptide, the column may then washed using standard buffers and conditions to remove unwanted materials but retaining the polypeptide on the HIC column. The polypeptide may be eluted with about 3 to about 10 column volumes of a standard buffer, such as a HEPES buffer containing EDTA and lower ammonium sulfate concentration than the equilibrating buffer, or an acetic acid/sodium chloride buffer, among others. A decreasing linear salt gradient using, for example, a gradient of potassium phosphate, may also be used to elute the polypeptides molecules. The eluent may then be concentrated, for example, by filtration such as diafiltration or ultrafiltration. Diafiltration may be utilized to remove the salt used to elute polypeptide.

[0340] Other Purification Techniques The techniques disclosed in this section can be applied to other purification techniques of the non-natural amino acid polypeptides described herein.

[0341] Yet another isolation step using, for example, gel filtration (GEL FILTRATION: PRINCIPLES AND METHODS (Cat. No. 18-1022-18, Amersham Biosciences, Piscataway, NJ). hydroxyapatite chromatography (suitable matrices include HA-Ultrogel, High Resolution (Calbiochem), CHT Ceramic Hydroxyapatite (BioRad), Bio-Gel HTP Hydroxyapatite (BioRad)), HPLC, expanded bed adsorption, ultrafiltration, diafiltration, lyophilization, and the like, may be performed on the first polypeptide mixture or any subsequent mixture thereof, to remove any excess salts and to replace the buffer with a suitable buffer for the next isolation step or even formulation of the final drug product. The yield of polypeptide, including substantially purified polypeptide, may be monitored at each step described herein using various techniques, including those described herein. Such techniques may also used to assess the yield of substantially purified polypeptide following the last isolation step. By way of example, the yield of polypeptide may be monitored using any of several reverse phase high pressure liquid chromatography columns, having a variety of alkyl chain lengths such as cyano RP-HPLC, C_{18}RP-HPLC; as well as cation exchange HPLC and gel filtration HPLC.

[0342] Purity may be determined using standard techniques, such as SDS-PAGE, or by measuring polypeptide using Western blot and ELISA assays. For example, polyclonal antibodies may be generated against proteins isolated from negative control yeast fermentation and the cation exchange recovery. The antibodies may also be used to probe for the presence of contaminating host cell proteins.

[0343] In certain embodiments, the yield of polypeptide after each purification step may be at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, at least about 99.9%, or at least about 99.99%, of the polypeptide in the starting material for each purification step.

[0344] RP-HPLC material Vydac C4 (Vydac) consists of silica gel particles, the surfaces of which carry C4-alkyl chains. The separation of polypeptide from the proteinaceous impurities is based on differences in the strength of hydrophobic interactions. Elution is performed with an acetonitrile gradient in diluted trifluoroacetic acid. Preparative HPLC is performed using a stainless steel column (filled with 2.8 to 3.2 liter of Vydac C4 silica gel). The Hydroxyapatite Ultrogel eluate is acidified by adding trifluoro-acetic acid and loaded onto the Vydac C4 column. For washing and elution an acetonitrile gradient in diluted trifluoroacetic acid is used. Fractions are collected and immediately neutralized with phosphate buffer. The polypeptide fractions which are within the IPC limits are pooled.

[0345] DEAE Sepharose (Pharmacia) material consists of diethylaminoethyl (DEAE)-groups which are covalently bound to the surface of Sepharose beads. The binding of polypeptide to the DEAE groups is mediated by ionic interactions. Acetonitrile and trifluoroacetic acid pass through the column without being retained. After these substances have been washed off, trace impurities are removed by washing the column with acetate buffer at a low pH. Then the column is washed with neutral phosphate buffer and polypeptide is eluted with a buffer with increased ionic strength. The column is packed with DEAE Sepharose fast flow. The column volume is adjusted to assure a polypeptide load in the range of 3-10 mg polypeptide/ml gel. The column is washed with water and equilibration buffer (sodium/potassium phosphate). The pooled fractions of the HPLC eluate are loaded and the column is washed with equilibration buffer. Then the column is washed with washing buffer (sodium acetate buffer) followed by washing with equilibration buffer. Subsequently, polypeptide is eluted from the column with elution buffer (sodium chloride, sodium/potassium phosphate) and collected in a single fraction in accordance with the master elution profile. The eluate of the DEAE Sepharose column is adjusted to the specified conductivity. The resulting drug substance is sterile filtered into Teflon bottles and stored at -70°C.

[0346] A wide variety of methods and procedures can be used to assess the yield and purity of a polypeptide one or more non-natural amino acids, including the Bradford assay, SDS-PAGE, silver stained SDS-PAGE, coomassie stained SDS-PAGE, mass spectrometry (including MALDI-TOF) and other methods for characterizing proteins known to one skilled in the art.

[0347] Additional methods include steps to remove endotoxins. Endotoxins are lipopoly-saccharides (LPSs) which are located on the outer membrane of Gram-negative host cells, such as, for example, Escherichia coli. Methods for reducing endotoxin levels include purification techniques using silica supports, glass powder or hydroxyapatite, reverse-phase, affinity, size-exclusion, anion-exchange chromatography, hydrophobic interaction chromatography, a combination of these methods, and the like. Modifications or additional methods may be required to remove contaminants such as co-migrating proteins from the polypeptide of interest. Methods for measuring endotoxin levels are known to one of ordinary skill in the art and include Limulus Amebocyte Lysate (LAL) assays.

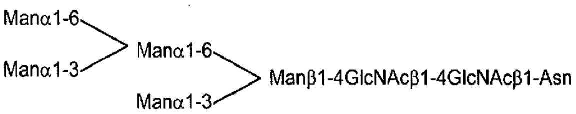
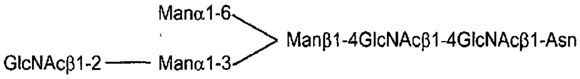
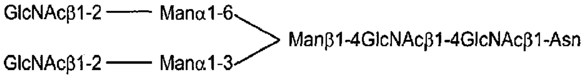
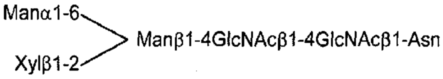
[0348] Additional methods and procedures include SDS-PAGE coupled with protein staining methods, immunoblotting, matrix assisted laser desorption/ionization-mass spectrometry (MALDI-MS), liquid chromatography/mass spectrometry, isoelectric focusing, analytical anion exchange, chromatofocusing, and circular dichroism

[0349] In certain aspects amino acids of Formulas I-XVIII, XXX-XXXIV(A&B), and XXXX-XXXXIII, including any sub-formulas or specific compounds that fall within the scope of Formulas I-XVIII, XXX-XXXIV(A&B), and XXXX-XXXXIII may be biosynthetically incorporated into polypeptides, thereby making non-natural amino acid polypeptides. In other embodiments, such amino acids are incorporated at a specific site within the polypeptide. In other aspects, such amino acids incorporated into the polypeptide using a translation system. In other aspects, such translation systems comprise: (i) a polynucleotide encoding the polypeptide, wherein the polynucleotide comprises a selector codon corresponding to the pre-designated site of incorporation of the above amino acids, and (ii) a tRNA comprising the amino acid, wherein the tRNA is specific to the selector codon. In other aspects of such translation systems, the polynucleotide is mRNA produced in the translation system. In other aspects of such translation systems, the translation system comprises a plasmid or a phage comprising the polynucleotide. In other aspects of such translation systems, the translation system comprises genomic DNA comprising the polynucleotide. In other aspects of such translation systems, the polynucleotide is stably integrated into the genomic DNA. In other aspects of such translation systems, the translation system comprises tRNA specific for a selector codon selected from the group consisting of an amber codon, ochre codon, opal codon, a unique codon, a rare codon, an unnatural codon, a five-base codon, and a four-base codon. In other aspects of such translation systems, the tRNA is a suppressor tRNA. In other aspects of such translation systems, the translation system comprises a tRNA that is aminoacylated to the amino acids above. In other aspects of such translation systems, the translation system comprises an aminoacyl synthetase specific for the tRNA. In other aspects of such translation systems, the translation system comprises an orthogonal tRNA and an orthogonal aminoacyl tRNA synthetase. In other aspects of such translation systems, the polypeptide is synthesized by a ribosome, and in further aspects the translation system is an *in vivo* translation system comprising a cell selected from the group consisting of a bacterial cell, archeaebacterial cell, and eukaryotic cell. In other aspects the cell is an *Escherichia coli* cell, yeast cell, a cell from a species of *Pseudomonas,* mammalian cell, plant cell, or an insect cell. In other aspects of such translation systems, the translation system is an *in vitro* translation system comprising cellular extract from a bacterial cell, archeaebacterial cell, or eukaryotic cell. In other aspects, the cellular extract is from an *Escherichia coli* cell, a cell from a species of *Pseudomonas,* yeast cell, mammalian cell, plant cell, or an insect cell. In other embodiments at least a portion of the polypeptide is synthesized by solid phase or solution phase peptide synthesis, or a combination thereof, while in other embodiments further comprise ligating the polypeptide to another polypeptide. In other embodiments amino acids of Formulas I-XVIII, XXX-XXXIV(A&B), and XXXX-XXXXIII, including any sub-formulas or specific compounds that fall within the scope of Formulas I-XVIII, XXX-XXXIV(A&B), and XXXX-XXXXIII may be biosynthetically incorporated into polypeptides, wherein the polypeptide is a protein homologous to a therapeutic protein selected from the group consisting of: alpha-1 antitrypsin, angiostatin, antihemolytic factor, antibody, apolipoprotein, apoprotein, atrial natriuretic factor, atrial natriuretic polypeptide, atrial peptide, C-X-C chemokine, T39765, NAP-2, ENA-78, gro-a, gro-b, gro-c, IP-10, GCP-2, NAP-4, SDF-1, PF4, MIG, calcitonin, c-kit ligand, cytokine, CC chemokine, monocyte chemoattractant protein-1, monocyte chemoattractant protein-2, monocyte chemoattractant protein-3, monocyte inflammatory protein-1 alpha, monocyte inflammatory protein-i beta, RANTES, 1309, R83915, R91733, HCC1, T58847, D31065, T64262, CD40, CD40 ligand, c-kit ligand, collagen, colony stimulating factor (CSF), complement factor 5a, complement inhibitor, complement receptor 1, cytokine, epithelial neutrophil activating peptide-78, MIP-16, MCP-1, epidermal growth factor (EGF), epithelial neutrophil activating peptide, erythropoietin (EPO), exfoliating toxin, Factor IX, Factor VII, Factor VIII, Factor X, fibroblast growth factor (FGF), fibrinogen, fibronectin, four-helical bundle protein, G-CSF, glp-1, GM-CSF, glucocerebrosidase, gonadotropin, growth factor, growth factor receptor, grf, hedgehog protein, hemoglobin, hepatocyte growth factor (hGF), hirudin, human growth hormone (hGH), human serum albumin, ICAM-1, ICAM-1 receptor, LFA-1, LFA-1 receptor, insulin, insulin-like growth factor (IGF), IGF-I, IGF-II, interferon (IFN), IFN-alpha, IFN-beta, IFN-gamma, interleukin (IL), IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, keratinocyte growth factor (KGF), lactoferrin, leukemia inhibitory factor, luciferase, neurturin, neutrophil inhibitory factor (NIF), oncostatin M, osteogenic protein, oncogene product, paracitonin, parathyroid hormone, PD-ECSF, PDGF, peptide hormone, pleiotropin, protein A, protein G, pth, pyrogenic exotoxin A, pyrogenic exotoxin B, pyrogenic exotoxin C, pyy, relaxin, renin, SCF, small biosynthetic protein, soluble complement receptor I, soluble I-CAM 1, soluble interleukin receptor, soluble TNF receptor, somatomedin, somatostatin, somatotropin, streptokinase, superantigens, staphylococcal enterotoxin, SEA, SEB, SEC1, SEC2, SEC3, SED, SEE, steroid hormone receptor, superoxide dismutase, toxic shock syndrome toxin, thymosin alpha 1, tissue plasminogen activator, tumor growth factor (TGF), tumor necrosis factor, tumor necrosis factor alpha, tumor necrosis factor beta, tumor necrosis factor receptor (TNFR), VLA-4 protein, VCAM-1 protein, vascular endothelial growth factor (VEGF), urokinase, mos, ras, raf, met, p53, tat, fos, myc, jun, myb, rel, estrogen receptor, progesterone receptor, testosterone receptor, aldosterone receptor, LDL receptor, and corticosterone.

8. In vivo Post-Translational Modifications

[0350] By producing polypeptides of interest with at least one non-natural amino acid in eukaryotic cells, such polypeptides may include eukaryotic post-translational modifications. In certain embodiments, a protein includes at least one non-natural amino acid and at least one post-translational modification that is made *in vivo* by a eukaryotic cell, where the post-translational modification is not made by a prokaryotic cell. By way of example, the post-translation modification includes 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 (including (GlcNAc-Man)$_2$-Man-GlcNAc-GlcNAc)) to an asparagine by a GlcNAc-asparagine linkage. See Table 1 which lists some examples of N-linked oligosaccharides of eukaryotic proteins (additional residues can also be present, which are not shown). In another aspect, the post-translational modification includes attachment of an oligosaccharide 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 1: 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

[0351] In yet another aspect, the post-translation modification includes proteolytic processing of precursors (including calcitonin precursor, calcitonin gene-related peptide precursor, preproparathyroid hormone, preproinsulin, proinsulin, prepro-opiomelanocortin, pro-opiomelanocortin and the like), assembly into a multisubunit protein or macromolecular assembly, translation to another site in the cell (including 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 aspects, the protein comprises a secretion or localization sequence, an epitope tag, a FLAG tag, a polyhistidine tag, a GST fusion, or the like.

[0352] One advantage of a non-natural amino acid is that it presents additional chemical moieties that can be used to add additional molecules. These modifications can be made *in vivo* in a eukaryotic or non-eukaryotic cell, or *in vitro*. Thus, in certain embodiments, the post-translational modification is through the non-natural 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, including the reaction of **d** -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 polypeptides described herein or produced using the methods described herein, other more selective reactions can be used, including the reaction of a non-natural 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 300:964-967. This allows the selective labeling of virtually any protein with a host of reagents including fluorophores, crosslinking agents, saccharide derivatives and cytotoxic molecules. *See also,* U.S.Patent Application Serial No. 10/686,944 entitled "Glycoprotein synthesis" filed January 16, 2003. Post-translational modifications, including through an azido amino acid, can also made through the Staudinger ligation (including 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(1):19-24.

IX. Alternate Systems For Producing Non-Natural Amino Acid Polypeptides

[0353] Several strategies have been employed to introduce non-natural amino acids into proteins in non-recombinant host cells, mutagenized host cells, or in cell-free systems. The alternate systems disclosed in this section can be applied to production of the non-natural amino acid polypeptides described herein. By way of example, derivatization of amino acids with reactive side-chains such as Lys, Cys and Tyr results in the conversion of lysine to N^2-acetyl-lysine. Chemical synthesis also provides a straightforward method to incorporate non-natural amino acids. With the recent development of enzymatic ligation and native chemical ligation of peptide fragments, it is possible to make larger proteins. *See, e.g.,* P. E. Dawson and S. B. H. Kent, Annu. Rev. Biochem., 69:923 (2000). Chemical peptide ligation and native chemical ligation are described in U.S. Patent No. 6,184,344, U.S. Patent Publication No. 2004/0138412, U.S. Patent Publication No. 2003/0208046, WO 02/098902, and WO 03/042235. A general *in vitro* biosynthetic method in which a suppressor tRNA chemically acylated with the desired non-natural amino acid is added to an *in vitro* extract capable of supporting protein biosynthesis, has been used to site-specifically incorporate over 100 non-natural amino acids into a variety of proteins of virtually any size. See, *e.g.,* V. W. Cornish, D. Mendel and P. G. Schultz, Angew. Chem. Int. Ed. Engl., 1995, 34:621-633 (1995); C.J. Noren, S.J. Anthony-Cahill, M.C. Griffith, P.G. Schultz, A general method for site-specific incorporation of unnatural amino acids into proteins, Science 244 182-188 (1989); and, J.D. Bain, C.G. Glabe, T.A. Dix, A.R. Chamberlin, E.S. Diala, Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide, J. Am. Chem. Soc. 111 8013-8014 (1989). A broad range of functional groups has been introduced into proteins for studies of protein stability, protein folding, enzyme mechanism, and signal transduction.

[0354] An *in vivo* method, termed selective pressure incorporation, was developed to exploit the promiscuity of wild-type synthetases. *See, e.g.,* N. Budisa, C. Minks, S. Alefelder, W. Wenger, F. M. Dong, L. Moroder and R. Huber, FASEB J., 13:41-51 (1999). An auxotrophic strain, in which the relevant metabolic pathway supplying the cell with a particular natural amino acid is switched off, is grown in minimal media containing limited concentrations of the natural amino acid, while transcription of the target gene is repressed. At the onset of a stationary growth phase, the natural amino acid is depleted and replaced with the non-natural amino acid analog. Induction of expression of the recombinant protein results in the accumulation of a protein containing the non-natural analog. For example, using this strategy, o, m and p-fluorophenylalanines have been incorporated into proteins, and exhibit two characteristic shoulders in the UV spectrum which can be easily identified, *see, e.g.,* C. Minks, R. Huber, L. Moroder and N. Budisa, Anal. Biochem., 284:29-34 (2000); trifluoromethionine has been used to replace methionine in bacteriophage T4 lysozyme to study its interaction with chitooligosaccharide ligands by ^{19}F NMR, *see, e.g.,* H. Duewel, E. Daub, V. Robinson and J. F. Honek, Biochemistry, 36:3404-3416 (1997); and trifluoroleucine has been incorporated in place of leucine, resulting in increased thermal and chemical stability of a leucine-zipper protein. *See, e.g.,* Y. Tang, G. Ghirlanda, W. A. Petka, T. Nakajima, W. F. DeGrado and D. A. Tirrell, Angew. Chem. Int. Ed. Engl., 40(8):1494-1496 (2001). Moreover, selenomethionine and telluromethionine are incorporated into various recombinant proteins to facilitate the solution of phases in X-ray crystallography. *See, e.g.,* W. A. Hendrickson, J. R. Horton and D. M. Lemaster, EMBO J., 9(5):1665-1672 (1990); J. O. Boles, K. Lewinski, M. Kunkle, J. D. Odom, B. Dunlap, L. Lebioda and M. Hatada, Nat. Struct. Biol., 1:283-284 (1994); N. Budisa, B. Steipe, P. Demange, C. Eckerskorn, J. Kellermann and R. Huber, Eur. J. Biochem., 230:788-796 (1995); and, N. Budisa, W. Karnbrock, S. Steinbacher, A. Humm, L. Prade, T. Neuefeind, L. Moroder and R. Huber, J. Mol. Biol., 270:616-623 (1997). Methionine analogs with alkene or alkyne functionalities have also been incorporated efficiently, allowing for additional modification of proteins by chemical means. *See, e.g.,* J. C. M. vanHest and D. A. Tirrell, FEBS Lett., 428:68-70 (1998); J. C. M. van Hest, K. L. Kiick and D. A. Tirrell, J. Am. Chem. Soc., 122:1282-1288 (2000); and, K. L. Kiick and D. A. Tirrell, Tetrahedron, 56:9487-9493 (2000); U.S.Patent No. 6,586,207; U.S.Patent Publication 2002/0042097.

[0355] The success of this method depends on the recognition of the non-natural amino acid analogs by aminoacyl-tRNA synthetases, which, in general, require high selectivity to insure the fidelity of protein translation. One way to expand the scope of this method is to relax the substrate specificity of aminoacyl-tRNA synthetases, which has been achieved in a limited number of cases. By way of example only, replacement of Ala^{294} by Gly in *Escherichia coli* phenylalanyl-tRNA synthetase (PheRS) increases the size of substrate binding pocket, and results in the acylation of tRNAPhe by p-Cl-phenylalanine (p-Cl-Phe). *See,* M. Ibba, P. Kast and H. Hennecke, Biochemistry, 33:7107-7112 (1994). An *Escherichia coli* strain harboring this mutant PheRS allows the incorporation of p-Cl-phenylalanine or p-Br-phenylalanine in place of phenylalanine. *See, e.g.,* M. Ibba and H. Hennecke, FEBS Lett., 364:272-275 (1995); and, N. Sharma, R. Furter, P. Kast and D. A. Tirrell, FEBS Lett., 467:37-40 (2000). Similarly, a point mutation Phe130Ser near the amino acid binding site of *Escherichia coli* tyrosyl-tRNA synthetase was shown to allow azatyrosine to be incorporated more efficiently than tyrosine. See, F. Hamano-Takaku, T. Iwama, S. Saito-Yano, K. Takaku, Y. Monden, M. Kitabatake, D. Soll and S. Nishimura, J. Biol. Chem., 275(51):40324-40328 (2000).

[0356] Another strategy to incorporate non-natural amino acids into proteins *in vivo* is to modify synthetases that have proofreading mechanisms. These synthetases cannot discriminate and therefore activate amino acids that are structurally similar to the cognate natural amino acids. This error is corrected at a separate site, which deacylates the mischarged amino acid from the tRNA to maintain the fidelity of protein translation. If the proofreading activity of the synthetase is disabled, structural analogs that are misactivated may escape the editing function and be incorporated. This approach has been demonstrated recently with the valyl-tRNA synthetase (ValRS). *See,* V. Doring, H. D. Mootz, L. A. Nangle, T. L. Hendrickson, V. de Crecy-Lagard, P. Schimmel and P. Marliere, Science, 292:501-504 (2001). ValRS can misaminoacylate tRNAVal with Cys, Thr, or aminobutyrate (Abu); these noncognate amino acids are subsequently hydrolyzed by the editing domain. After random mutagenesis of the *Escherichia coli* chromosome, a mutant *Escherichia coli* strain was selected that has a mutation in the editing site of ValRS. This edit-defective ValRS incorrectly charges tRNAVal with Cys. Because Abu sterically resembles Cys (-SH group of Cys is replaced with -CH3 in Abu), the mutant ValRS also incorporates Abu into proteins when this mutant *Escherichia coli* strain is grown in the presence of Abu. Mass spectrometric analysis shows that about 24% of valines are replaced by Abu at each valine position in the native protein.

[0357] Solid-phase synthesis and semisynthetic methods have also allowed for the synthesis of a number of proteins containing novel amino acids. For example, see the following publications and references cited within, which are as follows: Crick, F.J.C., Barrett, L. Brenner, S. Watts-Tobin, R. General nature of the genetic code for proteins. Nature, 192(4809):1227-1232 (1961); Hofmann, K., Bohn, H. Studies on polypeptides. XXXVI. The effect of pyrazole-imidazole replacements on the S-protein activating potency of an 5-peptide fragment, J. Am Chem, 88(24):5914-5919 (1966); Kaiser, E.T. Synthetic approaches to biologically active peptides and proteins including enyzmes, Acc Chem Res, 22(2):47-54 (1989); Nakatsuka, T., Sasaki, T., Kaiser, E.T. Peptide segment coupling catalyzed by the semisynthetic enzyme thiosubtilisin, J Am Chem Soc , 109, 3808-3810 (1987); Schnolzer, M., Kent, S B H. Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protedse, Science, 256, 221-225 (1992); Chaiken, I.M. Semisynthetic peptides and proteins, CRC Crit Rev Biochem, 255-301 (1981); Offord, R.E. Protein engineering by chemical means? Protein Eng., 1 (3):151-157 (1987); and, Jackson, D.Y., Bumier, J., Quan, C., Stanley, M., Tom, J., Wells, J.A. A Designed Peptide Ligase for Total Synthesis of Ribonuclease A with Unnatural Catalytic Residues, Science, 266, 243-247 (1994).

[0358] Chemical modification has been used to introduce a variety of non-natural side chains, including cofactors, spin labels and oligonucleotides into proteins *in vitro. See, e.g.,* Corey, D.R., Schultz, P.G. Generation of hybrid sequence-specific single-stranded deoxyribonuclease, Science, 238, 1401-1403 (1987) ; Kaiser, E.T., Laurence D.S., Rokita, S.E. The chemical modification of enzymatic specificity, Ann. Rev Biochem, 54, 565-595 (1985); Kaiser, E.T., Lawrence, D.S. Chemical mutation of enyzme active sites, Science, 226, 505-511 (1984); Neet, K.E., Nanci A, Koshland, D.E. Properties of thiol-subtilisin, J Biol. Chem, 243(24):6392-6401 (1968); Polgar, L.B., M.L. A new enzyme containing a synthetically formed active site. Thiol-subtilisin. J. Am Chem Soc, 88(13):3153-3154 (1966); and, Pollack, S.J., Nakayama, G. Schultz, P.G. Introduction of nucleophiles and spectroscopic probes into antibody combining sites, Science, 1(242):1038-1040 (1988).

[0359] Alternatively, biosynthetic methods that employ chemically modified aminoacyl-tRNAs have been used to incorporate several biophysical probes into proteins synthesized *in vitro.* See the following publications and references cited within: Brunner, J. New Photolabeling and crosslinking methods, Annu. Rev Biochem, 483-514 (1993); and, Krieg, U.C., Walter, P., Hohnson, A.E. Photocrosslinking of the signal sequence of nascent preprolactin of the 54-kilodalton polypeptide of the signal recognition particle, Proc. Natl. Acad. Sci, 83, 8604-8608 (1986).

[0360] Previously, it has been shown that non-natural amino acids can be site-specifically incorporated into proteins *in vitro* by the addition of chemically aminoacylated suppressor tRNAs to protein synthesis reactions programmed with a gene containing a desired amber nonsense mutation. Using these approaches, one can substitute a number of the common twenty amino acids with close structural homologues, e.g., fluorophenylalanine for phenylalanine, using strains auxotrophic for a particular amino acid. *See, e.g.,* Noren, C.J., Anthony-Cahill, Griffith, M.C., Schultz, P.G. A general method for site-specific incorporation of unnatural amino acids into proteins, Science, 244: 182-188 (1989); M.W. Nowak, et al., Science 268:439-42 (1995); Bain, J.D., Glabe, C.G., Dix, T.A., Chamberlin, A.R., Diala, E.S. Biosynthetic site-specific Incorporation of a non-natural amino acid into a polypeptide, J. Am Chem Soc, 111:8013-8014 (1989); N. Budisa et al., FASEB J. 13:41-51 (1999); Ellman, J.A., Mendel, D., Anthony-Cahill, S., Noren, C.J., Schultz, P.G. Biosynthetic method for introducing unnatural amino acids site-specifically into proteins, Methods in Enz., 202, 301-336 (1992); and, Mendel, D., Cornish, V.W. & Schultz, P.G. Site-Directed Mutagenesis with an Expanded Genetic Code, Annu Rev Biophys. Biomol Struct. 24, 435-62 (1995).

[0361] For example, a suppressor tRNA was prepared that recognized the stop codon UAG and was chemically aminoacylated with a non-natural amino acid. Conventional site-directed mutagenesis was used to introduce the stop codon TAG, at the site of interest in the protein *gene. See, e.g.,* Sayers, J.R., Schmidt, W. Eckstein, F. 5', 3' Exonuclease in phosphorothioate-based oligonucleotide-direted mutagenesis, Nucleic Acids Res, 16(3):791-802 (1988). When the acylated suppressor tRNA and the mutant gene were combined in an *in vitro* transcription/translation system, the non-natural amino acid was incorporated in response to the UAG codon which gave a protein containing that amino acid at the specified position. Experiments using [^{3}H]-Phe and experiments with α-hydroxy acids demonstrated that only the desired amino acid is incorporated at the position specified by the UAG codon and that this amino acid is not incorporated at any other site in the protein. *See, e.g.,* Noren, et al, *supra;* Kobayashi et al., (2003) Nature Structural Biology 10(6):425-432; and, Ellman, J.A., Mendel, D., Schultz, P.G. Site-specific incorporation of novel backbone structures into proteins, Science, 255, 197-200 (1992).

[0362] Microinjection techniques have also been used to incorporate non-natural amino acids into proteins. *See, e.g.,* M. W. Nowak, P. C. Kearney, J. R. Sampson, M. E. Saks, C. G. Labarca, S. K. Silverman, W. G. Zhong, J. Thorson, J. N. Abelson, N. Davidson, P. G. Schultz, D. A. Dougherty and H. A. Lester, Science, 268:439-442 (1995); and, D. A. Dougherty, Curr. Opin. Chem. Biol., 4:645 (2000). A Xenopus oocyte was coinjected with two RNA species made *in vitro:* an mRNA encoding the target protein with a UAG stop codon at the amino acid position of interest and an amber suppressor tRNA aminoacylated with the desired non-natural amino acid. The translational machinery of the oocyte then inserts the non-natural amino acid at the position specified by UAG. This method has allowed *in vivo* structure-function studies of integral membrane proteins, which are generally not amenable to *in vitro* expression systems. Examples include the incorporation of a fluorescent amino acid into tachykinin neurokinin-2 receptor to measure distances by fluorescence resonance energy transfer, see, e.g., G. Turcatti, K. Nemeth, M. D. Edgerton, U. Meseth, F. Talabot, M. Peitsch, J. Knowles, H. Vogel and A. Chollet, J. Biol. Chem., 271(33):19991-19998 (1996); the incorporation of biotinylated amino acids to identify surface-exposed residues in ion channels, see, e.g., J. P. Gallivan, H. A. Lester and D. A. Dougherty, Chem. Biol., 4(10):739-749 (1997); the use of caged tyrosine analogs to monitor conformational changes in an ion channel in real time, see, e.g., J. C. Miller, S. K. Silverman, P. M. England, D. A. Dougherty and H. A. Lester, Neuron, 20:619-624 (1998); and, the use of alpha hydroxy amino acids to change ion channel backbones for probing their gating mechanisms. *See, e.g.,* P. M. England, Y. Zhang, D. A. Dougherty and H. A. Lester, Cell, 96:89-98 (1999); and, T. Lu, A. Y. Ting, J. Mainland, L. Y. Jan, P. G. Schultz and J. Yang, Nat. Neurosci., 4(3):239-246 (2001).

[0363] The ability to incorporate non-natural amino acids directly into proteins *in vivo* offers the advantages of high yields of mutant proteins, technical ease, the potential to study the mutant proteins in cells or possibly in living organisms and the use of these mutant proteins in therapeutic treatments. The ability to include non-natural amino acids with various sizes, acidities, nucleophilicities, hydrophobicities, and other properties into proteins can greatly expand our ability to rationally and systematically manipulate the structures of proteins, both to probe protein function and create new proteins or organisms with novel properties.

[0364] In one attempt to site-specifically incorporate para-F-Phe, a yeast amber suppressor tRNAPheCUA /phenylalanyl-tRNA synthetase pair was used in a p-F-Phe resistant, Phe auxotrophic *Escherichia coli* strain. *See, e.g.,* R. Furter, Protein Sci., 7:419-426 (1998).

[0365] It may also be possible to obtain expression of a desired polynucleotide using a cell-free *(in-vitro)* translational system. In these systems, which can include either mRNA as a template *(in-vitro* translation) or DNA as a template (combined *in-vitro* transcription and translation), the *in vitro* synthesis is directed by the ribosomes. Considerable effort has been applied to the development of cell-free protein expression systems. See, e.g., Kim, D.-M. and J.R. Swartz, Biotechnology and Bioengineering, 74(4) :309-316 (2001); Kim, D.-M. and J.R. Swartz, Biotechnology Letters, 22, 1537-1542, (2000); Kim, D.-M., and J.R. Swartz, Biotechnology Progress, 16, 385-390, (2000); Kim, D.-M., and J.R. Swartz, Biotechnology and Bioengineering, 66(3): 180-188, (1999); and Patnaik, R. and J.R. Swartz, Biotechniques 24(5): 862-868, (1998) ; U.S. Patent No. 6,337,191; U.S. Patent Publication No. 2002/0081660; WO 00/55353; WO 90/05785. Another approach that may be applied to the expression of polypeptides comprising a non-natural amino acid includes the mRNA-peptide fusion technique. *See, e.g.,* R. Roberts and J. Szostak, Proc. Natl Acad. Sci. (USA) 94 12297-12302 (1997); A. Frankel, et al., Chemistry & Biology 10, 1043-1050 (2003). In this approach, an mRNA template linked to puromycin is translated into peptide on the ribosome. If one or more tRNA molecules has been modified, non-natural amino acids can be incorporated into the peptide as well. After the last mRNA codon has been read, puromycin captures the C-terminus of the peptide. If the resulting mRNA-peptide conjugate is found to have interesting properties in an *in vitro* assay, its identity can be easily revealed from the mRNA sequence. In this way, one may screen libraries of polypeptides comprising one or more non-natural amino acids to identify polypeptides having desired properties. More recently, *in vitro* ribosome translations with purified components have been reported that permit the synthesis of peptides substituted with non-natural amino *acids. See, e.g.,* A. Forster et al., Proc. Natl Acad. Sci. (USA) 100(11): 6353-6357 (2003).

X. Post-Translational Modifications of Non-Natural Amino Acid Components of a Polypeptide

[0366] Methods, compositions, techniques and strategies have been developed to site-specifically incorporate non-natural amino acids during the *in vivo* translation of proteins. By incorporating a non-natural amino acid with a sidechain chemistry that is orthogonal to those of the naturally-occurring amino acids, this technology makes possible the site-specific derivatization of recombinant proteins. As a result, a major advantage of the methods, compositions, techniques and strategies described herein is that derivatized proteins can now be prepared as defined homogeneous products. However, the methods, compositions, reaction mixtures, techniques and strategies described herein are not limited to non-natural amino acid polypeptides formed by *in vivo* protein translation techniques, but includes non-natural amino acid polypeptides formed by any technique, including by way of example only expressed protein ligation, chemical synthesis, ribozyme-based techniques (see, e.g., section herein entitled "Expression in Alternate Systems").

[0367] The ability to incorporate non-natural amino acids into recombinant proteins broadly expands the chemistries which may be implemented for post-translational derivatization, wherein such derivatization occurs either *in vivo* or *in vitro.* More specifically, protein derivatization to form an oxime linkage on a non-natural amino acid portion of a polypeptide offers several advantages. First, the naturally occurring amino acids generally do not form oxime linkages and thus reagents designed to form oxime linkages will react site-specifically with the non-natural amino acid component of the polypeptide (assuming of course that the non-natural amino acid and the corresponding reagent have been designed to form an oxime linkage), thus the ability to site-selectively derivatize proteins provides a single homogeneous product as opposed to the mixtures of derivatized proteins produced using prior art technology. Second, oxime adducts are stable under biological conditions, suggesting that proteins derivatized by oxime exchange are valid candidates for therapeutic applications. Third, the stability of the resulting oxime linkage can be manipulated based on the identity (i.e., the functional groups and/or structure) of the non-natural amino acid to which the oxime linkage has been formed. Thus, as shown in Example 16, the pH stability of the oxime linkage to a non-natural amino acid can vary from less than an hour to significantly more than a week. Thus, in some embodiments, the oxime linkage to the non-natural amino acid polypeptide has a decomposition half life less than one hour, in other embodiments less than 1 day, in other embodiments less than 2 days, in other embodiments less than 1 week and in other embodiments more than 1 week. In yet other embodiments, the resulting oxime is stable for at least two weeks under mildly acidic conditions, in other embodiments the resulting oxime is stable for at least 5 days under mildly acidic conditions. In other embodiments, the non-natural amino acid polypeptide is stable for at least 1 day in a pH between about 2 and about 8; in other embodiments, from a pH of about 2 to about 6; in other embodiment, in a pH of about 2 to about 4. In other embodiments, using the strategies, methods, compositions and techniques described herein, one of skill in the art will be able to synthesize an oxime linkage to a non-natural amino acid polypeptide with a decomposition half-life tuned to the needs of that skilled artisan (e.g., for a therapeutic use such as sustained release, or a diagnostic use, or an industrial use or a military use).

[0368] The non-natural amino acid polypeptides described above are useful for, including novel therapeutics, diagnostics, catalytic enzymes, industrial enzymes, binding proteins (including antibodies and antibody fragments), and including 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. Other uses for the non-natural amino acid polypeptides described above include, by way of example only, assay-based, cosmetic, plant biology, environmental, energy-production, and/or military uses. However, the non-natural amino acid polypeptides described above can undergo further modifications so as to incorporate new or modified functionalities, including manipulating the therapeutic effectiveness of the polypeptide, improving the safety profile of the polypeptide, adjusting the pharmacokinetics, pharmacologics and/or pharmacodynamics of the polypeptide (e.g., increasing water solubility, bioavailability, increasing serum half-life, increasing therapeutic half-life, modulating immunogenicity, modulating biological activity, or extending the circulation time), providing additional functionality to the polypeptide, incorporating a tag, label or detectable signal into the polypeptide, easing the isolation properties of the polypeptide, and any combination of the aforementioned modifications.

[0369] In certain embodiments are methods for easing the isolation properties of a polypeptide comprising utilizing a homologous biosynthetic non-natural amino acid polypeptide comprising at least one non-natural amino acid selected from the group consisting of an oxime-containing non-natural amino acid, a carbonyl-containing non-natural amino acid, and a hydroxylamine-containing non-natural amino acid. In other embodiments such non-natural amino acids have been biosynthetically incorporated into the polypeptide as described herein. In further or alternative embodiments such non-natural amino acid polypeptides comprise at least one non-natural amino acid selected from amino acids of Formula I-XVIII, XXX-XXXIV(A&B), or XXXX-XXXXIII.

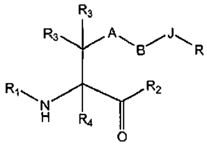
[0370] The methods, compositions, strategies and techniques described herein are not limited to a particular type, class or family of polypeptides. Indeed, virtually any polypeptide may include at least one non-natural amino acids described herein. By way of example only, the polypeptide can be homologous to a therapeutic protein selected from the group consisting of: alpha-1 antitrypsin, angiostatin, antihemolytic factor, antibody, apolipoprotein, apoprotein, atrial natriuretic factor, atrial natriuretic polypeptide, atrial peptide, C-X-C chemokine, T39765, NAP-2, ENA-78, gro-a, gro-b, gro-c, IP-10, GCP-2, NAP-4, SDF-1, PF4, MIG, calcitonin, c-kit ligand, cytokine, CC chemokine, monocyte chemoattractant protein-1, monocyte chemoattractant protein-2, monocyte chemoattractant protein-3, monocyte inflammatory protein-1 alpha, monocyte inflammatory protein-i beta, RANTES, 1309, R83915, R91733, HCC1, T58847, D31065, T64262, CD40, CD40 ligand, c-kit ligand, collagen, colony stimulating factor (CSF), complement factor 5a, complement inhibitor, complement receptor 1, cytokine, epithelial neutrophil activating peptide-78, MIP-16, MCP-1, epidermal growth factor (EGF), epithelial neutrophil activating peptide, erythropoietin (EPO), exfoliating toxin, Factor IX, Factor VII, Factor VIII, Factor X, fibroblast growth factor (FGF), fibrinogen, fibronectin, four-helical bundle protein, G-CSF, glp-1, GM-CSF, glucocerebrosidase, gonadotropin, growth factor, growth factor receptor, grf, hedgehog protein, hemoglobin, hepatocyte growth factor (hGF), hirudin, human growth hormone (hGH), human serum albumin, ICAM-1, ICAM-1 receptor, LFA-1, LFA-1 receptor, insulin, insulin-like growth factor (IGF), IGF-I, IGF-II, interferon (IFN), IFN-alpha, IFN-beta, IFN-gamma, interleukin (IL), IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, keratinocyte growth factor (KGF), lactoferrin, leukemia inhibitory factor, luciferase, neurturin, neutrophil inhibitory factor (NIF), oncostatin M, osteogenic protein, oncogene product, paracitonin, parathyroid hormone, PD-ECSF, PDGF, peptide hormone, pleiotropin, protein A, protein G, pth, pyrogenic exotoxin A, pyrogenic exotoxin B, pyrogenic exotoxin C, pyy, relaxin, renin, SCF, small biosynthetic protein, soluble complement receptor I, soluble I-CAM 1, soluble interleukin receptor, soluble TNF receptor, somatomedin, somatostatin, somatotropin, streptokinase, superantigens, staphylococcal enterotoxin, SEA, SEB, SEC1, SEC2, SEC3, SED, SEE, steroid hormone receptor, superoxide dismutase, toxic shock syndrome toxin, thymosin alpha 1, tissue plasminogen activator, tumor growth factor (TGF), tumor necrosis factor, tumor necrosis factor alpha, tumor necrosis factor beta, tumor necrosis factor receptor (TNFR), VLA-4 protein, VCAM-1 protein, vascular endothelial growth factor (VEGF), urokinase, mos, ras, raf, met, p53, tat, fos, myc, jun, myb, rel, estrogen receptor, progesterone receptor, testosterone receptor, aldosterone receptor, LDL receptor, and corticosterone. The non-natural amino acid polypeptide may also be homologous to any polypeptide member of the growth hormone supergene family.

[0371] Such modifications include the incorporation of further functionality onto the non-natural amino acid component of the polypeptide, including a label; a dye; a polymer; a water-soluble polymer; a derivative of polyethylene glycol; a photocrosslinker; a cytotoxic compound; a drug; an affinity label; a photoaffinity label; a reactive compound; a resin; a second protein or polypeptide or polypeptide analog; an antibody or antibody fragment; a metal chelator; a cofactor; a fatty acid; a carbohydrate; a polynucleotide; a DNA; a RNA; an antisense polynucleotide; a saccharide, a watersoluble dendrimer, a cyclodextrin, a biomaterial; a nanoparticle; a spin label; a fluorophore, a metal-containing moiety; a radioactive moiety; a novel functional group; a group that covalently or noncovalently interacts with other molecules; a photocaged moiety; an actinic radiation excitable moiety; a ligand; a photoisomerizable moiety; biotin; a biotin analogue; a moiety incorporating a heavy atom; a chemically cleavable group; a photocleavable group; an elongated side chain; a carbon-linked sugar; a redox-active agent; an amino thioacid; a toxic moiety; an isotopically labeled moiety; a biophysical probe; a phosphorescent group; a chemiluminescent group; an electron dense group; a magnetic group; an intercalating group; a chromophore; an energy transfer agent; a biologically active agent; a detectable label; a small molecule; an inhibitory ribonucleic acid, a radionucleotide; a neutron-capture agent; a derivative of biotin; quantum dot(s); a nanotransmitter; a radiotransmitter; an abzyme, an activated complex activator, a virus, an adjuvant, an aglycan, an allergan, an angiostatin, an antihormone, an antioxidant, an aptamer, a guide RNA, a saponin, a shuttle vector, a macromolecule, a mimotope, a receptor, a reverse micelle, and any combination thereof.

[0372] In addition, non-natural amino acid polypeptides may contain moieties which may be converted into other functional groups, such as, by way of example only, carbonyls, dicarbonyls or hydroxylamines. FIG. 63A illustrates the chemical conversion of non-natural amino acid polypeptides into carbonyl or dicarbonyl-containing non-natural amino acid polypeptides, while FIG. 63B illustrates the chemical conversion of non-natural amino acid polypeptides into hydroxylamine-containing non-natural amino acid polypeptides. The resulting carbonyl or dicarbonyl-containing non-natural amino acid polypeptides and hydroxylamine-containing non-natural amino acid polypeptides may be used in or incorporated into any of the methods, compositions, techniques and strategies for making, purifying, characterizing, and using non-natural amino acids, non-natural amino acid polypeptides and modified non-natural amino acid polypeptides described herein. The chemical conversion of chemical moieties into other functional groups, such as, by way of example only, carbonyls, di-carbonyls or hydroxylamines can be achieved using techniques and materials known to those of skill in the art, such as described, for example, in March, ADVANCED ORGANIC CHEMISTRY 5th Ed., (Wiley 2001); and Carey and Sundberg, ADVANCED ORGANIC CHEMISTRY 4th Ed., Vols. A and B (Plenum 2000, 2001).

[0373] Thus, by way of example only, a non-natural amino acid polypeptide containing any one of the following amino acids may be further modified using the methods and compositions described herein:

1. (a)



(I)

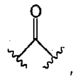
wherein:

A is optional, and when present is lower alkylene, substituted lower alkylene, lower cycloalkylene, substituted lower cycloalkylene, lower alkenylene, substituted lower alkenylene, alkynylene, lower heteroalkylene, substituted heteroalkylene, lower heterocycloalkylene, substituted lower heterocycloalkylene, arylene, substituted arylene, heteroarylene, substituted heteroarylene, alkarylene, substituted alkarylene, aralkylene, or substituted aralkylene;

B is optional, and when present is a linker selected from the group consisting of lower alkylene, substituted lower alkylene, lower alkenylene, substituted lower alkenylene, lower heteroalkylene, substituted lower heteroalkylene, -O-, -O-(alkylene or substituted alkylene)-, -S-, -S-(alkylene or substituted alkylene)-, -S(O)$_k$- where k is 1, 2, or 3, -S(O)$_k$(alkylene or substituted alkylene)-, -C(O)-, -C(O)-(alkylene or substituted alkylene)-, -C(S)-, -C(S)-(alkylene or substituted alkylene)-, -N(R')-, -NR'-(alkylene or substituted alkylene)-, -C(O)N(R')-, -CON(R')-(alkylene or substituted alkylene)-,-CSN(R')-, -CSN(R')-(alkylene or substituted alkylene)-, -N(R')CO-(alkylene or substituted alkylene)-, -N(R')C(O)O-, -S(O)$_k$N(R')-, -N(R')C(O)N(R')-, -N(R')C(S)N(R')-, -N(R')S(O)$_k$N(R')-, -N(R')-N=, -C(R')=N-, -C(R')=N-N(R')-, -C(R')=N-N=, -C(R')$_2$-N=N-, and

-C(R')$_2$-N(R')-N(R')-, where each R' is independently H, alkyl, or substituted alkyl;

J is



R is H, alkyl, substituted alkyl, cycloalkyl, or substituted cycloalkyl;

R$_1$ is H, an amino protecting group, resin; and

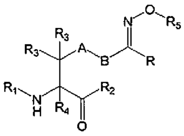
R$_2$ is OH, an ester protecting group, resin;

each of R$_3$ and R$_4$ is independently H, halogen, lower alkyl, or substituted lower alkyl;

or the -A-B-J-R groups together form a bicyclic or tricyclic cycloalkyl or heterocycloalkyl comprising at least one carbonyl group, including a dicarbonyl group, protected carbonyl group, including a protected dicarbonyl group, or masked carbonyl group, including a masked dicarbonyl group;

or the -J-R group together forms a monocyclic or bicyclic cycloalkyl or heterocycloalkyl comprising at least one carbonyl group, including a dicarbonyl group, protected carbonyl group, including a protected dicarbonyl group, or masked carbonyl group, including a masked dicarbonyl group;

2. (b)



(XI)

wherein:

A is optional, and when present is lower alkylene, substituted lower alkylene, lower cycloalkylene, substituted lower cycloalkylene, lower alkenylene, substituted lower alkenylene, alkynylene, lower heteroalkylene, substituted heteroalkylene, lower heterocycloalkylene, substituted lower heterocycloalkylene, arylene, substituted arylene, heteroarylene, substituted heteroarylene, alkarylene, substituted alkarylene, aralkylene, or substituted aralkylene;

B is optional, and when present is a linker selected from the group consisting of lower alkylene, substituted lower alkylene, lower alkenylene, substituted lower alkenylene, lower heteroalkylene, substituted lower heteroalkylene, -0-, -O-(alkylene or substituted alkylene)-, -S-, -S-(alkylene or substituted alkylene)-, -S(O)$_k$- where k is 1, 2, or 3, -S(O)$_k$(alkylene or substituted alkylene)-, -C(O)-, -C(O)-(alkylene or substituted alkylene)-, -C(S)-, -C(S)-(alkylene or substituted alkylene)-, -N(R')-, -NR'-(alkylene or substituted alkylene)-, -C(O)N(R')-, -CON(R')-(alkylene or substituted alkylene)-, - CSN(R')-, -CSN(R')-(alkylene or substituted alkylene)-, -N(R')CO-(alkylene or substituted alkylene)-, -N(R')C(O)O-, -S(O)$_k$N(R')-, -N(R')C(O)N(R')-, -N(R')C(S)N(R')-, -N(R')S(O)$_k$N(R')-, -N(R')-N=, -C(R')=N-, -C(R')=N-N(R')-, -C(R')=N-N=, -C(R')$_2$-N=N-, and -C(R')$_2$-N(R')-N(R')-, where each R' is independently H, alkyl, or substituted alkyl;

R is H, alkyl, substituted alkyl, cycloalkyl, or substituted cycloalkyl;

R$_1$ is H, an amino protecting group, resin; and

R$_2$ is OH, an ester protecting group, resin;

each of R$_3$ and R$_4$ is independently H, halogen, lower alkyl, or substituted lower alkyl;

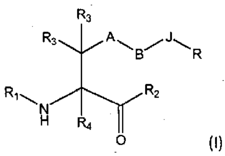
R$_5$ is L-X, where X is a derivative of polyethylene glycol.

A. Methods for Post-Translationally Modifying Non-Natural Amino Acid Polypeptides: Reactions of Carbonyl-Containing Non-Natural Amino Acids with Hydroxylamine-Containing Reagents

[0374] The sidechains of the naturally occurring amino acids lack highly electrophilic sites. Therefore, the incorporation of an unnatural amino acid with an electrophile-containing sidechain, including, by way of example only, an amino acid containing a carbonyl or dicarbonyl group such as ketones or aldehydes, makes possible the site-specific derivatization of this sidechain via nucleophilic attack of the carbonyl or dicarbonyl group. In the instance where the attacking nucleophile is a hydroxylamine, an oxime-derivatized protein will be generated. The methods for derivatizing and/or further modifying may be conducted with a polypeptide that has been purified prior to the derivatization step or after the derivatization step. In addition, the methods for derivatizing and/or further modifying may be conducted with on synthetic polymers, polysaccharides, or polynucleotides which have been purified before or after such modifications. Further, the derivatization step can occur under mildly acidic to slightly basic conditions, including by way of example, between a pH of about 2-8, or between a pH of about 4-8.

[0375] A polypeptide-derivatizing method based upon the reaction of carbonyl- or dicarbonyl-containing polypeptides with a hydroxylamine-substituted molecule has distinct advantages. First, hydroxylamines undergo condensation with carbonyl- or dicarbonyl-containing compounds in a pH range of 2-8 (and in further aspects in a pH range of 4-8) to generate oxime adducts. Under these conditions, the sidechains of the naturally occurring amino acids are unreactive. Second, such selective chemistry makes possible the site-specific derivatization of recombinant proteins: derivatized proteins can now be prepared as defined homogeneous products. Third, the mild conditions needed to effect the reaction of the hydroxylamines described herein with the carbonyl- or dicarbonyl-containing polypeptides described herein generally do not irreversibly destroy the tertiary structure of the polypeptide (excepting, of course, where the purpose of the reaction is to destroy such tertiary structure). Finally, although the hydroxylamine group amino appears to be metabolized by *E. coli,* the condensation of hydroxylamines with carbonyl- or dicarbonyl-containing molecules generates oxime adducts which are stable under biological conditions.

[0376] By way of example only, the following non-natural amino acids are the type of carbonyl- or dicarbonyl-containing amino acids that are reactive with the hydroxylamine-containing reagents described herein that can be used to further modify carbonyl- or dicarbonyl-containing non-natural amino acid polypeptides: wherein:

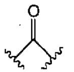


(I)

A is optional, and when present is lower alkylene, substituted lower alkylene, lower cycloalkylene, substituted lower cycloalkylene, lower alkenylene, substituted lower alkenylene, alkynylene, lower heteroalkylene, substituted heteroalkylene, lower heterocycloalkylene, substituted lower heterocycloalkylene, arylene, substituted arylene, heteroarylene, substituted heteroarylene, alkarylene, substituted alkarylene, aralkylene, or substituted aralkylene;

B is optional, and when present is a linker selected from the group consisting of lower alkylene, substituted lower alkylene, lower alkenylene, substituted lower alkenylene, lower heteroalkylene, substituted lower heteroalkylene, -0-, -0-(alkylene or substituted alkylene)-, -S-, -S-(alkylene or substituted alkylene)-, $-S(O)_k-$ where k is 1, 2, or 3, $-S(O)_k$(alkylene or substituted alkylene)-, -C(O)-, -C(O)-(alkylene or substituted alkylene)-, -C(S)-, -C(S)-(alkylene or substituted alkylene)-, -N(R')-, -NR'-(alkylene or substituted alkylene)-, -C(O)N(R')-, -CON(R')-(alkylene or substituted alkylene)-,-CSN(R')-, -CSN(R')-(alkylene or substituted alkylene)-, -N(R')CO-(alkylene or substituted alkylene)-, -N(R')C(O)O-, $-S(O)_kN(R')-$, -N(R')C(O)N(R')-, -N(R')C(S)N(R')-, $-N(R')S(O)_kN(R')-$, -N(R')-N=, -C(R')=N-, -C(R')=N-N(R')-, -C(R')=N-N=, $-C(R')_2-N=N-$, and $-C(R')_2-N(R')-N(R')-$, where each R' is independently H, alkyl, or substituted alkyl;

J is



,

R is H, alkyl, substituted alkyl, cycloalkyl, or substituted cycloalkyl;

R_1 is H, an amino protecting group, resin; and

R_2 is OH, an ester protecting group, resin;

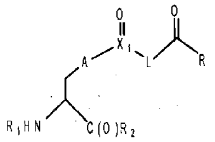
each of R_3 and R_4 is independently H, halogen, lower alkyl, or substituted lower alkyl;

or the -A-B-J-R groups together form a bicyclic or tricyclic cycloalkyl or heterocycloalkyl comprising at least one carbonyl group, including a dicarbonyl group, protected carbonyl group, including a protected dicarbonyl group, or masked carbonyl group, including a masked dicarbonyl group;

or the -J-R group together forms a monocyclic or bicyclic cycloalkyl or heterocycloalkyl comprising at least one carbonyl group, including a dicarbonyl group, protected carbonyl group, including a protected dicarbonyl group, or masked carbonyl group, including a masked dicarbonyl group.

[0377] In certain aspects, compound of Formula (I) are reactive with hydroxylamines in aqueous solution under mildly acidic conditions. In certain embodiments, such acidic conditions are pH 2 to 8.

[0378] By way of example only, for the aforementioned purposes, compounds of Formula (I) include compounds having the structure:



(XXX);

wherein;

A is optional, and when present is lower alkylene, substituted lower alkylene, lower cycloalkylene, substituted lower cycloalkylene, lower alkenylene, substituted lower alkenylene, alkynylene, lower heteroalkylene, substituted heteroalkylene, lower heterocycloalkylene, substituted lower heterocycloalkylene, arylene, substituted arylene, heteroarylene, substituted heteroarylene, alkarylene, substituted alkarylene, aralkylene, or substituted aralkylene;

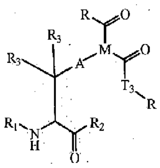
R is H, alkyl, substituted alkyl, cycloalkyl, or substituted cycloalkyl;

R_1 is H, an amino protecting group, resin, amino acid, polypeptide, or polynucleotide;

R_2 is OH, an ester protecting group, resin, amino acid, polypeptide, or polynucleotide;

X_1 is C, S, or S(O); and L is a bond, alkylene, substituted alkylene, N(R')(alkylene) or N(R')(substituted alkylene), where each R' is independently H, alkyl, or substituted alkyl.

[0379] By way of further example only, for the aforementioned purposes, compounds of Formula (I) include compounds, having the structure of Formula (XXXX):

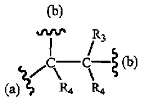


(XXXX),

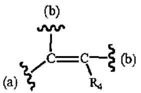
wherein:

A is optional, and when present is lower alkylene, substituted lower alkylene, lower cycloalkylene, substituted lower cycloalkylene, lower alkenylene, substituted lower alkenylene, alkynylene, lower heteroalkylene, substituted heteroalkylene, lower heterocycloalkylene, substituted lower heterocycloalkylene, arylene, substituted arylene; heteroarylene, substituted heteroarylene, alkarylene, substituted alkarylene, aralkylene, or substituted aralkylene;

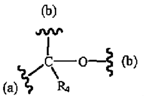
M is -C(R_3)-,



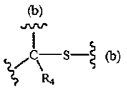
,



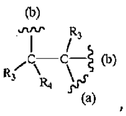
,



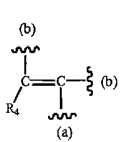
,



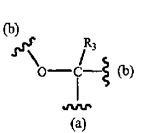
,



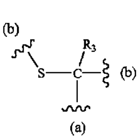
,



,



, or



,

where (a) indicates bonding to the A group and (b) indicates bonding to respective carbonyl groups, R_3 and R_4 are independently chosen from H, halogen, alkyl, substituted alkyl, cycloalkyl, or substituted cycloalkyl, or R_3 and R_4 or two R_3 groups or two R_4 groups optionally form a cycloalkyl or a heterocycloalkyl;

R is H, halogen, alkyl, substituted alkyl, cycloalkyl, or substituted cycloalkyl;

T_3 is a bond, C(R)(R), O, or S, and R is H, halogen, alkyl, substituted alkyl, cycloalkyl, or substituted cycloalkyl;

R_1 is H, an amino protecting group, resin, amino acid, polypeptide, or polynucleotide; and

R_2 is OH, an ester protecting group, resin, amino acid, polypeptide, or polynucleotide.

[0380] The types of polypeptides that comprise such carbonyl- or dicarbonyl-containing non-natural amino acids is practically unlimited as long as the carbonyl- or dicarbonyl-containing non-natural amino acid is located on the polypeptide so that the hydroxylamine reagent can react with the carbonyl or dicarbonyl group and not create a resulting modified non-natural amino acid that destroys the tertiary structure of the polypeptide (excepting, of course, if such destruction is the purpose of the reaction).

[0381] By way of example only, the following hydroxylamine-containing reagents are the type of hydroxylamine-containing reagents that are reactive with the carbonyl- or dicarbonyl-containing non-natural amino acids described herein and can be used to further modify carbonyl- or dicarbonyl-containing non-natural amino acid polypeptides:

$$\left[X - L\right]_n - L_1 - W \qquad \text{(XIX)}$$

wherein:

each X is independently H, alkyl, substituted alkyl, alkenyl, substituted alkenyl, alkynyl, substituted alkynyl, alkoxy, substituted alkoxy, alkylalkoxy, substituted alkylalkoxy, polyalkylene oxide, substituted polyalkylene oxide, aryl, substituted aryl, heteroaryl, substituted heteroaryl, alkaryl, substituted alkaryl, aralkyl, substituted aralkyl, -(alkylene or substituted alkylene)-ON(R")$_2$, -(alkylene or substituted alkylene)-C(O)SR", -(alkylene or substituted alkylene)-S-S-(aryl or substituted aryl), -C(O)R", -C(O)$_2$R", or-C(O)N(R")$_2$ wherein each R" is independently hydrogen, alkyl, substituted alkyl, alkenyl, substituted alkenyl, alkoxy, substituted alkoxy, aryl, substituted aryl, heteroaryl, alkaryl, substituted alkaryl, aralkyl, or substituted aralkyl;

or each X is independently selected from the group consisting of a label; a dye; a polymer; a water-soluble polymer; a derivative of polyethylene glycol; a photocrosslinker; a cytotoxic compound; a drug; an affinity label; a photoaffinity label; a reactive compound; a resin; a second protein or polypeptide or polypeptide analog; an antibody or antibody fragment; a metal chelator; a cofactor; a fatty acid; a carbohydrate; a polynucleotide; a DNA; a RNA; an antisense polynucleotide; a saccharide, a water-soluble dendrimer, a cyclodextrin, a biomaterial; a nanoparticle; a spin label; a fluorophore, a metal-containing moiety; a radioactive moiety; a novel functional group; a group that covalently or noncovalently interacts with other molecules; a photocaged moiety; an actinic radiation excitable moiety; a ligand; a photoisomerizable moiety; biotin; a biotin analogue; a moiety incorporating a heavy atom; a chemically cleavable group; a photocleavable group; an elongated side chain; a carbon-linked sugar; a redox-active agent; an amino thioacid; a toxic moiety; an isotopically labeled moiety; a biophysical probe; a phosphorescent group; a chemiluminescent group; an electron dense group; a magnetic group; an intercalating group; a chromophore; an energy transfer agent; a biologically active agent; a detectable label; a small molecule; an inhibitory ribonucleic acid, a radionucleotide; a neutron-capture agent; a derivative of biotin; quantum dot(s); a nanotransmitter; a radiotransmitter; an abzyme, an activated complex activator, a virus, an adjuvant, an aglycan, an allergan, an angiostatin, an antihormone, an antioxidant, an aptamer, a guide RNA, a saponin, a shuttle vector, a macromolecule, a mimotope, a receptor, a reverse micelle, and any combination thereof;

each L is independently selected from the group consisting of alkylene, substituted alkylene, alkenylene, substituted alkenylene, -O-, -O-(alkylene or substituted alkylene)-, -S-, -S-(alkylene or substituted alkylene)-, -S(O)$_k$- where k is 1, 2, or 3, -S(O)$_k$(alkylene or substituted alkylene)-, -C(O)-, -C(O)-(alkylene or substituted alkylene)-, -C(S)-, -C(S)-(alkylene or substituted alkylene)-, -N(R')-, -NR'-(alkylene or substituted alkylene)-, -C(O)N(R')-, -CON(R')-(alkylene or substituted alkylene)-, -(alkylene or substituted alkylene)NR'C(O)O-(alkylene or substituted alkylene)-, -O-CON(R')-(alkylene or substituted alkylene)-, -CSN(R')-, -CSN(R')-(alkylene or substituted alkylene)-, -N(R')CO-(alkylene or substituted alkylene)-, -N(R')C(O)O-, -N(R')C(O)O-(alkylene or substituted alkylene)-, -S(O)$_k$N(R')-, -N(R')C(O)N(R')-, -N(R')C(O)N(R')-(alkylene or substituted alkylene)-, -N(R')C(S)N(R')-, -N(R')S(O)$_k$N(R')-, -N(R')-N=, -C(R')=N-, -C(R')=N-N(R')-, -C(R')=N-N=, -C(R')$_2$-N=N-, and -C(R')$_2$-N(R')-N(R')-;

L_1 is optional, and when present, is -C(R')$_p$-NR'-C(O)O-(alkylene or substituted alkylene)- where p is 0,1, or 2;

each R' is independently H, alkyl, or substituted alkyl;

W is -N(R$_8$)$_2$, where each R_8 is independently H or an amino protecting group; and n is 1 to 3;

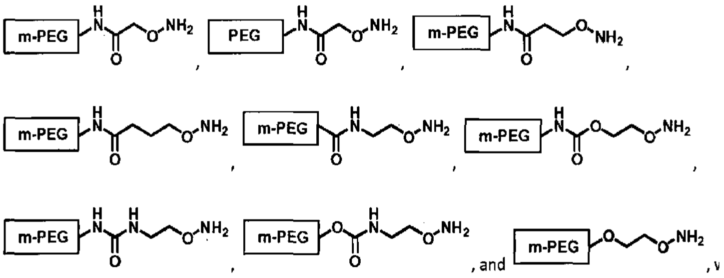
provided that L-L_1-W together provide at least one hydroxylamine group capable of reacting with a carbonyl (including a dicarbonyl) group on a non-natural amino acid or a "modified or unmodified" non-natural amino acid polypeptide.

[0382] In certain aspects of compounds of Formula (XIX), X is a polymer comprising alkyl, substituted alkyl, alkenyl, substituted alkenyl, alkynyl, substituted alkynyl, alkoxy, substituted alkoxy, alkylalkoxy, substituted alkylalkoxy, polyalkylene oxide, substituted polyalkylene oxide, aryl, substituted aryl, heteroaryl, substituted heteroaryl, alkaryl, substituted alkaryl, aralkyl, or substituted aralkyl. In certain aspects of compounds of Formula (XIX), X is a polymer comprising polyalkylene oxide or substituted polyalkylene oxide. In certain aspects of compounds of Formula (XIX), X is a polymer comprising - [(alkylene or substituted alkylene)-O-(hydrogen, alkyl, or substituted alkyl)]$_x$, wherein x is from 20-10,000. In certain aspects of compounds of Formula (XIX), X is m-PEG having a molecular weight ranging from 2 to 40 KDa. In certain aspects of compounds of Formula (XIX), X is a biologically active agent selected from the group consisting of a peptide, protein, enzyme, antibody, drug, dye, lipid, nucleoside, oligonucleotide, cell, virus, liposome, microparticle, and micelle. In certain aspects of compounds of Formula (XIX), X is a drug selected from the group consisting of an antibiotic, fungicide, anti-viral agent, anti-inflammatory agent, anti-tumor agent, cardiovascular agent, anti-anxiety agent, hormone, growth factor, and steroidal agent. In certain aspects of compounds of Formula (XIX), X is a an enzyme selected from the group consisting of horseradish peroxidase, alkaline phosphatase, β-galactosidase, and glucose oxidase. In certain aspects of compounds of Formula (XIX), X is a detectable label selected from the group consisting of a fluorescent, phosphorescent, chemiluminescent, chelating, electron dense, magnetic, intercalating, radioactive, chromophoric, and energy transfer moiety. In certain aspects of compounds of Formula (XIX), L is selected from the group consisting of -N(R')CO-(alkylene or substituted alkylene)-, -CON(R')-(alkylene or substituted alkylene)-, -N(R')C(O)N(R')-(alkylene or substituted alkylene)-, -O-CON(R')-(alkylene or substituted alkylene)-, -O-(alkylene or substituted alkylene)-, -C(O)N(R')-, and -N(R')C(O)O-(alkylene or substituted alkylene)-.

[0383] In certain embodiments of compounds of Formula (XIX), are compounds having the structure of Formula (XX):

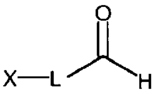
$$X—L—O—NH_2 \qquad (XX).$$

[0384] In certain embodiments of compounds of Formula (XX), are compounds selected from the group consisting of:



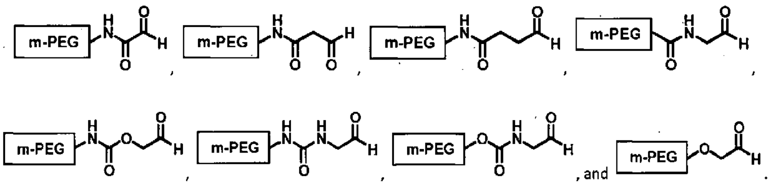
wherein other embodiments such m-PEG or PEG groups have a molecular weight ranging from 5 to 30 kDa.

[0385] In certain aspects of compounds of Formula (XIX), are compounds having the structure of Formula (XXI):

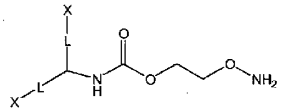


(XXI).

[0386] In certain embodiments of compounds of Formula (XXI), are compounds selected from the group consisting of:

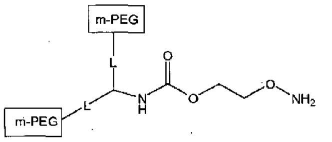


[0387] In certain embodiments of compounds of Formula (XIX), are compounds having the structure of Formula (XXII):



(XXII).

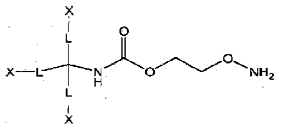
In certain aspects of compounds of Formula (XXII), L is -(alkylene or substituted alkylene)-N(R')C(0)0-(alkylene or substituted alkylene)-. In certain aspects of compounds of Formula (XXII), are compounds having the structure of Formula (XXIII):



(XXIII).

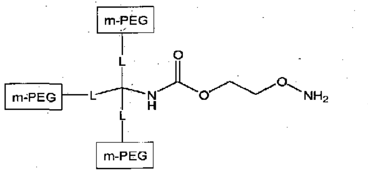
wherein other aspects of compounds of Formula (XXII) such m-PEG groups have a molecular weight ranging from 5 to 30 kDa.

[0388] In certain aspects of compounds of Formula (XIX), are compounds having the structure of Formula (XXIV):



(XXIV).

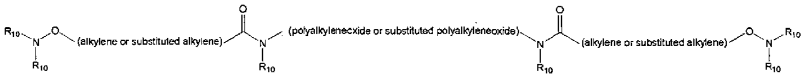
In certain aspects of compounds of Formula (XXIV), L is -(alkylene or substituted alkylene)-N(R')C(O)O-(alkylene or substituted alkylene)- or-N(R')C(O)O-(alkylene or substituted alkylene)-. In certain aspects of compounds of Formula (XXIV), are compounds having the structure of Formula (XXV):



(XXV).

wherein other aspects of compounds of Formula (XXIV) such m-PEG groups have a molecular weight ranging from 5 to 30 kDa.

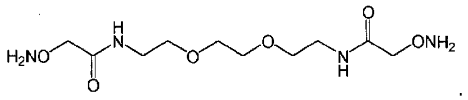
[0389] In certain aspects of compounds of Formula (XIX), are compounds having the structure of Formula (XXVI):



(XXVI)

wherein each R_{10} is independently H or an amino protecting group.

In certain aspects of compounds of Formula (XXVI), the polyalkylene oxide is PEG. In other aspects of compounds of Formula (XXVI), the PEG group has a molecular weight ranging from 5 to 30 kDa. In another embodiment of compounds of Formula (XXVI) is the compound corresponding to:

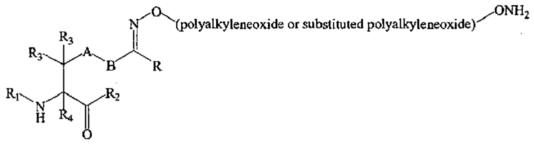


[0390] Three illustrative aspects of methods for coupling a hydroxylamine to a carbonyl-containing non-natural amino acid on a polypeptide are presented in FIG. 7. In these illustrative aspects, a hydroxylamine-derivatized reagent is added to a buffered solution (pH 2-8) of a carbonyl-containing non-natural amino acid polypeptide. The reaction proceeds at the ambient temperature for hours to days. To accelerate the conjugation, additives such as those presented in FIG. 8 are added; such compounds are also known herein as accelerants. In certain aspects, the accelerants or additives are capable of base catalysis. The resulting oxime-containing non-natural amino acid polypeptide is purified by HPLC, FPLC or size-exclusion chromatography.

[0391] In one aspect, multiple linker chemistries can react site-specifically with a carbonyl- or dicarbonyl-substituted non-natural amino acid polypeptide. In one embodiment, the linker methods described herein utilize linkers containing the hydroxylamine functionality on at least one linker termini (mono, bi- or multi-functional). The condensation of a hydroxylamine-derivatized linker with a keto-substituted protein generates a stable oxime linkage. Bi- and/or multi-functional linkers (e.g., hydroxylamine with one, or more, other linking chemistries) allow the site-specific connection of different molecules (e.g., other proteins, polymers or small molecules) to the non-natural amino acid polypeptide, while mono-functional linkers (hydroxylamine-substituted on all termini) facilitate the site-specific dimer- or oligomerization of the non-natural amino acid polypeptide. By combining this linker strategy with the *in vivo* translation technology described herein, it becomes possible to specify the three-dimensional structures of chemically-elaborated proteins.

[0392] In certain aspects are methods for derivatizing a polypeptide comprising amino acids of Formulas I-XVIII, XXX-XXXIV(A&B), or XXXX-XXXXIII, including any sub-formulas or specific compounds that fall within the scope of Formulas I-XVIII, XXX-XXXIV(A&B), or XXXX-XXXXIII, wherein the method comprises contacting the polypeptide comprising at least one amino acid of Formula I- XVIII, XXX-XXXIV(A&B), or XXXX-XXXXIII with a reagent of Formula (XIX). In certain aspects the polypeptide is purified prior to or after contact with the reagent of Formula (XIX). In other aspects are resulting derivatized polypeptide comprises at least one oxime containing amino acid corresponding to Formula (XI), while in other aspects such derivatized polypeptides are stable in aqueous solution for at least 1 month under mildly acidic conditions. In other aspects such derivatized polypeptides are stable for at least 2 weeks under mildly acidic conditions. In other aspects such derivatized polypeptides are stable for at least 5 days under mildly acidic conditions. In other aspects such conditions are pH 2 to 8. In certain aspects the tertiary structure of the derivatized polypeptide is preserved. In other aspects such derivatization of polypeptides further comprises ligating the derivatized polypeptide to another polypeptide. In other aspects such polypeptides being derivatized are homologous to a therapeutic protein selected from the group consisting of: alpha-1 antitrypsin, angiostatin, antihemolytic factor, antibody, apolipoprotein, apoprotein, atrial natriuretic factor, atrial natriuretic polypeptide, atrial peptide, C-X-C chemokine, T39765, NAP-2, ENA-78, gro-a, gro-b, gro-c, IP-10, GCP-2, NAP-4, SDF-1, PF4, MIG, calcitonin, c-kit ligand, cytokine, CC chemokine, monocyte chemoattractant protein-1, monocyte chemoattractant protein-2, monocyte chemoattractant protein-3, monocyte inflammatory protein-1 alpha, monocyte inflammatory protein-i beta, RANTES, 1309, R83915, R91733, HCC1, T58847, D31065, T64262, CD40, CD40 ligand, c-kit ligand, collagen, colony stimulating factor (CSF), complement factor 5a, complement inhibitor, complement receptor 1, cytokine, epithelial neutrophil activating peptide-78, MIP-16, MCP-1, epidermal growth factor (EGF), epithelial neutrophil activating peptide, erythropoietin (EPO), exfoliating toxin, Factor IX, Factor VII, Factor VIII, Factor X, fibroblast growth factor (FGF), fibrinogen, fibronectin, four-helical bundle protein, G-CSF, glp-1, GM-CSF, glucocerebrosidase, gonadotropin, growth factor, growth factor receptor, grf, hedgehog protein, hemoglobin, hepatocyte growth factor (hGF), hirudin, human growth hormone (hGH), human serum albumin, ICAM-1, ICAM-1 receptor, LFA-1, LFA-1 receptor, insulin, insulin-like growth factor (IGF), IGF-I, IGF-II, interferon (IFN), IFN-alpha, IFN-beta, IFN-gamma, interleukin (IL), IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, keratinocyte growth factor (KGF), lactoferrin, leukemia inhibitory factor, luciferase, neurturin, neutrophil inhibitory factor (NIF), oncostatin M, osteogenic protein, oncogene product, paracitonin, parathyroid hormone, PD-ECSF, PDGF, peptide hormone, pleiotropin, protein A, protein G, pth, pyrogenic exotoxin A, pyrogenic exotoxin B, pyrogenic exotoxin C, pyy, relaxin, renin, SCF; small biosynthetic protein, soluble complement receptor I, soluble I-CAM 1, soluble interleukin receptor, soluble TNF receptor, somatomedin, somatostatin, somatotropin, streptokinase, superantigens, staphylococcal enterotoxin, SEA, SEB, SEC1, SEC2, SEC3, SED, SEE, steroid hormone receptor, superoxide dismutase, toxic shock syndrome toxin, thymosin alpha 1, tissue plasminogen activator, tumor growth factor (TGF), tumor necrosis factor, tumor necrosis factor alpha, tumor necrosis factor beta, tumor necrosis factor receptor (TNFR), VLA-4 protein, VCAM-1 protein, vascular endothelial growth factor (VEGF), urokinase, mos, ras, raf, met, p53, tat, fos, myc, jun, myb, rel, estrogen receptor, progesterone receptor, testosterone receptor, aldosterone receptor, LDL receptor, and corticosterone.

[0393] In certain aspects are methods for producing a polypeptide dimer, wherein the method comprises: (i) derivatizing a first polypeptide comprising an amino acid of Formula (I) with a reagent of Formula (XXVI), and (ii) contacting the resulting derivatized protein of step (i) with a second protein comprising an amino acid of Formula (I), thereby forming a dimer comprising the first polypeptide and the second polypeptide. In other aspects are methods for producing a polypeptide dimer, wherein the first polypeptide and the second polypeptide comprise an amino acid of corresponding to Formula (II). In certain aspects the polypeptides are purified prior to or after contact with the reagent of Formula (XXVI). In other aspects the resulting derivatized protein of step (i) comprises at least one oxime containing amino acid corresponding to Formula (XXVIII):

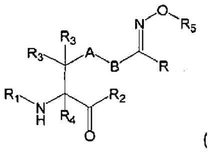


(XXVIII).

B. Methods for Post-Translationally Modifying Non-Naturol Amino Acid Polypeptides: Reactions of Oxime-Containing Non-Natural Amino Acids with Carbonyl-Containing Reagents

[0394] A protein-derivatizing method based upon the exchange reaction of an oxime-containing protein with a carbonyl- or dicarbonyl-substituted molecule has distinct advantages. First, studies indicate that amino acid-based oxime adducts undergo oxime exchange by equilibration with a more reactive carbonyl- or dicarbonyl-containing compound than the one used to generate the original oxime. This exchange reaction occurs in a pH range of 4-8: under these conditions, the side-chains of the naturally occurring amino acids are unreactive. Thus, a general method for the preparation of carbonyl-or dicarbonyl-substituted molecules suitable for reaction with oxime-containing proteins can provide access to a wide variety of site-specifically derivatized proteins. In the context of this *in vivo* translation technology, a general method to prepare carbonyl- or dicarbonyl-substituted versions of those molecules that are typically used to derivatize proteins (including, by way of example only, hydrophilic polymers such as polyethylene glycol) are valuable and will provide access to a wide variety of site-specifically derivatized non-natural amino acid polypeptides. Second, such selective chemistry makes possible the site-specific derivatization of recombinant proteins: derivatized proteins can now be prepared as defined homogeneous products. Third, the mild conditions needed to affect the exchange reactions described herein generally do not irreversibly destroy the tertiary structure of the polypeptide (excepting, of course, where the purpose of the reaction is to destroy such tertiary structure). Finally, the exchange reactions generate new oxime adducts which are stable under biological conditions.

[0395] By way of example only, the following non-natural amino acids are the type of oxime-containing amino acids that are reactive with the carbonyl- or dicarbonyl-containing reagents described herein that can be used to create new oxime-containing non-natural amino acid polypeptides:



(XI)

wherein:

A is optional, and when present is lower alkylene, substituted lower alkylene, lower cycloalkylene, substituted lower cycloalkylene, lower alkenylene, substituted lower alkenylene, alkynylene, lower heteroalkylene, substituted heteroalkylene, lower heterocycloalkylene, substituted lower heterocycloalkylene, arylene, substituted arylene, heteroarylene, substituted heteroarylene, alkarylene, substituted alkarylene, aralkylene, or substituted aralkylene;

B is optional, and when present is a linker selected from the group consisting of lower alkylene, substituted lower alkylene, lower alkenylene, substituted lower alkenylene, lower heteroalkylene, substituted lower heteroalkylene, -0-, -O-(alkylene or substituted alkylene)-, -S-, -S-(alkylene or substituted alkylene)-, $-S(O)_k-$ where k is 1, 2, or 3, $-S(O)_k$(alkylene or substituted alkylene)-, -C(O)-, -C(O)-(alkylene or substituted alkylene)-, -C(S)-, -C(S)-(alkylene or substituted alkylene)-, -N(R')-, -NR'-(alkylene or substituted alkylene)-, -C(O)N(R')-, -CON(R')-(alkylene or substituted alkylene)-, - CSN(R')-, -CSN(R')-(alkylene or substituted alkylene)-, -N(R')CO-(alkylene or substituted alkylene)-, -N(R')C(O)O-, $-S(O)_kN(R')-$, -N(R')C(O)N(R')-, -N(R')C(S)N(R')-, $-N(R')S(O)_kN(R')-$, -N(R')-N=, -C(R')=N-, -C(R')=N-N(R')-, -C(R')=N-N=, $-C(R')_2-N=N-$, and $-C(R')_2-N(R')-N(R')-$, where each R' is independently H, alkyl, or substituted alkyl;

R is H, alkyl, substituted alkyl, cycloalkyl, or substituted cycloalkyl;

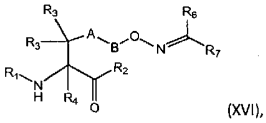
R_1 is H, an amino protecting group, resin, amino acid, polypeptide, or polynucleotide; and

R_2 is OH, an ester protecting group, resin, amino acid, polypeptide, or polynucleotide;

each of R_3 and R_4 is independently H, halogen, lower alkyl, or substituted lower alkyll;

R_5 is L-X, where X is a derivative of polyethylene glycol.

[0396] By way of further example only, the following non-natural amino acids are also the type of oxime-containing amino acids that are reactive with the carbonyl- or dicarbonyl-containing reagents described herein that can be used to create new oxime-containing non-natural amino acid polypeptides:



(XVI),

wherein:

A is optional, and when present is lower alkylene, substituted lower alkylene, lower cycloalkylene, substituted lower cycloalkylene, lower alkenylene, substituted lower alkenylene, alkynylene, lower heteroalkylene, substituted heteroalkylene, lower heterocycloalkylene, substituted lower heterocycloalkylene, arylene, substituted arylene, heteroarylene, substituted heteroarylene, alkarylene, substituted alkarylene, aralkylene, or substituted aralkylene;

B is optional, and when present is a linker selected from the group consisting of lower alkylene, substituted lower alkylene, lower alkenylene, substituted lower alkenylene, lower heteroalkylene, substituted lower heteroalkylene, -O-, -0-(alkylene or substituted alkylene)-, -S-, -S-(alkylene or substituted alkylene)-, -S(O)$_k$- where k is 1, 2, or 3, -S(O)$_k$(alkylene or substituted alkylene)-, -C(O)-, -NS(O)$_2$-, -OS(O)$_2$-, -C(O)-(alkylene or substituted alkylene)-, -C(S)-, -C(S)-(alkylene or substituted alkylene)-, -N(R')-, -NR'-(alkylene or substituted alkylene)-, -C(O)N(R')-, -CON(R')-(alkylene or substituted alkylene)-, -CSN(R')-, -CSN(R')-(alkylene or substituted alkylene)-, -N(R')CO-(alkylene or substituted alkylene)-, -N(R')C(O)O-, -S(O)$_k$N(R')-, -N(R')C(O)N(R')-, -N(R')C(S)N(R')-, -N(R')S(O)$_k$N(R')-, -N(R')-N=, -C(R')=N-, -C(R')=N-N(R')-, - C(R')=N-N=, -C(R')$_2$-N=N-, and -C(R')$_2$-N(R')-N(R')-, where each R' is independently H, alkyl, or substituted alkyl;

R_1 is H, an amino protecting group, resin, amino acid, polypeptide, or polynucleotide; and

R_2 is OH, an ester protecting group, resin, amino acid, polypeptide, or polynucleotide;

each of R_3 and R_4 is independently H, halogen, lower alkyl, or substituted lower alkyl, or R_3 and R_4 or two R_3 groups optionally form a cycloalkyl or a heterocycloalkyl;

each of R_6 and R_7 is independently selected from the group consisting of H, alkyl, substituted alkyl, alkenyl, substituted alkenyl, alkoxy, substituted alkoxy, polyalkylene oxide, substituted polyalkylene oxide, aryl, substituted aryl, heteroaryl, substituted heteroaryl, alkaryl, substituted alkaryl, aralkyl, and substituted aralkyl, -C(O)R", -C(O)$_2$R", -C(O)N(R")2, wherein each R" is independently hydrogen, alkyl, substituted alkyl, alkenyl, substituted alkenyl, alkoxy, substituted alkoxy, aryl, substituted aryl, heteroaryl, alkaryl, substituted alkaryl, aralkyl, or substituted aralkyl; or R_6 or R_7 is L-X, where

X is a selected from the group consisting of a label; a dye; a polymer; water-soluble polymer; a derivative of polyethylene glycol; a photocrosslinker; a cytotoxic compound; a drug; an affinity label; a photoaffinity label; a reactive compound; a resin; a second protein or polypeptide or polypeptide analog; an antibody or antibody fragment; a metal chelator; a cofactor; a fatty acid; a carbohydrate; a polynucleotide; a DNA; a RNA; an antisense polynucleotide; a saccharide, a water-soluble dendrimer, a cyclodextrin, a biomaterial; a nanoparticle; a spin label; a fluorophore, a metal-containing moiety; a radioactive moiety; a novel functional group; a group that covalently or noncovalently interacts with other molecules; a photocaged moiety; an actinic radiation excitable moiety; a ligand; a photoisomerizable moiety; biotin; a biotin analogue; a moiety incorporating a heavy atom; a chemically cleavable group; a photocleavable group; an elongated side chain; a carbon-linked sugar; a redox-active agent; an amino thioacid; a toxic moiety; an isotopically labeled moiety; a biophysical probe; a phosphorescent group; a chemiluminescent group; an electron dense group; a magnetic group; an intercalating group; a chromophore; an energy transfer agent; a biologically active agent; a detectable label; a small molecule; an inhibitory ribonucleic acid, a radionucleotide; a neutron-capture agent; a derivative of biotin; quantum dot(s); a nanotransmitter; a radiotransmitter; an abzyme, an activated complex activator, a virus, an adjuvant, an aglycan, an allergan, an angiostatin, an antihormone, an antioxidant, an aptamer, a guide RNA, a saponin, a shuttle vector, a macromolecule, a mimotope, a receptor, a reverse micelle, and any combination thereof; and L is optional, and when present is a linker selected from the group consisting of alkylene, substituted alkylene, alkenylene, substituted alkenylene, -O-, -O-(alkylene or substituted alkylene)-, -S-, -S-(alkylene or substituted alkylene)-, -S(O)$_k$- where k is 1, 2, or 3, -S(O)$_k$(alkylene or substituted alkylene)-, -C(O)-, -C(O)-(alkylene or substituted alkylene)-, -C(S)-, -C(S)-(alkylene or substituted alkylene)-, -N(R')-, -NR'-(alkylene or substituted alkylene)-, -C(O)N(R')-, -CON(R')-(alkylene or substituted alkylene)-, -CSN(R')-, -CSN(R')-(alkylene or substituted alkylene)-, -N(R')CO-(alkylene or substituted alkylene)-, -N(R')C(O)O-, -S(O)$_k$N(R')-, -N(R')C(O)N(R')-, -N(R')C(S)N(R')-, -N(R')S(O)$_k$N(R')-, -N(R')-N=, -C(R')=N-, -C(R')=N-N(R')-, -C(R')=N-N=, -C(R')$_2$-N=N-, and -C(R')$_2$-N(R')-N(R')-, where each R' is independently H, alkyl, or substituted alkyl.

[0397] By way of example only, the following carbonyl- or dicarbonyl-containing reagents are the type of carbonyl- or dicarbonyl-containing reagents that are reactive with the oxime-containing non-natural amino acids described herein and can be used to effect exchange reactions to form new oxime linkages and thus modify oxime-containing non-natural amino acid polypeptides:

$$\left[X - L\right]_n - L_1 - W \qquad (XIX)$$

wherein:

each X is independently H, alkyl, substituted alkyl, alkenyl, substituted alkenyl, alkynyl, substituted alkynyl, alkoxy, substituted alkoxy, alkylalkoxy, substituted alkylalkoxy, polyalkylene oxide, substituted polyalkylene oxide, aryl, substituted aryl, heteroaryl, substituted heteroaryl, alkaryl, substituted alkaryl, aralkyl, substituted aralkyl, -(alkylene or substituted alkylene)-ON(R")$_2$, -(alkylene or substituted alkylene)-C(O)SR", -(alkylene or substituted alkylene)-S-S-(aryl or substituted aryl), -C(O)R", -C(O)$_2$R", or-C(O)N(R")$_2$, wherein each R" is independently hydrogen, alkyl, substituted alkyl, alkenyl, substituted alkenyl, alkoxy, substituted alkoxy, aryl, substituted aryl, heteroaryl, alkaryl, substituted alkaryl, aralkyl, or substituted aralkyl;

or each X is independently selected from the group consisting of a label; a dye; a polymer; a water-soluble polymer; a derivative of polyethylene glycol; a photocrosslinker; a cytotoxic compound; a drug; an affinity label; a photoaffinity label; a reactive compound; a resin; a second protein or polypeptide or polypeptide analog; an antibody or antibody fragment; a metal chelator; a cofactor; a fatty acid; a carbohydrate; a polynucleotide; a DNA; a RNA; an antisense polynucleotide; a saccharide, a water-soluble dendrimer, a cyclodextrin, a biomaterial; a nanoparticle; a spin label; a fluorophore, a metal-containing moiety; a radioactive moiety; a novel functional group; a group that covalently or noncovalently interacts with other molecules; a photocaged moiety; an actinic radiation excitable moiety; a ligand; a photoisomerizable moiety; biotin; a biotin analogue; a moiety incorporating a heavy atom; a chemically cleavable group; a photocleavable group; an elongated side chain; a carbon-linked sugar; a redox-active agent; an amino thioacid; a toxic moiety; an isotopically labeled moiety; a biophysical probe; a phosphorescent group; a chemiluminescent group; an electron dense group; a magnetic group; an intercalating group; a chromophore; an energy transfer agent; a biologically active agent; a detectable label; a small molecule; an inhibitory ribonucleic acid, a radionucleotide; a neutron-capture agent; a derivative of biotin; quantum dot(s); a nanotransmitter; a radiotransmitter; an abzyme, an activated complex activator, a
virus, an adjuvant, an aglycan, an allergan, an angiostatin, an antihormone, an antioxidant, an aptamer, a guide RNA, a saponin, a shuttle vector, a macromolecule, a mimotope, a receptor, a reverse micelle, and any combination thereof;

each L is independently selected from the group consisting of alkylene, substituted alkylene, alkenylene, substituted alkenylene, -O-, -O-(alkylene or substituted alkylene)-, -S-, -S-(alkylene or substituted alkylene)-, -S(O)$_k$- where k is 1, 2, or 3, -S(O)$_k$(alkylene or substituted alkylene)-, -C(O)-, -C(O)-(alkylene or substituted alkylene)-, -C(S)-, -C(S)-(alkylene or substituted alkylene)-, -N(R')-, -NR'-(alkylene or substituted alkylene)-, -C(O)N(R')-, -CON(R')-(alkylene or substituted alkylene)-, -(alkylene or substituted alkylene)NR'C(O)O-(alkylene or substituted alkylene)-, -O-CON(R')-(alkylene or substituted alkylene)-, -CSN(R')-, -CSN(R')-(alkylene or substituted alkylene)-, -N(R')CO-(alkylene or substituted alkylene)-, -N(R')C(O)O-, -N(R')C(O)O-(alkylene or substituted alkylene)-, -S(O)$_k$N(R')-, -N(R')C(O)N(R')-, -N(R')C(O)N(R')-(alkylene or substituted alkylene)-, -N(R')C(S)N(R')-, -N(R')S(O)$_k$N(R')-, -N(R')-N=, -C(R')=N-, -C(R')=N-N(R')-, -C(R')=N-N=, -C(R')$_2$-N=N-, and -C(R')$_2$-N(R')-N(R')-;

L$_1$ is optional, and when present, is-C(R')$_p$-NR'-C(O)O-(alkylene or substituted alkylene)- where p is 0,1, or 2; each R' is independently H, alkyl, or substituted alkyl;

W is-C(=O)R$_9$, where R$_9$ is H or OR'; and n is 1 to 3;

or wherein L-L$_1$-W together forms a monocyclic or bicyclic cycloalkyl or heterocycloalkyl comprising at least one carbonyl group, including a dicarbonyl group, protected carbonyl group, including a protected dicarbonyl group, or masked carbonyl group, including a masked dicarbonyl group;

provided that L-L$_1$-W together provide at least one carbonyl group (including a dicarbonyl group) capable of undergoing an oxime exchange reaction with an oxime group on a non-natural amino acid or a "modified or unmodified" non-natural amino acid polypeptide.

[0398] Two illustrative examples of methods for effect an oxime exchange reaction between an oxime-containing amino acid on a polypeptide and a carbonyl-containing reagent are presented in FIG. 9. In these illustrative examples, a carbonyl-containing reagent is added to a buffered solution (pH 2-8) of an oxime-containing non-natural amino acid polypeptide. The reaction proceeds at the ambient temperature for hours to days. The modified oxime-containing non-natural amino acid polypeptide is purified by HPLC, FPLCorsize-exclusion chromatography.

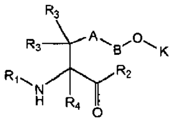
[0399] In one embodiment, multiple linker chemistries can react site-specifically with an oxime-substituted non-natural amino acid polypeptide. In one embodiment, the linker methods described herein utilize linkers containing the carbonyl or dicarbonyl functionality on at least one linker termini (mono, bi- or multi-functional). The condensation of a carbonyl- or d icarbonyl-derivatized linker with an oxime-substituted non-natural amino acid polypeptide generates a new stable oxime linkage. Bi- and/or multi-functional linkers (e.g., carbonyl or dicarbonyl with one, or more, other linking chemistries) allow the site-specific connection of different molecules (e.g., other proteins, polymers or small molecules) to the non-natural amino acid polypeptide, while mono-functional linkers (carbonyl- or dicarbonyl-substituted on all termini) facilitate the site-specific dimer- or oligomerization of the non-natural amino acid polypeptide. By combining this linker strategy with the *in vivo* translation technology described herein, it becomes possible to specify the three-dimensional structures of chemically-elaborated proteins.

C. Methods for Post-Translationally Modifying Non-Natural Amino Acid Polypeptides: Reactions of Hydroxylamine-Containing Non-Natural Amino Acids with Carbonyl-Containing Reagents

[0400] The post-translational modification techniques and compositions described above may also be used with hydroxylamine-containing non-natural amino acids reacting with carbonyl- or dicarbonyl-containing reagents to produce modified oxime-containing non-natural amino acid polypeptides.

[0401] A protein-derivatizing method based upon the reaction of a hydroxylamine-containing protein with a carbonyl- or dicarbonyl-substituted molecule has distinct advantages. First, hydroxylamines undergo condensation with carbonyl- or dicarbonyl-containing compounds in a pH range of 4-8 to generate oxime adducts. Under these conditions, the sidechains of the naturally occurring amino acids are unreactive. Second, such selective chemistry makes possible the site-specific derivatization of recombinant proteins: derivatized proteins can now be prepared as defined homogeneous products. Third, the mild conditions needed to effect the reaction of the carbonyl- or dicarbonyl-containing reagents described herein with the hydroxylamine-containing polypeptides described herein generally do not irreversibly destroy the tertiary structure of the polypeptide (excepting, of course, where the purpose of the reaction is to destroy such tertiary structure). Finally, although the hydroxylamine group amino appears to be metabolized by *E. coli,* the condensation of carbonyl- or dicarbonyl-containing reagents with hydroxylamine-containing amino acids generates oxime adducts which are stable under biological conditions.

[0402] By way of example only, the following non-natural amino acids are the type of hydroxylamine-containing amino acids that are reactive with the carbonyl- or dicarbonyl-containing reagents described herein that can be used to further modify hydroxylamine-containing non-natural amino acid polypeptides:



(XIV),

wherein:

A is optional, and when present is lower alkylene, substituted lower alkylene, lower cycloalkylene, substituted lower cycloalkylene, lower alkenylene, substituted lower alkenylene, alkynylene, lower heteroalkylene, substituted heteroalkylene, lower heterocycloalkylene, substituted lower heterocycloalkylene, arylene, substituted arylene, heteroarylene, substituted heteroarylene, alkarylene, substituted alkarylene, aralkylene, or substituted aralkylene;

B is optional, and when present is a linker selected from the group consisting of lower alkylene, substituted lower alkylene, lower alkenylene, substituted lower alkenylene, lower heteroalkylene, substituted lower heteroalkylene, -0-, -O-(alkylene or substituted alkylene)-, -S-,-S-(alkylene or substituted alkylene)-, -S(O)$_k$-where k is 1, 2, or 3, -S(O)$_k$(alkylene or substituted alkylene)-, -C(O)-, -NS(O)$_2$-, -OS(O)$_2$-, -C(O)-(alkylene or substituted alkylene)-, -C(S)-, -C(S)-(alkylene or substituted alkylene)-, -N(R')-, -NR'-(alkylene or substituted alkylene)-, -C(O)N(R')-, -CON(R')-(alkylene or substituted alkylene)-, -CSN(R')-, -CSN(R')-(alkylene or substituted alkylene)-, -N(R')CO-(alkylene or substituted alkylene)-, -N(R')C(O)O-, -S(O)$_k$N(R')-, -N(R')C(O)N(R')-, -N(R')C(S)N(R')-, -N(R')S(O)$_k$N(R')-, -N(R')-N=, -C(R')=N-, -C(R')=N-N(R')-,-C(R')=N-N=, -C(R')$_2$-N=N-, and -C(R')$_2$-N(R')-N(R')-, where each R' is independently H, alkyl, or substituted alkyl;

K is -NR$_6$R$_7$or-N=CR$_6$R$_7$;

R$_1$ is H, an amino protecting group, resin, amino acid, polypeptide, or polynucleotide; and

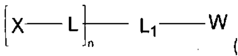
R$_2$ is OH, an ester protecting group, resin, amino acid, polypeptide, or polynucleotide;

each of R$_3$ and R$_4$ is independently H, halogen, lower alkyl, or substituted lower alkyl, or R$_3$ and R$_4$ or two R$_3$ groups optionally form a cycloalkyl or a heterocycloalkyl.

[0403] In certain aspects of compounds of Formula (XIV) K is NH$_2$.

[0404] The types of polypeptides that comprise such hydroxylamine-containing non-natural amino acids is practically unlimited as long as the hydroxylamine-containing non-natural amino acid is located on the polypeptide so that the carbonyl- or dicarbonyl-containing reagent can react with the hydroxylamine group and not create a resulting modified non-natural amino acid that destroys the tertiary structure of the polypeptide (excepting, of course, if such destruction is the purpose of the reaction).

[0405] By way of example only, the following carbonyl- or dicarbonyl-containing reagents are the type of carbonyl- or dicarbonyl-containing reagents that are reactive with the hydroxylamine-containing non-natural amino acids described herein and can be used to further modify hydroxylamine-containing non-natural amino acid polypeptides:



$$\left[X-L\right]_n-L_1-W$$

(XIX)

wherein:

each X is independently H, alkyl, substituted alkyl, alkenyl, substituted alkenyl, alkynyl, substituted alkynyl, alkoxy, substituted alkoxy, alkylalkoxy, substituted alkylalkoxy, polyalkylene oxide, substituted polyalkylene oxide, aryl, substituted aryl, heteroaryl, substituted heteroaryl, alkaryl, substituted alkaryl, aralkyl, substituted aralkyl, (alkylene or substituted alkylene)-ON(R")$_2$, -(alkylene or substituted alkylene)-C(O)SR", -(alkylene or substituted alkylene)-S-S-(aryl or substituted aryl), -C(O)R", -C(O)$_2$R", or-C(O)N(R")$_2$ wherein each R" is independently hydrogen, alkyl, substituted alkyl, alkenyl, substituted alkenyl, alkoxy, substituted alkoxy, aryl, substituted aryl, heteroaryl, alkaryl, substituted alkaryl, aralkyl, or substituted aralkyl;

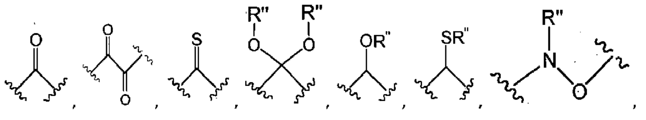
or each X is independently selected from the group consisting of a label; a dye; a polymer; a water-soluble polymer; a derivative of polyethylene glycol; a photocrosslinker; a cytotoxic compound; a drug; an affinity label; a photoaffinity label; a reactive compound; a resin; a second protein or polypeptide or polypeptide analog; an antibody or antibody fragment; a metal chelator; a cofactor; a fatty acid; a carbohydrate; a polynucleotide; a DNA; a RNA; an antisense polynucleotide; a saccharide, a water-soluble dendrimer, a cyclodextrin, a biomaterial; a nanoparticle; a spin label; a fluorophore, a metal-containing moiety; a radioactive moiety; a novel functional group; a group that covalently or noncovalently interacts with other molecules; a photocaged moiety; an actinic radiation excitable moiety; a ligand; a photoisomerizable moiety; biotin; a biotin analogue; a moiety incorporating a heavy atom; a chemically cleavable group; a photocleavable group; an elongated side chain; a carbon-linked sugar; a redox-active agent; an amino thioacid; a toxic moiety; an isotopically labeled moiety; a biophysical probe; a phosphorescent group; a chemiluminescent group; an electron dense group; a magnetic group; an intercalating group; a chromophore; an energy transfer agent; a biologically active agent; a detectable label; a small molecule; an inhibitory ribonucleic acid, a radionucleotide; a neutron-capture agent; a derivative of biotin; quantum dot(s); a nanotransmitter; a radiotransmitter; an abzyme, an activated complex activator, a virus, an adjuvant, an aglycan, an allergan, an angiostatin, an antihormone, an antioxidant, an aptamer, a guide RNA, a saponin, a shuttle vector, a macromolecule, a mimotope, a receptor, a reverse micelle, and any combination thereof;

each L is independently selected from the group consisting of alkylene, substituted alkylene, alkenylene, substituted alkenylene, -0-, -O-(alkylene or substituted alkylene)-, -S-, -S-(alkylene or substituted alkylene)-, -S(O)$_k$- where k is 1, 2, or 3, -S(O)$_k$(alkylene or substituted alkylene)-, -C(O)-, -C(O)-(alkylene or substituted alkylene)-, -C(S)-, -C(S)-(alkylene or substituted alkylene)-, -N(R')-, -NR'-(alkylene or substituted alkylene)-, -C(O)N(R')-, -CON(R')-(alkylene or substituted alkylene)-,-(alkylene or substituted alkylene)NR'C(O)O-(alkylene or substituted alkylene)-, -O-CON(R')-(alkylene or substituted alkylene)-, -CSN(R')-, -CSN(R')-(alkylene or substituted alkylene)-, -N(R')CO-(alkylene or substituted alkylene)-, -N(R')C(0)0-, -N(R')C(O)O-(alkylene or substituted alkylene)-, -S(O)$_k$N(R')-, -N(R')C(O)N(R')-, -N(R')C(O)N(R')-(alkylene or substituted alkylene)-, -N(R')C(S)N(R')-, -N(R')S(O)$_k$N(R')-, -N(R')-N=, -C(R')=N-, -C(R')=N-N(R')-, -C(R')=N-N=, -C(R')$_2$-N=N-, and -C(R')$_2$N(R')-N(R')-;

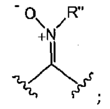
L$_1$ is optional, and when present, is -C(R')$_p$-NR'-C(O)O-(alkylene or substituted alkylene)- where p is 0,1, or 2;

each R' is independently H, alkyl, or substituted alkyl;

W is -J-R, where J is



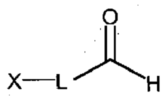
or



R is H, alkyl, substituted alkyl, cycloalkyl, or substituted cycloalkyl; each R" is independently H, alkyl, substituted alkyl, or a protecting group, or when more than one R" group is present, two R" optionally form a heterocycloalkyl; and n is 1 to 3;

provided that L-L_1-W together provide at least one carbonyl group (including a dicarbonyl group) capable of reacting with an hydroxylamine group on a non-natural amino acid or a "modified or unmodified" non-natural aminoacid polypeptide.

[0406] In certain aspects of compounds of Formula (XIX), are compounds having the structure of Formula (XXI):

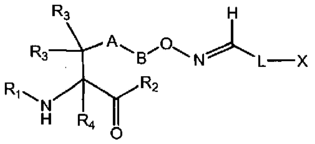


(XXI).

[0407] An illustrative aspect of methods for coupling a carbonyl-containing reagent to a hydroxylamine-containing non-natural amino acid on a polypeptide is presented in FIG. 10. In this illustrative aspect, a carbonyl-derivatized reagent is added to a buffered solution (pH 2-8) of a hydroxylamine-containing non-natural amino acid polypeptide. The reaction proceeds at the ambient temperature for hours to days. To accelerate the conjugation, additives such as those presented in FIG. 8 are added. The resulting oxime-containing non-natural amino acid polypeptide is purified by HPLC, FPLC or size-exclusion chromatography.

[0408] In one aspect , multiple linker chemistries can react site-specifically with a hydroxylamine-substituted non-natural amino acid polypeptide. In one aspect, the linker methods described herein utilize linkers containing the carbonyl or dicarbonyl functionality on at least one linker termini (mono, bi- or multi-functional). The condensation of a carbonyl- or dicarbonyl-derivatized linker with a hydroxylamine-substituted protein generates a stable oxime linkage. Bi- and/or multi-functional linkers (e.g., carbonyl or dicarbonyl with one, or more, other linking chemistries) allow the site-specific connection of different molecules (e.g., other proteins, polymers or small molecules) to the non-natural amino acid polypeptide, while mono-functional linkers (carbonyl- or dicarbonyl-substituted on all termini) facilitate the site-specific dimer- or oligomerization of the non-natural amino acid polypeptide. By combining this linker strategy with the *in vivo* translation technology described herein, it becomes possible to specify the three-dimensional structures of chemically-elaborated proteins.

[0409] In certain aspects are methods for derivatizing a polypeptide comprising amino acids of Formulas XIV-XVI, including any sub-formulas or specific compounds that fall within the scope of Formulas XIV-XVI, wherein the method comprises contacting the polypeptide comprising at least one amino acid of Formula XIV-XVI with a reagent of Formula (XIX). In certain aspects the polypeptide is purified prior to or after contact with the reagent of Formula (XIX). In other aspects are resulting derivatized polypeptide comprises at least one oxime containing amino acid corresponding to Formula (XXIX),



(XXIX)

wherein:

A is optional, and when present is lower alkylene, substituted lower alkylene, lower alkenylene, substituted lower alkenylene, arylene, substituted arylene, heteroarylene, substituted heteroarylene, alkarylene, substituted alkarylene, aralkylene, or substituted aralkylene;

B is optional, and when present is a linker selected from the group consisting of lower alkylene, substituted lower alkylene, lower alkenylene, substituted lower alkenylene, -O-, -O-(alkylene or substituted alkylene)-, -S-, -S-(alkylene or substituted alkylene)-, -$S(O)_k$- where k is 1, 2, or 3, -$S(O)_k$(alkylene or substituted alkylene)-, -C(O)-, -$NS(O)_2$-, -$OS(O)_2$-, -C(O)-(alkylene or substituted alkylene)-, -C(S)-, -C(S)-(alkylene or substituted alkylene)-, -N(R')-, -NR'-(alkylene or substituted alkylene)-, -C(O)N(R')-, -CON(R')-(alkylene or substituted alkylene)-, -CSN(R')-, -CSN(R')-(alkylene or substituted alkylene)-, -N(R')CO-(alkylene or substituted alkylene)-, -N(R')C(O)O-, -$S(O)_k$N(R')-, -N(R')C(O)N(R')-, -N(R')C(S)N(R')-, -N(R')$S(O)_k$N(R')-, -N(R')-N=, -C(R')=N-, -C(R')=N-N(R')-, -C(R')=N-N=, -$C(R')_2$-N=N-, and -$C(R')_2$-N(R')-N(R')-, where each R' is independently H, alkyl, or substituted alkyl;

L is a linker independently selected from the group consisting of alkylene, substituted alkylene, alkenylene, substituted alkenylene, -0-, -O-(alkylene or substituted alkylene)-, -S-, -S-(alkylene or substituted alkylene)-, -$S(O)_k$- where k is 1, 2, or 3, -$S(O)_k$(alkylene or substituted alkylene)-, -C(O)-, -C(O)-(alkylene or substituted alkylene)-, -C(S)-, -C(S)-(alkylene or substituted alkylene)-, -N(R')-, -NR'-(alkylene or substituted alkylene)-, -C(O)N(R')-, -CON(R')-(alkylene or substituted alkylene)-, -(alkylene or substituted alkylene)NR'C(O)O-(alkylene or substituted alkylene)-, -O-CON(R')-(alkylene or substituted alkylene)-, -CSN(R')-, -CSN(R')-(alkylene or substituted alkylene)-, -N(R')CO-(alkylene or substituted alkylene)-, -N(R')C(O)O-, -N(R')C(O)O-(alkylene or substituted alkylene)-, -$S(O)_k$N(R')-, -N(R')C(O)N(R')-, -N(R')C(O)N(R')-(alkylene or substituted alkylene)-, -N(R')C(S)N(R')-, -

$N(R')S(O)_kN(R')$-, -N(R')-N=, -C(R')=N-, -C(R')=N-N(R')-, -C(R')=N-N=, $-C(R')_2-N=N-$, and $-C(R')_2-N(R')-N(R')-$, where each R' is independently H, alkyl, or substituted alkyl;

R_1 is optional, and when present, is H, an amino protecting group, resin, amino acid, polypeptide, or polynucleotide; and

R_2 is optional, and when present, is OH, an ester protecting group, resin, amino acid, polypeptide, or polynucleotide; each R_3 and R_4 is independently H, halogen, lower alkyl, or substituted lower alkyl; and

X is independently a detectable label, biologically active agent, or polymer.

[0410] In other aspects such derivatized polypeptides are stable in aqueous solution for at least 1 month under mildly acidic conditions. In other aspects such derivatized polypeptides are stable for at least 2 weeks under mildly acidic conditions. In other aspects such derivatized polypeptides are stable for at least 5 days under mildly acidic conditions. In other embodiments such conditions are pH 2 to 8. In certain embodiments the tertiary structure of the derivatized polypeptide is preserved. In other embodiments such derivatization of polypeptides further comprises ligating the derivatized polypeptide to another polypeptide. In other aspects such polypeptides being derivatized are homologous to a therapeutic protein selected from the group consisting of: alpha-1 antitrypsin, angiostatin, antihemolytic factor, antibody, apolipoprotein, apoprotein, atrial natriuretic factor, atrial natriuretic polypeptide, atrial peptide, C-X-C chemokine, T39765, NAP-2, ENA-78, gro-a, gro-b, gro-c, IP-10, GCP-2, NAP-4, SDF-1, PF4, MIG, calcitonin, c-kit ligand, cytokine, CC chemokine, monocyte chemoattractant protein-1, monocyte chemoattractant protein-2, monocyte chemoattractant protein-3, monocyte inflammatory protein-1 alpha, monocyte inflammatory protein-i beta, RANTES, 1309, R83915, R91733, HCC1, T58847, D31065, T64262, CD40, CD40 ligand, c-kit ligand, collagen, colony stimulating factor (CSF), complement factor 5a, complement inhibitor, complement receptor 1, cytokine, epithelial neutrophil activating peptide-78, MIP-16, MCP-1, epidermal growth factor (EGF), epithelial neutrophil activating peptide, erythropoietin (EPO), exfoliating toxin, Factor IX, Factor VII, Factor VIII, Factor X, fibroblast growth factor (FGF), fibrinogen, fibronectin, four-helical bundle protein, G-CSF, glp-1, GM-CSF, glucocerebrosidase, gonadotropin, growth factor, growth factor receptor, grf, hedgehog protein, hemoglobin, hepatocyte growth factor (hGF), hirudin, human growth hormone (hGH), human serum albumin, ICAM-1, ICAM-1 receptor, LFA-1, LFA-1 receptor, insulin, insulin-like growth factor (IGF), IGF-I, IGF-II, interferon (IFN), IFN-alpha, IFN-beta, IFN-gamma, interleukin (IL), IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, keratinocyte growth factor (KGF), lactoferrin, leukemia inhibitory factor, luciferase, neurturin, neutrophil inhibitory factor (NIF), oncostatin M, osteogenic protein, oncogene product, paracitonin, parathyroid hormone, PD-ECSF, PDGF, peptide hormone, pleiotropin, protein A, protein G, pth, pyrogenic exotoxin A, pyrogenic exotoxin B, pyrogenic exotoxin C, pyy, relaxin, renin, SCF, small biosynthetic protein, soluble complement receptor I, soluble I-CAM 1, soluble interleukin receptor, soluble TNF receptor, somatomedin, somatostatin, somatotropin, streptokinase, superantigens, staphylococcal enterotoxin, SEA, SEB, SEC1, SEC2, SEC3, SED, SEE, steroid hormone receptor, superoxide dismutase, toxic shock syndrome toxin, thymosin alpha 1, tissue plasminogen activator, tumor growth factor (TGF), tumor necrosis factor, tumor necrosis factor alpha, tumor necrosis factor beta, tumor necrosis factor receptor (TNFR), VLA-4 protein, VCAM-1 protein, vascular endothelial growth factor (VEGF), urokinase, mos, ras, raf, met, p53, tat, fos, myc, jun, myb, rel, estrogen receptor, progesterone receptor, testosterone receptor, aldosterone receptor, LDL receptor, and corticosterone.

D Example of Adding Functionality: Macromolecular Polymers Coupled to Non-Natural Amino Acid Polypeptides

[0411] Various modifications to the non-natural amino acid polypeptides described herein can be effected using the compositions, methods, techniques and strategies described herein. These modifications include the incorporation of further functionality onto the non-natural amino acid component of the polypeptide, including a label; a dye; a polymer; a water-soluble polymer; a derivative of polyethylene glycol; a photocrosslinker; a cytotoxic compound; a drug; an affinity label; a photoaffinity label; a reactive compound; a resin; a second protein or polypeptide or polypeptide analog; an antibody or antibody fragment; a metal chelator; a cofactor; a fatty acid; a carbohydrate; a polynucleotide; a DNA; a RNA; an antisense polynucleotide; a saccharide, a water-soluble dendrimer, a cyclodextrin, a biomaterial; a nanoparticle; a spin label; a fluorophore, a metal-containing moiety; a radioactive moiety; a novel functional group; a group that covalently or noncovalently interacts with other molecules; a photocaged moiety; an actinic radiation excitable moiety; a ligand; a photoisomerizable moiety; biotin; a biotin analogue; a moiety incorporating a heavy atom; a chemically cleavable group; a photocleavable group; an elongated side chain; a carbon-linked sugar; a redox-active agent; an amino thioacid; a toxic moiety; an isotopically labeled moiety; a biophysical probe; a phosphorescent group; a chemiluminescent group; an electron dense group; a magnetic group; an intercalating group; a chromophore; an energy transfer agent; a biologically active agent; a detectable label; a small molecule; an inhibitory ribonucleic acid, a radionucleotide; a neutron-capture agent; a derivative of biotin; quantum dot(s); a nanotransmitter; a radiotransmitter; an abzyme, an activated complex activator, a virus, an adjuvant, an aglycan, an allergan, an angiostatin, an antihormone, an antioxidant, an aptamer, a guide RNA, a saponin, a shuttle vector, a macromolecule, a mimotope, a receptor, a reverse micelle, and any combination thereof. As an illustrative example of the compositions, methods, techniques and strategies described herein, the following description will focus on adding macromolecular polymers to the non-natural amino acid polypeptide with the understanding that the compositions, methods, techniques and strategies described thereto are also applicable (with appropriate modifications, if necessary and for which one of skill in the art could make with the disclosures herein) to adding other functionalities, including those listed above.

[0412] A wide variety of macromolecular polymers and other molecules can be coupled to the non-natural amino acid polypeptides described herein to modulate biological properties of the non-natural amino acid polypeptide (or the corresponding natural amino acid polypeptide), and/or provide new biological properties to the non-natural amino acid polypeptide (or the corresponding natural amino acid polypeptide). These macromolecular polymers can be coupled to the non-natural amino acid polypeptide via the non-natural amino acid, or any functional substituent of the non-natural amino acid, or any substituent or functional group added to the non-natural amino acid.

[0413] Water soluble polymers can be coupled to the non-natural amino acids incorporated into polypeptides (natural or synthetic), polynucleotides, poly saccharides or synthetic polymers described herein. The water soluble polymers may be coupled via a non-natural amino acid incorporated in the polypeptide or any functional group or substituent of a non-natural amino acid, or any functional group or substituent added to a non-natural amino acid. In some cases, the non-natural amino acid polypeptides described herein comprise one or more non-natural amino acid(s) coupled to water soluble polymers and one or more naturally-occurring amino acids linked to water soluble polymers. Covalent attachment of hydrophilic polymers to a biologically active molecule represents one approach to increasing water solubility (such as in a physiological environment), bioavailability, increasing serum half-life, increasing therapeutic half-life, modulating immunogenicity, modulating biological activity, or extending the circulation time of the biologically active molecule, including proteins, peptides, and particularly hydrophobic molecules. Additional important features of such hydrophilic polymers include biocompatibility, lack of toxicity, and lack of immunogenicity. Preferably, for therapeutic use of the end-product preparation, the polymer will be pharmaceutically acceptable.

[0414] Examples of hydrophilic polymers include: polyalkyl ethers and alkoxy-capped analogs thereof (e.g., polyoxyethylene glycol, polyoxyethylene/propylene glycol, and methoxy or ethoxy-capped analogs thereof, especially polyoxyethylene glycol, the latter is also known as polyethylene glycol or PEG); polyvinylpyrrolidones; polyvinylalkyl ethers; polyoxazolines, polyalkyl oxazolines and polyhydroxyalkyl oxazolines; polyacrylamides, polyalkyl acrylamides, and polyhydroxyalkyl acrylamides (e.g., polyhydroxypropylmethacrylamide and derivatives thereof); polyhydroxyalkyl acrylates; polysialic acids and analogs thereof; hydrophilic peptide sequences; polysaccharides and their derivatives, including dextran and dextran derivatives, e.g., carboxymethyldextran, dextran sulfates, aminodextran; cellulose and its derivatives, e.g., carboxymethyl cellulose, hydroxyalkyl celluloses; chitin and its derivatives, e.g., chitosan, succinyl chitosan, carboxymethylchitin, carboxymethylchitosan; hyaluronic acid and its derivatives; starches; alginates; chondroitin sulfate; albumin; pullulan and carboxymethyl pullulan; polyaminoacids and derivatives thereof, e.g., polyglutamic acids, polylysines, polyaspartic acids, polyaspartamides; maleic anhydride copolymers such as: styrene maleic anhydride copolymer, divinylethyl ether maleic anhydride copolymer; polyvinyl alcohols; copolymers thereof; terpolymers thereof; mixtures thereof; and derivatives of the foregoing. The water soluble polymer may be any structural form including linear, forked or branched. In some aspects, polymer backbones that are water-soluble, with from 2 to about 300 termini, are particularly useful. Multifunctional polymer derivatives include linear polymers having two termini, each terminus being bonded to a functional group which may be the same or different. In some embodiments, the water polymer comprises a poly(ethylene glycol) moiety. The molecular weight of the polymer may be of a wide range, including between about 100 Da and about 100,000 Da or more. The molecular weight of the polymer may be between about 100 Da and about 100,000 Da, including 100,000 Da, 95,000 Da, 90,000 Da, 85,000 Da, 80,000 Da, 75,000 Da, 70,000 Da, 65,000 Da, 60,000 Da, 55,000 Da, 50,000 Da, 45,000 Da, 40,000 Da, 35,000 Da, 30,000 Da, 25,000 Da, 20,000 Da, 15,000 Da, 10,000 Da, 9,000 Da, 8,000 Da, 7,000 Da, 6,000 Da, 5,000 Da, 4,000 Da, 3,000 Da, 2,000 Da, 1,000 Da, 900 Da, 800 Da, 700 Da, 600 Da, 500 Da, 400 Da, 300 Da, 200 Da, and 100 Da. In some embodiments, the molecular weight of the polymer is between about 100 Da and 50,000 Da. In some embodiments, the molecular weight of the polymer is between about 100 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 1,000 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 5,000 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 10,000 Da and 40,000 Da. In some embodiments, the poly(ethylene glycol) molecule is a branched polymer. The molecular weight of the branched chain PEG may be between about 1,000 Da and about 100,000 Da, including but not limited to, 100,000 Da, 95,000 Da, 90,000 Da, 85,000 Da, 80,000 Da, 75,000 Da, 70,000 Da, 65,000 Da, 60,000 Da, 55,000 Da, 50,000 Da, 45,000 Da, 40,000 Da, 35,000 Da, 30,000 Da, 25,000 Da, 20,000 Da, 15,000 Da, 10,000 Da, 9,000 Da, 8,000 Da, 7,000 Da, 6,000 Da, 5,000 Da, 4,000 Da, 3,000 Da, 2,000 Da, and 1,000 Da. In some embodiments, the molecular weight of the branched chain PEG is between about 1,000 Da and 50,000 Da. In some embodiments, the molecular weight of the branched chain PEG is between about 1,000 Da and 40,000 Da. In some embodiments, the molecular weight of the branched chain PEG is between about 5,000 Da and 40,000 Da. In some embodiments, the molecular weight of the branched chain PEG is between about 5,000 Da and 20,000 Da. Those of ordinary skill in the art will recognize that the foregoing list for substantially water soluble backbones is by no means exhaustive and is merely illustrative, and that all polymeric materials having the qualities described above are contemplated as being suitable for use in methods and compositions described herein.

[0415] As described above, one example of a hydrophilic polymer is poly(ethylene glycol), abbreviated PEG, which has been used extensively in pharmaceuticals, on artificial implants, and in other applications where biocompatibility, lack of toxicity, and lack of immunogenicity are of importance. The polymer:polypeptide embodiments described herein will use PEG as an example hydrophilic polymer with the understanding that other hydrophilic polymers may be similarly utilized in such embodiments.

[0416] PEG is a well-known, water soluble polymer that is commercially available or can be prepared by ring-opening polymerization of ethylene glycol according to methods well known in the art (Sandler and Karo, Polymer Synthesis, Academic Press, New York, Vol. 3, pages 138-161). PEG is typically clear, colorless, odorless, soluble in water, stable to heat, inert to many chemical agents, does not hydrolyze or deteriorate, and is generally non-toxic. Poly(ethylene glycol) is considered to be biocompatible, which is to say that PEG is capable of coexistence with living tissues or organisms without causing harm. More specifically, PEG is substantially non-immunogenic, which is to say that PEG does not tend to produce an immune response in the body. When attached to a molecule having some desirable function in the body, such as a biologically active agent, the PEG tends to mask the agent and can reduce or eliminate any immune response so that an organism can tolerate the presence of the agent. PEG conjugates tend not to produce a substantial immune response or cause clotting or other undesirable effects.

[0417] The term "PEG" is used broadly to encompass any polyethylene glycol molecule, without regard to size or to modification at an end of the PEG, and can be represented as linked to a non-natural amino acid polypeptide by the formula:

$$XO\text{-}(CH_2CH_2O)_n\text{-}CH_2CH_2\text{-}Y$$

where n is 2 to 10,000 and X is H or a terminal modification, including a C_{1-4} alkyl, a protecting group, or a terminal functional group. The term PEG includes poly(ethylene glycol) in any of its forms, including bifunctional PEG, multiarmed PEG, derivatized PEG, forked PEG, branched PEG (with each chain having a molecular weight of from about 1 kDa to about 100 kDa, from about 1 kDa to about 50 kDa, or from about 1 kDa to about 20 kDa), pendent PEG (i.e. PEG or related polymers having one or more functional groups pendent to the polymer backbone), or PEG with degradable linkages therein. In one embodiment, PEG in which n is from about 20 to about 2000 is suitable for use in the methods and compositions described herein. In some embodiments, the water polymer comprises a poly(ethylene glycol) moiety. The molecular weight of the PEG polymer may be of a wide range, including between about 100 Da and about 100,000 Da or more. The molecular weight of the PEG polymer may be between about 100 Da and about 100,000 Da, including 100,000 Da, 95,000 Da, 90,000 Da, 85,000 Da, 80,000 Da, 75,000 Da, 70,000 Da, 65,000 Da, 60,000 Da, 55,000 Da, 50,000 Da, 45,000 Da, 40,000 Da, 35,000 Da, 30,000 Da, 25,000 Da, 20,000 Da, 15,000 Da, 10,000 Da, 9,000 Da, 8,000 Da, 7,000 Da, 6,000 Da, 5,000 Da, 4,000 Da, 3,000 Da, 2,000 Da, 1,000 Da, 900 Da, 800 Da, 700 Da, 600 Da, 500 Da, 400 Da, 300 Da, 200 Da, and 100 Da. In some embodiments, the molecular weight of the polymer is between about 100 Da and 50,000 Da. In some embodiments, the molecular weight of the polymer is between about 100 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 1,000 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 5,000 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 10,000 Da and 40,000 Da. In some embodiments, the poly(ethylene glycol) molecule is a branched polymer. The molecular weight of the branched chain PEG may be between about 1,000 Da and about 100,000 Da, including but not limited to, 100,000 Da, 95,000 Da, 90,000 Da, 85,000 Da, 80,000 Da, 75,000 Da, 70,000 Da, 65,000 Da, 60,000 Da, 55,000 Da, 50,000 Da, 45,000 Da, 40,000 Da, 35,000 Da, 30,000 Da, 25,000 Da, 20,000 Da, 15,000 Da, 10,000 Da, 9,000 Da, 8,000 Da, 7,000 Da, 6,000 Da, 5,000 Da, 4,000 Da, 3,000 Da, 2,000 Da, and 1,000 Da. In some embodiments, the molecular weight of the branched chain PEG is between about 1,000 Da and 50,000 Da. In some embodiments, the molecular weight of the branched chain PEG is between about 1,000 Da and 40,000 Da. In some embodiments, the molecular weight of the branched chain PEG is between about 5,000 Da and 40,000 Da. In some embodiments, the molecular weight of the branched chain PEG is between about 5,000 Da and 20,000 Da. A wide range of PEG molecules are described in, including the Shearwater Polymers, Inc. catalog, Nektar Therapeutics catalog.

[0418] Specific examples of terminal functional groups in the literature include N-succinimidyl carbonate (see e.g., U.S. Pat. Nos. 5,281,698, 5,468,478), amine (see, e.g., Buckmann et al. Makromol. Chem. 182:1379 (1981) , Zalipsky et al. Eur. Polym. J. 19:1177 (1983)), hydrazide (See, e.g., Andresz et al. Makromol. Chem. 179:301 (1978)), succinimidyl propionate and succinimidyl butanoate (see, e.g., Olson et al. in Poly(ethylene glycol) Chemistry & Biological Applications, pp 170-181, Harris & Zalipsky Eds., ACS, Washington, D.C., 1997; see also U.S. Pat. No. 5,672,662), succinimidyl succinate (See, e.g., Abuchowski et al. Cancer Biochem. Biophys. 7:175 (1984) and Joppich et al. Makromol. Chem. 180:1381 (1979), succinimidyl ester (see, e.g., U.S. Pat. No.

4,670,417), benzotriazole carbonate (see, e.g., U.S. Pat. No. 5,650,234), glycidyl ether (see, e.g., Pitha et al. Eur. J Biochem. 94:11 (1979), Elling et al., Biotech. Appl. Biochem. 13:354 (1991), oxycarbonylimidazole (see, e.g., Beauchamp, et al., Anal. Biochem. 131:25 (1983), Tondelli et al. J. Controlled Release 1:251 (1985)), p-nitrophenyl carbonate (see, e.g., Veronese, et al., Appl. Biochem. Biotech., 11: 141 (1985); and Sartore et al., Appl. Biochem. Biotech., 27:45 (1991)), aldehyde (see, e.g., Harris et al. J. Polym. Sci. Chem. Ed. 22:341 (1984), U.S. Pat. No. 5,824,784, U.S. Pat. No. 5,252,714), maleimide (see, e.g., Goodson et al. Bio/Technology 8:343 (1990), Romani et al. in Chemistry of Peptides and Proteins 2:29 (1984)), and Kogan, Synthetic Comm. 22:2417 (1992)), orthopyridyl-disulfide (see, e.g., Woghiren, et al. Bioconj. Chem. 4:314(1993)), acrylol (see, e.g., Sawhney et al., Macromolecules, 26:581 (1993)), vinylsulfone (see, e.g., U.S. Pat. No. 5,900,461).

[0419] In some cases, a PEG terminates on one end with hydroxy or methoxy, i.e., X is H or CH_3 ("methoxy PEG"). Alternatively, the PEG can terminate with a reactive group, thereby forming a bifunctional polymer. Typical reactive groups can include those reactive groups that are commonly used to react with the functional groups found in the 20 common amino acids (including maleimide groups, activated carbonates (including p-nitrophenyl ester), activated esters (including N-hydroxysuccinimide, p-nitrophenyl ester) and aldehydes) as well as functional groups that are inert to the 20 common amino acids but that react specifically with complementary functional groups present in non-natural amino acids (including oxime, carbonyl or dicarbonyl and hydroxylamine groups).

[0420] It is noted that the other end of the PEG, which is shown in the above formula by Y, will attach either directly or indirectly to a polypeptide (synthetic or natural), polynucleotide, polysaccharide or synthetic polymer via a non-natural amino acid. When Y is a hydroxylamine group, then the hydroxylamine-containing PEG reagent can react with a carbonyl- or dicarbonyl-containing non-natural amino acid in a polypeptide to form a PEG group linked to the polypeptide via an oxime linkage. When Y is a carbonyl or dicarbonyl group, then the carbonyl- or dicarbonyl-containing PEG reagent can react with a hydroxylamine-containing non-natural amino acid in a polypeptide to form a PEG group linked to the polypeptide via an oxime linkage. When Y is a carbonyl or dicarbonyl group, then the carbonyl- or dicarbonyl-containing PEG reagent also can react with an oxime-containing non-natural amino acid in a polypeptide to form a PEG group linked to the polypeptide via a new oxime linkage. Examples of appropriate reaction conditions, purification methods and reagents are described throughout this specification and the accompanying Figures. For example, FIG. 7 presents three examples of the reaction of a carbonyl-containing non-natural amino acid polypeptide with a hydroxylamine-containing PEG reagent to form an oxime-containing non-natural amino acid polypeptide linked to a PEG group. Further, FIG. 9 presents two examples of the reaction of an oxime-containing non-natural amino acid polypeptide with a carbonyl-containing PEG reagent to form a new oxime-containing non-natural amino acid polypeptide linked to a PEG group. And FIG. 10 presents one example of the reaction of a hydroxylamine-containing non-natural amino acid polypeptide with a carbonyl-containing PEG reagent to form an oxime-containing non-natural amino acid polypeptide linked to a PEG group.

[0421] By way of example only and not as a limitation on the types or classes of PEG reagents that may be used with the compositions, methods, techniques and strategies described herein, FIG. 11 presents further illustrative examples of hydroxylamine-containing PEG reagents that can react with carbonyl-containing non-natural amino acid polypeptides to form oxime-containing non-natural amino acid polypeptides linked to the PEG group, as well as examples of carbonyl-containing PEG reagents that can react with react with oxime-containing non-natural amino acid polypeptides or hydroxylamine-containing non-natural amino acid polypeptides to form new oxime-containing non-natural amino acid polypeptides linked to PEG groups. FIG. 12 presents four illustrative examples of synthetic methods for forming hydroxylamine-containing PEG reagents, or protected forms of hydroxylamine-containing PEG reagents, or masked forms of hydroxylamine-containing PEG reagents. FIG. 13 presents an illustrative example of synthetic methods for forming amide-linked hydroxylamine-containing PEG reagents, or protected forms of amide-linked hydroxylamine-containing PEG reagents, or masked forms of amide-linked hydroxylamine-containing PEG reagents. FIG. 14 and FIG. 15 present an illustrative examples of synthetic methods for forming carbamate-linked hydroxylamine-containing PEG reagents, or protected forms of carbamate-linked hydroxylamine-containing PEG reagents, or masked forms of carbamate-linked hydroxylamine-containing PEG reagents. FIG. 16 presents illustrative examples of synthetic methods for forming simple hydroxylamine-containing PEG reagents, or protected forms of simple hydroxylamine-containing PEG reagents, or masked forms of simple hydroxylamine-containing PEG reagents. Further, FIG. 17 presents illustrative examples of hydroxylamine-containing reagents that have multiple branches of linked PEG groups, and further shows the reaction of one such hydroxylamine-containing multi-PEG-branched reagents with a carbonyl-containing non-natural amino acid polypeptide to form an oxime-containing non-natural amino acid polypeptide with a linked multi-PEG-branched group.

[0422] Heterobifunctional derivatives are also particularly useful when it is desired to attach different molecules to each terminus of the polymer. For example, the omega-N-amino-N-azido PEG would allow the attachment of a molecule having an activated electrophilic group, such as an aldehyde, ketone, activated ester, activated carbonate and so forth, to one terminus of the PEG and a molecule having an acetylene group to the other terminus of the PEG.

[0423] In some aspects, a strong nucleophile (including but not limited to hydroxylamine) can be reacted with an aldehyde or ketone group present in a non-natural amino acid to form an oxime, which in some cases can be further reduced by treatment with an appropriate reducing agent. Alternatively, the strong nucleophile can be incorporated into the polypeptide via a non-natural amino acid and used to react preferentially with a ketone or aldehyde group present in the water soluble polymer. Generally, at least one terminus of the PEG molecule is available for reaction with the non-natural amino acid.

[0424] Thus, in some embodiments, the polypeptide comprising the non-natural amino acid is linked to a water soluble polymer, such as polyethylene glycol (PEG), via the side chain of the non-natural amino acid. The non-natural amino acid methods and compositions described herein provide a highly efficient method for the selective modification of proteins with PEG derivatives, which involves the selective incorporation of non-natural amino acids, those amino acids containing functional groups or substituents not found in the 20 naturally incorporated amino acids, into proteins in response to a selector codon and the subsequent modification of those amino acids with a suitably reactive PEG derivative. Known chemistry methodologies of a wide variety are suitable for use with the non-natural amino acid methods and compositions described herein to incorporate a water soluble polymer into the protein.

[0425] The polymer backbone can be linear or branched. Branched polymer backbones are generally known in the art. Typically, a branched polymer has a central branch core moiety and a plurality of linear polymer chains linked to the central branch core. PEG is used in branched forms that can be prepared by addition of ethylene oxide to various polyols, such as glycerol, glycerol oligomers, pentaerythritol and sorbitol. The central branch moiety can also be derived from several amino acids, such as lysine. The branched poly(ethylene glycol) can be represented in general form as $R(\text{-PEG-OH})_m$ in which R is derived from a core moiety, such as glycerol, glycerol oligomers, or pentaerythritol, and m represents the number of arms. Multi-armed PEG molecules, such as those described in U.S. Pat. Nos. 5,932,462 5,643,575; 5,229,490; 4,289,872; U.S. Pat. Appl. 2003/0143596; WO 96/21469; and WO 93/21259, can also be used as the polymer backbone.

[0426] Branched PEG can also be in the form of a forked PEG represented by $\text{PEG}(\text{-YCHZ}_2)_n$, where Y is a linking group and Z is an activated terminal group linked to CH by a chain of atoms of defined length. Yet another branched form, the pendant PEG, has reactive groups, such as carboxyl, along the PEG backbone rather than at the end of PEG chains.

[0427] In addition to these forms of PEG, the polymer can also be prepared with weak or degradable linkages in the backbone. For example, PEG can be prepared with ester linkages in the polymer backbone that are subject to hydrolysis. As shown herein, this hydrolysis results in cleavage of the polymer into fragments of lower molecular weight:

$$\text{-PEG-CO}_2\text{-PEG-}+\text{H}_2\text{O} \rightarrow \text{PEG-CO}_2\text{H}+\text{HO-PEG-}$$

It is understood by those skilled in the art that the term polyethylene glycol or PEG represents or includes all the forms known in the art including those disclosed herein. The molecular weight of the polymer may be of a wide range, including between about 100 Da and about 100,000 Da or more. The molecular weight of the polymer may be between about 100 Da and about 100,000 Da, including 100,000 Da, 95,000 Da, 90,000 Da, 85,000 Da, 80,000 Da, 75,000 Da, 70,000 Da, 65,000 Da, 60,000 Da, 55,000 Da, 50,000 Da, 45,000 Da, 40,000 Da, 35,000 Da, 30,000 Da, 25,000 Da, 20,000 Da, 15,000 Da, 10,000 Da, 9,000 Da, 8,000 Da, 7,000 Da, 6,000 Da, 5,000 Da, 4,000 Da, 3,000 Da, 2,000 Da, 1,000 Da, 900 Da, 800 Da, 700 Da, 600 Da, 500 Da, 400 Da, 300 Da, 200 Da, and 100 Da. In some embodiments, the molecular weight of the polymer is between about 100 Da and 50,000 Da. In some embodiments, the molecular weight of the polymer is between about 100 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 1,000 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 5,000 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 10,000 Da and 40,000 Da.

[0428] In order to maximize the desired properties of PEG, the total molecular weight and hydration state of the PEG polymer or polymers attached to the biologically active molecule must be sufficiently high to impart the advantageous characteristics typically associated with PEG polymer attachment, such as increased water solubility and circulating half life, while not adversely impacting the bioactivity of the parent molecule.

[0429] The methods and compositions described herein may be used to produce substantially homogenous preparations of polymer:protein conjugates. "Substantially homogenous" as used herein means that polymer:protein conjugate molecules are observed to be greater than half of the total protein. The polymer:protein conjugate has biological activity and the present "substantially homogenous" PEGylated polypeptide preparations provided herein are those which are homogenous enough to display the advantages of a homogenous preparation, e.g., ease in clinical application in predictability of lot to lot pharmacokinetics.

[0430] One may also choose to prepare a mixture of polymer:protein conjugate molecules, and the advantage provided herein is that one may select the proportion of mono-polymer:protein conjugate to include in the mixture. Thus, if desired, one may prepare a mixture of various proteins with various numbers of polymer moieties attached (i.e., di-, tri-, tetra-, etc.) and combine said conjugates with the mono-polymer:protein conjugate prepared using the methods described herein, and have a mixture with a predetermined proportion of mono-polymer:protein conjugates.

[0431] The proportion of polyethylene glycol molecules to protein molecules will vary, as will their concentrations in the reaction mixture. In general, the optimum ratio (in terms of efficiency of reaction in that there is minimal excess unreacted protein or polymer) may be determined by the molecular weight of the polyethylene glycol selected and on the number of available reactive groups available. As relates to molecular weight, typically the higher the molecular weight of the polymer, the fewer number of polymer molecules which may be attached to the protein. Similarly, branching of the polymer should be taken into account when optimizing these parameters. Generally, the higher the molecular weight (or the more branches) the higher the polymer:protein ratio.

[0432] As used herein, and when contemplating hydrophilic polymer:polypeptide/protein conjugates, the term "therapeutically effective amount" further refers to an amount which gives an increase in desired benefit to a patient. The amount will vary from one individual to another and will depend upon a number of factors, including the overall physical condition of the patient and the underlying cause of the disease, disorder or condition to be treated. A therapeutically effective amount of the present compositions may be readily ascertained by one skilled in the art using publicly available materials and procedures.

[0433] The number of water soluble polymers linked to a "modified or unmodified" non-natural amino acid polypeptide (i.e., the extent of PEGylation or glycosylation) described herein can be adjusted to provide an altered (including increased or decreased) pharmacologic, pharmacokinetic or pharmacodynamic characteristic such *as in vivo* half-life. In some aspects, the half-life of the polypeptide is increased at least about 10, 20, 30, 40, 50, 60, 70, 80, 90 percent, two fold, five-fold, 10-fold, 50-fold, or at least about 100-fold over an unmodified polypeptide.

[0434] In one embodiment, a polypeptide comprising a carbonyl- or dicarbonyl-containing non-natural amino acid is modified with a PEG derivative that contains a terminal hydroxylamine moiety that is linked directly to the PEG backbone.

[0435] In some aspects, the hydroxylamine-terminal PEG derivative will have the structure:

$$RO\text{-}(CH_2CH_2O)_n\text{-}O\text{-}(CH_2)_m\text{-}O\text{-}NH_2$$

where R is a simple alkyl (methyl, ethyl, propyl, etc.), m is 2-10 and n is 100-1,000 (i.e., average molecular weight is between 5-40 kDa). The molecular weight of the polymer may be of a wide range, including between about 100 Da and about 100,000 Da or more. The molecular weight of the polymer may be between about 100 Da and about 100,000 Da, including 100,000 Da, 95,000 Da, 90,000 Da, 85,000 Da, 80,000 Da, 75,000 Da, 70,000 Da, 65,000 Da, 60,000 Da, 55,000 Da, 50,000 Da, 45,000 Da, 40,000 Da, 35,000 Da, 30,000 Da, 25,000 Da, 20,000 Da, 15,000 Da, 10,000 Da, 9,000 Da, 8,000 Da, 7,000 Da, 6,000 Da, 5,000 Da, 4,000 Da, 3,000 Da, 2,000 Da, 1,000 Da, 900 Da, 800 Da, 700 Da, 600 Da, 500 Da, 400 Da, 300 Da, 200 Da, and 100 Da. In some embodiments, the molecular weight of the polymer is between about 100 Da and 50,000 Da. In some embodiments, the molecular weight of the polymer is between about 100 Da and 40,000 Da. In some embodiments, the

[0436] molecular weight of the polymer is between about 1,000 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 5,000 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 10,000 Da and 40,000 Da.

[0437] In another embodiment, a polypeptide comprising a carbonyl- or dicarbonyl-containing amino acid is modified with a PEG derivative that contains a terminal hydroxylamine moiety that is linked to the PEG backbone by means of an amide linkage.

[0438] In some aspects, the hydroxylamine-terminal PEG derivatives have the structure:

$$RO\text{-}(CH_2CH_2O)_n\text{-}O\text{-}(CH_2)_2\text{-}NH\text{-}C(O)(CH_2)_m\text{-}O\text{-}NH_2$$

where R is a simple alkyl (methyl, ethyl, propyl, etc.), m is 2-10 and n is 100-1,000 (i.e., average molecular weight is between 5-40 kDa). The molecular weight of the polymer may be of a wide range, including between about 100 Da and about 100,000 Da or more. The molecular weight of the polymer may be between about 100 Da and about 100,000 Da, including 100,000 Da, 95,000

Da, 90,000 Da, 85,000 Da, 80,000 Da, 75,000 Da, 70,000 Da, 65,000 Da, 60,000 Da, 55,000 Da, 50,000 Da, 45,000 Da, 40,000 Da, 35,000 Da, 30,000 Da, 25,000 Da, 20,000 Da, 15,000 Da, 10,000 Da, 9,000 Da, 8,000 Da, 7,000 Da, 6,000 Da, 5,000 Da, 4,000 Da, 3,000 Da, 2,000 Da, 1,000 Da, 900 Da, 800 Da, 700 Da, 600 Da, 500 Da, 400 Da, 300 Da, 200 Da, and 100 Da. In some embodiments, the molecular weight of the polymer is between about 100 Da and 50,000 Da. In some embodiments, the molecular weight of the polymer is between about 100 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 1,000 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 5,000 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 10,000 Da and 40,000 Da.

[0439] In another embodiment, a polypeptide comprising a carbonyl- or dicarbonyl-containing amino acid is modified with a branched PEG derivative that contains a terminal hydroxylamine moiety, with each chain of the branched PEG having a MW ranging from 10-40 kDa and, in other embodiments, from 5-20 kDa. The molecular weight of the branched polymer may be of a wide range, including but not limited to, between about 100 Da and about 100,000 Da or more. The molecular weight of the polymer may be between about 100 Da and about 100,000 Da, including but not limited to, 100,000 Da, 95,000 Da, 90,000 Da, 85,000 Da, 80,000 Da, 75,000 Da, 70,000 Da, 65,000 Da, 60,000 Da, 55,000 Da, 50,000 Da, 45,000 Da, 40,000 Da, 35,000 Da, 30,000 Da, 25,000 Da, 20,000 Da, 15,000 Da, 10,000 Da, 9,000 Da, 8,000 Da, 7,000 Da, 6,000 Da, 5,000 Da, 4,000 Da, 3,000 Da, 2,000 Da, 1,000 Da, 900 Da, 800 Da, 700 Da, 600 Da, 500 Da, 400 Da, 300 Da, 200 Da, and 100 Da. In some embodiments, the molecular weight of the polymer is between about 100 Da and 50,000 Da. In some embodiments, the molecular weight of the polymer is between about 100 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 1,000 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 5,000 Da and 40,000 Da. In some embodiments, the molecular weight of the polymer is between about 10,000 Da and 40,000 Da.

[0440] In another embodiment, a polypeptide comprising a non-natural amino acid is modified with at least one PEG derivative having a branched structure. In some embodiments, the PEG derivatives containing a hydroxylamine group will have the structure:

$$[RO\text{-}(CH_2CH_2O)_n\text{-}O\text{-}(CH_2)_2\text{-}C(O)\text{-}NH\text{-}CH_2\text{-}CH_2]_2CH\text{-}X\text{-}(CH_2)_m\text{-}O\text{-}NH_2$$

where R is a simple alkyl (methyl, ethyl, propyl, etc.), X is optionally NH, O, S, C(O) or not present, m is 2-10 and n is 100-1,000. The molecular weight of the polymer may be between about 1,000 Da and about 100,000 Da, including 100,000 Da, 95,000 Da, 90,000 Da, 85,000 Da, 80,000 Da, 75,000 Da, 70,000 Da, 65,000 Da, 60,000 Da, 55,000 Da, 50,000 Da, 45,000 Da, 40,000 Da, 35,000 Da, 30,000 Da, 25,000 Da, 20,000 Da, 15,000 Da, 10,000 Da, 9,000 Da, 8,000 Da, 7,000 Da, 6,000 Da, 5,000 Da, 4,000 Da, 3,000 Da, 2,000 Da, and 1,000 Da. In some embodiments, the molecular weight of the branched chain PEG is between about 1,000 Da and 50,000 Da. In some embodiments, the molecular weight of the branched chain PEG is between about 1,000 Da and 40,000 Da. In some embodiments, the molecular weight of the branched chain PEG is between about 5,000 Da and 40,000 Da. In some embodiments, the molecular weight of the branched chain PEG is between about 5,000 Da and 20,000 Da.

[0441] Several reviews and monographs on the functionalization and conjugation of PEG are available. *See*, for example, Harris, Macromol. Chem. Phys. C25: 325-373 (1985); Scouten, Methods in Enzymology 135: 30-65 (1987); Wong et al., Enzyme Microb. Technol. 14: 866-874 (1992); Delgado et al., Critical Reviews in Therapeutic Drug Carrier Systems 9: 249-304 (1992); Zalipsky, Bioconjugate Chem. 6:150-165 (1995).

[0442] Methods for activation of polymers can also be found in WO 94/17039, U.S. Pat. No. 5,324,844, WO 94/18247, WO 94/04193, U.S. Pat. No. 5,219,564, U.S. Pat. No. 5,122,614, WO 90/13540, U.S. Pat. No. 5,281,698, and more WO 93/15189, and for conjugation between activated polymers and enzymes including Coagulation Factor VIII (WO 94/15625), haemoglobin (WO 94/09027), oxygen carrying molecule (U.S. Pat. No. 4,412,989), ribonuclease and superoxide dismutase (Veronese at al., App. Biochem. Biotech. 11: 141-152 (1985)).

[0443] If necessary, the PEGylated non-natural amino acid polypeptides described herein obtained from the hydrophobic chromatography can be purified further by one or more procedures known to those skilled in the art including affinity chromatography; anion- or cation-exchange chromatography (using, including DEAE SEPHAROSE); chromatography on silica; reverse phase HPLC; gel filtration (using, including SEPHADEX G-75); hydrophobic interaction chromatography; size-exclusion chromatography, metal-chelate chromatography; ultrafiltration/diafiltration; ethanol precipitation; ammonium sulfate precipitation; chromatofocusing; displacement chromatography; electrophoretic procedures (including preparative isoelectric focusing), differential solubility (including ammonium sulfate precipitation), or extraction. Apparent molecular weight may be estimated by GPC by comparison to globular protein standards (Preneta AZ, PROTEIN PURIFICATION METHODS, A PRACTICAL APPROACH (Harris & Angal, Eds.) IRL Press 1989, 293-306). The purity of the non-natural amino acid polypeptide:PEG conjugate can be assessed by proteolytic degradation (including trypsin cleavage) followed by mass spectrometry analysis. Pepinsky RB., et.al., J.

Pharmacol. & Exp. Ther. 297(3):1059-66 (2001).

[0444] A water soluble polymer linked to a non-natural amino acid of a polypeptide described herein can be further derivatized or substituted without limitation.

E. Enhancing affinity for serum albumin

[0445] Various molecules can also be fused to the non-natural amino acid polypeptides described herein to modulate the half-life in serum. In some aspects, molecules are linked or fused to the "modified or unmodified" non-natural amino acid polypeptides described herein to enhance affinity for endogenous serum albumin in an animal.

[0446] For example, in some cases, a recombinant fusion of a polypeptide and an albumin binding sequence is made. Exemplary albumin binding sequences include the albumin binding domain from streptococcal protein G (*see. e.g.,* Makrides et al., J. Pharmacol. Exp. Ther. 277(1):534-542 (1996) and Sjolander et al., J, Immunol. Methods 201:115-123 (1997)), or albumin-binding peptides such as those described in, e.g., Dennis, et al., J. Biol. Chem. 277(38):35035-35043 (2002).

[0447] In other aspects, the "modified or unmodified" non-natural amino acid polypeptides described herein are acylated with fatty acids. In some cases, the fatty acids promote binding to serum albumin. *See, e.g.,* Kurtzhals, et al., Biochem. J. 312:725-731 (1995).

[0448] In other aspects, the "modified or unmodified" non-natural amino acid polypeptides described herein are fused directly with serum albumin (including human serum albumin). Those of skill in the art will recognize that a wide variety of other molecules can also be linked to non-natural amino acid polypeptides, modified or unmodified, as described herein, to modulate binding to serum albumin or other serum components.

F. Glycosylation of non-natural amino acid polypeptides described herein

[0449] The methods and compositions described herein include polypeptides incorporating one or more non-natural amino acids bearing saccharide residues. The saccharide residues may be either natural (including N-acetylglucosamine) or non-natural (including 3-fluorogalactose). The saccharides may be linked to the non-natural amino acids either by an N- or O-linked glycosidic linkage (including N-acetylgalactose-L-serine) or a non-natural linkage (including an oxime or the corresponding C- or S-linked glycoside).

[0450] The saccharide (including glycosyl) moieties can be added to the non-natural amino acid polypeptides either *in vivo* or *in vitro.* In some aspects, a polypeptide comprising a carbonyl- or dicarbonyl-containing non-natural amino acid is modified with a saccharide derivatized with an aminooxy group to generate the corresponding glycosylated polypeptide linked via an oxime linkage. Once attached to the non-natural amino acid, the saccharide may be further elaborated by treatment with glycosyltransferases and other enzymes to generate an oligosaccharide bound to the non-natural amino acid polypeptide. *See, e.g.,* H. Liu, et al. J. Am. Chem. Soc. 125: 1702-1703 (2003).

G. Use of Linking Groups and Applications, Including Polypeptide Dimers and Multimers

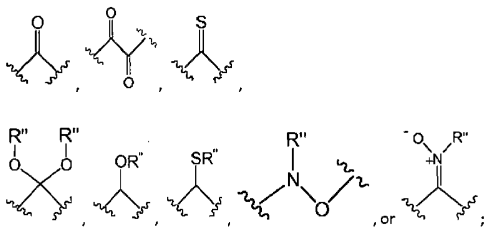
[0451] In addition to adding functionality directly to the non-natural amino acid polypeptide, the non-natural amino acid portion of the polypeptide may first be modified with a multifunctional (e.g., bi-, tri, tetra-) linker molecule that then subsequently is further modified. That is, at least one end of the multifunctional linker molecule reacts with at least one non-natural amino acid in a polypeptide and at least one other end of the multifunctional linker is available for further functionalization. If all ends of the multifunctional linker are identical, then (depending upon the stoichiometric conditions) homomultimers of the non-natural amino acid polypeptide may be formed. If the ends of the multifunctional linker have distinct chemical reactivities, then at least one end of the multifunctional linker group will be bound to the non-natural amino acid polypeptide and the other end can subsequently react with a different functionality, including by way of example only: a label; a dye; a polymer; a water-soluble polymer; a derivative of polyethylene glycol; a photocrosslinker; a cytotoxic compound; a drug; an affinity label; a photoaffinity label; a reactive compound; a resin; a second protein or polypeptide or polypeptide analog; an antibody or antibody fragment; a metal chelator; a cofactor; a fatty acid; a carbohydrate; a polynucleotide; a DNA; a RNA; an antisense polynucleotide; a saccharide, a water-soluble dendrimer, a cyclodextrin, a biomaterial; a nanoparticle; a spin label; a fluorophore, a metal-containing moiety; a radioactive moiety; a novel functional group; a group that covalently or noncovalently interacts with other molecules; a photocaged moiety; an actinic radiation excitable moiety; a ligand; a photoisomerizable moiety; biotin; a biotin analogue; a moiety incorporating a heavy atom; a chemically cleavable group; a photocleavable group; an elongated side chain; a carbon-linked sugar; a redox-active agent; an amino thioacid; a toxic moiety; an isotopically labeled moiety; a biophysical probe; a phosphorescent group; a chemiluminescent group; an electron dense group; a magnetic group; an intercalating group; a chromophore; an energy transfer agent; a biologically active agent; a detectable label; a small molecule; an inhibitory ribonucleic acid, a radionucleotide; a neutron-capture agent; a derivative of biotin; quantum dot(s); a nanotransmitter; a radiotransmitter, an abzyme, an activated complex activator, a virus, an adjuvant, an aglycan, an allergan, an angiostatin, an antihormone, an antioxidant, an aptamer, a guide RNA, a saponin, a shuttle vector, a macromolecule, a mimotope, a receptor, a reverse micelle, and any combination thereof.

[0452] The multifunctional linker group has the general structure:

$$[X-L]_n-L_1-W \quad \text{(XIX)}$$

wherein:

each X is independently NH_2, $-C(=O)R_9$, -SR' or-J-R, where R_9 is H or OR', where J is



R is H, alkyl, substituted alkyl, cycloalkyl, or substituted cycloalkyl; each R" is independently H, alkyl, substituted alkyl, or a protecting group, or when more than one R" group is present, two R" optionally form a heterocycloalkyl;

each R' is independently H, alkyl, or substituted alkyl;

each L is independently selected from the group consisting of alkylene, substituted alkylene, alkenylene, substituted alkenylene, -O-, -O-(alkylene or substituted alkylene)-, -S-, -S-(alkylene or substituted alkylene)-, $-S(O)_k-$where k is 1, 2, or 3, $-S(O)_k$(alkylene or substituted alkylene)-, -C(O)-, -C(O)-(alkylene or substituted alkylene)-, -C(S)-, -C(S)-(alkylene or substituted alkylene)-, -N(R')-, -NR'-(alkylene or substituted alkylene)-, -C(O)N(R')-, -CON(R')-(alkylene or substituted alkylene)-, -(alkylene or substituted alkylene)NR'C(O)O-(alkylene or substituted alkylene)-, -O-CON(R')-(alkylene or substituted alkylene)-, -CSN(R')-, -CSN(R')-(alkylene or substituted alkylene)-, -N(R')CO-(alkylene or substituted $\text{alkylene})_7$, -N(R')C(O)O-, -N(R')C(O)O-(alkylene or substituted alkylene)-, $-S(O)_kN(R')-$, -N(R')C(O)N(R')-, -N(R')C(O)N(R')-(alkylene or substituted alkylene)-, -N(R')C(S)N(R')-, $-N(R')S(O)_kN(R')-$, -N(R')-N=, -C(R')=N-, -C(R')=N-N(R')-, -C(R')=N-N=, $-C(R')_2-N=N-$, and $-C(R')_2-N(R')-N(R')-$;

L_1 is optional, and when present, is $-C(R')_p-NR'-C(O)O-$(alkylene or substituted alkylene)- where p is 0, 1, or 2; W is NH_2, $-C(=O)R_9$, -SR' or-J-R; and n is 1 to 3

provided that X and $L-L_1-W$ together independently each provide at least one of the following (a) a hydroxylamine group capable of reacting with a carbonyl (including a dicarbonyl) group on a non-natural amino acid or a "modified or unmodified" non-natural amino acid polypeptide; (b) a carbonyl group (including a dicarbonyl group) capable of reacting with an hydroxylamine group on a non-natural amino acid or a "modified or unmodified" non-natural amino acid polypeptide; or (c) a carbonyl group (including a dicarbonyl group) capable of undergoing an exchange reaction with an oxime group on a non-natural amino acid or a "modified or unmodified" non-natural amino acid polypeptide.

[0453] FIG. 18 presents an illustrative, example of the synthesis of a bifunctional homolinker in which the linker has two identical ends, i.e., hydroxylamine groups. Such a linker may be used to form a homodimer of a carbonyl- or dicarbonyl-containing non-natural amino acid polypeptide to form two oxime linkages. Alternatively, if one end of such a linker is protected, then such a partially protected linker can be used to bind the unprotected hydroxylamine end to a carbonyl- or dicarbonyl-containing non-natural amino acid polypeptide via an oxime linkage, leaving the other protected end available for further linking reactions following deprotection. Alternatively, careful manipulation of the stoichiometry of the reagents may provide a similar result (a heterodimer), albeit a result in which the desired heterodimer will likely be contaminated with some homodimer.

[0454] FIG. 19 presents illustrative, examples of two multifunctional heterolinkers in which each linker has more than one type of terminal reactive group, i.e., hydroxylamine, oxime and thioester groups. Such a linker may be used to form a heterodimer of a non-natural amino acid polypeptide using the oxime-based chemistry discussed throughout this specification.

[0455] FIG. 20 presents a schematic illustrative, example of the use of a heterobifunctional linker to attach a PEG group to a non-natural amino acid polypeptide in a multi-step synthesis. In the first step, as depicted in this illustrative figure, a carbonyl-containing non-natural amino acid polypeptide reacts with a hydroxylamine-containing bifunctional linker to form a modified oxime-containing non-natural amino acid polypeptide. However, the bifunctional linker still retains a functional group (here illustrated by a shaped object) that is capable of reacting with a reagent with appropriate reactivity (illustrated in the figure by a matching shaped object) to form a modified oxime-containing functionalized non-natural amino acid polypeptide. In this particular illustrative figure, the functionalization is a PEG group, but may also include any of the aforementioned functionalities, or in this case of a tri- or tetra-functional linker, more than one type of functionality or multiple types of the same functionality. FIG. 21 presents illustrative examples of four types of linker groups used to link a hydroxylamine-containing non-natural amino acid polypeptide to a PEG group. As before, the PEG functionality is provided for illustrative purposes only. Thus, the linker groups described herein provide an additional means to further modify a non-natural amino acid polypeptide in a site-selective fashion.

[0456] The methods and compositions described herein also provide for polypeptide combinations, such as homodimers, heterodimers, homomultimers, or heteromultimers (i.e., trimers, tetramers, etc.). By way of example only, the following description focuses on the GH supergene family members, however, the methods, techniques and compositions described in this section can be applied to virtually any other polypeptide which can provide benefit in the form of dimers and multimers, including by way of example only: alpha-1 antitrypsin, angiostatin, antihemolytic factor, antibody, apolipoprotein, apoprotein, atrial natriuretic factor, atrial natriuretic polypeptide, atrial peptide, C-X-C chemokine, T39765, NAP-2, ENA-78, gro-a, gro-b, gro-c, IP-10, GCP-2, NAP-4, SDF-1, PF4, MIG, calcitonin, c-kit ligand, cytokine, CC chemokine, monocyte chemoattractant protein-1, monocyte chemoattractant protein-2, monocyte chemoattractant protein-3, monocyte inflammatory protein-1 alpha, monocyte inflammatory protein-i beta, RANTES, 1309, R83915, R91733, HCC1, T58847, D31065, T64262, CD40, CD40 ligand, c-kit ligand, collagen, colony stimulating factor (CSF), complement factor 5a, complement inhibitor, complement receptor 1, cytokine, epithelial neutrophil activating peptide-78, MIP-16, MCP-1, epidermal growth factor (EGF), epithelial neutrophil activating peptide, erythropoietin (EPO), exfoliating toxin, Factor IX, Factor VII, Factor VIII, Factor X, fibroblast growth factor (FGF), fibrinogen, fibronectin, four-helical bundle protein, G-CSF, glp-1, GM-CSF, glucocerebrosidase, gonadotropin, growth factor, growth factor receptor, grf, hedgehog protein, hemoglobin, hepatocyte growth factor (hGF), hirudin, human growth hormone (hGH), human serum albumin, ICAM-1, ICAM-1 receptor, LFA-1, LFA-1 receptor, insulin, insulin-like growth factor (IGF), IGF-I, IGF-II, interferon (IFN), IFN-alpha, IFN-beta, IFN-gamma, interleukin (IL), IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, keratinocyte growth factor (KGF), lactoferrin, leukemia inhibitory factor, luciferase, neurturin, neutrophil inhibitory factor (NIF), oncostatin M, osteogenic protein, oncogene product, paracitonin, parathyroid hormone, PD-ECSF, PDGF, peptide hormone, pleiotropin, protein A, protein G, pth, pyrogenic exotoxin A, pyrogenic exotoxin B, pyrogenic exotoxin C, pyy, relaxin, renin, SCF, small biosynthetic protein, soluble complement receptor I, soluble I-CAM 1, soluble interleukin receptor, soluble TNF receptor, somatomedin, somatostatin, somatotropin, streptokinase, superantigens, staphylococcal enterotoxin, SEA, SEB, SEC1, SEC2, SEC3, SED, SEE, steroid hormone receptor, superoxide dismutase, toxic shock syndrome toxin, thymosin alpha 1, tissue plasminogen activator, tumor growth factor (TGF), tumor necrosis factor, tumor necrosis factor alpha, tumor necrosis factor beta, tumor necrosis factor receptor (TNFR), VLA-4 protein, VCAM-1 protein, vascular endothelial growth factor (VEGF), urokinase, mos, ras, raf, met, p53, tat, fos, myc, jun, myb, rel, estrogen receptor, progesterone receptor, testosterone receptor, aldosterone receptor, LDL receptor, and corticosterone.

[0457] Thus, encompassed within the methods, techniques and compositions described herein are a GH supergene family member polypeptide containing one or more non-natural amino acids bound to another GH supergene family member or variant thereof or any other polypeptide that is a non-GH supergene family member or variant thereof, either directly to the polypeptide backbone or via a linker. Due to its increased molecular weight compared to monomers, the GH supergene family member dimer or multimer conjugates may exhibit new or desirable properties, including different pharmacological, pharmacokinetic, pharmacodynamic, modulated therapeutic half-life, or modulated plasma half-life relative to the monomeric GH supergene family member. In some aspects, the GH supergene family member dimers described herein will modulate the dimerization of the GH supergene family member receptor. In other aspects, the GH supergene family member dimers or multimers described herein will act as a GH supergene family member receptor antagonist, agonist, or modulator.

[0458] In some aspects, the GH supergene family member polypeptides are linked directly, including via an Asn-Lys amide linkage or Cys-Cys disulfide linkage. In some aspects, the linked GH supergene family member polypeptides, and/or the linked non-GH supergene family member, will comprise different non-natural amino acids to facilitate dimerization, including a first GH supergene family member, and/or the linked non-GH supergene family member, polypeptide comprising a ketone-containing non-natural amino acid conjugated to a second GH supergene family member polypeptide comprising a hydroxylamine-containing non-natural amino acid and the polypeptides are reacted via formation of the corresponding oxime.

[0459] Alternatively, the two GH supergene family member polypeptides, and/or the linked non-GH supergene family member, are linked via a linker. Any hetero- or homo-bifunctional linker can be used to link the two GH supergene family member, and/or the linked non-GH supergene family member, polypeptides, which can have the same or different primary sequence. In some cases, the linker used to tether the GH supergene family member, and/or the linked non-GH supergene family member, polypeptides together can be a bifunctional PEG reagent.

[0460] In some aspects, the methods and compositions described herein provide for water-soluble bifunctional linkers that have a dumbbell structure that includes: a) an azide, an alkyne, a hydrazine, a hydrazide, a hydroxylamine, or a carbonyl- or dicarbonyl-containing moiety on at least a first end of a polymer backbone; and b) at least a second functional group on a second end of the polymer backbone. The second functional group can be the same or different as the first functional group. The second functional group, in some aspects, is not reactive with the first functional group. The methods and compositions described herein provide, in some aspects, water-soluble compounds that comprise at least one arm of a branched molecular structure. For example, the branched molecular structure can be dendritic.

[0461] In some aspects, the methods and compositions described herein provide multimers comprising one or more GH supergene family member formed by reactions with water soluble activated polymers that have the structure:

$$R\text{-}(CH_2CH_2O)_n\text{-}O\text{-}(CH_2)_m\text{-}X$$

wherein n is from about 5 to 3,000, m is 2-10, X can be an azide, an alkyne, a hydrazine, a hydrazide, an aminooxy group, a hydroxylamine, a acetyl, or carbonyl- or dicarbonyl-containing moiety, and R is a capping group, a functional group, or a leaving group that can be the same or different as X. R can be, for example, a functional group selected from the group consisting of hydroxyl, protected hydroxyl, alkoxyl, N-hydroxysuccinimidyl ester, 1-benzotriazolyl ester, N-hydroxysuccinimidyl carbonate, 1-benzotriazolyl carbonate, acetal, aldehyde, aldehyde hydrates, alkenyl, acrylate, methacrylate, acrylamide, active sulfone, amine, aminooxy, protected amine, hydrazide, protected hydrazide, protected thiol, carboxylic acid, protected carboxylic acid, isocyanate, isothiocyanate, maleimide, vinylsulfone, dithiopyridine, vinylpyridine, iodoacetamide, epoxide, glyoxals, diones, mesylates, tosylates, and tresylate, alkene, and ketone.

[0462] FIG. 22 presents an illustrative, example of the use of a linker group described herein to form a homodimer of a non-natural amino acid polypeptide described herein. In this illustrative example, a carbonyl-containing non-natural amino acid polypeptide is reacted with a bifunctional linker group having two available hydroxylamine groups under conditions suitable for the formation of linked homodimer oxime-containing non-natural amino acid polypeptides. The method presented in the figure is not limited to carbonyl-containing non-natural amino acid polypeptides coupled to hydroxylamine-containing bifunctional linkers. The non-natural amino acid polypeptide may further contain an oxime group that is capable of undergoing an exchange reaction with a carbonyl-containing multifunctional linker group to form a homomultimer linked by a structure containing multiple oxime groups, or the non-natural amino acid polypeptide may further contain a hydroxylamine group that is capable of undergoing a reaction with a carbonyl-containing multifunctional linker group to form a homomultimer linked by a structure containing multiple oxime groups. Of course, the homomultimer may be a homodimer, a homotrimer or a homotetramer.

[0463] FIG. 23 presents an illustrative, example of the use of a heterofunctional linker group to form a heterodimer of polypeptides, in which at least one of the members of the heterodimer is a non-natural amino acid polypeptide described herein and the other members are optionally non-natural amino acid polypeptides as described herein, other types of non-natural amino acid polypeptides, or naturally-occurring amino acid polypeptides. In the example presented in this figure, the linker group contains two identical hydroxylamine groups, by controlling the stoichiometry, temperature and other parameters of the reaction, the dominant product from the reaction of the linker with a carbonyl-containing non-natural amino acid polypeptide is a modified oxime-containing non-natural amino acid polypeptide attached to a linker with an available hydroxylamine group. This latter group can further react with another carbonyl or dicarbonyl containing non-natural amino acid polypeptide to form a bifunctional heterodimer of oxime-containing non-natural amino acid polypeptides. Of course, the functional groups on the linker do not have to be identical, nor do they have to be hydroxylamine groups. Using the chemistry detailed throughout this specification, one of skill in the art could design a linker in which at least one functional group can form an oxime group with a non-natural amino acid polypeptide; the other functional groups on the linker could utilize other known chemistry, including the nucleophile/electrophile based chemistry well known in the art of organic chemistry.

H. Example of Adding Functionality: Easing the Isolation Properties of a Polypeptide

[0464] A naturally-occurring or non-natural amino acid polypeptide may be difficult to isolate from a sample for a number of reasons, including the solubility or binding characteristics of the polypeptide. For example, in the preparation of a polypeptide for therapeutic use, such a polypeptide may be isolated from a recombinant system that has been engineered to overproduce the polypeptide. However, because of the solubility or binding characteristics of the polypeptide, achieving a desired level of purity often proves difficult. The methods, compositions, techniques and strategies described herein provide a solution to this situation.

[0465] Using the methods, compositions, techniques and strategies described herein, one of skill in the art can produce an oxime-containing non-natural amino acid polypeptide that is homologous to the desired polypeptide, wherein the oxime-containing non-natural amino acid polypeptide has improved isolation characteristics. In one embodiment, a homologous non-natural amino acid polypeptide is produced biosynthetically. In a further or additional embodiment, the non-natural amino acid has incorporated into its structure one of the non-natural amino acids described herein. In a further or additional embodiment, the non-natural amino acid is incorporated at a terminal or internal position and is further incorporated site specifically.

[0466] In one embodiment, the resulting non-natural amino acid, as produced biosynthetically, already has the desired improved isolation characteristics. In further or additional embodiments, the non-natural amino acid comprises an oxime linkage to a group that provides the improved isolation characteristics. In further or additional embodiments, the non-natural amino acid is further modified to form a modified oxime-containing non-natural amino acid polypeptide, wherein the modification provides an oxime linkage to a group that provides the improved isolation characteristics. In some embodiments, such a group is directly linked to the non-natural amino acid, and in other embodiments, such a group is linked via a linker group to the non-natural amino acid. In certain embodiments, such a group is connected to the non-natural amino acid by a single chemical reaction, in other embodiments a series of chemical reactions is required to connect such a group to the non-natural amino acid. Preferably, the group imparting improved isolation characteristics is linked site specifically to the non-natural amino acid in the non-natural amino acid polypeptide and is not linked to a naturally occurring amino acid under the reaction conditions utilized.

[0467] In further or additional aspects the resulting non-natural amino acid polypeptide is homologous to the GH supergene family members, however, the methods, techniques and compositions described in this section can be applied to virtually any other polypeptide which can benefit from improved isolation characteristics, including by way of example only: alpha-1 antitrypsin, angiostatin, antihemolytic factor, antibody, apolipoprotein, apoprotein, atrial natriuretic factor, atrial natriuretic polypeptide, atrial peptide, C-X-C chemokine, T39765, NAP-2, ENA-78, gro-a, gro-b, gro-c, IP-10, GCP-2, NAP-4, SDF-1, PF4, MIG, calcitonin, c-kit ligand, cytokine, CC chemokine, monocyte chemoattractant protein-1, monocyte chemoattractant protein-2, monocyte chemoattractant protein-3, monocyte inflammatory protein-1 alpha, monocyte inflammatory protein-i beta, RANTES, 1309, R83915, R91733, HCC1, T58847, D31065, T64262, CD40, CD40 ligand, c-kit ligand, collagen, colony stimulating factor (CSF), complement factor 5a, complement inhibitor, complement receptor 1, cytokine, epithelial neutrophil activating peptide-78, MIP-16, MCP-1, epidermal growth factor (EGF), epithelial neutrophil activating peptide, erythropoietin (EPO), exfoliating toxin, Factor IX, Factor VII, Factor VIII, Factor X, fibroblast growth factor (FGF), fibrinogen, fibronectin, four-helical bundle protein, G-CSF, glp-1, GM-CSF, glucocerebrosidase, gonadotropin, growth factor, growth factor receptor, grf, hedgehog protein, hemoglobin, hepatocyte growth factor (hGF), hirudin, human growth hormone (hGH), human serum albumin, ICAM-1, ICAM-1 receptor, LFA-1, LFA-1 receptor, insulin, insulin-like growth factor (IGF), IGF-I, IGF-II, interferon (IFN), IFN-alpha, IFN-beta, IFN-gamma, interleukin (IL), IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, keratinocyte growth factor (KGF), lactoferrin, leukemia inhibitory factor, luciferase, neurturin, neutrophil inhibitory factor (NIF), oncostatin M, osteogenic protein, oncogene product, paracitonin, parathyroid hormone, PD-ECSF, PDGF, peptide hormone, pleiotropin, protein A, protein G, pth, pyrogenic exotoxin A, pyrogenic exotoxin B, pyrogenic exotoxin C, pyy, relaxin, renin, SCF, small biosynthetic protein, soluble complement receptor I, soluble I-CAM 1, soluble interleukin receptor, soluble TNF receptor, somatomedin, somatostatin, somatotropin, streptokinase, superantigens, staphylococcal enterotoxin, SEA, SEB, SEC1, SEC2, SEC3, SED, SEE, steroid hormone receptor, superoxide dismutase, toxic shock syndrome toxin, thymosin alpha 1, tissue plasminogen activator, tumor growth factor (TGF), tumor necrosis factor, tumor necrosis factor alpha, tumor necrosis factor beta, tumor necrosis factor receptor (TNFR), VLA-4 protein, VCAM-1 protein, vascular endothelial growth factor (VEGF), urokinase, mos, ras, raf, met, p53, tat, fos, myc, jun, myb, rel, estrogen receptor, progesterone receptor, testosterone receptor, aldosterone receptor, LDL receptor, and corticosterone.

[0468] In further or additional aspects, the group imparting improved isolation characteristics improves the water solubility of the polypeptide; in other aspects, the group improves the binding properties of the polypeptide; in other aspects, the group provides new binding properties to the polypeptide (including, by way of example only, a biotin group or a biotin-binding group). In aspects wherein the group improves the water solubility of the polypeptide, the group is selected from the water soluble polymers described herein, including by way of example only, any of the PEG polymer groups described herein.

A. Example of Adding Functionality: Detecting the Presence of a Polypeptide

[0469] A naturally-occurring or non-natural amino acid polypeptide may be difficult to detect in a sample (including an *in vivo* sample and an *in vitro* sample) for a number of reasons, including the lack of a reagent or label that can readily bind to the polypeptide. The methods, compositions, techniques and strategies described herein provide a solution to this situation.

[0470] Using the methods, compositions, techniques and strategies described herein, one of skill in the art can produce an oxime-containing non-natural amino acid polypeptide that is homologous to the desired polypeptide, wherein the oxime-containing non-natural amino acid polypeptide allows the detection of the polypeptide in an *in vivo* sample and an *in vitro* sample. In one embodiment, a homologous non-natural amino acid polypeptide is produced biosynthetically. In a further or additional embodiment, the non-natural amino acid has incorporated into its structure one of the non-natural amino acids described herein. In a further or additional embodiment, the non-natural amino acid is incorporated at a terminal or internal position and is further incorporated site specifically.

[0471] In one embodiment, the resulting non-natural amino acid polypeptide, as produced biosynthetically, already has the desired detection characteristics. In further or additional embodiments, the non-natural amino acid polypeptide comprises at least one non-natural amino acid selected from the group consisting of an oxime-containing non-natural amino acid, a carbonyl-containing non-natural amino acid, and a hydroxylamine-containing non-natural amino acid. In other embodiments such non-natural amino acids have been biosynthetically incorporated into the polypeptide as described herein. In further or alternative embodimentsnon-natural amino acid polypeptide comprises at least one non-natural amino acid selected from amino acids of Formula I-XVIII, XXX-XXXIV(A&B), or XXXX-XXXXIII. In further or additional embodiments, the non-natural amino acid comprises an oxime linkage to a group that provides the improved detection characteristics. In further or additional embodiments, the non-natural amino acid is further modified to form a modified oxime-containing non-natural amino acid polypeptide, wherein the modification provides an oxime linkage to a group that provides the improved detection characteristics. In some embodiments, such a group is directly linked to the non-natural amino acid, and in other embodiments, such a group is linked via a linker group to the non-natural amino acid. In certain embodiments, such a group is connected to the non-natural amino acid by a single chemical reaction, in other embodiments a series of chemical reactions is required to connect such a group to the non-natural amino acid. Preferably, the group imparting improved detection characteristics is linked site specifically to the non-natural amino acid in the non-natural amino acid polypeptide and is not linked to a naturally occurring amino acid under the reaction conditions utilized.

[0472] In further or additional aspects the resulting non-natural amino acid polypeptide is homologous to the GH supergene family members, however, the methods, techniques and compositions described in this section can be applied to virtually any other polypeptide which needs to be detected in an *in vivo* sample and an *in vitro* sample, including by way of example only: alpha-1 antitrypsin, angiostatin, antihemolytic factor, antibody, apolipoprotein, apoprotein, atrial natriuretic factor, atrial natriuretic polypeptide, atrial peptide, C-X-C chemokine, T39765, NAP-2, ENA-78, gro-a, gro-b, gro-c, IP-10, GCP-2, NAP-4, SDF-1, PF4, MIG, calcitonin, c-kit ligand, cytokine, CC chemokine, monocyte chemoattractant protein-1, monocyte chemoattractant protein-2, monocyte chemoattractant protein-3, monocyte inflammatory protein-1 alpha, monocyte inflammatory protein-i beta, RANTES, 1309, R83915, R91733, HCC1, T58847, D31065, T64262, CD40, CD40 ligand, c-kit ligand, collagen, colony stimulating factor (CSF), complement factor 5a, complement inhibitor, complement receptor 1, cytokine, epithelial neutrophil activating peptide-78, MIP-16, MCP-1, epidermal growth factor (EGF), epithelial neutrophil activating peptide, erythropoietin (EPO), exfoliating toxin, Factor IX, Factor VII, Factor VIII, Factor X, fibroblast growth factor (FGF), fibrinogen, fibronectin, four-helical bundle protein, G-CSF, glp-1, GM-CSF, glucocerebrosidase, gonadotropin, growth factor, growth factor receptor, grf, hedgehog protein, hemoglobin, hepatocyte growth factor (hGF), hirudin, human growth hormone (hGH), human serum albumin, ICAM-1, ICAM-1 receptor, LFA-1, LFA-1 receptor, insulin, insulin-like growth factor (IGF), IGF-I, IGF-II, interferon (IFN), IFN-alpha, IFN-beta, IFN-gamma, interleukin (IL), IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, keratinocyte growth factor (KGF), lactoferrin, leukemia inhibitory factor, luciferase, neurturin, neutrophil inhibitory factor (NIF), oncostatin M, osteogenic protein, oncogene product, paracitonin, parathyroid hormone, PD-ECSF, PDGF, peptide hormone, pleiotropin, protein A, protein G, pth, pyrogenic exotoxin A, pyrogenic exotoxin B, pyrogenic exotoxin C, pyy, relaxin, renin, SCF, small biosynthetic protein, soluble complement receptor I, soluble I-CAM 1, soluble interleukin receptor, soluble TNF receptor, somatomedin, somatostatin, somatotropin, streptokinase, superantigens, staphylococcal enterotoxin, SEA, SEB, SEC1, SEC2, SEC3, SED, SEE, steroid hormone receptor, superoxide dismutase, toxic shock syndrome toxin, thymosin alpha 1, tissue plasminogen activator, tumor growth factor (TGF), tumor necrosis factor, tumor necrosis factor alpha, tumor necrosis factor beta, tumor necrosis factor receptor (TNFR), VLA-4 protein, VCAM-1 protein, vascular endothelial growth factor (VEGF), urokinase, mos, ras, raf, met, p53, tat, fos, myc, jun, myb, rel, estrogen receptor, progesterone receptor, testosterone receptor, aldosterone receptor, LDL receptor, and corticosterone.

[0473] In further or additional aspects, the group imparting improved detection characteristics is selected from the group consisting of a label; a dye; an affinity label; a photoaffinity label; a spin label; a fluorophore; a radioactive moiety; a moiety incorporating a heavy atom; an isotopically labeled moiety; a biophysical probe; a phosphorescent group; a chemiluminescent group; an electron dense group; a magnetic group; a chromophore; an energy transfer agent; a detectable label, and any combination thereof.

J. Example of Adding Functionality: Improving the Therapeutic Properties of a Polypeptide

[0474] A naturally-occurring or non-natural amino acid polypeptide will be able to provide a certain therapeutic benefit to a patient with a particular disorder, disease or condition. Such a therapeutic benefit will depend upon a numbers of factors, including by way of example only: the safety profile of the polypeptide, and the pharmacokinetics, pharmacologics and/or pharmacodynamics of the polypeptide (e.g., water solubility, bioavailability, serum half-life, therapeutic half-life, immunogenicity, biological activity, or circulation time). In addition, it may be advantageous to provide additional functionality to the polypeptide, such as an attached cytotoxic compound or drug, or it may be desirable to attach additional polypeptides to form the homo- and heteromultimers described herein. Such modifications preferably do not destroy the activity and/or tertiary structure of the original polypeptide. The methods, compositions, techniques and strategies described herein provide solutions to these issues.

[0475] Using the methods, compositions, techniques and strategies described herein, one of skill in the art can produce an oxime-containing non-natural amino acid polypeptide that is homologous to the desired polypeptide, wherein the oxime-containing non-natural amino acid polypeptide has improved therapeutic characteristics. In one embodiment, a homologous non-natural amino acid polypeptide is produced biosynthetically. In a further or additional embodiment, the non-natural amino acid has incorporated into its structure one of the non-natural amino acids described herein. In a further or additional embodiment, the non-natural amino acid is incorporated at a terminal or internal position and is further incorporated site specifically.

[0476] In one embodiment, the resulting non-natural amino acid, as produced biosynthetically, already has the desired improved therapeutic characteristics. In further or additional embodiments, the non-natural amino acid comprises an oxime linkage to a group that provides the improved therapeutic characteristics. In further or additional embodiments, the non-natural amino acid is further modified to form a modified oxime-containing non-natural amino acid polypeptide, wherein the modification provides an oxime linkage to a group that provides the improved therapeutic characteristics. In some embodiments, such a group is directly linked to the non-natural amino acid, and in other embodiments, such a group is linked via a linker group to the non-natural amino acid. In certain embodiments, such a group is connected to the non-natural amino acid by a single chemical reaction, in other embodiments a series of chemical reactions is required to connect such a group to the non-natural amino acid. Preferably, the group imparting improved therapeutic characteristics is linked site specifically to the non-natural amino acid in the non-natural amino acid polypeptide and is not linked to a naturally occurring amino acid under the reaction conditions utilized.

[0477] In further or additional aspects the resulting non-natural amino acid polypeptide is homologous to the GH supergene family members, however, the methods, techniques and compositions described in this section can be applied to virtually any other polypeptide which can benefit from improved therapeutic characteristics, including by way of example only: alpha-1 antitrypsin, angiostatin, antihemolytic factor, antibody, apolipoprotein, apoprotein, atrial natriuretic factor, atrial natriuretic polypeptide, atrial peptide, C-X-C chemokine, T39765, NAP-2, ENA-78, gro-a, gro-b, gro-c, IP-10, GCP-2, NAP-4, SDF-1, PF4, MIG, calcitonin, c-kit ligand, cytokine, CC chemokine, monocyte chemoattractant protein-1, monocyte chemoattractant protein-2, monocyte chemoattractant protein-3, monocyte inflammatory protein-1 alpha, monocyte inflammatory protein-i beta, RANTES, 1309, R83915, R91733, HCC1, T58847, D31065, T64262, CD40, CD40 ligand, c-kit ligand, collagen, colony stimulating factor (CSF), complement factor 5a, complement inhibitor, complement receptor 1, cytokine, epithelial neutrophil activating peptide-78, MIP-16, MCP-1, epidermal growth factor (EGF), epithelial neutrophil activating peptide, erythropoietin (EPO), exfoliating toxin, Factor IX, Factor VII, Factor VIII, Factor X, fibroblast growth factor (FGF), fibrinogen, fibronectin, four-helical bundle protein, G-CSF, glp-1, GM-CSF, glucocerebrosidase, gonadotropin, growth factor, growth factor receptor, grf, hedgehog protein, hemoglobin, hepatocyte growth factor (hGF), hirudin, human growth hormone (hGH), human serum albumin, ICAM-1, ICAM-1 receptor, LFA-1, LFA-1 receptor, insulin, insulin-like growth factor (IGF), IGF-I, IGF-II, interferon (IFN), IFN-alpha, IFN-beta, IFN-gamma, interleukin (IL), IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, keratinocyte growth factor (KGF), lactoferrin, leukemia inhibitory factor, luciferase, neurturin, neutrophil inhibitory factor (NIF), oncostatin M, osteogenic protein, oncogene product, paracitonin, parathyroid hormone, PD-ECSF, PDGF, peptide hormone, pleiotropin, protein A, protein G, pth, pyrogenic exotoxin A, pyrogenic exotoxin B, pyrogenic exotoxin C, pyy, relaxin, renin, SCF, small biosynthetic protein, soluble complement receptor I, soluble I-CAM 1, soluble interleukin receptor, soluble TNF receptor, somatomedin, somatostatin, somatotropin, streptokinase, superantigens, staphylococcal enterotoxin, SEA, SEB, SEC1, SEC2, SEC3, SED, SEE, steroid hormone receptor, superoxide dismutase, toxic shock syndrome toxin, thymosin alpha 1, tissue plasminogen activator, tumor growth factor (TGF), tumor necrosis factor, tumor necrosis factor alpha, tumor necrosis factor beta, tumor necrosis factor receptor (TNFR), VLA-4 protein, VCAM-1 protein, vascular endothelial growth factor (VEGF), urokinase, mos, ras, raf, met, p53, tat, fos, myc, jun, myb, rel, estrogen receptor, progesterone receptor, testosterone receptor, aldosterone receptor, LDL receptor, and corticosterone.

[0478] In further or additional aspects, the group imparting improved therapeutic characteristics improves the water solubility of the polypeptide; in other aspects, the group improves the binding properties of the polypeptide; in other aspects, the group provides new binding properties to the polypeptide (including, by way of example only, a biotin group or a biotin-binding group). In aspects wherein the group improves the water solubility of the polypeptide, the group is selected from the water soluble polymers described herein, including by way of example only the PEG polymer groups. In further or additional aspects the group is a cytotoxic compound, whereas in other aspects the group is a drug. In further aspects the linked drug or cytotoxic compound can be cleaved from the non-natural amino acid polypeptide so as to deliver the drug or cytotoxic compound to a desired therapeutic location. In other aspects, the group is a second polypeptide, including by way of example, an oxime-containing non-natural amino acid polypeptide, further including by way of example, a polypeptide that has the same amino acid structure as the first non-natural amino acid polypeptide.

[0479] In further or additional embodiments, the oxime-containing non-natural amino acid polypeptide is a modified oxime-containing non-natural amino acid polypeptide. In further or additional aspects, the oxime-containing non-natural amino acid polypeptide increases the bioavailability of the polypeptide relative to the homologous naturally-occurring amino acid polypeptide. In further or additional aspects, the oxime-containing non-natural amino ecid polypeptide increases the safety profile of the polypeptide relative to the homologous naturally-occurring amino acid polypeptide. In further or additional aspects, the oxime-containing non-natural amino acid polypeptide increases the water solubility of the polypeptide relative to the homologous naturally-occurring amino acid polypeptide. In further or additional aspects, the oxime-containing non-natural amino acid polypeptide increases the therapeutic half-life of the polypeptide relative to the homologous naturally-occurring amino acid polypeptide. In further or additional aspects, the oxime-containing non-natural amino acid polypeptide increases the serum half-life of the polypeptide relative to the homologous naturally-occurring amino acid polypeptide. In further or additional aspects, the oxime-containing non-natural amino acid polypeptide extends the circulation time of the polypeptide relative to the homologous naturally-occurring amino acid polypeptide. In further or additional aspects, the oxime-containing non-natural amino acid polypeptide modulates the activity of the polypeptide relative to the homologous naturally-occurring amino acid polypeptide. In further or additional aspects, the oxime-containing non-natural amino acid polypeptide modulates the immunogenicity of the polypeptide relative to the homologous naturally-occurring amino acid polypeptide.

XI. Therapeutic Uses of Modified Polypeptides

[0480] For convenience, the "modified or unmodified" non-natural polypeptides described in this section have been described generically and/or with specific examples. However, the "modified or unmodified" non-natural polypeptides described in this section should not be limited to just the generic descriptions or specific example provided in this section, but rather the "modified or unmodified" non-natural polypeptides described in this section apply equally well to all "modified or unmodified" non-natural polypeptides comprising at least one amino acid which falls within the scope of Formulas I-XVIII, XXX-XXXIV(A&B), and XXXX-XXXXIII, including any sub-formulas or specific compounds that fall within the scope of Formulas I-XVIII, XXX-XXXIV(A&B), and XXXX-XXXXIII that are described in the specification, claims and figures herein.

[0481] The "modified or unmodified" non-natural amino acid polypeptides described herein, including homo- and hetero-multimers thereof find multiple uses, including: therapeutic, diagnostic, assay-based, industrial, cosmetic, plant biology, environmental, energy-production, and/or military uses. As an illustration, the following therapeutic uses of "modified or unmodified" non-natural amino acid polypeptides are provided.

[0482] The "modified or unmodified" non-natural amino acid polypeptides described herein are useful for treating a wide range of disorders, conditions or diseases. Administration of the "modified or unmodified" non-natural amino acid polypeptide products described herein results in any of the activities demonstrated by commercially available polypeptide preparations in humans. Average quantities of the "modified or unmodified" non-natural amino acid polypeptide product may vary and in particular should be based upon the recommendations and prescription of a qualified physician. The exact amount of the "modified or unmodified" non-natural amino acid polypeptide is a matter of preference subject to such factors as the exact type of condition being treated, the condition of the patient being treated, as well as the other ingredients in the composition. The amount to be given may be readily determined by one skilled in the art based upon therapy with the "modified or unmodified" non-natural amino acid polypeptide.

A. Administration and Pharmaceutical Compositions

[0483] The "modified or unmodified" non-natural amino acid polypeptides described herein, including homo- and hetero-multimers thereof find multiple uses, including: therapeutic, diagnostic, assay-based, industrial, cosmetic, plant biology, environmental, energy-production, and/or military uses. As an illustration, the following therapeutic uses of "modified or unmodified" non-natural amino acid polypeptides are provided.

[0484] The "modified or unmodified" non-natural amino acid polypeptides described herein are useful for treating a wide range of disorders. Administration of the "modified or unmodified" non-natural amino acid polypeptide products described herein results in any of the activities demonstrated by commercially available polypeptide preparations in humans. Average quantities of the "modified or unmodified" non-natural amino acid polypeptide product may vary and in particular should be based upon the recommendations and prescription of a qualified physician. The exact amount of the "modified or unmodified" non-natural amino acid polypeptide is a matter of preference subject to such factors as the exact type of condition being treated, the condition of the patient being treated, as well as the other ingredients in the composition. The amount to be given may be readily determined by one skilled in the art based upon therapy with the "modified or unmodified" non-natural amino acid polypeptide.

[0485] The non-natural amino acid polypeptides, modified or unmodified, as described herein (including synthetases, proteins comprising one or more non-natural amino acid, etc.) are optionally employed for therapeutic uses, including in combination with a suitable pharmaceutical carrier. Such compositions, for example, comprise a therapeutically effective amount of the non-natural amino acid polypeptides, modified or unmodified, as described herein, and a pharmaceutically acceptable carrier or excipient. Such a carrier or excipient includes 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 non-natural amino acid polypeptides, modified or unmodified, as described herein.

[0486] Therapeutic compositions comprising one or more of the non-natural amino acid polypeptides, modified or unmodified, as described herein 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 non-natural to natural amino acid homologues (including comparison of a polypeptide modified to include one or more non-natural amino acids to a natural amino acid polypeptide), i.e., in a relevant assay.

[0487] Administration is by any of the routes normally used for introducing a molecule into ultimate contact with blood or tissue cells. The non-natural amino acid polypeptides, modified or unmodified, as described herein, are administered in any suitable manner, optionally with one or more pharmaceutically acceptable carriers. Suitable methods of administering the non-natural amino acid polypeptides, modified or unmodified, as described herein, 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.

[0488] 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 described herein.

[0489] The non-natural amino acid polypeptides described herein and compositions comprising such polypeptides may be administered by any conventional route suitable for proteins or peptides, including parenterally, e.g. injections including subcutaneously or intravenously or any other form of injections or infusions. Polypeptide pharmaceutical compositions (including the various non-natural amino acid polypeptides described herein) can be administered by a number of routes including oral, intravenous, intraperitoneal, intramuscular, transdermal, subcutaneous, topical, sublingual, or rectal means. Compositions comprising non-natural amino acid polypeptides, modified or unmodified, as described herein, can also be administered via liposomes. Such administration routes and appropriate formulations are generally known to those of skill in the art. The non-natural amino acid polypeptides described herein may be used alone or in combination with other suitable components, including a pharmaceutical carrier.

[0490] The non-natural amino acid polypeptides, modified or unmodified, as described herein, 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.

[0491] 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.

[0492] 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 (including those typically used for EPO, IFN, GH, G-CSF, GM-CSF, 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 non-natural amino acid polypeptides, modified or unmodified, as described herein.

[0493] The dose administered to a patient, in the context compositions and methods described herein, is sufficient to have a beneficial therapeutic response in the patient over time. The dose is determined by the efficacy of the particular formulation, and the activity, stability or serum half-life of the non-natural amino acid polypeptides, modified or unmodified, 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 formulation, or the like in a particular patient.

[0494] In determining the effective amount of the formulation to be administered in the treatment or prophylaxis of disease (including 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-non-natural amino acid polypeptide antibodies.

[0495] The dose administered, for example, 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 pharmaceutical formulations described herein 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.

[0496] For administration, the pharmaceutical formulations described herein are administered at a rate determined by the LD-50 or ED-50 of the relevant formulation, and/or observation of any side-effects of the non-natural amino acid polypeptides, modified or unmodified, at various concentrations; including as applied to the mass and overall health of the patient. Administration can be accomplished via single or divided doses.

[0497] 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, including diphenhydramine. Meperidine is used for more severe chills and muscle aches that do not quickly respond to antipyretics and antihistamines. Cell infusion is slowed or discontinued depending upon the severity of the reaction.

[0498] Non-natural amino acid polypeptides, modified or unmodified, as described herein, can be administered directly to a mammalian subject. Administration is by any of the routes normally used for introducing a polypeptide to a subject. The non-natural amino acid polypeptides, modified or unmodified, as described herein, include those suitable for oral, rectal, topical, inhalation (including via an aerosol), buccal (including sub-lingual), vaginal, parenteral (including subcutaneous, intramuscular, intradermal, intraarticular, intrapleural, intraperitoneal, inracerebral, intraarterial, or intravenous), topical (i.e., both skin and mucosal surfaces, including airway surfaces) and transdermal administration, although the most suitable route in any given case will depend on the nature and severity of the condition being treated. Administration can be either local or systemic. The formulations can be presented in unit-dose or multi-dose sealed containers, such as ampoules and vials. The non-natural amino acid polypeptides, modified or unmodified, as described herein, can be prepared in a mixture in a unit dosage injectable form (including solution, suspension, or emulsion) with a pharmaceutically acceptable carrier. The non-natural amino acid polypeptides, modified or unmodified, as described herein, can also be administered by continuous infusion (using, including minipumps such as osmotic pumps), single bolus or slow-release depot formulations.

[0499] Formulations suitable for administration include aqueous and non-aqueous solutions, isotonic sterile solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. Solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described.

[0500] Freeze-drying is a commonly employed technique for presenting proteins which serves to remove water from the protein preparation of interest. Freeze-drying, or lyophilization, is a process by which the material to be dried is first frozen and then the ice or frozen solvent is removed by sublimation in a vacuum environment. An excipient may be included in pre-lyophilized formulations to enhance stability during the freeze-drying process and/or to improve stability of the lyophilized product upon storage. Pikal, M. Biopharm. 3(9)26-30 (1990) and Arakawa et al. Pharm. Res. 8(3):285-291 (1991).

[0501] The spray drying of pharmaceuticals is also known to those of ordinary skill in the art. For example, see Broadhead, J. et al., "The Spray Drying of Pharmaceuticals," in Drug Dev. Ind. Pharm, 18 (11 & 12), 1169-1206 (1992). In addition to small molecule pharmaceuticals, a variety of biological materials have been spray dried and these include: enzymes, sera, plasma, micro-organisms and yeasts. Spray drying is a useful technique because it can convert a liquid pharmaceutical preparation into a fine, dustless or agglomerated powder in a one-step process. The basic technique comprises the following four steps: a) atomization of the feed solution into a spray; b) spray-air contact; c) drying of the spray; and d) separation of the dried product from the drying air. U.S. Patent Nos. 6,235,710 and 6,001,800, describe the preparation of recombinant erythropoietin by spray drying.

[0502] The pharmaceutical compositions described herein may comprise a pharmaceutically acceptable carrier, excipient or stabilizer. 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 (including optional pharmaceutically acceptable carriers, excipients, or stabilizers) for the non-natural amino acid polypeptides, modified or unmodified, described herein, (see, for example, in Remington: The Science and Practice of Pharmacy, Nineteenth Ed (Easton, Pa.: Mack Publishing Company, 1995) ; Hoover, John E., Remington's Pharmaceutical Sciences, Mack Publishing Co., Easton, Pennsylvania 1975; Liberman, H.A. and Lachman, L., Eds., Pharmaceutical Dosage Forms, Marcel Decker, New York, N.Y., 1980 ; and Pharmaceutical Dosage Forms and Drug Delivery Systems, Seventh Ed. (Lippincott Williams & Wilkins, 1999)). Suitable carriers include buffers containing succinate, phosphate, borate, HEPES, citrate, imidazole, acetate, bicarbonate, and other organic acids; antioxidants including ascorbic acid; low molecular weight polypeptides including those less than about 10 residues; proteins, including serum albumin, gelatin, or immunoglobulins; hydrophilic polymers including polyvinylpyrrolidone; amino acids including glycine, glutamine, asparagine, arginine, histidine or histidine derivatives, methionine, glutamate, or lysine; monosaccharides, disaccharides, and other carbohydrates, including trehalose, sucrose, glucose, mannose, or dextrins; chelating agents including EDTA; divalent metal ions including zinc, cobalt, or copper; sugar alcohols including mannitol or sorbitol; salt-forming counter ions including to, sodium; and/or nonionic surfactants, including Tween™ (including Tween 80 (polysorbate 80) and Tween 20 (polysorbate 20), Pluronics™ and other pluronic acids, including and other pluronic acids, including pluronic acid F68 (poloxamer 188), or PEG. Suitable surfactants include for example polyethers based upon poly(ethylene oxide)-poly(propylene oxide)-poly(ethylene oxide), i.e., (PEO-PPO-PEO), or poly(propylene oxide)-poly(ethylene oxide)-poly(propylene oxide), i.e., (PPO-PEO-PPO), or a combination thereof. PEO-PPO-PEO and PPO-PEO-PPO are commercially available under the trade names PluronicsTM, R-PluronicsTM, TetronicsTM and R-TetronicsTM (BASF Wyandotte Corp., Wyandotte, Mich.) and are further described in U.S. Pat. No. 4,820,352. Other ethylene/polypropylene block polymers may be suitable surfactants. A surfactant or a combination of surfactants may be used to stabilize PEGylated non-natural amino acid polypeptides against one or more stresses including stress that results from agitation. Some of the above may be referred to as "bulking agents." Some may also be referred to as "tonicity modifiers."

[0503] The non-natural amino acid polypeptides, modified or unmodified, as described herein, including those linked to water soluble polymers such as PEG can also be administered by or as part of sustained-release systems. Sustained-release compositions include, including semi-permeable polymer matrices in the form of shaped articles, including films, or microcapsules. Sustained-release matrices include from biocompatible materials such as poly(2-hydroxyethyl methacrylate) (Langer et al., J. Biomed. Mater. Res., 15: 167-277 (1981); Langer, Chem. Tech., 12: 98-105 (1982), ethylene vinyl acetate (Langer *et al.*, *supra*) or poly-D-(-)-3-hydroxybutyric acid (EP 133,988), polylactides (polylactic acid) (U.S. Patent No. 3,773,919; EP 58,481), polyglycolide (polymer of glycolic acid), polylactide co-glycolide (copolymers of lactic acid and glycolic acid) polyanhydrides, copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (U. Sidman et al., Biopolymers, 22, 547-556 (1983), poly(ortho)esters, polypeptides, hyaluronic acid, collagen, chondroitin sulfate, carboxylic acids, fatty acids, phospholipids, polysaccharides, nucleic acids, polyamino acids, amino acids such as phenylalanine, tyrosine, isoleucine, polynucleotides, polyvinyl propylene, polyvinylpyrrolidone and silicone. Sustained-release compositions also include a liposomally entrapped compound. Liposomes containing the compound are prepared by methods known per se: DE 3,218,121; Epstein et al., Proc. Natl. Acad. Sci. U.S.A., 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. U.S.A., 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appln. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324:

[0504] Liposomally entrapped polypeptides can be prepared by methods described in, e.g., DE 3,218,121; Epstein et al., Proc. Natl. Acad. Sci. U.S.A., 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. U.S.A., 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appln. 83-118008; U.S. Patent Nos. 4,485,045 and 4,544,545; and EP 102,324. Composition and size of liposomes are well known or able to be readily determined empirically by one skilled in the art. Some examples of liposomes as described in, e.g., Park JW, et al., Proc. Natl. Acad. Sci. USA 92:1327-1331 (1995); Lasic D and Papahadjopoulos D (eds): MEDICAL APPLICATIONS OF LIPOSOMES (1998); Drummond DC, et al., Liposomal drug delivery systems for cancer therapy, in Teicher B (ed): CANCER DRUG DISCOVERY AND DEVELOPMENT (2002); Park JW, et al., Clin. Cancer Res. 8:1172-1181 (2002); Nielsen UB, et al., Biochim. Biophys. Acta 1591(1-3):109-118 (2002); Mamot C, et al., Cancer Res. 63: 3154-3161 (2003).

[0505] The dose administered to a patient in the context of the compositions, formulations and methods described herein, should be sufficient to cause a beneficial response in the subject over time. Generally, the total pharmaceutically effective amount of the non-natural amino acid polypeptides, modified or urunodified, as described herein, administered parenterally per dose is in the range of about 0.01 µg/kg/day to about 100 µg/kg, or about 0.05 mg/l:g to about 1 mg/kg, of patient body weight, although this is subject to therapeutic discretion. The frequency of dosing is also subject to therapeutic discretion, and may be more frequent or less frequent than the commercially available products approved for use in humans. Generally, a polymer:polypeptide conjugate, including by way of example only, a PEGylated polypeptide, as described herein, can be administered by any of the routes of administration described above.

EXAMPLES

Example 1

[0506] This example details the synthesis of the carbonyl-containing amino acid presented in FIG. 4. The carbonyl-containing non-natural amino acid was produced as described in FIG. 4.

Example 2

[0507] This example details the synthesis of the protected hydroxylamine-containing amino acid presented in FIG. 5a. The protected hydroxylamine-containing non-natural amino acid was produced as described in FIG. 5a.

Example 3

[0508] This example details the synthesis of the hydroxylamine-containing amino acid presented in FIG. 5b. The hydroxylamine-containing non-natural amino acid was produced as described in FIG. 5b

Example 4

[0509] This example details the synthesis of the hydroxylamine-containing amino acid presented in FIG. 5c. The hydroxylamine-containing non-natural amino acid was produced as described in FIG. 5c.

Example 5

[0510] This example details the synthesis of the oxime-containing amino acid presented in FIG. 5d. The oxime-containing non-natural amino acid was produced as described in FIG. 5d.

Example 6

[0511] This example details the synthesis of the oxime-containing amino acid presented in FIG. 6a. The oxime-containing non-natural amino acid was produced as described in FIG. 6a.

Example 7

[0512] This example details the synthesis of the oxime-containing amino acid presented in FIG. 6b. The oxime-containing non-natural amino acid was produced as described in FIG. 6b.

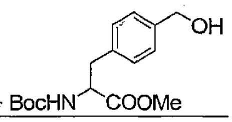
Example 8

[0513] This example details the synthesis of the oxime-containing amino acid presented in FIG. 6c. The oxime-containing non-natural amino acid was produced as described in FIG. 6c.

Example 9

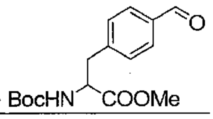
[0514] This example details the synthesis of the carbonyl-containing amino acid presented in FIG. 24.

Synthesis of



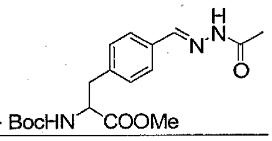
[0515] To a solution NaOH (40 mL, 25 % vol.) at 0 °C was added ether (60 mL). A blast shield was placed in front of the reaction flask. To the resultant mixture was added *N*-nitroso-*N*-methyl urea (6.0 g, 57.9 mmol) in 3 portions over 3 minutes. The reaction was stirred at 0 °C for 10 minutes. The diethyl ether and sodium hydroxide layers were then allowed to separate. The organic layer was added to the solution of *N*-Boc-4-hydroxymethylphenylalanine (7.5 g, 25.4 mmol) in anhydrous THF (20 mL) potionwise (approximately 6 additions) over 5 minutes until the starting material had completely disappeared (monitored by TLC). 5 Drops of glacial acetic acid were then added to quench the reaction. After the organic solvents were removed by rotary evaporation, ethyl acetate was added. The organic layer was washed successively with saturated $NaHCO_3$ solution, H_2O and brine, then dried over anhydrous $MgSO_4$, filtered and concentrated to yield the product (5.9 g, 75%) as a white powder.

Synthesis of

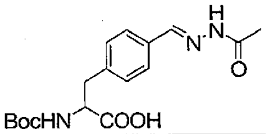


[0516] To a stirred solution of alcohol (6.0 g, 19.4 mmol) and pyridine (12 mL, 150 mmol) in CH_2Cl_2 (400 mL) at 0 °C was added Dess-Martin periodinane (14.2 g, 33.4 mmol). The mixture was stirred at room temperature overnight. The reaction was then quenched with saturated aqueous $Na_2S_2O_3$-$NaHCO_3$ (1:1, 300 mL) and extracted with CH_2Cl_2. The organic layers were combined and washed with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo.* Purification of the residue by flash chromatography (silica, 1:100-1:1 hexane:EtOAc) afforded the aldehyde product (5.48 g, 92%) as a white solid.

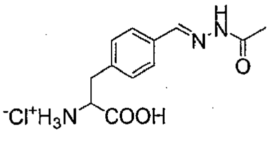
Synthesis of



[0517] To a solution of aldehyde (3.07 g, 10 mmol) in EtOH (40 mL) was added acetic hydrazide (1.7 g, 20 mmol). The reaction mixture was stirred at room temperature for 30 minutes and concentrated. To the residue was added H_2O (200 mL) followed by CH_2Cl_2. The organic layer was separated and concentrated *in vacuo.* Purification of the residue by flash chromatography (silica, 3:7-1:9 hexane:EtOAc) yielded the product (3.29 g, 90%) as a white solid.

Synthesis of 

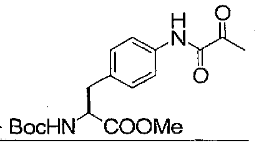
[0518] To a solution of the above methyl ester (3.29 g, 9.1 mmol) in dioxane (10 mL) at 0 °C was added LiOH (10 mL, 1 N). The mixture was stirred at the same temperature for 1 h and then quenched by the addition of citric acid (5 g) and diluted with H_2O. The mixture was extracted with EtOAc. The organic layer was washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated to afford a white solid (3.05 g, 96%).

Synthesis of 

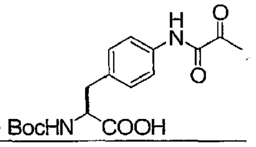
[0519] To a solution of the above acid (3.02 g, 8.6 mmol) in CH_2Cl_2 (20 mL) at 0 °C was added trifluoroacetic acid (20 mL). The reaction mixture was stirred at 0 °C for 2 h and concentrated. To the residue was added MeOH (1mL) followed by the addition of HCl (2.0 mL, 4 N in dioxane). Ether (200 mL) was then added to precipitate the product (2.07 g, 83%) as a yellow solid.

Example 10

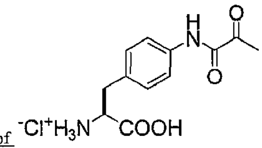
[0520] This example details the synthesis of the dicarbonyl-containing amino acid presented in FIG. 25.

Synthesis of 

[0521] To a stirred solution of amine (10 g, 34 mmol) in DMF (70 mL) at 0 °C were added pyruvate acid (5 mL, 72 mmol), 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochoride (EDC, 20 g, 104 mmol), 1-hydroxybenzotriazole hydrate (HOBt, 85 g, 71 mmol) and *N, N*-diisopropylethylamine (DIEA, 35 mL, 200 mmol). The mixture was stirred at room temperature for 6 h and then quenched with aqueous citric acid solution (5%, 500 mL) and extracted with EtOAc (500 mL). The organic layer was washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated *in vacuo*. The residue was purified by flash chromatography (silica, 3:1-1:1 hexane: EtOAc) to afford product as a solid (4.78 g, 40%).

Synthesis of 

[0522] To a solution of the above methyl ester (2.96 g, 8.1 mmol) in dioxane (10 mL) at 0 °C was added LiOH (10 mL, 1 N). The mixture was stirred at the same temperature for 3 hours. The reaction was then quenched with aqueous citric acid solution (5%) and diluted with EtOAc. The organic layer was separated and washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated to afford product as a yellow solid (2.87 g, 100%).

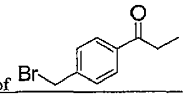
Synthesis of 

[0523] To a solution of the above acid (2.05 g, 5.9 mmol) in CH_2Cl_2 (10 mL) at 0 °C was added trifluoroacetic acid (10 mL). The mixture was stirred for 2 h and concentrated *in vacuo.* To the residue was added HCl (1 mL, 4 N in dioxane) followed by ether (400 mL). The precipitate was collected as a white solid (1.38 g, 82%).

Example 11

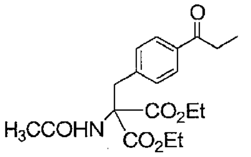
[0524] This example details the synthesis of the dicarbonyl-containing amino acid presented in FIG. 26.

Synthesis of



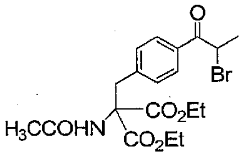
[0525] To a solution of 4'-methylpropiophenone (20 g, 122 mmol) and *N*-bromosuccinimde (NBS, 23 g, 130 mmol) in benzene (300 mL) at 90 °C was added 2, 2'-azobisisobutyronitrile (AIBN, 0.6 g, 3.6 mmol). The resultant solution was heated to reflux overnight. The reaction was then cooled to room temperature. The brown solution was washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated *in vacuo.* The residue was crystallized from hexanes to afford product as a light yellow solid (27 g, 87%).

Synthesis of



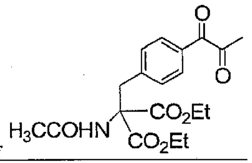
[0526] To a solution of EtONa (14.5 g, 203 mmol) in EtOH (400 mL) at 0 °C was added diethyl acetamidomalonate (39 g, 180 mmol) followed by the solution of the above bromide (27 g, 119 mmol) in EtOH (100 mL). The resultant mixture was heated to reflux for 1 h and quenched with citric acid (30 g) and diluted with H_2O (300 mL). After most solvent was removed *in vacuo,* the residue was extracted with EtOAc. The organic layer was washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated *in vacuo.* The residue was purified by flash chromatography (silica, 10:1-3:1 hexane: EtOAc) to afford product (37 g, 88%) as a yellow solid.

Synthesis of

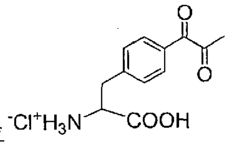


[0527] To a solution of the ketone (5 g, 13.8 mmol) in ether (100 mL) at 0 °C was added Br_2 (0.8 mL, 15.6 mmol). The mixture was stirred at room temperature for 3 h and then quenched with saturated aqueous $NaHCO_3$. The mixture was extracted with Et_2O. The organic layer was washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated *in vacuo* to afford product as a yellow solid (5.4 g, 88%) which was directly used for the next step with further purification.

Synthesis of



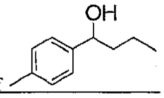
[0528] To the solution of α-bromo ketone (5.4 g, 12.2 mmol) and Na_2CO_3 (2.0 g, 18.9 mmol) in DMSO (20 mL) was added KI (2.1 g, 13.2 mmol). The mixture was stirred at 90 °C under a nitrogen atmosphere for 28 hours. The reaction was then quenched with H_2O and diluted with EtOAc. The organic layer was separated and washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated *in vacuo.* The residue was purified by flash chromatography (silica, 6:1-1:10 hexane: EtOAc) to afford product as a solid (1.12 g, 24%).

Synthesis of 

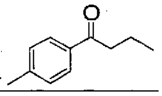
[0529] The solution of diketone (1.12 g, 3.0 mmol) in conc. HCl (10 mL) and dioxane (10 mL) was heated to reflux overnight. After the solvent was removed *in vacuo,* MeOH (3 mL) was added to dissolve the residue. Ether (300 mL) was then added to precipitate the product (302 mg, 42%) as a light yellow solid.

Example 12

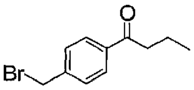
[0530] This example details the synthesis of the dicarbonyl-containing amino acid presented in FIG. 27.

Synthesis of 

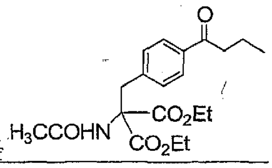
[0531] To a solution of C_3H_7MgCl (2 M, 50 mmol) in ether (25 mL) at 0 °C was added benzaldehyde (5 mL, 42.5 mmol) in ether (50 mL). The resultant solution was stirred at 0 °C for 30 minutes. The reaction was then quenched with saturated NH_4Cl and diluted with ether. The organic layer was separated and washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated *in vacuo* to afford the crude product (7.2 g) which was directly used for the next reaction without purification.

Synthesis of 

[0532] To a solution of the above alcohol (7.2 g, 43.9 mmol) and pyridine (7 mL, 86.7 mmol) in CH_2Cl_2 (300 mL) at 0 °C was added Dess-Martin periodinane (19.2 g, 45.3 mmol). The resultant mixture was stirred overnight and quenched with saturated aqueous $Na_2S_2O_3$ and saturated aqueous $NaHCO_3$ (1:1). The organic layer was washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated *in vacuo.* The residue was purified by flash chromatography (silica, 8:1-4:1 hexane: EtOAc) to afford product as a colorless oil (6.28 g, 91% for two steps).

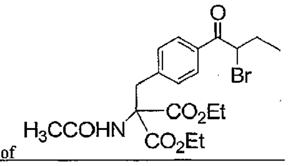
Synthesis of 

[0533] To a solution of the above ketone (4.43 g, 27.3 mmol) and *N*-bromosuccinimde (NBS, 5.5 g, 30.9 mmol) in benzene (150 mL) was added 2, 2'-azobisisobutyronitrile (AIBN, 0.2 g, 1.2 mmol) at 90 °C. The resultant solution was heated to reflux overnight and then cooled to room temperature. The brown solution was washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated *in vacuo.* The residue was crystallized from hexanes to afford product as a white solid (6.21 g, 95%).

Synthesis of 

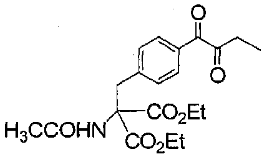
[0534] To a solution of EtONa (2.5 g, 34.9 mmol) in EtOH (200 mL) at 0 °C was added diethyl acetamidomalonate (6.7 g, 30.9 mmol) followed by the solution of the above bromide (6.2 g, 25.8 mmol) in EtOH (100 mL). The resultant mixture was heated to reflux for 1 h and then quenched with citric acid (9 g) and diluted with H_2O. After most solvent was removed, the residue was extracted with EtOAc. The organic layer was washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated *in vacuo*. The residue was purified by flash chromatography (silica, 4:1-2:1 hexane: EtOAc) to afford product as a light yellow solid (8.92 g, 92%).

Synthesis of



[0535] To a solution of the above ketone (1.4 g, 3.71 mmol) in HOAc (50 mL) was added Br_2 (0.7 mL, 13.6 mmol). The mixture was stirred at room temperature overnight and then quenched with saturated aqueous $NaHCO_3$. The mixture was extracted with Et_2O. The organic layer was washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated *in vacuo*. The residue was purified by flash chromatography (silica, 5:1-3:2 hexane: EtOAc) to afford product as a yellow solid (1.23 g, 73%).

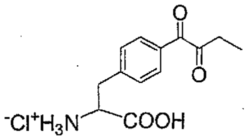
Synthesis of



[0536] To a solution of α-bromo ketone (1.12 g, 2.46 mmol) and Na_2CO_3 (0.4 g, 3.77 mmol) in DMSO (30 mL) was added KI (0.45 g, 13.2 mmol). The mixture was stirred at 90 °C overnight and then quenched with citric acid (2 g) and H_2O (200 mL). The mixture was extracted with EtOAc. The organic layer was washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated *in vacuo*. The residue was purified by flash chromatography (silica, 6:1-1:10 hexane: EtOAc) to afford α-hydroxyl ketone as an oil (0.62 g, 64%).

[0537] To a solution of the above alcohol (0.62 g, 1.58 mmol) and pyridine (0.5 mL, 6.19 mmol) in CH_2Cl_2 (100 mL) at 0 °C was added Dess-Martin periodinane (0.9 g, 2.12 mmol). The resultant mixture was stirred overnight and then quenched with saturated aqueous $Na_2S_2O_3$ and saturated aqueous $NaHCO_3$ (1:1). The organic layer was washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated *in vacuo*. The residue was purified by flash chromatography (silica, 9:1-3:2 hexane: EtOAc) to afford product as a yellow oil (287 mg, 30 % for two steps).

Synthesis of



[0538] The mixture of the above diketone (272 mg, 0.7 mmol) in conc. HCl (10 mL) and dioxane (10 mL) was heated to reflux overnight. After the solvent was removed *in vacuo,* MeOH (1 mL) was added to dissolve the residue. Ether (200 mL) was then added to precipitate the product as a yellow solid (162 mg, 81%).

Example 13

[0539] This example details the synthesis of the dicarbonyl-containing amino acid presented in FIG. 28. The compounds were synthesized as presented in FIG. 28.

Example 14

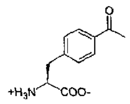
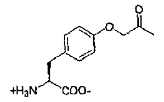
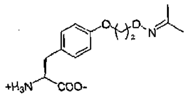
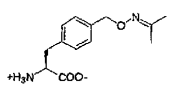
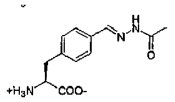
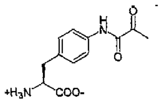
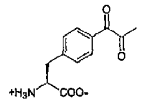
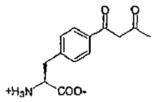
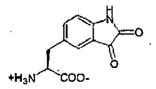
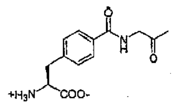
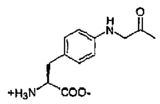
[0540] This example details cloning and expression of a modified polypeptide in *E. coli.* An introduced translation system that comprises an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl tRNA synthetase (O-RS) is used to express the polypeptide containing a non-natural amino acid. The O-RS preferentially aminoacylates the Q-tRNA with a non-natural amino acid. In turn the translation system inserts the non-natural amino acid into the polypeptide, in response to an encoded selector codon. Amino acid and polynucleotide sequences of O-tRNA and O-RS useful for the incorporation of non-natural amino acids are described in U.S. Patent application serial no. 10/126,927 entitled "*In Vivo* Incorporation of Unnatural Amino Acids" and U.S. Patent application serial no. 10/126,931 entitled "Methods and Compositions for the Production of Orthogonal tRNA-Aminoacyl tRNA Synthetase Pairs". The following O-RS and O-tRNA sequences may also be used:

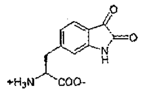
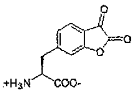
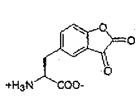
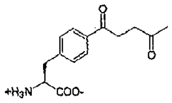
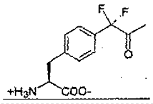
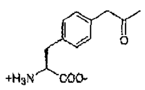
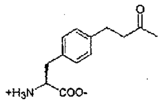
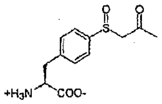
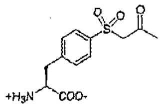
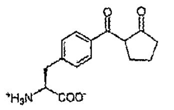
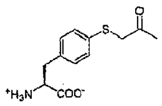
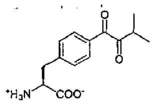
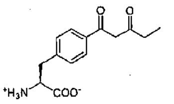
SEQ ID NO:1	*M. jannaschii* $mtRNA_{CUA}^{Tyr}$	tRNA
SEQ ID NO:2	*HLAD03; an optimized amber suppressor tRNA*	tRNA
SEQ ID NO:3	*HL325A; an optimized AGGA frameshift suppressor tRN$_A$*	tRNA
SEQ ID NO:4	*Aminoacyl tRNA synthetase for the incorporation of p-azido-L-phenylalanine*	RS
	p-Az-PheRS(6)	
SEQ ID NO:5	*Aminoacyl tRNA synthetase for the incorporation of p-benzoyl-L-phenylalanine*	RS
	p-BpaRS(1)	
SEQ ID NO:6	*Aminoacyl tRNA synthetase for the incorporation of propargyl-phenylalanine*	RS
	Propargyl-PheRS	
SEQID NO:7	*Aminoacyl tRNA synthetase for the incorporation of propargyl-phenylalanine Propargyl-PheRS*	RS
SEQ ID NO:8	*Aminoacyl tRNA synthetase for the incorporation of propargyl-phenylalanine*	RS
	Propargyl-PheRS	
SEQ ID NO:9	*Aminoacyl tRNA synthetase for the incorporation of p-azido-phenylalanine*	RS
	p-Az-PheRS(1)	
SEQ ID NO:10	*Aminoacyl tRNA synthetase for the incorporation of p-azido-phenylalanine*	RS
	p-Az-PheRS(3)	
SEQ ID NO:11	*Aminoacyl tRNA synthetase for the incorporation of p-azido-phenylalanine*	RS
	p-Az-PheRS(4)	
SEQ ID NO:12	*Aminoacyl tRNA synthetase for the incorporation of p-azido-phenylalanine*	RS
	p-Az-PheRS(2)	
SEQ ID NO:13	*Aminoacyl tRNA synthetase for the incorporation of p-azido-phenylalanine (LW1)*	RS
SEQ ID NO:14	*Aminoacyl tRNA synthetase for the incorporation of p-azido-phenylalanine (LW5)*	RS
SEQ ID NO:15	*Aminoacyl tRNA synthetase for the incorporation of p-azido-phenylalanine (LW6)*	RS
SEQ ID NO:16	*Amiraoacyl tRNA synthetase for the incorporation of p-azido-phenylalanine (AzPheRS-5)*	RS
SEQ ID NO:17	*Aminoacyl tRNA synthetase for the incorporation of p-azido-phenylalanine (AzPheRS-6)*	RS

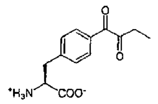
[0541] The transformation of *E. coli* with plasmids containing the modified gene and the orthogonal aminoacyl tRNA synthetase/tRNA pair (specific for the desired non-natural amino acid) allows the site-specific incorporation of non-natural amino acid into the polypeptide. The transformed *E. coli,* grown at 37° C in media containing between 0.01 - 100 mM of the particular non-natural amino acid, expresses modified polypeptide with high fidelity and efficiency. The His-tagged polypeptide containing a non-natural amino acid is produced by the E. coli host cells as inclusion bodies or aggregates. The aggregates are solubilized and affinity purified under denaturing conditions in 6M guanidine HCl. Refolding is performed by dialysis at 4°C overnight in 50mM TRIS-HCl, pH8.0, 40μM $CuSO_4$, and 2% (w/v) Sarkosyl. The material is then dialyzed against 20mM TRIS-HCl, pH 8.0, 100mM NaCl, 2mM $CaCl_2$, followed by removal of the His-tag. See Boissel et al., J. Biol. Chem., (1993) 268:15983-93. Methods for purification of polypeptides are well known in the art and are confirmed by SDS-PAGE, Western Blot analyses, or electrosprayionization ion trap mass spectrometry and the like.

Example 15: Testing Non-Natural Amino Acids

[0542] This example provides results of four tests that were conducted on certain illustrative non-natural amino acids as an aid for predicting their properties for incorporation into non-natural amino acid polypeptides.

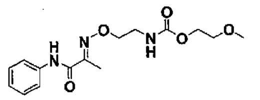
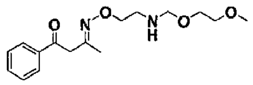
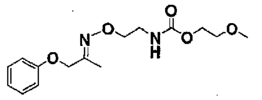
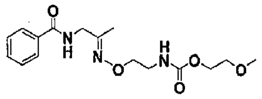
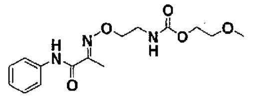
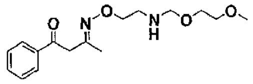
Structure	Oxime Formation at pH6.5	Oxime Stability pH 4-8	Intracellular Concentration	RS test
	+	***	600 µM	√
	+++	*	1000-1800 µM	
	++	***	60 µM	X
	++	***	300 µM	
	+++	***	30 µM	√
	++	***	376 µM	
	+++	***	Reduced [M+2]	
	+++	***		Meta-bolized
	+++	***		
	+++	** (6.5-7.4)	628 µM (little reduced)	
	+++			

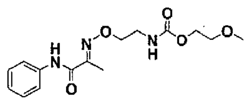
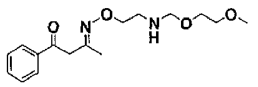
Structure	Oxime Formation at pH6.5	Oxime Stability pH 4-8	Intracellular Concentration	RS test
	+++	- -		
	+++	***		
	+++	***		
	+++	**		
	++	**	279 µM	
	+++	*		
	+++	**		
	+++	***		
	+++	***		
	+++	***		
	+++	***		
	+++	***		
	+++	***		

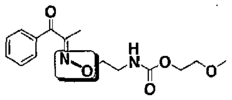
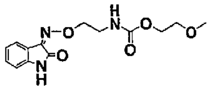
Structure	Oxime Formation at pH6.5	Oxime Stability pH 4-8	Intracellular Concentration	RS test
	+++	***		

Example 16: Testing Non-Natural Amino Acids

[0543] This example provides the results of pH stability tests that were conducted on certain illustrative non-natural amino acids as an aid for predicting their properties for incorporation into non-natural amino acid polypeptides.

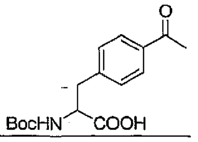
										
pH	**4.0**	**5.0**	**6.5**	**7.4**	**8.0**	**4.0**	**5.0**	**6.5**	**7.4**	**8.0**
1 hour	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**
1 day	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**
2 days	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**
3 days	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**
4 days	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**
5 days	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**<1**	**3%**	**<1**
6 days	**<1**	**<1**	**<1**	**<1**	**<1**	**2%**	**<1**	**<1**	**4%**	**<1**
7 days	**<1**	**<1**	**<1**	**<1**	**<1**	**2%**	**<1**	**<1**	**7%**	**<1**
10 days	**<1**	**<1**	**<1**	**<1**	**<1**	**3%**	**<1**	**<1**	**11%**	**<1**
										
pH	**4.0**	**5.0**	**6.5**	**7.4**	**8.0**	**4.0**	**5.0**	**6.5**	**7.4**	**8.0**
1 hour	**2%**	**<1**	**<1**	**<1**	**<1**	**5%**	**<1**	**<1**	**<1**	**9%**
1 day	**5%**	**5%**	**<1**	**15%**	**35%**	**7%**	**3%**	**<1**	**<1**	**24%**
2 days	**5%**	**5%**	**<1**	**15%**	**38%**	**8%**	**3%**	**<1**	**<1**	**30%**
3 days	**5%**	**5%**	**<1**	**15%**	**40%**	**10%**	**5%**	**2%**	**<1**	**30%**
4 days	**8%**	**5%**	**<1**	**15%**	**45%**	**11%**	**7%**	**2%**	**<1**	**30%**
5 days	**8%**	**5%**	**<1**	**15%**	**45%**	**11%**	**7%**	**2%**	**<1**	**30%**
6 days	**8%**	**5%**	**<1**	**15%**	**45%**	**11%**	**7%**	**2%**	**<1**	**30%**
7 days	**9%**	**5%**	**<1**	**15%**	**45%**	**11%**	**8%**	**2%**	**<1**	**33%**
10 days	**10%**	**5%**	**<1**	**15%**	**45%**	**11%**	**8%**	**2%**	**<1**	**34%**
										
1 hour	4.0	5.0	6.5	7.4	8.0	4.0	5.0	6.5	7.4	8.0
1 day	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
2 days	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1

										
1 hour	**4.0**	**5.0**	**6.5**	**7.4**	**8.0**	**4.0**	**5.0**	**6.5**	**7.4**	**8.0**
3 days	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
4 days	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
5 days	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
6 days	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
7 days	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
10 days	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1

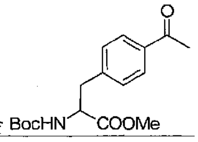
										
1 hour	**4.0**	**5.0**	**6.5**	**7.4**	**8.0**	**4.0**	**5.0**	**6.5**	**7.4**	**8.0**
1 hour	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
1 day	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
2 days	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
3 days	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
4 days	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
5 days	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
6 days	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
7 days	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
10 days	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1

Example 17

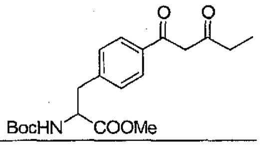
[0544] This example details the synthesis of the dicarbonyl-containing amino acid presented in FIG. 29.

Synthesis of 

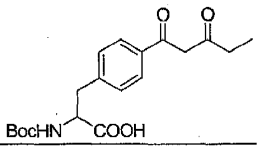
[0545] To a solution of amino acid pAF (10 g, 41.1 mmol) in H_2O-dioxane (300 mL, 1:1) was added $NaHCO_3$ (12 g, 142.9 mmol) and Boc_2O (12 g, 55.0 mmol). The mixture was stirred at room temperature for 7 hours and then quenched with citric acid. The mixture was extracted with EtOAc. The organic layer was washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered and concentrated to afford the *N*-Boc-pAF as a white solid (13.7 g, quant.).

Synthesis of 

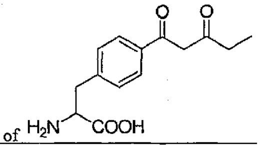
[0546] To NaOH (40 mL, 25 % vol.) at 0 °C was added ether (60 mL). A blast shield was placed in front of the reaction flask. To the resultant mixture was added *N*-nitroso-*N*-methyl urea (6.0 g, 57.9 mmol) in 3 portions over 3 minutes. The reaction was stirred at 0 °C for 10 minutes. The diethyl ether and sodium hydroxide layers were then allowed to separate. The organic layer was added to the solution of N-Boc-pAF (5.0 g, 16.2 mmol) in anhydrous THF (20 mL) potionwise (approximately 6 additions) over 5 minutes until the starting material had completely disappeared (monitored by TLC). 5 Drops of glacial acetic acid were then added to quench the reaction. After the organic solvents were removed by rotary evaporation, ethyl acetate was added. The organic layer was washed successively with saturated $NaHCO_3$ solution, H_2O and brine, then dried over anhydrous $MgSO_4$, filtered and concentrated to yield a white powder (4.1 g, 80%).

Synthesis of


[0547] To *t*-BuOK (60 mL, 1.0 M in THF) was slowly added the solution of the protected pAF (3.82 g, 11.9 mmol) in freshly distilled methyl propionate (20 mL, 208 mmol). The resultant mixture was stirred at room temperature for 30 minutes and quenched with citric acid solution (10%, 300 mL). The mixture was extracted with EtOAc. The organic layer was washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered and concentrated. The residue was purified by flash chromatography (silica, 4:1-1:1 hexane: EtOAc) to afford product as a white solid (3.89 g, 87%).

Synthesis of


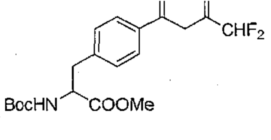
[0548] To a solution of the above methyl ester (1.12 g, 2.97 mmol) in dioxane (4 mL) at 0 °C was added LiOH (4 mL, 1 N). The mixture was stirred at 0 °C for 3 h and quenched with aqueous citric acid solution (5%, 200 mL) and diluted with EtOAc. The organic layer was separated and washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated to afford a white solid (1.02 g, 94%).

Synthesis of


[0549] To a solution of the above acid (1.0 g, 2.75 mmol) in CH_2Cl_2 (10 mL) at 0 °C was added trifluoroacetic acid (10 mL). The mixture was stirred at 0 °C for 2 h and then concentrated. To the residue was added MeOH (1 mL) followed by HCl (1.5 mL, 4 N in dioxane). Ether (200 mL) was then added to precipitate the product (701 mg, 96%) as a white solid.

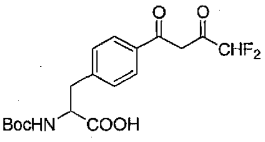
Example 18

[0550] This example details the synthesis of the dicarbonyl-containing amino acid presented in FIG. 30.

Synthesis of


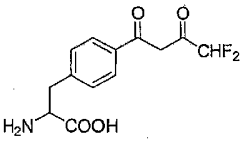
[0551] To t-BuOK (15 mL, 1.0 M in THF) was slowly added the solution of the protected pAF (1.09 g, 3.4 mmol) in methyl difluoroacetate (6 mL, 68.7 mmol). The resultant mixture was stirred at room temperature for 30 minutes and quenched with citric acid (5 g, 25.4 mmol) and diluted with H_2O. The mixture was extracted with EtOAc. The organic layer was washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered and concentrated. The residue was purified by flash chromatography (silica, 20:1-3:2 hexane: EtOAc) to afford product as a light brown solid (1.27 g, 94%).

Synthesis of



[0552] To a solution of the above methyl ester (1.26 g, 3.17 mmol) in dioxane (30 mL) at 0 °C was added LiOH (30 mL, 1 N). The mixture was stirred at 0 °C for 0.5 h and quenched with citric acid (10 g, 51 mmol) and diluted with H_2O. The mixture was extracted with EtOAc. The organic layer was washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated. The residue was purified by flash chromatography (silica, 100:1-10:1 CH_2Cl_2: MeOH, 0.5% HOAc) to afford a brown oil (1.19 g, 98%).

Synthesis of



[0553] To a solution of the above acid (1.19 g, 3.1 mmol) in CH_2Cl_2 (15 mL) at 0 °C was added trifluoroacetic acid (15 mL). The mixture was stirred for 0.5 h and concentrated. To the residue was added MeOH (2 mL) followed by HCl (2 mL, 4 N in dioxane). Ether (200 mL) was then added to precipitate product (0.82 mg, 82%) as a white solid.

Example 19

[0554] This example details the synthesis of the hydroxylamine-containing PEG reagent presented in FIG. 12a. The hydroxylamine-containing PEG reagent was produced as described in FIG. 12a.

Example 20

[0555] This example details the synthesis of the hydroxylamine-containing PEG reagent presented in FIG. 12b. The hydroxylamine-containing PEG reagent was produced as described in FIG. 12b.

Example 21

[0556] This example details the synthesis of the hydroxylamine-containing PEG reagent presented in FIG. 12c. The hydroxylamine-containing PEG reagent was produced as described in FIG. 12c.

Example 22

[0557] This example details the synthesis of the hydroxylamine-containing PEG reagent presented in FIG. 12d. The hydroxylamine-containing PEG reagent was produced as described in FIG. 12d.

Example 23

[0558] This example details the synthesis of the hydroxylamine-containing PEG reagent presented in FIG. 13. The hydroxylamine-containing PEG reagent was produced as described in FIG. 13.

Example 24

[0559] This example details the synthesis of the hydroxylamine-containing PEG reagent presented in FIG. 14. The hydroxylamine-containing PEG reagent was produced as described in FIG. 14.

Example 25

[0560] This example details the synthesis of the hydroxylamine-containing PEG reagent presented in FIG. 15. The hydroxylamine-containing PEG reagent was produced as described in FIG. 15.

Example 26

[0561] This example details the synthesis of the hydroxylamine-containing PEG reagent presented in FIG. 16a. The hydroxylamine-containing PEG reagent was produced as described in FIG. 16a.

Example 27

[0562] This example details the synthesis of the hydroxylamine-containing PEG reagent presented in FIG. 16b. The hydroxylamine-containing PEG reagent was produced as described in FIG. 16b.

Example 28

[0563] This example details the synthesis of a hydroxylamine-containing linker reagent presented in FIG. 18. The hydroxylamine-containing linker reagent was produced as described in FIG. 18.

Example 29

[0564] This example details the synthesis of **1,2-bis(4-(bromomethyl)phenyl)disulfane (1)** presented in Figure 36. To an oven-dried roundbottom flask with stirbar under nitrogen gas pressure was added p-Tolyl disulfide (5.0 g, 20.3 mmol), N-bromo succinimide (8.6 g, 48.4 mmol) and 60 mL anhydrous benzene. The solution was heated to 95 C. Azobisisobutylnitrile (.106 g, .64 mmol) was added in one portion. The reaction was refluxed for 16 hours. The solvent was removed by rotary evaporation and the brown solid dissolved in 100 mL ethyl acetate. The reaction mixture was washed successively with saturated aqueous sodium bicarbonate solution (2 X 50 mL), deionized water (1 X 50 mL) and brine (1 X 50 mL). The organic layer was separated and dried over anhydrous magnesium sulfate, filtered and concentrated under reduced pressure. The crude product was purified by silica chromatography using a Biotage Inc. HORIZON™ chromatography system to afford, after concentration of the appropriate fractions and removal of traces of solvent (vacuum pump), 1,2-bis(4-(bromomethyl)phenyl)disulfane (2.1 g, 25 %) as a white solid. ^{1}H NMR spectral data, and a mass spectrum were obtained. The reaction was repeated to yield (2.0 g, 23%) of the product.

Example 30

[0565] This example details the synthesis of **1,2-bis(4-diethyl-2-acetamidomalonate)phenyldisulfane (2)** presented in Figure 36. To an oven-dried roundbottom flask with stirbar under nitrogen gas pressure was added Diethyl acetamidomalonate (6.48 g, 30 mmol) and 50 mL anhydrous EtOH. To the solution was added sodium ethoxide (2.6 g, 38 mmol) in one portion. The reaction was cooled to 0 °C. 1,2-bis(4-(bromomethyl)phenyl)disulfane (4.1g, 10.1 mmol) was dissolved in 20 mL 1:1 EtOH/THF and added via addition funnel to the cold solution over the course of 1 hour. The ice-bath was removed and the reaction allowed to stir at room temperature for 6 hours. The solvent was removed by rotary evaporation and the red solid dissolved in 100 mL ethyl acetate. The reaction mixture was washed successively with 5 % citric acid solution (2 X 50 mL), deionized water (1 X 50 mL) and brine (1 X 50 mL). The organic layer was separated and dried over anhydrous magnesium sulfate, filtered and concentrated under reduced pressure. The crude product was purified by silica chromatography using a Biotage Inc. HORIZON™ chromatography system to afford, after concentration of the appropriate fractions and removal of traces of solvent (vacuum pump), 1,2-bis(4-diethyl-2-acetamidomalonate)phenyldisulfane (5.0 g, 73 %) as a yellow solid. ^{1}H NMR spectral data, a HPLC trace and a mass spectrum were obtained.

Example 31

[0566] This example details the synthesis of **1,2-bis(4-(2-amino-3-propanoic acid)phenyldisulfane (3)** presented in Figure 36. To an oven-dried roundbottom flask with stirbar under nitrogen gas pressure was added 1,2-bis(4-diethyl-2-acetamidomalonate)phenyldisulfane (1.0 g, 1.4 mmol), HCl (8 mL, 12 M) and 8 mL 1,4 Dioxane. The reaction was stirred at reflux for 16 hours. The solvent was removed by rotary evaporation and vacuum pump to yield crude I,2-bis(4-(2-amino-3-propanoic acid)phenyldisulfane (0.75 g, 135 %) as a clear oil. ^{1}H NMR spectral data and a mass spectrum were obtained.

Example 32

[0567] This example details the synthesis of **N,N'-diBoc-1,2-bis(4-(2-amino-3-propanoic acid)phenyldisulfane (4)** presented in Figure 36. To 1,2-bis(4-(2-amino-3-propanoic acid)phenyldisulfane (0.75 g, 1.9 mmol) in a dry roundbottom flask was added 5 mL 1,4 dioxane, 5 mL deionized water, Di-t-butyl dicarbonate (.65 g, 3.0 mmol) and sodium bicarbonate (0.98 g, 12 mmol). The reaction was stirred at room temperature for 16 hours. The solvent was removed by rotary evaporation and the clear oil dissolved in 100 mL ethyl acetate. The reaction mixture was washed successively with 5 % citric acid solution (5 mL X 2), deionized water (50 mL) and brine (50 mL). The organic layer was separated and dried over anhydrous magnesium sulfate, filtered and concentrated under reduced pressure. The crude product was purified by silica chromatography using a Biotage Inc. HORIZON™ chromatography system to afford, after concentration of the appropriate fractions and removal of traces of solvent (vacuum pump), N,N'-diBoc-1,2-bis(4-(2-amino-3-propanoic acid)phenyldisulfane (0.5 g, 44 % from crude, 52 % over 2 steps) as a white solid. ^{1}H NMR spectral data, HPLC trace and a mass spectrum were obtained.

Example 33

[0568] This example details the synthesis of **N-BOC-2-amino-3-(4-mercaptophenyl)propanoic acid (5)** presented in Figure 36. To an oven-dried roundbottom flask with stirbar under nitrogen gas pressure was added N,N'-diBoc-1,2-bis(4-(2-amino-3-propanoic acid)phenyldisulfane (0.5 g, 0.84 mmol), n-Butyl phosphine (0.6 mL, 2.44 mmol) and 15 mL anhydrous THF. The reaction was stirred at room temperature for 2 hours. The solvent was removed by rotary evaporation and the clear oil dissolved in 50 mL ethyl acetate. The reaction mixture was washed successively with 5 % citric acid solution (2 X 25 mL), deionized water (25 mL) and brine (25 mL). The organic layer was separated and dried over anhydrous magnesium sulfate, filtered and concentrated under reduced pressure. The crude product was purified by silica chromatography using a Biotage Inc. HORIZON™ chromatography system to afford, after concentration of the appropriate fractions and removal of traces of solvent (vacuum pump), N-BOC-2-amino-3-(4-mercaptophenyl)propanoic acid (0.5 g, 100 %) as a white solid. ^{1}H NMR spectral data, HPLC trace and a mass spectrum were obtained.

Example 34

[0569] This example details the synthesis of **2-amino-3-(4-mercaptophenyl)propanoic acid hydrochloride (6)** presented in Figure 36. To an oven-dried roundbottom flask with stirbar was added N-BOC-2-amino-3-(4-mercaptophenyl)propanoic acid (.5g, 1.6 mmol), 10 mL anhydrous dichloromethane and 3 mL trifluoroacetic acid. The reaction was stirred at room temperature for 2 hours. The solvent was removed by rotary evaporation and 1 mL 4.0 M Hydrogen chloride in 1,4-dioxane was added. The roundbottom flask was briefly swirled, then 100 mL anhydrous diethyl ether was added to precipitate 2-amino-3-(4-mercaptophenyl)propanoic acid hydrochloride (0.39 g, 100%) was a white solid. ^{1}H NMR spectral data, HPLC trace and a mass spectrum were obtained.

Example 35

[0570] This example details the synthesis of **diethyl 2-(4-(2-oxopropylthio)benzyl)-2-acetamidomalonate (7)** presented in Figure 37. To an oven-dried roundbottom flask with stirbar under nitrogen gas pressure was added (2) (1.1 g, 1.6 mmol), n-Bu_3P

(1.2 mL, 4.8 mmol) and 25 mL anhydrous THF (25 mL). The reaction was stirred at room temperature for 2hours. To the reaction was added chloroacetone (0.16 mL, 2.0 mmol) and $NaHCO_3$ (0.98 g, 12 mmol). The reaction was stirred at room temperature for 2 hours. The solvent was removed by rotary evaporation and the white solid dissolved in 100 mL ethyl acetate. The reaction mixture was washed successively with saturated aqueous sodium bicarbonate solution (50 mL X 2), deionized water (50 mL) and brine (50 mL). The organic layer was separated and dried over anhydrous magnesium sulfate, filtered and concentrated under reduced pressure. The crude product was purified by silica chromatography using a Biotage Inc. HORIZON™ chromatography system to afford, after concentration of the appropriate fractions and removal of traces of solvent (vacuum pump), diethyl 2-(4-(2-oxopropylthio)benzyl)-2-acetamidomalonate (0.62 g, 98 %) as a white solid. ^1H NMR spectral data, a HPLC trace and a mass spectrum were obtained.

Example 36

[0571] This example details the synthesis of **3-(4-(2-oxopropylthio)phenyl)-2-aminopropanoic acid (8)** presented in Figure 37. To an oven-dried roundbottom flask with stirbar under nitrogen gas pressure was added 7 (0.62 g, 1.5 mmol), 10 mL 1,4-dioxane and 10 mL 12M HCl. The reaction was brought to reflux and allowed to stir overnight. The solvent was removed by rotary evaporation to yield 3-(4-(2-oxopropylthio)phenyl)-2-aminopropanoic acid (0.40 g, 99 % crude).

Example 37

[0572] This example details the synthesis of **N-BOC-3-(4-(2-oxopropylthio)phenyl)-2-aminopropanoic acid (9)** presented in Figure 37. To **8** (0.35 g, 1.3 mmol) in an oven-dried roundbottom flask with stirbar under nitrogen gas pressure was added Di-t-butyl dicarbonate (0.63 g, 3.0 mmol), sodium bicarbonate (0.98 g, 12 mmol), 8 mL 1,4-dioxane and 8 mL deionized water. The reaction was stirred at room temperature for 16 hours. The solvent was removed by rotary evaporation and the clear oil dissolved in 100 mL ethyl acetate. The reaction mixture was washed successively with 5 % citric acid solution (50 mL X 2), deionized water (50 mL) and brine (50 mL). The organic layer was separated and dried over anhydrous magnesium sulfate, filtered and concentrated under reduced pressure. The crude product was purified by silica chromatography using a Biotage Inc. HORIZON™ chromatography system to afford, after concentration of the appropriate fractions and removal of traces of solvent (vacuum pump), N-BOC-3-(4-(2-oxopropylthio)phenyl)-2-aminopropanoic acid (0.30 g, 66 % from crude) as a white solid. ^1H NMR spectral data, HPLC trace and a mass spectrum were obtained.

Example 38

[0573] This example details the synthesis of **N-BOC-3-(4-(2-oxopropylsulfinyl)phenyl)-2-aminopropanoic acid (10)** presented in Figure 37. To an oven-dried roundbottom flask with stirbar under nitrogen gas pressure was added 9 (150 mg, .4 mmol), 8 mL glacial acetic acid and 2 mL 30 % v/v hydrogen peroxide in water. The reaction was stirred for 2 hours at room temperature. The solvent was removed by rotary evaporation and the clear oil dissolved in 50 mL ethyl acetate. The reaction mixture was washed successively with 5 % citric acid solution (25 mL X 2), deionized water (25 mL) and brine (25 mL). The organic layer was separated and dried over anhydrous magnesium sulfate, filtered and concentrated under reduced pressure. The crude product was purified by silica chromatography using a Biotage Inc. HORIZON™ chromatography system to afford, after concentration of the appropriate fractions and removal of traces of solvent (vacuum pump), N-BOC-3-(4-(2-oxopropylsulfinyl)phenyl)-2-aminopropanoic acid (0.13 g, 86% crude). A HPLC trace and a mass spectrum were obtained.

Example 39

[0574] This example details the synthesis of **3-(4-(2-oxopropylsulfinyl)phenyl)-2-aminopropanoic acid (11)** presented in Figure 37. To an oven-dried roundbottom flask with stirbar was added N-BOC-3-(4-(2-oxopropylsulfinyl)phenyl)-2-aminopropanoic acid **10** (0.13 g, 0.35 mmol), 10 mL anhydrous dichloromethane and 3 mL trifluoroacetic acid. The reaction was stirred at room temperature for 2 hours. The solvent was removed by rotary evaporation and 1 mL 4.0 M Hydrogen chloride in 1,4-dioxane was added. The roundbottom flask was briefly swirled, then 100 mL anhydrous diethyl ether was added to precipitate 3-(4-(2-oxopropylsulfinyl)phenyl)-2-aminopropanoic acid (0.072 g, 74 % from crude) as a white solid. ^1H NMR spectral data, HPLC trace and a mass spectrum were obtained.

Example 40

[0575] This example details the synthesis of **N-BOC-3-(4-(2-oxopropylsulfonyl)phenyl)-2-aminopropanoic acid (12)** presented in Figure 37. To an oven-dried roundbottom flask with stirbar under nitrogen gas pressure was added **9** (150 mg, 0.4 mmol), 8 mL glacial acetic acid and 2 mL 30 % v/v hydrogen peroxide in water. The reaction was stirred at room temperature for 24 hours. The solvent was removed by rotary evaporation and the clear oil dissolved in 50 mL ethyl acetate. The reaction mixture was washed successively with 5 % citric acid solution (25 mL X 2), deionized water (25 mL) and brine (25 mL). The organic layer was separated and dried over anhydrous magnesium sulfate, filtered and concentrated under reduced pressure. The crude product was purified by silica chromatography using a Biotage Inc. HORIZON™ chromatography system to afford, after concentration of the appropriate fractions and removal of traces of solvent (vacuum pump), N-BOC-3-(4-(2-oxopropylsulfonyl)phenyl)-2-aminopropanoic acid (**12**) (0.13 g, 86% crude). A HPLC trace and a mass spectrum were obtained.

Example 41

[0576] This example details the synthesis of **3-(4-(2-oxopropylsulfonyl)phenyl)-2-aminopropanoic acid** (**13**) presented in Figure 37. To an oven-dried roundbottom flask with stirbar was added N-BOC-3-(4-(2-oxopropylsulfonyl)phenyl)-2-aminopropanoic acid **12** (0.13 g, 0.35 mmol), 10 mL anhydrous dichloromethane and 3 mL trifluoroacetic acid. The reaction was stirred at room temperature for 2 hours. The solvent was removed by rotary evaporation and 1 mL 4.0 M Hydrogen chloride in 1,4-dioxane was added. The roundbottom flask was briefly swirled, then 100 mL anhydrous diethyl ether was added to precipitate 3-(4-(2-oxopropylsulfonyl)phenyl)-2-aminopropanoic acid (0.067 g, 65 % from crude) as a white solid. ^{1}H NMR spectral data, HPLC trace and a mass spectrum were obtained.

Example 42

[0577] This example details the synthesis of **3-(4-(2-oxopropylthio)phenyl)-2-aminopropanoic acid (14)** presented in Figure 37. To an oven-dried roundbottom flask with stirbar was added N-BOC-3-(4-(2-oxopropylthio)phenyl)-2-aminopropanoic acid 9 (0.10 g, .28 mmol), 10 mL anhydrous dichloromethane and 3 mL trifluoroacetic acid. The reaction was stirred at room temperature for 2 hours. The solvent was removed by rotary evaporation and 1 mL 4.0 M Hydrogen chloride in 1,4-dioxane was added. The roundbottom flask was briefly swirled, then 100 mL anhydrous diethyl ether was added to precipitate 3-(4-(2-oxopropylthio)phenyl)-2-aminopropanoic acid (0.062 g, 85 % from crude) as a white solid. ^{1}H NMR spectral data, HPLC trace and a mass spectrum were obtained.

Example 43

[0578] This example details the synthesis of N-**BOC-3-(4-(2-oxocyclopentylthio)phenyl)-2-aminopropanoic acid (15)** presented in Figure 38. To an oven-dried roundbottom flask with stirbar under nitrogen gas pressure was added 5 (0.15 g, 0.76 mmol), 2-chlorocyclopentanone (0.12 mL, 1.25 mmol), sodium bicarbonate (0.98 g, 12 mmol), 15 mL anhydrous THF. The reaction was stirred at room temperature for 16 hours. The solvent was removed by rotary evaporation and the white solid dissolved in 100 mL ethyl acetate. The reaction mixture was washed successively with saturated aqueous sodium bicarbonate solution (50 mL X 2), deionized water (50 mL) and brine (50 mL). The organic layer was separated and dried over anhydrous magnesium sulfate, filtered and concentrated under reduced pressure. The crude product was purified by silica chromatography using a Biotage Inc. HORIZON™ chromatography system to afford, after concentration of the appropriate fractions and removal of traces of solvent (vacuum pump), N-BOC-3-(4-(2-oxocyclopentylthio)phenyl)-2-aminopropanoic acid (0.15 g, 51 %) as a white solid. A HPLC trace and a mass spectrum were obtained.

Example 44

[0579] This example details the synthesis of **3-(4-(2-oxocyclopentylthio)phenyl)-2-aminopropanoic acid (16)** presented in Figure 38. To an oven-dried roundbottom flask with stirbar was added N-BOC-3-(4-(2-oxocyclopentylthio)phenyl)-2-aminopropanoic acid 15 (0.15 g, 0.39 mmol), 10 mL anhydrous dichloromethane and 3 mL trifluoroacetic acid. The reaction was stirred at room temperature for 2 hours. The solvent was removed by rotary evaporation and 1 mL 4.0 M Hydrogen chloride in 1,4-dioxane was added. The roundbottom flask was briefly swirled, then 100 mL anhydrous diethyl ether was added to precipitate 3-(4-(2-oxocyclopentylthio)phenyl)-2-aminopropanoic acid (0.108 g, 100 %) as a white solid. ^{1}H NMR spectral data, HPLC trace and a mass spectrum were obtained.

Example 45

[0580] This example details the synthesis of **N-BOC-3-(4-(2-oxobutylthio)phenyl)-2-aminopropanoic acid (18)** presented in Figure 38. To an oven-dried roundbottom flask with stirbar under nitrogen gas pressure was added 5 (0.15 g, 0.76 mmol), 1-bromo-2-butanone (0.12 mL, 1.25 mmol), sodium bicarbonate (0.98 g, 12 mmol), 15 mL anhydrous THF. The reaction was stirred at room temperature for 16 hours. The solvent was removed by rotary evaporation and the white solid dissolved in 100 mL ethyl acetate. The reaction mixture was washed successively with saturated aqueous sodium bicarbonate solution (50 mL X 2), deionized water (50 mL) and brine (50 mL). The organic layer was separated and dried over anhydrous magnesium sulfate, filtered and concentrated under reduced pressure. The crude product was purified by silica chromatography using a Biotage Inc. HORIZON™ chromatography system to afford, after concentration of the appropriate fractions and removal of traces of solvent (vacuum pump), N-BOC-3-(4-(2-oxobutylthio)phenyl)-2-aminopropanoic acid (0.15 g, 51 %) as a white solid. A HPLC trace and a mass spectrum were obtained.

Example 46

[0581] This example details the synthesis of **Compounds 19-22** presented in Figure 38. Compounds 19-22 are synthesized using methodology analogous to that described for **Compounds 10-14.**

[0582] This example details the synthesis of **3-(4-(2-oxobutylthio)phenyl)-2-aminopropanoic acid (23)** presented in Figure 38. To an oven-dried roundbottom flask with stirbar was added N-BOC-3-(4-(2-oxobutylthio)phenyl)-2-aminopropanoic acid 18 (0.15 g, .56 mmol), 10 mL anhydrous dichloromethane and 3 mL trifluoroacetic acid. The reaction was stirred at room temperature for 2 hours. The solvent was removed by rotary evaporation and 1 mL 4.0 M Hydrogen chloride in 1,4-dioxane was added. The roundbottom flask was briefly swirled, then 100 mL anhydrous diethyl ether was added to precipitate 3-(4-(2-oxobutylthio)phenyl)-2-aminopropanoic acid (0.149 g, 100 %) as a white solid. ^{1}H NMR spectral data, HPLC trace and a mass spectrum were obtained.

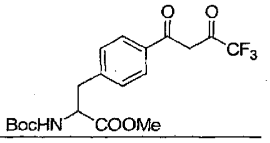
Example 48

[0583] This example details the synthesis of **Compounds 24-27** presented in Figure 39. **Compounds 24-27** are synthesized using methodology analogous to that described for **Compounds 10-14.**

Example 49

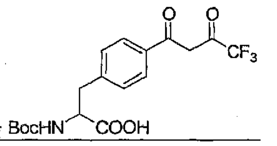
[0584] This example details the synthesis of the dicarbonyl-containing amino acid presented in FIG. 31.

Synthesis of



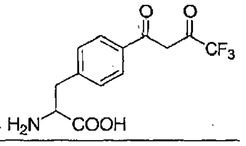
[0585] To *t*-BuOK (15 mL, 1.0 M in THF) is slowly added the solution of the protected pAF (1.0 g, 3.1 mmol) in methyl trifluoroacetate (5 mL, 50 mmol). The reaction mixture is stirred at room temperature for 30 minutes and quenched with citric acid (5 g, 25.4 mmol) and diluted with EtOAc. The organic layer is washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered and concentrated. The residue is purified by flash chromatography (silica, 20:1—3:2 hexane: EtOAc) to afford product (1.07 g, 83%) as a light brown solid.

Synthesis of



[0586] To a solution of the above methyl ester (1.0 g, 2.4 mmol) in dioxane (30 mL) at 0 °C is added LiOH (30 mL, 1 N). The mixture is stirred at 0 °C for 0.5 h and quenched with citric acid (10 g, 51 mmol) and diluted with H_2O. The mixture is extracted with EtOAc. The organic layer is washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated. The residue is purified by flash chromatography (silica, 100:1-10:1 CH_2Cl_2: MeOH, 0.5% HOAc) to afford a brown oil (0.87 g, 98%).

Synthesis of

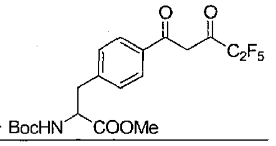


[0587] To a solution of the above acid (1.0 g, 2.5 mmol) in CH_2Cl_2 (15 mL) at 0 °C is added trifluoroacetic acid (15 mL). The resultant mixture is stirred for 0.5 h and concentrated *in vacuo*. To the residue was added MeOH (2 mL) followed by HCl (2 mL, 4 N in dioxane). Ether (200 mL) was then added to precipitate product (0.56 g, 75%) as a white solid.

Example 50

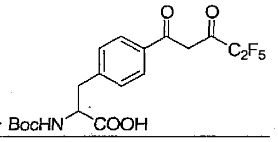
[0588] This example details the synthesis of the dicarbonyl-containing amino acid presented in FIG. 32.

Synthesis of



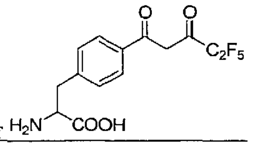
[0589] To *t*-BuOK (15 mL, 1.0 M in THF) is slowly added the solution of the protected pAF (1.0 g, 3.1 mmol) in methyl pentafluoropropionate (8 mL, 62 mmol). The resultant mixture is stirred at room temperature for 1 h and quenched with citric acid (5 g, 25.4 mmol) and diluted with H_2O (100 mL). After most solvent is removed, the residue is extracted with EtOAc. The organic layer is washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered and concentrated. The residue is purified by flash chromatography (silica, 20:1-3:2 hexane: EtOAc) to afford product (1.1 g, 76%) as a light brown solid.

Synthesis of



[0590] To a solution of the above methyl ester (1.0 g, 2.1 mmol) in dioxane (30 mL) at 0 °C is added LiOH (30 mL, 1 N). The resultant mixture is stirred at 0 °C for 0.5 h and quenched with citric acid (10 g, 51 mmol) and H_2O. The mixture is extracted with EtOAc. The organic layer is washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered and concentrated. The residue is purified by flash chromatography (silica, 100:1-10:1 CH_2Cl_2: MeOH, 0.5% HOAc) to afford product (0.8 g, 84%) as a yellow solid.

Synthesis of

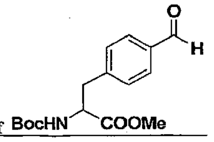


[0591] To a solution of the above acid (0.7 g, 2.5 mmol) in CH_2Cl_2 (15 mL) at 0 °C is added trifluoroacetic acid (15 mL). The mixture is stirred at the same temperature for 0.5 h and concentrated. To the residue is added MeOH (2 mL) followed by HCl (2 mL, 4 N in dioxane). Ether (200 mL) is then added to precipitate the product (0.62 g, 70%) as a white solid.

Example 51

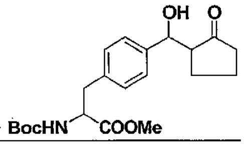
[0592] This example details the synthesis of the dicarbonyl-containing amino acid presented in FIG. 33.

Synthesis of



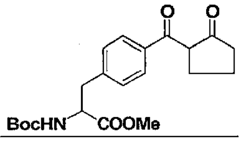
[0593] To a stirred solution ot alcohol **1** (2.35 g, 7.6 mmol) and pyridine (1.5 mL, 18.6 mmol) in CH_2Cl_2 (150 mL) at 0 °C was added Dess-Martin periodinane (3.5 g, 8.3 mmol). The mixture was stirred at room temperature overnight and quenched with saturated aqueous $Na_2S_2O_3$-$NaHCO_3$ (1:1, 100 mL) and diluted with CH_2Cl_2. The organic layer was separated and washed with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. Purification of the residue by flash chromatography (silica, 20:1-3:1 hexane:EtOAc) to afford aldehyde **2** as a white solid (2.15 g, 92%). ESI-MS, m/z: 292 (M^+-OH), 232, 204, 175, 131, 115 (100).

Synthesis of



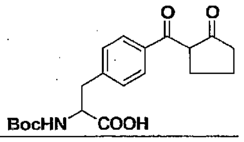
[0594] To a stirred solution of diisopropylamine (0.33 mL, 2.33 mmol) in THF (60 mL) at 0 °C was added *n*-butyllithium (1.46 mL, 2.34 mmol). The mixture was stirred at 0 °C for 20 minutes and cooled to -78 °C and cyclopentanone was then added. After the mixture was stirred for 20 minutes at -78 °C, the solution of aldehyde **2** (0.5 g, 1.63 mmol) in THF (20 mL, washed with 20 mL) was added. The resultant mixture was stirred at -78 °C for 1.0 h and quenched with saturated aqueous NH_4Cl solution. After most of the solvent was removed, the residue was extracted with EtOAc. The organic layer was washed with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. Purification of the residue by flash chromatography (silica, 10:1-1:1 hexane:EtOAc) afforded 3 as a colorless oil (482 mg, 80%).

Synthesis of



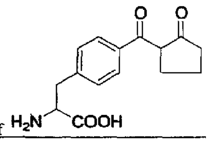
[0595] To a stirred solution of alcohol 3 (0.44 g, 1.13 mmol) and pyridine (0.6 mL, 7.44 mmol) in CH_2Cl_2 (150 mL) at 0 °C was added Dess-Martin periodinane (0.6 g, 1.41 mmol). The mixture was stirred at room temperature overnight. The reaction was quenched with saturated aqueous $Na_2S_2O_3$-$NaHCO_3$ (1:1, 100 mL) and extracted with CH_2Cl_2. The organic layers were combined and washed with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. Purification of the residue by flash chromatography (silica, 20:1-2:1 hexane:EtOAc) gave the diketone **4** (342 mg, 78%) as a colorless oil. ESI-MS, m/z: 412 (M^++ Na), 356, 312, 230, 212, 184, 146 (100).

Synthesis of



[0596] To a solution of methyl ester **4** (330 mg, 0.85 mmol) in dioxane (4 mL) at 0 °C was added LiOH (4 mL, 1 N). The resultant mixture was stirred at 0 °C for 30 minutes and quenched with aqueous citric acid solution (5%, 100 mL). The mixture was extracted with EtOAc. The organic layer was washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered and concentrated to afford a white solid (310 mg, 97%). ESI-MS, m/z: 342, 330 (M^+- COOH), 298, 230, 185, 119 (100).

Synthesis of

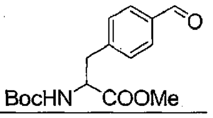


[0597] To a solution of acid 5 (310 mg, 0.83 mmol) in CH_2Cl_2 (4 mL) at 0 °C was added trifluoroacetic acid (4 mL). The mixture was stirred at 0 °C for 30 minutes and concentrated *in vacuo*. To the residue was added MeOH (1 mL) followed by HCl (1 mL, 4 N in dioxane). Ether (100 mL) was then added to precipitate product (241 mg, 94%) as a white solid. ESI-MS, m/z: 298 (M^+ + Na), 276 (M^+ + 1), 230 (M^+ - COOH), 184, 119 (100).

Example 52

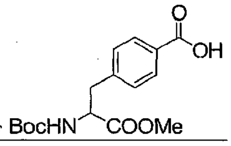
[0598] This example details the synthesis of the dicarbonyl-containing amino acid presented in FIG. 34.

Synthesis of



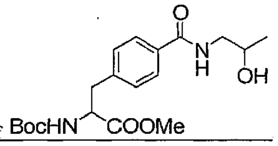
[0599] To a stirred solution of alcohol (6.0 g, 19.4 mmol) and pyridine (12 mL, 150 mmol) in CH_2Cl_2 (400 mL) at 0 °C was added Dess-Martin periodinane (14.2 g, 33.4 mmol). The mixture was stirred at room temperature overnight. The reaction was quenched with saturated aqueous $Na_2S_2O_3$-$NaHCO_3$ (1:1, 300 mL) and extracted with CH_2Cl_2. The organic layers were combined and washed with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. Purification of the residue by flash chromatography (silica, 1:100-1:1 hexane:EtOAc) afforded the aldehyde (5.48 g, 92%) as a white solid.

Synthesis of



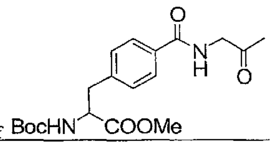
[0600] To a solution of the above aldehyde (3.41 g, 11.1 mmol) in acetone (70 mL) was added $KMnO_4$ (2.5 g, 15.8 mmol) in H_2O (10 mL). The resultant mixture was stirred at room temperature overnight. After most solvent was removed, the residue was dissolved in citric acid aqueous solution (5%, 300 mL) and extracted with EtOAc. The organic layers were combined and washed with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo* to afford product as a white solid (2.83 g, 79%) which was directly used for the next step without further purification.

Synthesis of



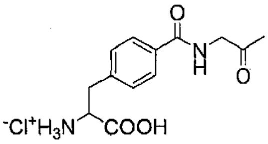
[0601] To a solution of the above acid (2.83 g, 8.76 mmol) in DMF (60 mL) at 0 °C were added 1-amino-2-propanol (1.4 mL, 17.9 mmol), 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide hydrochoride (EDC, 4.1 g, 21.4 mmol), 1-hydroxybenzotriazole hydrate (HOBt, 2.2 g, 18.5 mmol) and *N,N*-diisopropylethylamine (DIEA, 9 mL, 51.6 mmol). The mixture was stirred at room temperature overnight and then quenched with citric acid aqueous solution (5%, 200 mL) and extracted with EtOAc. The organic layers were combined and washed with H_2O and brine, then dried over anhydrous Na_2SO_4, intered and concentrated *in vacuo*. Purification of the residue by flash chromatography (silica, 10:1-1:1 hexane:EtOAc) afforded product (2.45 g, 74%) as a white foam.

Synthesis of



[0602] To a stirred solution of the above alcohol (2.44 g, 6.4 mmol) and pyridine (4 mL, 49.6 mmol) in CH_2Cl_2 (100 mL) at 0 °C was added Dess-Martin periodinane (4.1 g, 9.7 mmol). The mixture was stirred at room temperature overnight. The reaction was quenched with saturated aqueous $Na_2S_2O_3$-$NaHCO_3$ (1:1, 300 mL) and extracted with CH_2Cl_2. The organic layer was washed with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. Purification of the residue by flash chromatography (silica, 1:1-1:3 hexane:EtOAc) afforded product (1.84 g, 76%) as a yellow solid.

Synthesis of



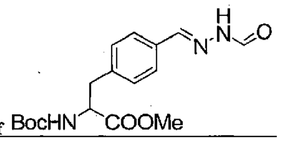
[0603] To a solution of the above methyl ester (1.72 g, 4.6 mmol) in dioxane (10 mL) at 0 °C was added LiOH (10 mL, 1 N). The mixture was stirred at the same temperature for 3 h and quenched with citric acid aqueous solution (5%). The mixture was extracted with EtOAc. The organic layer was washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated to afford product (1.7 g) as a solid which was used directly for the next step without purification.

[0604] To a solution of the above acid (1.7 g, 4.7 mmol) in CH_2Cl_2 (15 mL) at 0 °C was added trifluoroacetic acid (15 mL). The mixture was stirred at the same temperature for 2 h and concentrated *in vacuo*. To the residue was added HCl (1.5 mL, 4 N in dioxane) followed by ether (400 mL). The precipitated product (1.52 g, 90% for 2 steps) was collected as a white solid.

Example 53

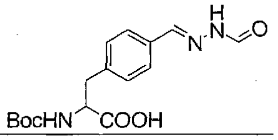
[0605] This example details the synthesis of the hydrazide-containing amino acid presented in FIG. 44.

Synthesis of



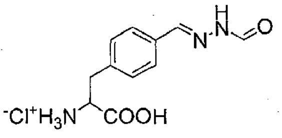
[0606] To a solution of aldehyde (410 mg, 1.34 mmol) in EtOH (15 mL) was added formic hydrazide (170 mg, 2.83 mmol). The reaction mixture was stirred at room temperature for 1 hour. After most solvent was removed, the residue was extracted with CH_2Cl_2. The organic layers were combined and concentrated in vacuo. Purification of the residue by flash chromatography (silica, 1:6-1:1 hexane:EtOAc) afforded a white solid (390 mg, 83%).

Synthesis of



[0607] To a solution of the above methyl ester (349 mg, 1 mmol) in dioxane (7 mL) was added LiOH (7 mL, 1 N) at 0 °C. The mixture was stirred at the same temperature for 10 minutes and quenched with citric acid (2.5 g) and diluted with H_2O. The mixture was extracted with EtOAc. The organic layer was washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated to give a white solid (290 mg, 87%).

Synthesis of

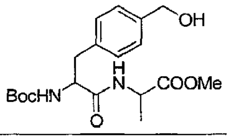


[0608] To a solution of the above acid (290 mg, 0.87 mmol) in CH_2Cl_2 (20 mL) at 0 °C was added trifluoroacetic acid (20 mL). The mixture was stirred for 20 minutes and concentrated. To the residue was added MeOH (1mL) followed by HCl (1.0 mL, 4 N in dioxane). Ether (200 mL) was then added to precipitate product (195 mg, 83%) as a light yellow solid.

Example 54

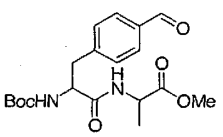
[0609] This example details the synthesis of the hydrazide-containing amino acid presented in FIG. 45.

Synthesis of



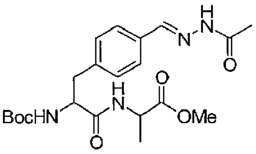
[0610] To a solution of N-Boc-4-hydroxymethylphenylalanine (11.73 g, 39.8 mmol) in DMF (100 mL) was added alanine methyl ester hydrochloride (9.0 g, 64.5 mmol), 1-(3-dimethylaminopropyl)3-ethylcarbodiimide hydrochloride (EDC, 15.4 g, 80.3mmol), N, N-diisopropylethylamine (DIEA, 30 mL, 172 mmol) and 1-hydroxybenzotriazole hydrate (HOBt, 8.4 g, 70.6 mmol). The reaction mixture was stirred at room temperature overnight and then diluted with EtOAc. The organic layer was separated and washed successively with H_2O, citric acid (5%), H_2O, $NaHCO_3$, H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated to afford the protected dipeptide as a white solid (13.74 g, 91%).

Synthesis of



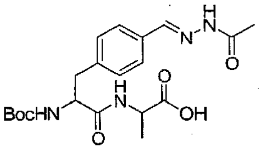
[0611] To a solution of the above protected dipeptide (10.33 g, 27.2 mmol) in CH_2Cl_2 (300 mL) at 0 °C were added pyridine (δ mL, 99.1 mmol) and Dess-Martin periodinane (14 g, 33.0 mmol). The reaction mixture was stirred overnight and then quenched with saturated aqueous NaHCO3/Na2S2O3 (1:1). The organic layer was washed successively with H2O, citric acid, H20 and brine, then dried over anhydrous Na_2SO_4, filtered and concentrated. The residue was purified by flash chromatography (silica, 9:1-1:1 hexane: EtOAc) to afford product (10.12 g, 98%) as a white solid.

Synthesis of



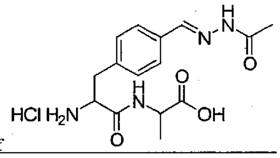
[0612] To a solution of dipeptide aldehyde (10.1 g, 26.7 mmol) in EtOH (200 mL) was added acetic hydrazide (3.7 g, 45 mmol). The reaction mixture was stirred at room temperature for 30 minutes and concentrated. To the residue were added H20 (1 L) and CH2Cl2 (500 mL). The organic layer was separated and concentrated to afford a white solid (11.21 g, 97%).

Synthesis of



[0613] To a solution of the above methyl ester (11.1 g, 25.6 mmol) in dioxane (50 mL) at 0 °C was added LiOH (50 mL, 1 N). The mixture was stirred at the same temperature for 30 minutes and then quenched with citric acid (20 g) and diluted with H20 (200 mL). The mixture was extracted with EtOAc. The organic layer was washed successively with H2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated to afford a white solid (9.52 g, 88%).

Synthesis of

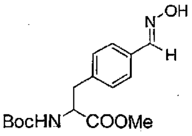


[0614] To a solution of the above acid (9.5 g, 22.6 mmol) in CH2Cl2 (50 mL) at 0 °C was added trifluoroacetic acid (50 mL). The mixture was stirred at 0 °C for 1 h and concentrated in vacuo. To the residue was added HCl (7 mL, 4 N in dioxane) followed by ether (500 mL). The precipitate was collected as a white solid (7.25 g, 90%).

Example 55

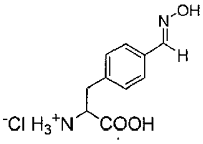
[0615] This example details the synthesis of the oxime-containing amino acid presented in FIG. 46A.

Synthesis of



[0616] To a solution of the aldehyde (3.0 g) in MeOH/H$^+$ was added 2 equivalents of hydroxylamine hydrochloride. The reaction mixture was stirred at room temperature for 2 hours and concentrated. To the residue was added H_2O (200 mL) followed by CH_2Cl_2. The organic layer was separated and concentrated *in vacuo.* Purification of the residue by flash chromatography (silica, 3:7-1:9 hexane:EtOAc) yielded the product (96%) as a solid.

Synthesis of

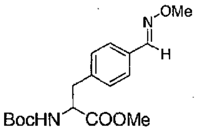


[0617] To a solution of the above methyl ester (3.0 g) in dioxane (10 mL) at 0 °C was added LiOH (10 mL, 1 N). The mixture was stirred at the same temperature for 3 hours and then quenched by the addition of citric acid (5 g) and diluted with H_2O. The mixture was extracted with EtOAc. The organic layer was washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated to afford a solid. Then, to a solution of the resulting acid in CH_2Cl_2 (20 mL) at 0 °C was added trifluoroacetic acid (20 mL). The reaction mixture was stirred at 0 °C for 1 hour and concentrated. To the residue was added MeOH (1mL) followed by the addition of HCl (2.0 mL, 4 N in dioxane). Ether (200 mL) was then added to precipitate the product (87%) as a solid.

Example 56

[0618] This example details the synthesis of the oxime-containing amino acid presented in FIG. 46B.

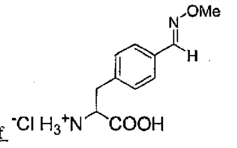
Synthesis of



[0619] To a solution of the aldehyde (3.0 g) in MeOH/H$^+$ was added 2 equivalents of methoxyamine hydrochloride. The reaction mixture was stirred at room temperature for 2 hours and concentrated. To the residue was added H_2O (200 mL) followed by

CH_2Cl_2. The organic layer was separated and concentrated *in vacuo*. Purification of the residue by flash chromatography (silica, 3:7-1:9 hexane:EtOAc) yielded the product (93%) as a solid.

Synthesis of

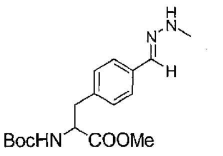


[0620] To a solution of the above methyl ester (3.0 g) in dioxane (10 mL) at 0 °C was added LiOH (10 mL, 1 N). The mixture was stirred at the same temperature for 3 hours and then quenched by the addition of citric acid (5 g) and diluted with H_2O. The mixture was extracted with EtOAc. The organic layer was washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated to afford a solid. Then, to a solution of the resulting acid in CH_2Cl_2 (20 mL) at 0 °C was added trifluoroacetic acid (20 mL). The reaction mixture was stirred at 0 °C for 1 hour and concentrated. To the residue was added MeOH (1mL) followed by the addition of HCl (2.0 mL, 4 N in dioxane). Ether (200 mL) was then added to precipitate the product (89%) as a solid.

Example 57

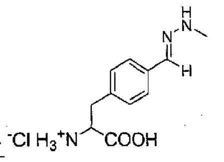
[0621] This example details the synthesis of the hydrazine-containing amino acid presented in FIG. 46C.

Synthesis of



[0622] To a solution of the aldehyde (3.0 g) in MeOH/H^+ was added 2 equivalents of methylhydrazine. The reaction mixture was stirred at room temperature for 2 hours and concentrated. To the residue was added H_2O (200 mL) followed by CH_2Cl_2. The organic layer was separated and concentrated *in vacuo*. Purification of the residue by flash chromatography (silica, 3:7-1:9 hexane:EtOAc) yielded the product (93%) as a solid.

Synthesis of

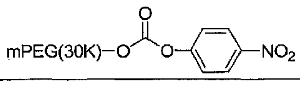


[0623] To a solution of the above methyl ester (3.0 g) in dioxane (10 mL) at 0 °C was added LiOH (10 mL, 1 N). The mixture was stirred at the same temperature for 3 hours and then quenched by the addition of citric acid (5 g) and diluted with H_2O. The mixture was extracted with EtOAc. The organic layer was washed successively with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered, and concentrated to afford a solid. Then, to a solution of the resulting acid in CH_2Cl_2 (20 mL) at 0 °C was added trifluoroacetic acid (20 mL). The reaction mixture was stirred at 0 °C for 1 hour and concentrated. To the residue was added MeOH (1mL) followed by the addition of HCl (2.0 mL, 4 N in dioxane). Ether (200 mL) was then added to precipitate the product (89%) as a solid.

Example 58

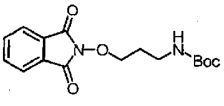
[0624] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 48A.

Synthesis of



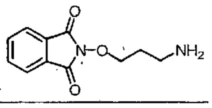
[0625] To a solution of mPEG(30K)-OH (1.0 g, 0.033 mmol) in anhydrous CH_2Cl_2 (10 mL) was added p-nitrophenol chloroformate (60 mg, 0.28 mmol). The mixture was stirred at room temperature for 15 hours. Ether (200 mL) was added. The precipitate was filtered, washed with ether and dried in vacuo to afford product (1.0 g, 100%) as a white powder.

Synthesis of



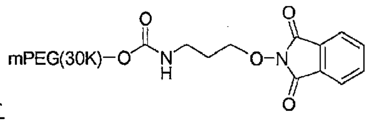
[0626] To a solution of t-butyl 3-hydroxyethylcarbamate (1.75 g, 10 mmol) in THF (60 mL) were added N-hydroxyphthalimide (3.2 g, 20 mmol), triphenylphosphine (2.0 g, 15 mmol). The reaction mixture was stirred at room temperature for 10 minutes and then cooled to 0 °C. Diisopropylazodicarboxylate (DIAD, 2.0 mL, 10.5 mmol) was added dropwise via syringe over 1 hour. The icebath was removed and the mixture was stirred overnight and concentrated. The white solid dissolved in ethyl acetate (100 mL). The reaction mixture was washed successively with saturated aqueous sodium bicarbonate solution (100 mL), H_2O (100 mL) and brine (100 mL), then dried over anhydrous $MgSO_4$, filtered and concentrated in vacuo. The crude product was purified by flash chromatography (silica, 100:1-10:1 hexane: EtOAc) to afford the title compound (2.6 g, 81 %) as a white solid.

Synthesis of



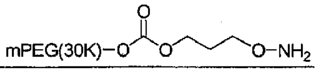
[0627] To a solution of the Boc-protected linker (2.0 g, 9.1 mmol) in CH_2Cl_2 (5 mL) was added trifluoroacetic acid (5 mL). The resultant mixture was stirred at room temperature for 1 hour and concentrated. To the residue was added HCl (4 N in dioxane, 1.5 mL) followed by the addition of Et_2O (150 mL). The precipitate was filtered, washed with ether and dried in vacuo to afford the amine linker (1.1 g, 85 %) as a white solid.

Synthesis of



[0628] To a mixture of mPEG(30 K) p-nitrophenolcarbonate (1.0 g, 0.033 mmol) and amine linker (53 mg, 0.21 mmol) in DMF-CH_2Cl_2 (10 mL, 1:2) were added diisopropylethylamine (50 µL, 0.28 mmol) and DMAP (5 mg, 0.041 mmol). The resultant mixture was stirred at room temperature for 15 hours. Ether (200 mL) was added. The precipitate was filtered, washed with ether and dried in vacuo to afford product (0.83 g, 83%) as a white powder.

Synthesis of

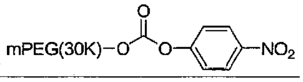


[0629] To a solution of mPEG phthalimide (30K, 0.8 g, 0.0266 mmol) in MeOH (5 mL) was added hydrazine (8.5 µL, 0.27 mmol). The resultant mixture was stirred at 45°C for 1.0 hours. After the reaction was cooled to room temperature, CH_2Cl_2 (150 mL) was added and the solution was washed with aqueous HCl solution (0.1 N, 100 mL). The aqueous layer was extracted with CH_2Cl_2 (150 mL). The organic layers were combined and washed with H_2O (100 mL), then dried over anhydrous Na_2SO_4, filtered and concentrated. The residue was dissolved in CH_2Cl_2 (5 mL). Ether (200 mL) was added to precipitate the hydroxylamine product (0.72 g, 90%) as a white powder.

Example 59

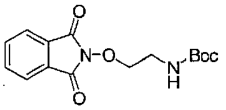
[0630] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 48B.

Synthesis of



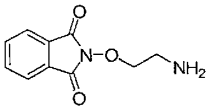
[0631] To a solution of mPEG(30K)-OH (1.0 g, 0.033 mmol) in anhydrous CH_2Cl_2 (10 mL) was added p-nitrophenol chloro formate (60 mg, 0.28 mmol). The mixture was stirred at room temperature for 15 hours. Ether (200 mL) was added. The precipitate was filtered, washed with ether and dried in vacuo to afford product (1.0 g, 100%) as a white powder.

Synthesis of



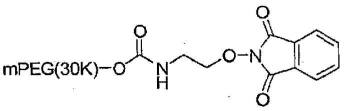
[0632] To a solution of t-butyl 2-hydroxyethylcarbamate (2.8 mL, 18 mmol) in THF (60 mL) were added N-hydroxyphthalimide (5.8 g, 36 mmol), triphenylphosphine (3.6 g, 27 mmol). The reaction mixture was stirred at room temperature for 10 minutes and then cooled to 0 °C. Diisopropylazodicarboxylate (DIAD, 3.6 mL, 19 mmols) was added dropwise via syringe over 1 hour. The icebath was removed and the mixture was stirred overnight and concentrated. The white solid dissolved in ethyl acetate (100 mL). The reaction mixture was washed successively with saturated aqueous sodium bicarbonate solution (2 X 50 mL), H_2O (50 mL) and brine (50 mL), then dried over anhydrous $MgSO_4$, filtered and concentrated in vacuo. The crude product was purified by flash chromatography using a Biotage Inc. HORIZON™ chromatography system to afford the title compound with impurities (12 g, 206 %) as a white solid.

Synthesis of



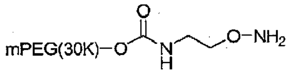
[0633] To a solution of the crude Boc-protected linker (12 g) in CH_2Cl_2 (5 mL) was added trifluoroacetic acid (5 mL). The resultant mixture was stirred at room temperature for 1 hour and concentrated. To the residue was added HCl (4 N in dioxane, 1.5 mL) followed by the addition of Et_2O (150 mL). The precipitate was filtered, washed with ether and dried in vacuo to afford the amine linker (3.0 g, 68 % for two steps) as a white solid.

Synthesis of



[0634] To a mixture of mPEG(30 K) p-nitrophenolcarbonate (1.0 g, 0.033 mmol) and amine linker (50 mg, 0.21 mmol) in DMF-CH_2Cl_2 (10 mL, 1:2) were added diisopropylethylamine (50 µL, 0.28 mmol) and 4-dimethylaminopyridine (4 mg, 0.033 mmol). The resultant mixture was stirred at room temperature for 15 hours. Ether (200 mL) was added. The precipitate was filtered, washed with ether and dried in vacuo to afford product (0.81 g, 81%) as a white powder.

Synthesis of

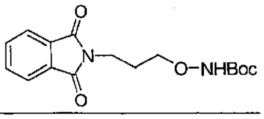


[0635] To a solution of mPEG(30K) phthalimide (0.8 g, 0.0266 mmol) in MeOH (5 mL) was added hydrazine (8.5 µL, 0.27 mmol). The resultant mixture was stirred at 45 °C for 1.0 hour. After the reaction was cooled to room temperature, CH_2Cl_2 (150 mL) was added and the solution was washed with aqueous HCl solution (0.1 N, 100 mL). The aqueous layer was extracted with CH_2Cl_2 (150 mL). The organic layers were combined and washed with H_2O (100 mL), then dried over anhydrous Na_2SO_4, filtered and concentrated. The residue was dissolved in CH_2Cl_2 (5 mL). Ether (200 mL) was added to precipitate the hydroxylamine product (0.68 g, 85%) as a white powder.

Example 60

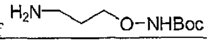
[0636] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 49A.

Synthesis of



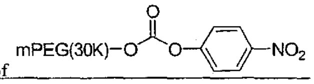
[0637] To a solution of N-(3-bromopropyl)phthalimide (4.0 g, 15.0 mmol) in DMF (50 mL) at 0 °C were added K_2CO_3 (10 g, 73 mmol) and t-butyl N-hydroxycarbamate (2.5 g, 18.8 mmol). The reaction mixture was stirred at room temperature for 3 hours. The mixture was diluted with H_2O (200 mL) and extracted with EtOAc (200 mL). The organic layer was washed with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered and concentrated in vacuo. Purification of the residue by flash chromatography (silica, 20:1-3:1 hexane:EtOAc) afforded product (3.5 g, 72%) as a colorless oil.

Synthesis of



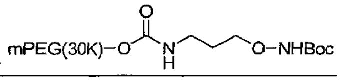
[0638] To a solution of the above phthalimide (500 mg, 1.6 mmol) in EtOH (10 mL) was added hydrazine (0.25 mL, 8.0 mmol). The resultant mixture was stirred at room temperature for 3 hours. After removal of the precipitate, the filtrate was concentrated. The residue was left in high vacuum overnight. Purification of the residue by flash chromatography (silica, 3:1-1:1 EtOAc:MeOH) afforded the amine linker (252 mg, 85%) as a white solid.

Synthesis of



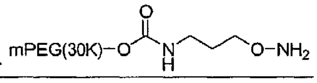
[0639] To a solution of mPEG(30K)-OH (1.0 g, 0.033 mmol) in anhydrous CH_2Cl_2 (10 mL) was added p-nitrophenol chloroformate (60 mg, 0.28 mmol). The mixture was stirred at room temperature for 15 hours. Ether (200 mL) was added. The precipitate was filtered, washed with ether and dried in vacuo to afford product (1.0 g, 100%) as a white solid.

Synthesis of



[0640] To a solution of the above activated mPEG(30K) (3.0 g, 0.1 mmol) in anhydrous CH_2Cl_2 (30 mL) were added diisopropylethylamine (88 (µL, 0.5 mmol) and the amine linker (76 mg, 0.4 mmol). The resultant mixture was stirred at room temperature for 15 hours. Ether (700 mL) was added to the reaction mixture. The precipitate was filtered, washed with ether and dried in vacuo to afford a white powder (2.8 g, 93%).

Synthesis of

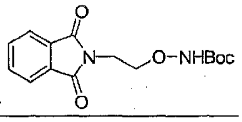


[0641] To a solution of Boc protected mPEG(30 K) (2.0 g, 0.067 mmol) in anhydrous CH_2Cl_2 (10 mL) at 0 °C was added trifluoroacetic acid (10 mL). The resultant mixture was stirred at room temperature for 5 hours. Ether (500 mL) was added to the reaction mixture. The precipitate was filtered, washed with ether and dried in vacuo to afford a white powder (1.8 g, 90%).

Example 61

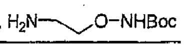
[0642] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 49B.

Synthesis of



[0643] To a solution of t-butyl N-hydroxycarbamate (5.0 g, 37.6 mmol) in DMF (30 mL) at 0 °C were added K_2CO_3 (12 g, 87.6 mmol) and N-(2-bromoethyl)phthalimide (10.0 g, 39.7 mmol). The reaction mixture was stirred at room temperature for 3 hours. The mixture was diluted with H_2O (200 mL) and extracted with EtOAc (200 mL). The organic layer was washed with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered and concentrated in vacuo. Purification of the residue by flash chromatography (silica, 20:1-1:1 hexane:EtOAc) afforded product (5.2 g, 55%) as a white solid.

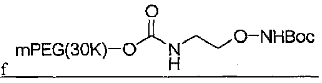
Synthesis of



[0644] To a solution of the above phthalimide (500 mg, 1.6 mmol) in EtOH (10 mL) was added hydrazine (0.25 mL, 8.0 mmol). The resultant mixture was stirred at room temperature for 3 hours. After removal of the precipitate, the filtrate was concentrated.

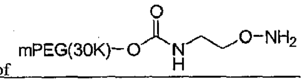
The residue was left in high vacuum overnight. Purification of the residue by flash chromatography (silica, 3:1-1:1 EtOAc:MeOH) afforded the amine linker (301 mg, 86%) as a white solid.

Synthesis of



[0645] To a solution of the above activated mPEG(30K) (1.0 g, 0.033 mmol) in anhydrous CH_2Cl_2 (10 mL) were added diisopropylethylamine (58 µL, 0.33 mmol), 4-dimethylaminopyridine (4 mg, 0.033 mmol) and the above amine linker (64 mg, 0.31 mmol). The resultant mixture was stirred at room temperature for 15 hours. Ether (200 mL) was added to the reaction mixture. The precipitate was filtered, washed with ether and dried in vacuo to afford a white powder (0.85 g, 85%).

Synthesis of

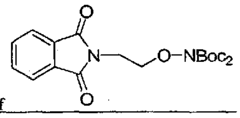


[0646] To a solution of Boc protected mPEG(30K) (0.85 g, 0.028 mmol) in anhydrous CH_2Cl_2 (10 mL) at 0 °C was added trifluoroacetic acid (10 mL). The resultant mixture was stirred at room temperature for 5 hours. Ether (200 mL) was added to the reaction mixture. The precipitate was filtered, washed with ether and dried in vacuo to afford a white powder (0.68 g, 80%).

Example 62

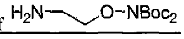
[0647] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 50A.

Synthesis of



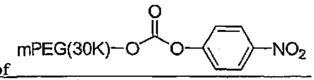
[0648] To a solution of mono-Boc phthalimide (2.1 g, 6.9 mmol) in pyridine (50 mL) at 0 °C was added Boc_2O (3.3 g, 15.1 mmol). The resultant mixture was heated to 60 °C overnight. After the solvent was removed in vacuo, the residue was diluted with EtOAc (200 mL) and washed with citric acid (5%, 200 mL), water (200 mL) and brine (200 mL), then dried over anhydrous Na_2SO_4, filtered and concentrated in vacuo. Purification of the residue by flash chromatography (silica, 20:1-3:1 hexane:EtOAc) afforded product (2.37 g, 85%) as a yellow oil.

Synthesis of



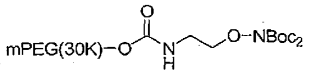
[0649] To a solution of di-Boc phthalimide (1.21 g, 2.98 mmol) in MeOH (15 mL) at 0 °C was added ammonia in MeOH (15 mL, 7 N, 105 mmol). The resultant mixture was stirred at room temperature overnight. The precipitate was filtered off and the filtrate was concentrated in vacuo. Purification of the residue by flash chromatography (silica, 10:1-6:4 EtOAc:MeOH) afforded the amine linker (0.61 g, 74%) as a white solid.

Synthesis of



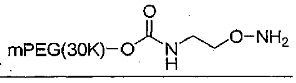
[0650] To a solution of mPEG(30K)-OH (1.0 g, 0.033 mmol) in anhydrous CH_2Cl_2 (10 mL) was added p-nitrophenol chloro formate (60 mg, 0.28 mmol). The mixture was stirred at room temperature for 15 hours. Ether (200 mL) was added to precipitate mPEG(30K) product. The product was filtered, washed with ether and dried in vacuo (1.0 g, 100%).

Synthesis of



[0651] To a solution of the above activated mPEG(30K) (1.0 g, 0.033 mmol) in anhydrous CH_2Cl_2 (10 mL) were added diisopropylethylamine (58 µL, 0.33 mmol), 4-dimethylaminopyridine (5 mg, 0.041mmol) and the above diBoc amine linker (90 mg, 0.33 mmol). The resultant mixture was stirred at room temperature for 15 hours. Ether (200 mL) was added to the reaction mixture. The precipitate was filtered, washed with ether and dried in vacuo to afford a white powder (0.82 g, 82%).

Synthesis of

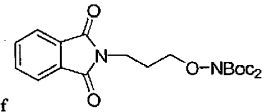


[0652] To a solution of Boc protected mPEG(30K) (0.82 g, 0.027 mmol) in anhydrous CH_2Cl_2 (8 mL) at 0 °C was added trifluoroacetic acid (8 mL). The resultant mixture was stirred at room temperature for 5 hours. Ether (200 mL) was added to the reaction mixture. The precipitate was filtered, washed with ether and dried in vacuo to afford a white powder (0.57 g, 70%).

Example 63

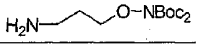
[0653] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 50B.

Synthesis of



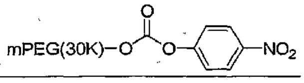
[0654] To a solution of morio-Boc phthalimide (1.5 g, 4.7 mmol) in pyridine at 0 °C was added Boc_2O (2.2 g, 10.0 mmol). The resultant mixture was heated to 60 °C overnight. After the solvent was removed in vacuo, the residue was diluted with EtOAc (200 mL) and washed with citric acid (5%, 200 mL), water (200 mL) and brine (200 mL), then dried over anhydrous Na_2SO_4, filtered and concentrated in vacuo. Purification of the residue by flash chromatography (silica, 20:1-3:1 hexane:EtOAc) afforded product (1.6 g, 81%) as an oil.

Synthesis of



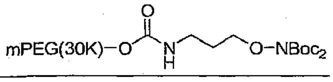
[0655] To a solution of di-Boc phthalimide (1.5 g, 3.6 mmol) in MeOH (15 mL) at 0 °C was added ammonia in MeOH (15 mL, 7 N, 105 mmol). The resultant mixture was stirred at room temperature overnight. The precipitate was filtered off and the filtrated was concentrated in vacuo. Purification of the residue by flash chromatography (silica, 10:1-6:4 EtOAc:MeOH) afforded the amine linker (0.85 g, 82%) as a white solid.

Synthesis of



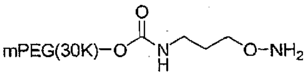
[0656] To a solution of mPEG(30K)-OH (1.0 g, 0.033 mmol) in anhydrous CH_2Cl_2 (10 mL) was added p-nitrophenol chloroformate (60 mg, 0.28 mmol). The mixture was stirred at room temperature for 15 hours. Ether (200 mL) was added. The precipitate was filtered, washed with ether and dried in vacuo to afford product (1.0 g, 100%) as a white powder.

Synthesis of



[0657] To a solution of the above activated mPEG(30K) (1.0 g, 0.033 mmol) in anhydrous GH_2Cl_2 (10 mL) were added diisopropylethylamine (58 µL, 0.33 mmol), 4-dimethylaminopyridine (5 mg, 0.041mmol) and the above di-Boc amine linker (100 mg, 0.34 mmol). The resultant mixture was stirred at room temperature for 15 hours. Ether (200 mL) was added to the reaction mixture. The precipitate was filtered, washed with ether and dried *in vacuo* to afford a white powder (0.89 g, 89%).

Synthesis of



[0658] To a solution of Boc protected mPEG(30K) (0.89 g, 0.030 mmol) in anhydrous CH_2Cl_2 (8 mL) at 0 °C was added trifluoroacetic acid (8 mL). The resultant mixture was stirred at room temperature for 5 hours. Ether (200 mL) was added to the reaction mixture. The precipitate was filtered, washed with ether and dried *is vacuo* to afford a white powder (0.65 g, 73%).

Example 64

[0659] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 51A.

Synthesis of 

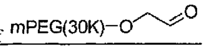
[0660] To a mixture of mPEG(30K) propionaldehyde (0.5 g, 0.0166 mmol) and amine linker (73 mg, 0.25 mmol) in MEOH (10 mL) was added $NaCNBH_3$ (20 mg, 0.30 mmol). The resultant mixture was stirred at room temperature for 48 hours. After most solvent was removed, ether (200 mL) was added. The precipitate was filtered, washed with ether and dried *in vacuo* to afford a white powder (0.43 g, 86%).

Synthesis of 

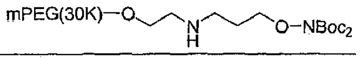
[0661] To a solution of di-Boc protected mPEG(30K) (0.42 g, 0.014 mmol) in anhydrous CH_2Cl_2 (4 mL) at 0 °C was added trifluoroacetic acid (4 mL). The resultant mixture was stirred at room temperature for 8 hours. Ether (100 mL) was added to the reaction mixture. The precipitate was filtered, washed with ether and dried *in vacuo* to afford a white powder (0.35 g, 83%).

Example 65

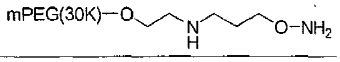
[0662] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 51B.

Synthesis of 

[0663] To a stirred solution of mPEG(30K) (6.0 g, 0.2 mmol) and pyridine (0.1 mL, 1.2 mmol) in CH_2Ch (60 mL) at 0 °C is added Dess-Martin periodinane (0.2 g, 0.47 mmol). The mixture is stirred at room temperature overnight. The reaction is then quenched with saturated aqueous $Na_2S_2O_3$-$NaHCO_3$ (1:1, 100 mL) and extracted with CH_2Cl_2 (500 mL x 2). The organic layers are combined and washed with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The residue is dissolved in CH_2Cl_2 (50 mL). Ether (1 L) is added to the solution. The precipitate is filtered, washed with ether and dried *in vacuo* to afford a white powder (4.9 g, 82%).

Synthesis of 

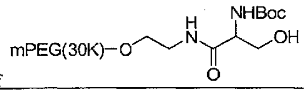
[0664] To a mixture of mPEG(30K) aldehyde (0.5 g, 0.0166 mmol) and amine linker (73 mg, 0.25 mmol) in MeOH (10 mL) is added $NaCNBH_3$ (20 mg, 0.30 mmol). The resultant mixture is stirred at room temperature for 48 hours. After most solvent is removed, ether (200 mL) is added. The precipitate is filtered, washed with ether and dried *in vacuo* to afford a white powder (0.43 g, 85%).

Synthesis of 

[0665] To a solution of di-Boc protected mPEG(30K) (0.42 g, 0.014 mmol) in anhydrous CH_2Cl_2 (4 mL) at 0 °C is added trifluoroacetic acid (4 mL). The resultant mixture is stirred at room temperature for 8 hours. Ether (100 mL) is added to the reaction mixture. The precipitate is filtered, washed with ether and dried *in vacuo* to afford a white powder (0.35 g, 83%).

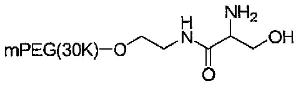
Example 66

[0666] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 52A.

Synthesis of 

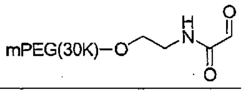
[0667] To a solution of mPEG(30K)-NH, (6.0 g, 0.2 mmol) in DMF (60 mL) are added Boc-Ser-OH (205 mg, 1.0 mmol), 1-ethyl-3-(3-dimethylammopropyl)carbodiimide hydrochloride (EDC, 190 mg, 1.0 mmol) and *N,N'*-diisopropylethylamine (0.17 mL, 1.0 mmol). The mixture is stirred at room temperature for 10 h and diluted with EtOAc (500 mL). The resultant mixture is washed successively with saturated aqueous $NaHCO_3$ (300 mL), H_2O (300 mL) and brine (300 mL), then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The residue is dissolved in CH_2Cl_2 (50 mL). Ether (700 mL) is added. The precipitate is filtered, washed with ether and dried *in vacuo* to afford a white powder (5.1 g, 82%).

Synthesis of



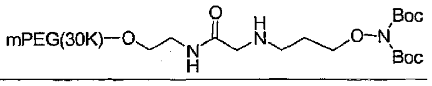
[0668] To a solution of the above mPEG(30K) (3.0 g, 0.1 mmol) in CH_2Cl_2 (15 mL) at 0 °C is added trifluoroacetic acid (15 mL). The resultant mixture is stirred at room temperature for 3 h and concentrated. To the residue is added CH_2Cl_2 (5 mL) followed by the addition of HCl (4 N in dioxane, 2 mL). Ether (400 mL) is added to precipitate dihydroxylamine (2.6 g, 85%) as a white solid.

Synthesis of



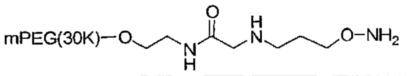
[0669] To a solution of the above mPEG(30K) (2.0 g, 0.067 mmol) in H_2O-CH_3CN (1:1, 20 mL) is added $NaIO_4$ (15 mg, 0.07 mmol). The mixture is stirred at room temperature for 4.0 h and diluted with CH_2Cl_2 (500 mL). The resultant mixture is washed with H_2O (100 mL) and brine (100 mL), then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The residue is dissolved in CH_2Cl_2 (50 mL). Ether (700 mL) is added. The precipitate is filtered, washed with ether and dried *in vacuo* to afford a white powder (1.8 g, 90%).

Synthesis of



[0670] To a mixture of mPEG(30K) aldehyde (0.5 g, 0.0166 mmol) and amine linker (73 mg, 0.25 mmol) in MeOH (10 mL) is added $NaCNBH_3$ (20 mg, 0.30 mmol). The resultant mixture is stirred at room temperature for 48 h. After most solvent is removed, ether (200 mL) was added. The precipitate is filtered, washed with ether and dried *in vacuo* to afford a white powder (0.43 g, 86%).

Synthesis of

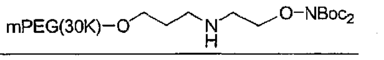


[0671] To a solution of diBoc protected mPEG(30K) (0.42 g, 0.014 mmol) in anhydrous CH_2Cl_2 (4 mL) at 0 °C is added trifluoroacetic acid (4 mL). The resultant mixture was stirred at room temperature for 8 h. Ether (100 mL) is added to the reaction mixture. The precipitate is filtered, washed with ether and dried *in vacuo* to afford a white powder (0.35 g, 83%).

Example 67

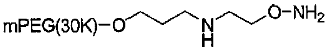
[0672] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 52B.

Synthesis of



[0673] To a mixture of mPEG propionaldehyde (30 K, 0.5 g, 0.0166 mmol) and amine linker (70 mg, 0.25 mmol) in MeOH (10 mL) was added $NaCNBH_3$ (20 mg, 0.30 mmol). The resultant mixture was stirred at room temperature for 48 hours. After most solvent was removed, ether (200 mL) was added. The precipitate was filtered, washed with ether and dried *in vacuo* to afford a white powder (0.40 g, 80%).

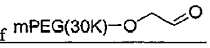
Synthesis of



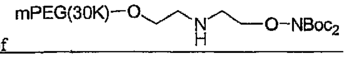
[0674] To a solution of di-Boc protected mPEG(30K) (0.40 g, 0.013 mmol) in anhydrous CH_2Cl_2 (4 mL) at 0 °C was added trifluoroacetic acid (4 mL). The resultant mixture was stirred at room temperature for 8 hours. Ether (100 mL) was added to the reaction mixture. The precipitate was filtered, washed with ether and dried *in vacuo* to afford a white powder (0.35 g, 87%).

Example 68

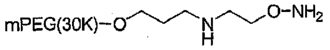
[0675] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 53A.

Synthesis of 

[0676] To a stirred solution of mPEG(30K) (6.0 g, 0.2 mmol) and pyridine (0.1 mL, 1.2 mmol) in CH_2Cl_2 (60 mL) at 0 °C is added Dess-Martin periodinane (0.2 g, 0.47 mmol). The mixture is stirred at room temperature overnight. The reaction is then quenched with saturated aqueous $Na_2S_2O_3$-$NaHCO_3$ (1:1, 100 mL) and extracted with CH_2Cl_2 (500 mL x 2). The organic layers are combined and washed with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The residue is dissolved in CH_2Cl_2 (50 mL). Ether (1 L) was added to the solution. The precipitate was filtered, washed with ether and dried *in vacuo* to afford a white powder (4.9 g, 82%).

Synthesis of 

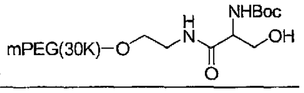
[0677] To a mixture of mPEG(30K) aldehyde (0.5 g, 0.0166 mmol) and amine linker (70 mg, 0.25 mmol) in MeOH (10 mL) is added $NaCNBH_3$ (20 mg, 0.30 mmol). The resultant mixture is stirred at room temperature for 48 hours. After most solvent is removed, ether (200 mL) is added. The precipitate is filtered, washed with ether and dried *in vacuo* to afford a white powder (0.40 g, 80%).

Synthesis of 

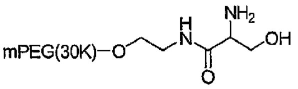
[0678] To a solution of di-Boc protected mPEG(30K) (0.40 g, 0.013 mmol) in anhydrous CH_2Cl_2 (4 mL) at 0 °C is added trifluoroacetic acid (4 mL). The resultant mixture is stirred at room temperature for 8 hours. Ether (100 mL) is added to the reaction mixture. The precipitate is filtered, washed with ether and dried *in vacuo* to afford a white powder (0.35 g, 87%).

Example 69

[0679] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 53B.

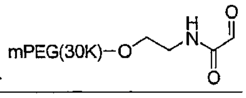
Synthesis of 

[0680] To a solution of mPEG(30K)-NH_2 (6.0 g, 0.2 mmol) in DMF (60 mL) are added Boc-Ser-OH (205 mg , 1.0 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC, 190 mg, 1.0 mmol) and *N,N'*-diisopropylethylamine (0.17 mL, 1.0 mmol). The mixture is stirred at room temperature for 10 h and diluted with EtOAc (500 mL). The resultant mixture is washed successively with saturated aqueous $NaHCO_3$ (300 mL), H_2O (300 mL) and brine (300 mL), then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The residue is dissolved in CH_2Cl_2 (50 mL). Ether (700 mL) is added. The precipitate is filtered, washed with ether and dried *in vacuo* to afford a white powder (5.1 g, 82%).

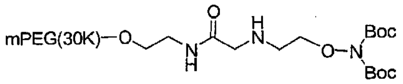
Synthesis of 

[0681] To a solution of the above mPEG(30K) (3.0 g, 0.1 mmol) in CH_2Cl_2 (15 mL) at 0 °C is added trifluoroacetic acid (15 mL).

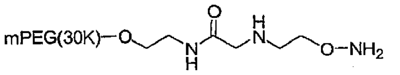
The resultant mixture is stirred at room temperature for 3 h and concentrated. To the residue is added CH_2Cl_2 (5 mL) followed by the addition of HCl (4 N in dioxane, 2 mL). Ether (400 mL) is added to precipitate dihydroxylamine (2.6 g, 85%) as a white solid.

Synthesis of 

[0682] To a solution of the above mPEG(30K) (2.0 g, 0.067 mmol) in H_2O-CH_3CN (1:1, 20 mL) is added $NaIO_4$ (15 mg, 0.07 mmol). The mixture is stirred at room temperature for 4.0 h and diluted with CH_2Cl_2 (500 mL). The resultant mixture is washed with H_2O (100 mL) and brine (100 mL), then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The residue is dissolved in CH_2Cl_2 (50 mL). Ether (700 mL) is added. The precipitate is filtered, washed with ether and dried *in vacuo* to afford a white powder (1.8 g, 90%).

Synthesis of 

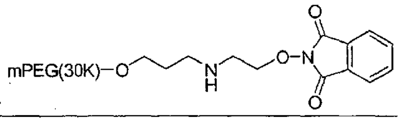
[0683] To a mixture of mPEG(30K) aldehyde (0.5 g, 0.0166 mmol) and amine linker (70 mg, 0.25 mmol) in MeOH (10 mL) is added $NaCNBH_3$ (20 mg, 0.30 mmol). The resultant mixture is stirred at room temperature for 48 h. After most solvent is removed, ether (200 mL) is added. The precipitate is filtered, washed with ether and dried *in vacuo* to afford a white powder (0.40 g, 80%).

Synthesis of 

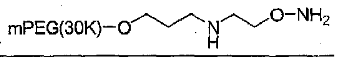
[0684] To a solution of di-Boc protected mPEG(30K) (0.40 g, 0.013 mmol) in anhydrous CH_2Cl_2 (4 mL) at 0 °C is added trifluoroacetic acid (4 mL). The resultant mixture is stirred at room temperature for 8 h. Ether (100 mL) is added to the reaction mixture. The precipitate is filtered, washed with ether and dried *in vacuo* to afford a white powder (0.35 g, 87%).

Example 70

[0685] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 54A.

Synthesis of 

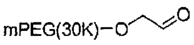
[0686] To a mixture of mPEG(30K) propionaldehyde (0.5 g, 0.0166 mmol) and amine linker (40 mg, 0.16 mmol) in MeOH (10 mL) was added $NaCNBH_3$ (12 mg, 0.17 mmol). The resultant mixture was stirred at room temperature for 60 hours. After most solvent was removed, the residue was dissolved in CH_2Cl_2 (200 mL) and washed with citric acid (5%, 100 mL). The organic layer was dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo* to afford a white solid (0.42 g, 84%).

Synthesis of 

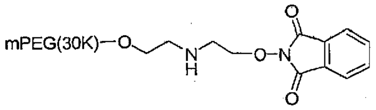
[0687] To a mixture of mPEG(30K) phthalimide (0.4 g, 0.013mmol) in MeOH (4 mL) was added H_2NNH_2 (4.2 µL, 0.13 mmol). The mixture was stirred at 45 °C for 1.0 hour and concentrated. The residue was dissolved in CH_2Cl_2 (100 mL) and washed with HCl (0.1 N, 100 mL). The aqueous layer was extracted with CH_2Cl_2 (100 mL). The organic layers were combined and washed with H_2O, then dried over anhydrous Na_2SO_4, filtered and concentrated. The residue was dissolved in CH_2Cl_2 (5 mL). Ether (200 mL) was added to precipitate the hydroxylamine product (0.34 g, 85%) as a white powder.

Example 71

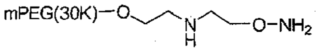
[0688] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 54B.

Synthesis of 

[0689] To a stirred solution of mPEG(30K) (6.0 g, 0.2 mmol) and pyridine (0.1 mL, 1.2 mmol) in CH_2Cl_2 (60 mL) at 0 °C is added Dess-Martin periodinane (0.2 g, 0.47 mmol). The mixture is stirred at room temperature overnight. The reaction is then quenched with saturated aqueous $Na_2S_2O_3$-$NaHCO_3$ (1:1, 100 mL) and extracted with CH_2Cl_2 (500 mL x 2). The organic layers are combined and washed with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The residue is dissolved in CH_2Cl_2 (50 mL). Ether (1 L) is added to the solution. The precipitate is filtered, washed with ether and dried *in vacuo* to afford a white powder (4.9 g, 82%).

Synthesis of 

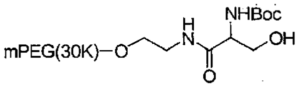
[0690] To a mixture of mPEG(30K) aldehyde (0.5 g, 0.0166 mmol) and amine linker (40 mg, 0.16 mmol) in MeOH (10 mL) is added $NaCNBH_3$ (12 mg, 0.17 mmol). The resultant mixture is stirred at room temperature for 60 hours. After most solvent is removed, the residue is dissolved in CH_2Cl_2 (200 mL) and washed with citric acid (5%, 100 mL). The organic layer is dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo* to afford a white solid (0.42 g, 84%).

Synthesis of 

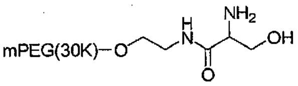
[0691] To a mixture of mPEG(30K) phthalimide (0.4 g, 0.013mmol) in MeOH (4 mL) is added H_2NNH_2 (4.2 µL, 0.13 mmol). The mixture is stirred at 45 °C for 1.0 hour and concentrated. The residue is dissolved in CH_2Cl_2 (100 mL) and washed with HCl (0.1 N, 100 mL). The aqueous layer is extracted with CH_2Cl_2 (100 mL). The organic layers are combined and washed with H_2O, then dried over anhydrous Na_2SO_4, filtered and concentrated. The residue is dissolved in CH_2Cl_2 (5 mL). Ether (200 mL) is added to precipitate the hydroxylamine product (0.34 g, 85%) as a white powder.

Example 72

[0692] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 55.

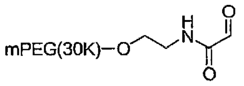
Synthesis of 

[0693] To a solution of mPEG(30K)-NH_2 (6.0 g, 0.2 mmol) in DMF (60 mL) are added Boc-Ser-OH (205 mg , 1.0 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC, 190 mg, 1.0 mmol) and *N,N'*-diisopropylethylamine (0.17 mL, 1.0 mmol). The mixture is stirred at room temperature for 10 h and diluted with EtOAc (500 mL). The resultant mixture is washed successively with saturated aqueous $NaHCO_3$ (300 mL), H_2O (300 mL) and brine (300 mL), then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The residue is dissolved in CH_2Cl_2 (50 mL). Ether (700 mL) is added. The precipitate is filtered, washed with ether and dried *in vacuo* to afford a white powder (5.1 g, 82%).

Synthesis of 

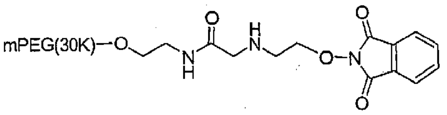
[0694] To a solution of the above mPEG(30K) (3.0 g, 0.1 mmol) in CH_2Cl_2 (15 mL) at 0 °C is added trifluoroacetic acid (15 mL). The resultant mixture is stirred at room temperature for 3 h and concentrated. To the residue is added CH_2Cl_2 (5 mL) followed by the addition of HCl (4 N in dioxane, 2 mL). Ether (400 mL) is added to precipitate dihydroxylamine (2.6 g, 85%) as a white solid.

Synthesis of



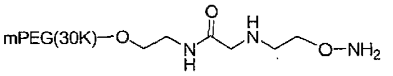
[0695] To a solution of the above mPEG(30K) (2.0 g, 0.067 mmol) in H_2O-CH_3CN (1:1, 20 mL) is added $NaIO_4$ (15 mg, 0.07 mmol). The mixture is stirred at room temperature for 4.0 h and diluted with CH_2Cl_2 (500 mL). The resultant mixture is washed with H_2O (100 mL) and brine (100 mL), then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The residue is dissolved in CH_2Cl_2 (50 mL). Ether (700 mL) is added. The precipitate is filtered, washed with ether and dried *in vacuo* to afford a white powder (1.8 g, 90%).

Synthesis of



[0696] To a mixture of mPEG(30K) aldehyde (0.5 g, 0.0166 mmol) and amine linker (40 mg, 0.16 mmol) in MeOH (10 mL) is added $NaCNBH_3$ (12 mg, 0.17 mmol). The resultant mixture is stirred at room temperature for 60 h. After most solvent is removed, the residue is dissolved in CH_2Cl_2 (200 mL) and washed with citric acid (5%, 100 mL). The organic layer is dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo* to afford a white solid (0.42 g, 84%).

Synthesis of

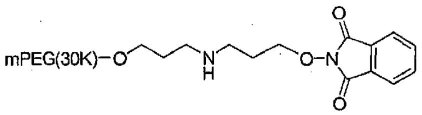


[0697] To a mixture of mPEG(30K) phthalimide (0.4 g, 0.013mmol) in MeOH (4 mL) is added H_2NNH_2 (4.2 µL, 0.13 mmol). The mixture is stirred at 45 °C for 1.0 h and concentrated. The residue is dissolved in CH_2Cl_2 (100 mL) and washed with HCl (0.1 N, 100 mL). The aqueous layer is extracted with CH_2Cl_2 (100 mL). The organic layers are combined and washed with H_2O, then dried over anhydrous Na_2SO_4, filtered and concentrated. The residue is dissolved in CH_2Cl_2 (5 mL). Ether (200 mL) is added to precipitate the hydroxylamine product (0.34 g, 85%) as a white powder.

Example 73

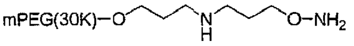
[0698] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 56A.

Synthesis of



[0699] To a mixture of mPEG(30K) propionaldehyde (0.5 g, 0.0166 mmol) and amine linker (40 mg, 0.16 mmol) in MeOH (10 mL) is added $NaCNBH_3$ (12 mg, 0.17 mmol). The resultant mixture is stirred at room temperature for 60 hours. After most solvent is removed, the residue is dissolved in CH_2Cl_2 (200 mL) and washed with citric acid (5%, 100 mL). The organic layer is dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo* to afford a white solid (0.41 g, 82%).

Synthesis of

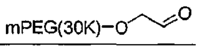


[0700] To a mixture of mPEG(30K) phthalimide (0.4 g, 0.013mmol) in MeOH (4 mL) is added H_2NNH_2 (4.2 µL, 0.13 mmol). The mixture is stirred at 45 °C for 1.0 hour and concentrated. The residue is dissolved in CH_2Cl_2 (100 mL) and washed with HCl (0.1 N, 100 mL). The aqueous layer is extracted with CH_2Cl_2 (100 mL). The organic layers are combined and washed with H_2O, then dried over anhydrous Na_2SO_4, filtered and concentrated. The residue was dissolved in CH_2Cl_2 (5 mL). Ether (200 mL) is added to precipitate the hydroxylamine product (0.34 g, 83%) as a white powder.

Example 74

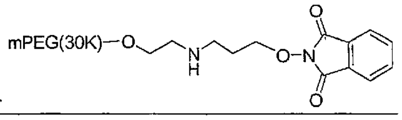
[0701] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 56B.

Synthesis of



[0702] To a stirred solution of mPEG(30K) (6.0 g, 0.2 mmol) and pyridine (0.1 mL, 1.2 mmol) in CH_2Cl_2 (60 mL) at 0 °C is added Dess-Martin periodinane (0.2 g, 0.47 mmol). The mixture is stirred at room temperature overnight. The reaction is then quenched with saturated aqueous $Na_2S_2O_3$-$NaHCO_3$ (1:1, 100 mL) and extracted with CH_2Cl_2 (500 mL x 2). The organic layers are combined and washed with H_2O and brine, then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The residue is dissolved in CH_2Cl_2 (50 mL). Ether (1 L) is added to the solution. The precipitate is filtered, washed with ether and dried *in vacuo* to afford a white powder (4.9 g, 82%).

Synthesis of



[0703] To a mixture of mPEG(30K) aldehyde (0.5 g, 0.0166 mmol) and amine linker (40 mg, 0.16 mmol) in MeOH (10 mL) is added $NaCNBH_3$ (12 mg, 0.17 mmol). The resultant mixture is stirred at room temperature for 60 hours. After most solvent is removed, the residue is dissolved in CH_2Cl_2 (200 mL) and washed with citric acid (5%, 100 mL). The organic layer is dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo* to afford a white solid (0.41 g, 82%).

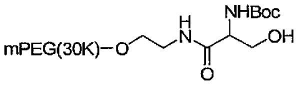
Synthesis of

[0704] To a mixture of mPEG(30K) phthalimide (0.4 g, 0.013mmol) in MeOH (4 mL) is added H_2NNH_2 (4.2 µL, 0.13 mmol). The mixture is stirred at 45 °C for 1.0 hour and concentrated. The residue is dissolved in CH_2Cl_2 (100 mL) and washed with HCl (0.1 N, 100 mL). The aqueous layer is extracted with CH_2Cl_2 (100 mL). The organic layers are combined and washed with H_2O, then dried over anhydrous Na_2SO_4, filtered and concentrated. The residue is dissolved in CH_2Cl_2 (5 mL). Ether (200 mL) is added to precipitate the hydroxylamine product (0.34 g, 83%) as a white powder.

Example 75

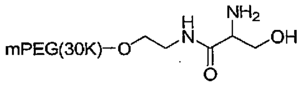
[0705] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 57.

Synthesis of



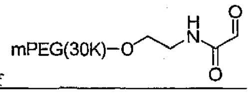
[0706] To a solution of mPEG(30K)-NH_2 (6.0 g, 0.2 mmol) in DMF (60 mL) are added Boc-Ser-OH (205 mg, 1.0 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC, 190 mg, 1.0 mmol) and *N,N'*-diisopropylethylamine (0.17 mL, 1.0 mmol). The mixture is stirred at room temperature for 10 h and diluted with EtOAc (500 mL). The resultant mixture is washed successively with saturated aqueous $NaHCO_3$ (300 mL), H_2O (300 mL) and brine (300 mL), then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The residue is dissolved in CH_2Cl_2 (50 mL). Ether (700 mL) is added. The precipitate is filtered, washed with ether and dried *in vacuo* to afford a white powder (5.1 g, 82%).

Synthesis of



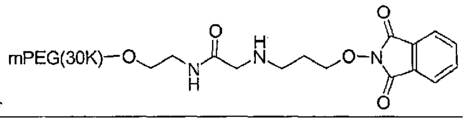
[0707] To a solution of the above mPEG(30K) (3.0 g, 0.1 mmol) in CH_2Cl_2 (15 mL) at 0 °C is added trifluoroacetic acid (15 mL). The resultant mixture is stirred at room temperature for 3 h and concentrated. To the residue is added CH_2Cl_2 (5 mL) followed by the addition of HCl (4 N in dioxane, 2 mL). Ether (400 mL) is added to precipitate dihydroxylamine (2.6 g, 85%) as a white solid.

Synthesis of



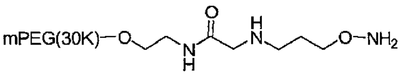
[0708] To a solution of the above mPEG(30K) (2.0 g, 0.067 mmol) in H_2O-CH_3CN (1:1, 20 mL) is added $NaIO_4$ (15 mg, 0.07 mmol). The mixture is stirred at room temperature for 4.0 h and diluted with CH_2Cl_2 (500 mL). The resultant mixture is washed with H_2O (100 mL) and brine (100 mL), then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The residue is dissolved in CH_2Cl_2 (50 mL). Ether (700 mL) is added. The precipitate is filtered, washed with ether and dried *in vacuo* to afford a white powder (1.8 g, 90%).

Synthesis of



[0709] To a mixture of mPEG(30K) aldehyde (0.5 g, 0.0166 mmol) and amine linker (40 mg, 0.16 mmol) in MeOH (10 mL) is added $NaCNBH_3$ (12 mg, 0.17 mmol). The resultant mixture is stirred at room temperature for 60 h. After most solvent is removed, the residue is dissolved in CH_2Cl_2 (200 mL) and washed with citric acid (5%, 100 mL). The organic layer is dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo* to afford a white solid (0.41 g, 82%).

Synthesis of



[0710] To a mixture of mPEG(30K) phthalimide (0.4 g, 0.013mmol) in MeOH (4 mL) is added H_2NNH_2 (4.2 µL, 0.13 mmol). The mixture is stirred at 45 °C for 1.0 h and concentrated. The residue is dissolved in CH_2Cl_2 (100 mL) and washed with HCl (0.1 N, 100 mL). The aqueous layer is extracted with CH_2Cl_2 (100 mL). The organic layers are combined and washed with H_2O, then dried over anhydrous Na_2SO_4, filtered and concentrated. The residue is dissolved in CH_2Cl_2 (5 mL). Ether (200 mL) is added to precipitate the hydroxylamine product (0.34 g, 83%) as a white powder.

Example 76

[0711] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 58A.

Synthesis of mPEG(5K)-OMs

[0712] To a solution of mPEG(5K)-OH (1.0 g, 0.2 mmol) in CH_2Cl_2 (40 mL) were added triethylamine (110 µL, 0.79 mmol) and MsCl (50 µL, 0.64 mmol). The mixture was stirred at room temperature for 10 hours and concentrated. The crude product (1.0 g) was directly used for the next step without purification.

Synthesis of mPEG(5K)-O-NHBoc

[0713] To a solution of the crude mPEG(5K)-OMs (1.0 g, 0.2 mmol) in CH_2Cl_2 (10 mL) were added *t*-butyl-N-hydroxycarbamate (0.3 g, 2.2 mmol) and triethylamine (0.4 mL, 2.9 mmol). The resultant mixture was stirred at 45 °C for 10 hours and cooled to the room temperature. Ether (200 mL) was added. The precipitate was filtered, washed and dried *in vacuo* to afford product (0.42, 42%) as a white solid.

Synthesis of mPEG(5K)-O-$NH_3^+Cl^-$

[0714] To a solution of mPEG(5K)-ONHBoc (0.2 g, 0.04 mmol) in CH_2Cl_2 (3 mL) at °C was added trifluoroacetic acid (7 mL). The resultant mixture was stirred at room temperature for 1 hour and concentrated. To the residue is added CH_2Cl_2 (5 mL) followed by the addition of HCl (4 N in dioxane, 2 mL). Ether (300 mL) is added to precipitate PEG dihydroxylamine derivative (170 mg, 85%) as a white solid.

Example 77

[0715] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 58B.

Synthesis of mPEG(30K)-OTf

[0716] To a solution of mPEG(30K)-OH (3.0 g, 0.1 mmol) in CH_2Cl_2 (30 mL) are added 2,6-lutidine (60 µL, 0.5 mmol) and Tf_2O (65 µL, 0.4 mmol). The mixture is stirred at room temperature for 10 hours. Ether (400 mL) is added to the mixture. The precipitate is filtered, washed with ether and dried *in vacuo* to afford a white powder (2.7 g, 90%).

Synthesis of mPEG(30K)-O-NHBoc

[0717] To a solution of mPEG(30K)-OTf (2.5 g, 0.083 mmol) in CH_2Cl_2 (25 mL) are added *t*-butyl *N*-hydroxycarbamate (110mg, 0.84 mmol) and diisopropylethylamine (0.2 mL, 1,1 mmol). The mixture is stirred at room temperature overnight. Ether (200 mL) is added to precipitate the product (2.2 g, 88%) as a white powder.

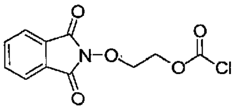
Synthesis of mPEG(30K)-O-$NH_3^+Cl^-$

[0718] To a solution of the above mPEG(30K)-ONHBoc (2.0 g, 0.067 mmol) in CH_2Cl_2 (15 mL) at 0 °C is added trifluoroacetic acid (15 mL). The resultant mixture is stirred at room temperature for 3 hours and concentrated. To the residue is added CH_2Cl_2 (5 mL) followed by the addition of HCl (4 N in dioxane, 2 mL). Ether (300 mL) is added to precipitate PEG dihydroxylamine derivative (1.72 g, 86%) as a white solid.

Example 78

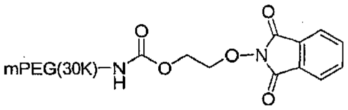
[0719] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 59A.

Synthesis of



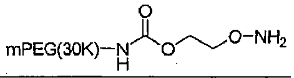
[0720] To a solution of 2-(2-hydroxyethoxy)phthalimide (0.5 g, 2.4 mmol) in CH_2Cl_2 (20 mL) was added phosgene (20% in toluene, 8.0 mL, 15.0 mmol). The reaction mixture was stirred at room temperature for 10 hour and concentrated *in vacuo*. The residue (0.45 g, 70%) was used directly for the next reaction without purification.

Synthesis of



[0721] To a mixture of mPEG(30K)-NH_2 (3 g, 0.1mmol) and chloroformate linker (0.27 g, 1.0 mmol) in CH_2Cl_2 (30 mL) was added diisopropylethylamine (0.2 mL, 1.1 mmol). The resultant mixture was stirred at room temperature for 15 hours. Ether (500 mL) was added to the mixture. The precipitate was filtered, washed and dried *in vacuo* to afford product (2.7 g, 90%) as a white solid.

Synthesis of

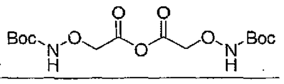


[0722] To a solution of mPEG(30K) phthalimide (2.1 g, 0.07 mmol) in MeOH (15 mL) was added ammonia (7 N in methanol, 15 mL). The resultant mixture was stirred at room temperature for 15 hours and concentrated. To the residue was added CH_2Cl_2 (5 mL) followed by the addition of HCl (4 N in dioxane, 1 mL). Ether (300 mL) was added to precipitate the product (2.4 g, 89%) as a white solid.

Example 79

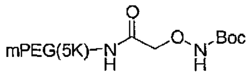
[0723] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 59B.

Synthesis of



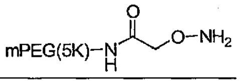
[0724] To a solution of (Boc-aminooxy)acetic acid (3.0 g, 16 mmol) in CH_2Cl_2 (80 mL) was added N-N'-diisopropylcarbodiimide (DIC, 1.3 mL, 8 mmol). The mixture was stirred at room temperature for 1 hour and concentrated *in vacuo*. The crude product (4.9 g, 84%) was directly used for the next step without further purification.

Synthesis of



[0725] To a solution of anhydride (7.3 g, 20 mmol) in DMF (20 mL) was added mPEG(5K)-NH_2 (20 g, 4 mmol). The mixture was stirred at room temperature for 10 hours and diluted with H_2O (200 mL). The mixture was extracted with CH_2Cl_2 (500 mL). The organic layer was washed with H_2O and brine (100 mL), then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. To the residue was added CH_2Cl_2 (10 mL) followed by ether (500 mL). The precipitate was filtered, washed with ether and dried *in vacuo* to afford product (19.8 g. 99%) as a white powder.

Synthesis of

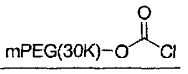


[0726] To a solution of Boc protected mPEG(5K) (1.0 g, 0.2 mmol) in CH_2Cl_2 (5 mL) was added trifluoroacetic acid (5 mL). The resultant mixture was stirred at room temperature for 1 hour and concentrated. To the residue was added CH_2Cl_2 (2 mL) followed by the addition of HCl (4 N in dioxane, 0.1 mL) and ether (40 mL). The precipitate was filtered, washed with ether and dried *in vacuo* to afford product (0.75 g, 75%).

Example 80

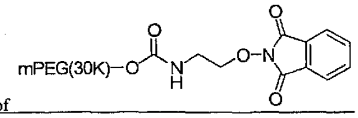
[0727] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 60A.

Synthesis of



[0728] To a solution of mPEG(30K)-OH (4.5 g, 0.15 mmol) in CH_2Cl_2 (50 mL) was added phosgene (20% in toluene, 1.6 mL, 3.0 mmol). The reaction mixture was stirred at room temperature for 10 hours and concentrated *in vacuo*. The residue (4.2 g, 93%) was used directly for the next reaction without purification.

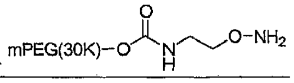
Synthesis of



[0729] To a mixture of the above activated mPEG(30 K) **II** (4.2 g, 0.14 mmol) and the amine linker **VIII** (67 mg, 0.28 mmol) in

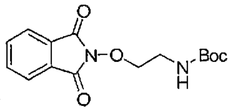
DMF-CH_2Cl_2 (10 mL, 1:2) were added diisopropylethylamine (75 µL, 0.42 mmol). After the resultant mixture was stirred at room temperature for 15 hours, Ether (200 mL) was added. The precipitate was filtered, washed with ether and dried *in vacuo* to afford product (3.8 g, 90%) as a white powder.

Synthesis of



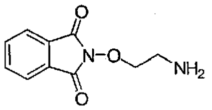
[0730] To a solution of mPEG(30K) phthalimide **III** (3.5 g, 0.12 mmol) in MeOH (15 mL) was added ammonia (7 N in methanol, 15 mL). The resultant mixture was stirred at room temperature for 15 hours and concentrated. To the residue was added CH_2Cl_2 (5 mL) followed by the addition of HCl (4 N in dioxane, 1 mL). Ether (300 mL) was added to precipitate the product (3.0 g, 86%) as a white solid.

Synthesis of



[0731] To a solution of *t*-butyl 2-hydroxyethylcarbamate (2.8 mL, 18 mmol) in THF (60 mL) were added N-hydroxyphthalimide (5.8 g, 36 mmol), triphenylphosphine (3.6 g, 27 mmol). The reaction mixture was stirred at room temperature for 10 minutes and then cooled to 0 °C. Diisopropylazodicarboxylate (DIAD, 3.6 mL, 19 mmol) was added dropwise *via* syringe over 1 hour. The icebath was removed and the mixture was stirred overnight and concentrated. The white solid dissolved in ethyl acetate (100 mL). The reaction mixture was washed successively with saturated aqueous sodium bicarbonate solution (2 X 50 mL), deionized water (50 mL) and brine (50 mL), then dried over anhydrous $MgSO_4$, filtered and concentrated *in vacuo*. The crude product was purified by flash chromatography using a Biotage Inc. HORIZON™ chromatography system to afford the title compound with impurities (12 g, 206 %) as a white solid.

Synthesis of

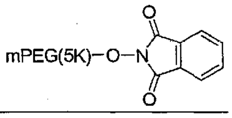


[0732] To a solution of the crude Boc-protected linker (12 g) in CH_2Cl_2 (5 mL) was added trifluoroacetic acid (5 mL). The resultant mixture was stirred at room temperature for 1 hour and concentrated. To the residue was added HCl (4 N in dioxane, 1.5 mL) followed by the addition of Et_2O (150 mL). The precipitate was filtered, washed with ether and dried *in vacuo* to afford the amine linker (3.0 g, 68 % for two steps) as a white solid.

Example 81

[0733] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 60B.

Synthesis of

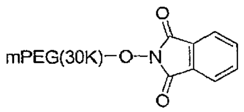


[0734] To a solution of mPEG(5K)-OH (10 g, 2.0 mmol), Ph_3P (790 mg, 3.0 mmol) and *N*-hydroxyphthalimide (0.49 g, 3.0 mmol) in CH_2Cl_2-THF (2:3, 45 mL) at 0 °C was added diisopropyl azodicarboxylate (DIAD, 409 µL, 2.0 mmol). After the mixture was stirred at room temperature for 15 hours, ether (1 L) was added to the mixture. The precipitate was filtered, washed with ether and dried *in vacuo* to afford mPEG(5K)-phthalimide (9.8 g, 98%) as a white powder.

Synthesis of mPEG(5K)-O-NH_2

[0735] To a solution of mPEG(5K) phthalimide (3 g, 0.6 mmol) in MeOH (25 mL) was added ammonia (7 N in methanol, 25 mL). The resultant mixture was stirred at room temperature for 15 hours and concentrated. To the residue was added CH_2Cl_2 (5 mL) followed by the addition of HCl (4 N in dioxane, 1 mL). Ether (300 mL) was added to precipitate hydroxylamine (2.5 g, 83%) as a white solid.

Synthesis of



[0736] To a solution of mPEG(30K)-OH (6 g, 0.2 mmol), Ph_3P (80 mg, 0.3 mmol) and *N*-hydroxyphthalimide (49 mg, 0.3 mmol) in CH_2Cl_2-THF (2:3, 45 mL) at 0 °C was added diisopropyl azodicarboxylate (DIAD, 41 µL, 0.2 mmol). The mixture was stirred at room temperature for 15 hours. Ether (200 mL) was added to the mixture. The precipitate was washed with ether and dried *in vacuo* to afford mPEG(30K)-phthalimide product (5.8 g, 96%) as a white powder.

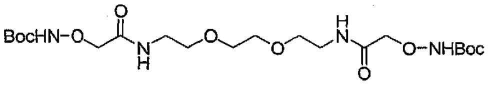
Synthesis of mPEG(30K)-O-NH_2

[0737] To a solution of mPEG(30K) phthalimide (3 g, 0.1 mmol) in MeOH (20 mL) was added ammonia (7 N in methanol, 20 mL). The resultant mixture was stirred at room temperature for 15 hours and concentrated. To the residue was added CH_2Cl_2 (5 mL) followed by the addition of HCl (4 N in dioxane, 1 mL). Ether (300 mL) was added to precipitate mPEG(30K)-ONH_2 (2.6 g, 87%) as a white solid.

Example 82

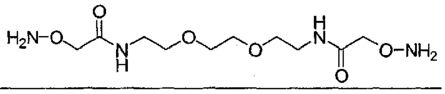
[0738] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 61A.

Synthesis of



[0739] To a solution of 2-(2-(2-aminoethoxy)ethoxy)ethanamine (5.0 g, 33.8 mmol) in DMF (100 mL) were added (Boc-aminooxy)acetic acid (14.2 g, 74.2 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC, 28.5 g, 0.15 mol) and *N,N'*-diisopropylethylamine (26 mL, 0.15 mol). The mixture was stirred at room temperature for 10 hours and diluted with EtOAc (500 mL). The resultant mixture was washed successively with saturated aqueous $NaHCO_3$ (300 mL), H_2O (300 mL) and brine (300 mL), then dried over anhydrous Na_2SO_4, filtered and concentrated *in vacuo*. The crude product (14.2 g, 85%) was used directly for the next reaction without purification.

Synthesis of

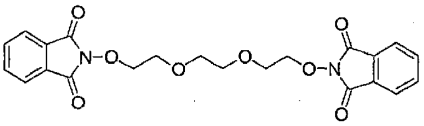


[0740] To a solution of the above diBoc linker (3.0 g, 6.1 mmol) in CH_2Cl_2 (15 mL) at 0 °C was added trifluoroacetic acid (15 mL). The resultant mixture was stirred at room temperature for 3 hours and concentrated. To the residue was added CH_2Cl_2 (5 mL) followed by the addition of HCl (4 N in dioxane, 2 mL). Ether (400 mL) was added to precipitate dihydroxylamine (1.47 g, 82%) as a white solid.

Example 83

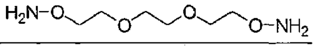
[0741] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 61B.

Synthesis of



[0742] To a solution of tri(ethylene glycol) (1.5 g, 10 mmol) in THF (100 mL) at 0 °C were added Ph_3P (8.0 g, 30 mmol) and N-hydroxylphthalimide (4.9 g, 30 mmol). To the mixture was slowly added diisopropyl azodicarboxylate (DIAD, 4.08 mL, 20 mmol). The resultant mixture was stirred at 0 °C for 4 hours and room temperature for 2 days. Ether (25 mL) was added to the reaction mixture. The precipitate was washed with ether and dried *in vacuo* to afford di-phthalimide product (3.72 g, 82%) as a white powder.

Synthesis of

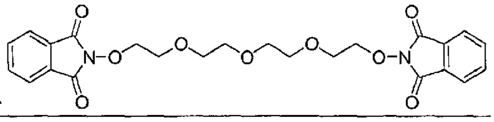


[0743] To a solution of diphthalimide (2.2 g, 5.0 mmol) in MeOH (20 mL) was added ammonia (7 N in methanol, 20 mL). The resultant mixture was stirred at room temperature for 15 hours and concentrated. To the residue was added CH_2Cl_2 (5 mL) followed by the addition of HCl (4 N in dioxane, 1 mL). Ether (300 mL) was added to precipitate tri(ethylene glycol) dihydroxylamine linker (1.1 g, 87%) as a white solid.

Example 84

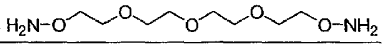
[0744] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 61C.

Synthesis of



[0745] To a solution of tetra(ethylene glycol (1.94 g, 10 mmol) in THF (100 mL) at 0 °C were added Ph_3P (8.0 g, 30 mmol) and N-hydroxylphthalinude (4.9 g, 30 mmol). To the mixture was slowly added diisopropyl azodicarboxylate (DIAD, 4.08 mL, 20 mmol). The resultant mixture was stirred at 0 °C for 4 hours and room temperature for 2 days. Ether (25 mL) was added to the reaction mixture. The precipitate was washed with ether and dried *in vacuo* to afford di-phthalimide product (3.58 g, 74%) as a white powder.

Synthesis of

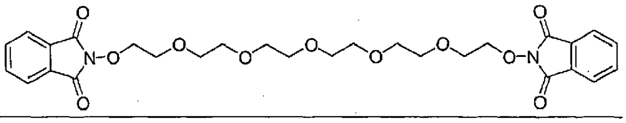


[0746] To a solution of diphthalimide (2.42 g, 5.0 mmol) in MeOH (20 mL) was added ammonia (7 N in methanol, 20 mL). The resultant mixture was stirred at room temperature for 15 hours and concentrated. To the residue was added CH_2Cl_2 (5 mL) followed by the addition of HCl (4 N in dioxane, 1 mL). Ether (300 mL) was added to precipitate tetra(ethylene glycol) dihydroxylamine linker (1.27 g, 85%) as a white solid.

Example 85

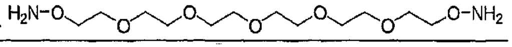
[0747] This example details the synthesis of the mPEG-hydroxylamine presented in FIG. 62A.

Synthesis of



[0748] To a solution of hexa(ethylene glycol (2.82 g, 10 mmol) in THF (100 mL) at 0 °C were added Ph_3P (8.0 g, 30 mmol) and N-hydroxylphthalimide (4.9 g, 30 mmol). To the mixture was slowly added diisopropyl azodicarboxylate (DIAD, 4.08 mL, 20 mmol). The resultant mixture was stirred at 0 °C for 4 hours and room temperature for 2 days. Ether (25 mL) was added to the reaction mixture. The precipitate was washed with ether and dried *in vacuo* to afford di-phthalimide product (4.40 g, 77%) as a white powder.

Synthesis of

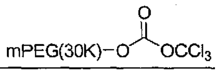


[0749] To a solution of diphthalimide (2.86 g, 5.0 mmol) in MeOH (20 mL) was added ammonia (7 N in methanol, 20 mL). The resultant mixture was stirred at room temperature for 15 hours and concentrated. To the residue was added CH_2Cl_2 (5 mL) followed by the addition of HCl (4 N in dioxane, 1 mL). Ether (300 mL) was added to precipitate hexa(ethylene glycol) dihydroxylamine linker (1.68 g, 87%) as a white solid.

Example 86

[0750] This example details the synthesis of the mPEG compound presented in FIG. 62B.

Synthesis of



$$\text{mPEG(30K)–O–C(=O)–OCCl}_3$$

[0751] To a solution of mPEG-OH (30 K, 3.0 g, 0.1 mmol) in anhydrous CH_2Cl_2 (30 mL) was added diphosgene (63 µL, 0.5 mmol). The mixture was stirred at room temperature overnight. Ether (700 mL) was added to precipitate mPEG. The product was filtered, washed with ether and dried *in vacuo* (3.0 g, 100%).

[0752] The following examples describe methods to measure and compare the *in vitro* and *in vivo* activity of a modified therapeutically active non-natural amino acid polypeptide to the in vitro and in vivo activity of a therapeutically active natural amino acid polypeptide.

Example 87: Cell Binding Assays

[0753] Cells ($3x10^6$) are incubated in duplicate in PBS/1% BSA (100 µl) in the absence or presence of various concentrations (volume: 10 µl) of unlabeled GH, hGH or GM-CSF and in the presence of ^{125}I-GH (approx. 100,000 cpm or 1. ng) at 0°C for 90 minutes (total volume: 120 µl). Cells are then resuspended and layered over 200 µl ice cold FCS in a 350 µl plastic centrifuge tube and centrifuged (1000 g; 1 minute). The pellet is collected by cutting off the end of the tube and pellet and supernatant counted separately in a gamma counter (Packard).

[0754] Specific binding (cpm) is determined as total binding in the absence of a competitor (mean of duplicates) minus binding (cpm) in the presence of 100-fold excess of unlabeled GH (non-specific binding). The non-specific binding is measured for each of the cell types used. Experiments are run on separate days using the same preparation of ^{125}I-GH and should display internal consistency. ^{125}I-GH demonstrates binding to the GH receptor-producing cells. The binding is inhibited in a dose dependent manner by unlabeled natural GH or hGH, but not by GM-CSF or other negative control. The ability of hGH to compete for the binding of natural ^{125}I-GH, similar to natural GH, suggests that the receptors recognize both forms equally well.

Example 88: In Vivo Studies of PEGylated hGH

[0755] PEG-hGH, unmodified hGH and buffer solution are administered to mice or rats. The results will show superior activity and prolonged half life of the PEGylated hGH of the present invention compared to unmodified hGH which is indicated by significantly increased bodyweight.

Example 89: Measurement of the in vivo Half-life of Conjugated and Non-conjugated hGH and Variants Thereof.

[0756] All animal experimentation is conducted in an AAALAC accredited facility and under protocols approved by the Institutional Animal Care and Use Committee of St. Louis University. Rats are housed individually in cages in rooms with a 12-hour light/dark cycle. Animals are provided access to certified Purina rodent chow 5001 and water ad libitum. For hypophysectomized rats, the drinking water additionally contains 5% glucose.

Example 90: Pharmacokinetic studies

[0757] The quality of each PEGylated mutant hGH was evaluated by three assays before entering animal experiments. The purity of the PEG-hGH was examined by running a 4-12% acrylamide NuPAGE Bis-Tris gel with MES SDS running buffer under non-reducing conditions (Invitrogen, Carlsbad, CA). The gels were stained with Coomassie blue. The PEG-hGH band was greater than 95% pure based on densitometry scan. The endotoxin level in each PEG-hGH was tested by a kinetic LAL assay using the KTA^2 kit from Charles River Laboratories (Wilmington, MA), and was less than 5 EU per dose. The biological activity of the PEG-hGH was assessed with a IM-9 pSTAT5 bioassay, and the EC_{50} value confirmed to be less than 15 nM.

[0758] Pharmacokinetic properties of PEG-modified growth hormone compounds were compared to each other and to nonPEGylated growth hormone in male Sprague-Dawley rats (261-425g) obtained from Charles River Laboratories. Catheters were surgically installed into the carotid artery for blood collection. Following successful catheter installation, animals were assigned to treatment groups (three to six per group) prior to dosing. Animals were dosed subcutaneously with 1 mg/kg of compound in a dose volume of 0.41-0.55 ml/kg. Blood samples were collected at various time points via the indwelling catheter and into EDTA-coated microfuge tubes. Plasma was collected after centrifugation, and stored at -80°C until analysis. Compound concentrations were measured using antibody sandwich growth hormone ELISA kits from either BioSource International (Camarillo, CA) or Diagnostic Systems Laboratories (Webster, TX). Concentrations were calculated using standards corresponding to the analog that was dosed. Pharmacokinetic parameters were estimated using the modeling program WinNonlin (Pharsight, version 4.1). Noncompartmental analysis with linear-up/log-down trapezoidal integration was used, and concentration data was uniformly weighted.

[0759] Plasma concentrations were obtained at regular intervals following a single subcutaneous dose in rats. Rats (n=3-6 per group) were given a single bolus dose of 1 mg/kg protein. hGH wild-type protein (WHO hGH), His-tagged hGH polypeptide (his-hGH), or His-tagged hGH polypeptide comprising non-natural amino acid p-acetyl-phenylalanine covalently linked to 30 kDa PEG at each of six different positions were compared to WHO hGH and (his)-hGH. Plasma samples were taken over the regular time intervals and assayed for injected compound as described. The table below shows the pharmacokinetic parameter values for single-dose administration of the various hGH polypeptides. Concentration vs time curves were evaluated by noncompartmental analysis (Pharsight, version 4.1). Values shown are averages (+/- standard deviation). Cmax: maximum concentration; terminal $t_{1/2}$: terminal half-life; $AUC_{0->inf}$: area under the concentration-time curve extrapolated to infinity; MRT: mean residence time; Cl/f: apparent total, plasma clearance; Vz/f apparent volume of distribution during terminal phase. 30KPEG-pAF92 (his)hGH was observed to dramatically extended circulation, increase serum half-life, and increase bioavailability compared to control hGH

Table : Pharmacokinetic parameter values for single-dose 1 mg/kg bolus s.c. administration in normal male Sprague-Dawley rats.

Compound (n)	Parameter					
	Cmax (ng/ml)	Terminal $t_{1/2}$ (h)	$AUC_{0->inf}$ (ngXhr/ ml)	MRT (h)	Cl/f (ml/hr/ kg)	Vz/f (ml/kg)
WHO hGH (3)	529 (±127)	0.53 (±0.07)	759(±178)	1.29 (±0.05)	1,368 (±327)	1051 (±279)
(his)hGH (4)	680 (±167)	0.61 (±0.05)	1,033 (±92)	1.30 (±0.17)	974 (±84)	853 (±91)
30KPEG-pAF35(his)hGH (4)	1,885 (±1,011)	4.85 (±0.80)	39,918 (±22,683)	19.16 (±4.00)	35 (±27)	268 (±236)
30KPEG-pAF92(his)hGH (6)	663 (±277)	4.51 (±0.90)	10,539 (±6,639)	15.05 (±2.07)	135 (±90)	959 (±833)
30KPEG-pAF13 (his)hGH (5)	497 (±187)	4.41 (±0.27)	6,978 (±2,573)	14.28 (±0.92)	161 (±61)	1,039 (±449)
30KPEG-pAF134(his)hGH (3)	566 (±204)	4.36 (±0.33)	7,304 (±2,494)	12.15 (±1.03)	151 (±63)	931 (±310)
30KPEG-pAF143(his)hGH (5)	803 (±149)	6.02 (±1.43)	17,494 (±3,654)	18.83 (±1.59)	59 (±11)	526 (±213)
30KPEG-pAF145(his)hGH (5)	634 (±256)	5.87 (±0.09)	13,162 (±6,726)	17.82 (±0.56)	88 (±29)	743 (±252)

Example 91: **Pharmacodynamic studies**

[0760] Hypophysectomized male Sprague-Dawley rats were obtained from Charles River Laboratories. Pituitaries were surgically removed at 3-4 weeks of age. Animals were allowed to acclimate for a period of three weeks, during which time bodyweight was monitored. Animals with a bodyweight gain of 0-8g over a period of seven days before the start of the study were included and randomized to treatment groups. Rats were administered either a bolus dose or daily dose subcutaneously. Throughout the study rats were daily and sequentially weighed, anesthetized, bled, and dosed (when applicable). Blood was collected from the orbital sinus using a heparinized capillary tube and placed into an EDTA coated microfuge tube. Plasma was isolated by centrifugation and stored at -80°C until analysis. The mean (+/- S.D.) plasma concentrations were plotted versus time intervals.

[0761] The peptide IGF-1 is a member of the family of somatomedins or insulin-like growth factors. IGF-1 mediates many of the growth-promoting effects of growth hormone. IGF-1 concentrations were measured using a competitive binding enzyme immunoassay kit against the provided rat/mouse IGF-1 standards (Diagnosic Systems Laboratories). Hypophysectomized rats. Rats (n= 5-7 per group) were given either a single dose or daily dose subcutaneously. Animals were sequentially weighed, anesthetized, bled, and dosed (when applicable) daily. Bodyweight results are taken for placebo treatments, wild type hGH (hGH), His-tagged hGH ((his)hGH), and hGH polypeptides comprising p-acetyl-phenylalanine covalently-linked to 30 kDa PEG at positions 35 and 92. The bodyweight gain at day 9 for 30KPEG-pAF35(his)hGH compound was observed to be statistically different ($p<0.0005$) from the 30KPEG-pAF92(his)hGH compound, in that greater weight gain was observed. The effect on circulating plasma IGF-1 levels after administration of a single dose of hGH polypeptides comprising a non-naturally encoded amino acid that is PEGylated, with significant difference determined by t-test using two-tailed distribution, unpaired, equal variance.

Example 92: Human Clinical Trial of the Safety and/or Efficacy of PEGylated hGH Comprising a Non-Naturally Encoded Amino Acid.

[0762] Objective To compare the safety and pharmacokinetics of subcutaneously administered PEGylated recombinant human hGH comprising a non-naturally encoded amino acid with one or more of the commercially available hGH products (including Humatrope™ (Eli Lilly & Co.), Nutropin™ (Genentech), Norditropin™ (Novo-Nordisk), Genotropin™ (Pfizer) and Saizen/Serostim™ (Serono)).

[0763] Patients Eighteen healthy volunteers ranging between 20-40 years of age and weighing between 60-90 kg are enrolled in the study. The subjects will have no clinically significant abnormal laboratory values for hematology or serum chemistry, and a negative urine toxicology screen, HIV screen, and hepatitis B surface antigen. They should not have any evidence of the following: hypertension; a history of any primary hematologic disease; history of significant hepatic, renal, cardiovascular, gastrointestinal, genitourinary, metabolic, neurologic disease; a history of anemia or seizure disorder; a known sensitivity to bacterial or mammalian-derived products, PEG, or human serum albumin; habitual and heavy consumer to beverages containing caffeine; participation in any other clinical trial or had blood transfused or donated within 30 days of study entry; had exposure to hGH within three months of study entry; had an illness within seven days of study entry; and have significant abnormalities on the pre-study physical examination or the clinical laboratory evaluations within 14 days of study entry. All subjects are evaluated for safety and all blood collections for pharmacokinetic analysis are collected as scheduled. All studies are performed with institutional ethics committee approval and patient consent.

[0764] Study Design This will be a Phase I, single-center, open-label, randomized, two-period crossover study in healthy male volunteers. Eighteen subjects are randomly assigned to one of two treatment sequence groups (nine subjects/group). GH is administered over two separate dosing periods as a bolus s.c. injection in the upper thigh using equivalent doses ot tne PEGylated hGH comprising a non-naturally encoded amino acid and the commercially available product chosen. The dose and frequency of administration of the commercially available product is as instructed in the package label. Additional dosing, dosing frequency, or other parameter as desired, using the commercially available products may be added to the study by including additional groups of subjects. Each dosing period is separated by a 14-day washout period. Subjects are confined to the study center at least 12 hours prior to and 72 hours following dosing for each of the two dosing periods, but not between dosing periods. Additional groups of subjects may be added if there are to be additional dosing, frequency, or other parameter, to be tested for the PEGylated hGH as well. Multiple formulations of GH that are approved for human use may be used in this study. Humatrope™ (Eli Lilly & Co.), Nutropin™ (Genentech), Norditropin™ (Novo-Nordisk), Genotropin™ (Pfizer) and Saizen/Serostim™ (Serono)) are commercially available GH products approved for human use. The experimental formulation of hGH is the PEGylated hGH comprising a non-naturally encoded amino acid.

[0765] Blood Sampling Serial blood is drawn by direct vein puncture before and after administration of hGH. Venous blood samples (5 mL) for determination of serum GH concentrations are obtained at about 30, 20, and 10 minutes prior to dosing (3 baseline samples) and at approximately the following times after dosing: 30 minutes and at 1, 2, 5, 8, 12, 15, 18, 24, 30, 36, 48, 60 and 72 hours. Each serum sample is divided into two aliquots. All serum samples are stored at -20°C. Serum samples are shipped on dry ice. Fasting clinical laboratory tests (hematology, serum chemistry, and urinalysis) are performed immediately prior to the initial dose on day 1, the morning of day 4, immediately prior to dosing on day 16, and the morning of day 19.

[0766] Bioanalytical Methods An ELISA kit procedure (Diagnostic Systems Laboratory [DSL], Webster TX), is used for the determination of serum GH concentrations.

[0767] Safety Determinations Vital signs are recorded immediately prior to each dosing (Days 1 and 16), and at 6, 24, 48, and 72 hours after each dosing. Safety determinations are based on the incidence and type of adverse events and the changes in clinical laboratory tests from baseline. In addition, changes from pre-study in vital sign measurements, including blood pressure, and physical examination results are evaluated.

[0768] Data Analysis Post-dose serum concentration values are corrected for pre-dose baseline GH concentrations by subtracting from each of the post-dose values the mean baseline GH concentration determined from averaging the GH levels from the three samples collected at 30, 20, and 10 minutes before dosing. Pre-dose serum GH concentrations are not included in the calculation of the mean value if they are below the quantification level of the assay. Pharmacokinetic parameters are determined from serum concentration data corrected for baseline GH concentrations. Pharmacokinetic parameters are calculated by model independent methods on a Digital Equipment Corporation VAX 8600 computer system using the latest version of the BIOAVL software. The following pharmacokinetics parameters are determined: peak serum concentration (C_{max}); time to peak serum concentration (t_{max}); area under the concentration-time curve (AUC) from time zero to the last blood sampling time (AUC_{0-72}) calculated with the use of the linear trapezoidal rule; and terminal elimination half-life ($t_{1/2}$), computed from the elimination rate constant. The elimination rate constant is estimated by linear regression of consecutive data points in the terminal linear region of the log-linear concentration-time plot. The mean, standard deviation (SD), and coefficient of variation (CV) of the pharmacokinetic parameters are calculated for each treatment. The ratio of the parameter means (preserved formulation/non-preserved formulation) is calculated.

[0769] Safety Results The incidence of adverse events is equally distributed across the treatment groups. There are no clinically significant changes from baseline or pre-study clinical laboratory tests or blood pressures, and no notable changes from pre-study in physical examination results and vital sign measurements. The safety profiles for the two treatment groups should appear similar.

[0770] Pharmacokinetic Results Mean serum GH concentration-time profiles (uncorrected for baseline GH levels) in all 18 subjects after receiving a single dose of one or more of commercially available hGH products (including Humatrope™ (Eli Lilly & Co.), Nutropin™ (Genentech), Norditropin™ (Novo-Nordisk), Genotropin™ (Pfizer) and Saizen/Serostim™ (Serono)) are compared to the PEGylated hGH comprising a non-naturally encoded amino acid at each time point measured. All subjects should have pre-dose baseline GH concentrations within the normal physiologic range. Pharmacokinetic parameters are determined from serum data corrected for pre-dose mean baseline GH concentrations and the C_{max} and t_{max} are determined. The mean t_{max} for the clinical comparator(s) chosen (Humatrope™ (Eli Lilly & Co.), Nutropin™ (Genentech), Norditropin™ (Novo-Nordisk), Genotropin™ (Pfizer), Saizen/Serostim™ (Serono)) is significantly shorter than the t_{max} for the PEGylated hGH comprising the non-naturally encoded amino acid. Terminal half-life values are significantly shorter for the commercially available hGH products tested compared with the terminal half-life for the PEGylated hGH comprising a non-naturally encoded amino acid.

[0771] Although the present study is conducted in healthy male subjects, similar absorption characteristics and safety profiles would be anticipated in other patient populations; such as male or female patients with cancer or chronic renal failure, pediatric renal failure patients, patients in autologous predeposit programs, or patients scheduled for elective surgery.

[0772] In conclusion, subcutaneously administered single doses of PEGylated hGH comprising non-naturally encoded amino acid will be safe and well tolerated by healthy male subjects. Based on a comparative incidence of adverse events, clinical laboratory values, vital signs, and physical examination results, the safety profiles of the commercially available forms of hGH and PEGylated hGH comprising non-naturally encoded amino acid will be equivalent. The PEGylated hGH comprising non-naturally encoded amino acid potentially provides large clinical utility to patients and health care providers.

Example 93: Comparison of water solubility of PEGylated hGH and non-PEGylated hGH

[0773] The water solubility of hGH wild-type protein (WHO hGH), His-tagged hGH polypeptide (his-hGH), or His-tagged hGH polypeptide comprising non-natural amino acid p-acetyl-phenylalanine covalently linked to 30 kDa PEG at position 92 are obtained by determining the quantity of the respective polypeptides which can dissolve on 100 d L of water. The quantity of PEGylated hGH is larger than the quantities for WHO hGH and hGH which shows a that PEGylation of non-natural amino acid polypeptides increases the water solubility.

Example 94: *In Vivo* Studies of modified therapeutically active non-natural amino acid polypeptide

[0774] Prostate cancer tumor xenografts are implanted into mice which are then separated into two groups. One group is treated daily with a modified therapeutically active non-natural amino acid polypeptide and the other group is treated daily with therapeutically active natural amino acid polypeptide. The tumor size is measured daily and the modified therapeutically active non-natural amino acid polypeptide has improved therapeutic effectiveness compared to the therapeutically active natural amino acid polypeptide as indicated by a decrease in tumor size for the group treated with the modified therapeutically active non-natural amino acid polypeptide.

Example 95: *In Vivo* Studies of modified therapeutically active non-natural amino acid polypeptide

[0775] Prostate cancer tumor xenografts are implanted into mice which are then separated into two groups. One group is treated daily with a modified therapeutically active non-natural amino acid polypeptide and the other group is treated daily with therapeutically active natural amino acid polypeptide. The tumor size is measured daily and the modified therapeutically active non-natural amino acid polypeptide has improved therapeutic effectiveness compared to the therapeutically active natural amino acid polypeptide as indicated by a decrease in tumor size for the group treated with the modified therapeutically active non-natural amino acid polypeptide.

SEQUENCE LISTING

[0776]

<110> AMBRX, INC.

<120> COMPOSITIONS CONTAINING, METHODS INVOLVING, AND USES OF NON-NATURAL AMINO ACIDS AND POLYPEPTIDES

<130> P047064EP

<140> EP 05855215.9
<141> 2005-12-21

<150> US 60/638,418
<151> 2004-12-22

<150> US 60/638,527
<151> 2004-12-22

<150> US 60/639,195
<151> 2004-12-22

<150> US 60/696,210
<151> 2005-07-01

<150> US 60/696,302
<151> 2005-07-01

<150> US 60/696,068
<151> 2005-07-01

<160> 17

<170> PatentIn ver. 3.3

<210> 1
<211> 77
<212> DNA
<213> Methanococcus jannaschii

<400> 1

```
ccggcggtag ttcagcaggg cagaacggcg gactctaaat ccgcatggcg ctggttcaaa 60
tccggcccgc cggacca                                                77
```

<210> 2
<211> 88
<212> DNA
<213> Halobacterium species NRC-I

<400> 2

```
cccagggtag ccaagctcgg ccaacggcga cggactctaa atccgttctc gtaggagttc 60
gagggttcga atcccttccc tgggacca                                    88
```

<210> 3
<211> 89
<212> DNA
<213> Halobacterium species NRC-I

<400> 3

```
gcgagggtag ccaagctcgg ccaacggcga cggacttcct aatccgttct cgtaggagtt 60
cgagggttcg aatccctccc ctcgcacca                                   89
```

<210> 4
<211> 306
<212> PRT
<213> Methanococcus jannaschii

<400> 4

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

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

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

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

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

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

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Thr Phe Gln Leu Asp Lys
            100                 105                 110

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

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

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

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

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
            180                 185                 190

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

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

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

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

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

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

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

Arg Leu
305
```

<210> 5
<211> 306
<212> PRT
<213> Methanococcus jannaschii

<400> 5

```
Met Asp Glu Phe Glu Met Ile Lys Arg Asn Thr Ser Glu Ile Ile Ser
```

```
 1               5                   10                  15
Glu Glu Glu Leu Arg Glu Val Leu Lys Lys Asp Glu Lys Ser Ala Gly
             20                  25                  30
Ile Gly Phe Glu Pro Ser Gly Lys Ile His Leu Gly His Tyr Leu Gln
         35                  40                  45
Ile Lys Lys Met Ile Asp Leu Gln Asn Ala Gly Phe Asp Ile Ile Ile
     50                  55                  60
Leu Leu Ala Asp Leu His Ala Tyr Leu Asn Gln Lys Gly Glu Leu Asp
 65                  70                  75                  80
Glu Ile Arg Lys Ile Gly Asp Tyr Asn Lys Lys Val Phe Glu Ala Met
                 85                  90                  95
Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Ser Phe Gln Leu Asp Lys
            100                 105                 110
Asp Tyr Thr Leu Asn Val Tyr Arg Leu Ala Leu Lys Thr Thr Leu Lys
        115                 120                 125
Arg Ala Arg Arg Ser Met Glu Leu Ile Ala Arg Glu Asp Glu Asn Pro
    130                 135                 140
Lys Val Ala Glu Val Ile Tyr Pro Ile Met Gln Val Asn Thr Ser His
145                 150                 155                 160
Tyr Leu Gly Val Asp Val Ala Val Gly Gly Met Glu Gln Arg Lys Ile
                165                 170                 175
His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
            180                 185                 190
Asn Pro Val Leu Thr Gly Leu Asp Gly Glu Gly Lys Met Ser Ser Ser
        195                 200                 205
Lys Gly Asn Phe Ile Ala Val Asp Asp Ser Pro Glu Glu Ile Arg Ala
    210                 215                 220
Lys Ile Lys Lys Ala Tyr Cys Pro Ala Gly Val Val Glu Gly Asn Pro
225                 230                 235                 240
Ile Met Glu Ile Ala Lys Tyr Phe Leu Glu Tyr Pro Leu Thr Ile Lys
                245                 250                 255
Arg Pro Glu Lys Phe Gly Gly Asp Leu Thr Val Asn Ser Tyr Glu Glu
            260                 265                 270
Leu Glu Ser Leu Phe Lys Asn Lys Glu Leu His Pro Met Asp Leu Lys
        275                 280                 285
Asn Ala Val Ala Glu Glu Leu Ile Lys Ile Leu Glu Pro Ile Arg Lys
    290                 295                 300
Arg Leu
305
```

<210> 6

<211> 304

<212> PRT

<213> Methanococcus jannaschii

<400> 6

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

```
Glu Glu Glu Leu Arg Glu Val Leu Lys Lys Asp Glu Lys Ala Ala Ile
             20                  25                  30

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

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

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

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

Leu Lys Ala Lys Tyr Val Tyr Gly Ser Pro Phe Gln Leu Asp Lys Asp
            100                 105                 110

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

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

Val Ala Glu Val Ile Tyr Pro Ile Met Gln Val Asn Ala Ile Tyr Leu
145                 150                 155                 160

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

Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His Asn Pro
            180                 185                 190

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

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

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

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

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

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

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

<210> 7
<211> 305
<212> PRT
<213> Methanococcus jannaschii

<400> 7

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

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

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

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

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

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

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Pro Phe Gln Leu Asp Lys
            100                 105                 110

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

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

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

Leu Pro Val Asp Val Ala Val Gly Gly Met Glu Gln Arg Lys Ile His
                165                 170                 175

Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His Asn
            180                 185                 190

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

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

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

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

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

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

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

Leu
305
```

<210> 8
<211> 305
<212> PRT
<213> Methanococcus jannaschii

<400> 8

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

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

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

Ile Lys Lys Met Ile Asp Leu Gln Asn Ala Gly Phe Asp Ile Ile Ile
```

```
     50                  55                  60

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

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

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Lys Phe Gln Leu Asp Lys
            100                 105                 110

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

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

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

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

Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His Asn
            180                 185                 190

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

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

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

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

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

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

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

Leu
305
```

<210> 9

<211> 306

<212> PRT

<213> Methanococcus jannaschii

<400> 9

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

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

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

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

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

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

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Asn Phe Gln Leu Asp Lys
100 105 110

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

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

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

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

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
180 185 190

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

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

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

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

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

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

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

Arg Leu
305

<210> 10
<211> 306
<212> PRT
<213> Methanococcus jannaschii

<400> 10
Met Asp Glu Phe Glu Met Ile Lys Arg Asn Thr Ser Glu Ile Ile Ser
1 5 10 15

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

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

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

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

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

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Ser Phe Gln Leu Asp Lys
            100                 105                 110

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

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

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

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

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
            180                 185                 190

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

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

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

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

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

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

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

Arg Leu
305
```

<210> 11

<211> 306

<212> PRT

<213> Methanococcus jannaschii

<400> 11

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

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

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

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

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

Glu Ile Arg Lys Ile Gly Asp Tyr Asn Lys Lys Val Phe Glu Ala Met
```

85 90 95

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Thr Phe Gln Leu Asp Lys
100 105 110

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

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

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

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

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
180 185 190

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

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

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

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

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

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

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

Arg Leu
305

<210> 12

<211> 306

<212> PRT

<213> Methanococcus jannaschii

<400> 12

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

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

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

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

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

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

```
Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Ser Phe Gln Leu Asp Lys
            100                 105                 110

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

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

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

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

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
            180                 185                 190

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

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

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

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

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

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

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

Arg Leu
305
```

<210> 13

<211> 306

<212> PRT

<213> Methanococcus jannaschii

<400> 13

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

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

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

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

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

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

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Glu Phe Gln Leu Asp Lys
            100                 105                 110
```

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

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

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

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

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
            180                 185                 190

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

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

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

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

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

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

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

Arg Leu
305
```

<210> 14
<211> 306
<212> PRT
<213> Methanococcus jannaschii

<400> 14

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

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

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

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

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

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

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

Asp Tyr Thr Leu Asn Val Tyr Arg Leu Ala Leu Lys Thr Thr Leu Lys
```

```
        115                 120                 125

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

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

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

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
            180                 185                 190

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

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

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

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

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

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

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

Arg Leu
305
```

<210> 15
<211> 306
<212> PRT
<213> Methanococcus jannaschii

<400> 15

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

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

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

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

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

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

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

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

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

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

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

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
            180                 185                 190

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

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

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

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

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

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

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

Arg Leu
305
```

<210> 16
<211> 306
<212> PRT
<213> Methanococcus jannaschii

<400> 16

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

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

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

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

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

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

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Arg Phe Gln Leu Asp Lys
            100                 105                 110

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

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

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

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

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
            180                 185                 190

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

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

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

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

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

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

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

Arg Leu
305
```

<210> 17
<211> 306
<212> PRT
<213> Methanococcus jannaschii

<400> 17

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

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

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

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

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

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

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Thr Phe Gln Leu Asp Lys
            100                 105                 110

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

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

Lys Val Ala Glu Val Ile Tyr Pro Ile Met Gln Val Asn Thr Tyr Tyr

145                 150                 155                 160

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

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
            180                 185                 190

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

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

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

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

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

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

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

Arg Leu
305
```

SEQUENCE LISTING

<110> AMBRX, INC.

<120> COMPOSITIONS CONTAINING, METHODS INVOLVING, AND USES OF NON-NATURAL AMINO ACIDS AND POLYPEPTIDES

<130> P047064EP

<140> EP 05855215.9
<141> 2005-12-21

<150> US 60/638,418
<151> 2004-12-22

<150> US 60/638,527
<151> 2004-12-22

<150> US 60/639,195
<151> 2004-12-22

<150> US 60/696,210

<151> 2005-07-01

<150> US 60/696,302
<151> 2005-07-01

<150> US 60/696,068
<151> 2005-07-01

<160> 17

<170> PatentIn ver. 3.3

<210> 1
<211> 77
<212> DNA
<213> Methanococcus jannaschii

<400> 1

```
ccggcggtag ttcagcaggg cagaacggcg gactctaaat ccgcatggcg ctggttcaaa 60
tccggcccgc cggacca                                                77
```

<210> 2
<211> 88
<212> DNA
<213> Halobacterium species NRC-I

<400> 2

```
cccagggtag ccaagctcgg ccaacggcga cggactctaa atccgttctc gtaggagttc 60
gagggttcga atcccttccc tgggacca                                    88
```

<210> 3
<211> 89
<212> DNA
<213> Halobacterium species NRC-I

<400> 3

```
gcgagggtag ccaagctcgg ccaacggcga cggacttcct aatccgttct cgtaggagtt 60
cgagggttcg aatccctccc ctcgcacca                                   89
```

<210> 4
<211> 306
<212> PRT
<213> Methanococcus jannaschii

<400> 4

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

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

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

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

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

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

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Thr Phe Gln Leu Asp Lys
            100                 105                 110

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

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

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

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

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
            180                 185                 190

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

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

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

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

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

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

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

Arg Leu
305
```

<210> 5
<211> 306
<212> PRT
<213> Methanococcus jannaschii

<400> 5

```
Met Asp Glu Phe Glu Met Ile Lys Arg Asn Thr Ser Glu Ile Ile Ser
```

```
 1               5                  10                  15

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

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

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

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

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

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Ser Phe Gln Leu Asp Lys
            100                 105                 110

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

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

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

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

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
            180                 185                 190

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

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

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

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

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

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

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

Arg Leu
305
```

<210> 6
<211> 304
<212> PRT
<213> Methanococcus jannaschii

<400> 6

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

```
Glu Glu Glu Leu Arg Glu Val Leu Lys Lys Asp Glu Lys Ala Ala Ile
             20                  25                  30

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

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

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

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

Leu Lys Ala Lys Tyr Val Tyr Gly Ser Pro Phe Gln Leu Asp Lys Asp
            100                 105                 110

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

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

Val Ala Glu Val Ile Tyr Pro Ile Met Gln Val Asn Ala Ile Tyr Leu
145                 150                 155                 160

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

Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His Asn Pro
            180                 185                 190

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

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

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

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

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

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

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

<210> 7
<211> 305
<212> PRT
<213> Methanococcus jannaschii

<400> 7

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

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

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

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

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

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

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Pro Phe Gln Leu Asp Lys
            100                 105                 110

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

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

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

Leu Pro Val Asp Val Ala Val Gly Gly Met Glu Gln Arg Lys Ile His
                165                 170                 175

Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His Asn
            180                 185                 190

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

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

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

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

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

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

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

Leu
305
```

<210> 8

<211> 305

<212> PRT

<213> Methanococcus jannaschii

<400> 8

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

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

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

Ile Lys Lys Met Ile Asp Leu Gln Asn Ala Gly Phe Asp Ile Ile Ile
```

```
     50                  55                  60

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

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

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Lys Phe Gln Leu Asp Lys
            100                 105                 110

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

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

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

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

Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His Asn
            180                 185                 190

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

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

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

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

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

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

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

Leu
305
```

<210> 9
<211> 306
<212> PRT
<213> Methanococcus jannaschii

<400> 9

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

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

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

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

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

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

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Asn Phe Gln Leu Asp Lys
            100                 105                 110

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

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

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

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

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
            180                 185                 190

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

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

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

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

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

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

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

Arg Leu
305
```

<210> 10
<211> 306
<212> PRT
<213> Methanococcus jannaschii

<400> 10

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

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

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

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

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

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

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Ser Phe Gln Leu Asp Lys
            100                 105                 110

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

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

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

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

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
            180                 185                 190

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

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

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

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

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

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

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

Arg Leu
305
```

<210> 11
<211> 306
<212> PRT
<213> Methanococcus jannaschii

<400> 11

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

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

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

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

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

Glu Ile Arg Lys Ile Gly Asp Tyr Asn Lys Lys Val Phe Glu Ala Met
```

```
                 85                  90                  95

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Thr Phe Gln Leu Asp Lys
            100                 105                 110

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

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

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

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

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
            180                 185                 190

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

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

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

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

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

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

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

Arg Leu
305
```

<210> 12

<211> 306

<212> PRT

<213> Methanococcus jannaschii

<400> 12

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

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

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

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

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

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

```
Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Ser Phe Gln Leu Asp Lys
            100                 105                 110

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

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

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

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

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
            180                 185                 190

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

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

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

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

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

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

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

Arg Leu
305
```

<210> 13

<211> 306

<212> PRT

<213> Methanococcus jannaschii

<400> 13

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

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

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

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

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

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

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Glu Phe Gln Leu Asp Lys
            100                 105                 110
```

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

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

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

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

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
            180                 185                 190

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

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

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

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

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

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

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

Arg Leu
305
```

<210> 14

<211> 306

<212> PRT

<213> Methanococcus jannaschii

<400> 14

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

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

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

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

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

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

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

Asp Tyr Thr Leu Asn Val Tyr Arg Leu Ala Leu Lys Thr Thr Leu Lys
```

```
        115                 120                 125

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

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

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

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
            180                 185                 190

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

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

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

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

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

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

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

Arg Leu
305
```

<210> 15

<211> 306

<212> PRT

<213> Methanococcus jannaschii

<400> 15

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

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

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

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

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

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

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

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

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

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

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

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
            180                 185                 190

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

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

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

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

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

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

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

Arg Leu
305
```

<210> 16

<211> 306

<212> PRT

<213> Methanococcus jannaschii

<400> 16

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

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

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

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

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

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

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Arg Phe Gln Leu Asp Lys
            100                 105                 110

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

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

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

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

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
            180                 185                 190

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

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

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

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

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

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

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

Arg Leu
305
```

<210> 17

<211> 306

<212> PRT

<213> Methanococcus jannaschii

<400> 17

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

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

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

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

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

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

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Thr Phe Gln Leu Asp Lys
            100                 105                 110

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

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

Lys Val Ala Glu Val Ile Tyr Pro Ile Met Gln Val Asn Thr Tyr Tyr


145                 150                 155                 160

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

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
            180                 185                 190

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

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

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

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

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

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

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

Arg Leu
305
```

REFERENCES CITED IN THE DESCRIPTION

Patent documents cited in the description

- EP188256A **[0048]**
- US4671958A **[0048]**
- US4659839A **[0048]**
- US4414148A **[0048]**
- US4699784A **[0048]**
- US4680338A **[0048]**
- US4569789A **[0048]**
- US5252714A **[0153]** **[0418]**
- US6608183B **[0192]** **[0243]**
- US2004198637A **[0206]**
- WO2002085923A **[0208]**
- US6521427B **[0224]**
- US20030082575A **[0245]** **[0247]** **[0247]** **[0248]** **[0251]** **[0251]**
- US10126927B **[0245]** **[0247]** **[0248]**
- US20030108885A **[0245]** **[0247]** **[0247]** **[0248]** **[0251]**
- US10126931B **[0245]** **[0247]** **[0248]** **[0251]**
- WO10126927A **[0251]**
- USNO10126927A **[0251]**
- WO04035743A **[0251]**
- WO04094593A **[0251]**
- US5324639A **[0278]** **[0284]** **[0285]**
- US4929555A **[0278]**
- US4837148A **[0278]**
- EP0244234A **[0278]**
- WO9100357A **[0278]**
- EP0284044A **[0279]**
- EP0329203A **[0279]** **[0283]**
- EP0073657A **[0279]**
- US4880734A **[0280]**
- US4876197A **[0280]**
- EP0164556A **[0280]** **[0283]**
- US20020055169A **[0283]**
- US6361969B **[0283]**
- US6312923B **[0283]**
- US6183985B **[0283]**
- US6083723A **[0283]**
- US6017731A **[0283]**
- US5674706A **[0283]**
- US5629203A **[0283]**
- US5602034A **[0283]**
- US5089398A **[0283]**
- USRE37343E **[0283]**
- USRE35749E **[0283]**
- WO9907862A **[0283]**

- WO9837208A **[0283]**
- WO9826080A **[0283]**
- EP0946736A **[0283]**
- EP0732403A **[0283]**
- EP0480480A **[0283]**
- WO9010277A **[0283]**
- EP0460071A **[0283]**
- EP0340986A **[0283]**
- EP0324274A **[0283]**
- US5231178A **[0285]**
- US6368825B **[0290]**
- US6342216B **[0290]**
- US6338846B **[0290]**
- US6261805B **[0290]**
- US6245528B **[0290]**
- US6225060B **[0290]**
- US6183987B **[0290]**
- US6168932B **[0290]**
- US6126944A **[0290]**
- US6096304A **[0290]**
- US6013433A **[0290]**
- US5965393A **[0290]**
- US5939285A **[0290]**
- US5891676A **[0290]**
- US5871986A **[0290]**
- US5861279A **[0290]**
- US5858368A **[0290]**
- US5843733A **[0290]**
- US5762939A **[0290]**
- US5753220A **[0290]**
- US5605827A **[0290]**
- US5583023A **[0290]**
- US5571709A **[0290]**
- US5516657A **[0290]**
- US5290686A **[0290]**
- WO0206305A **[0290]**
- WO0190390A **[0290]**
- WO0127301A **[0290]**
- WO0105956A **[0290]**
- WO0055345A **[0290]**
- WO0020032A **[0290]**
- WO9951721A **[0290]**
- WO9945130A **[0290]**
- WO9931257A **[0290]**
- WO9910515A **[0290]**
- WO9909193A **[0290]**
- WO9726332A **[0290]**
- WO9629400A **[0290]**
- WO9625496A **[0290]**
- WO9606161A **[0290]**
- WO9520672A **[0290]**
- WO9303173A **[0290]**
- WO9216619A **[0290]**
- WO9203628A **[0290]**
- WO9201801A **[0290]**
- WO9014428A **[0290]**
- WO9010078A **[0290]**
- WO9002566A **[0290]**

- WO9002186A **[0290]**
- WO9001556A **[0290]**
- WO8901038A **[0290]**
- WO8901037A **[0290]**
- WO8807082A **[0290]**
- EP0127839A **[0297]**
- EP0155476A **[0297]**
- WO89046699A **[0299]**
- US4738921A **[0303]**
- US036776A **[0303]**
- US121775A **[0303]**
- US4689406A **[0303]**
- WO9905297A **[0303]**
- US4551433A **[0304]**
- US267851A **[0304]**
- US4511502A **[0322]**
- US4511503A **[0322]**
- US4512922A **[0322]**
- US686944A **[0352]**
- US6184344B **[0353]**
- US20040138412A **[0353]**
- US20030208046A **[0353]**
- WO02098902A **[0353]**
- WO03042235A **[0353]**
- US6586207B **[0354]**
- US20020042097A **[0354]**
- US6337191B **[0365]**
- US20020081660A **[0365]**
- WO0055353A **[0365]**
- WO9005785A **[0365]**
- US5281698A **[0418]** **[0442]**
- US5468478A **[0418]**
- US5672662A **[0418]**
- US4670417A **[0418]**
- US5650234A **[0418]**
- US5824784A **[0418]**
- US5900461A **[0418]**
- US5932462A **[0425]**
- US5643575A **[0425]**
- US5229490A **[0425]**
- US4289872A **[0425]**
- US20030143596A **[0425]**
- WO9621469A **[0425]**
- WO9321259A **[0425]**
- WO9417039A **[0442]**
- US5324844A **[0442]**
- WO9418247A **[0442]**
- WO9404193A **[0442]**
- US5219564A **[0442]**
- US5122614A **[0442]**
- WO9013540A **[0442]**
- WO9315189A **[0442]**
- WO9415625A **[0442]**
- WO9409027A **[0442]**
- US4412989A **[0442]**
- US6235710B **[0501]**
- US6001800A **[0501]**

- US4820352A [0502]
- EP133988A [0503]
- US3773919A [0503]
- EP58481A [0503]
- DE3218121 [0503] [0504]
- EP52322A [0503] [0504]
- EP36676A [0503] [0504]
- EP88046A [0503] [0504]
- EP143949A [0503] [0504]
- EP142641A [0503] [0504]
- JP58118008A [0503] [0504]
- US4485045A [0503] [0504]
- US4544545A [0503] [0504]
- EP102324A [0503] [0504]
- US126927A [0540]
- US126931A [0540]
- EP05855215A [0776] [0776]
- US60638418B [0776] [0776]
- US60638527B [0776] [0776]
- US60639195B [0776] [0776]
- US60696210B [0776] [0776]
- US60696302B [0776] [0776]
- US60696068B [0776] [0776]

Non-patent literature cited in the description

- **WANG LEI et al.**PNAS, 2003, [0003]
- **MAYNARDGEORGIOU**Annu. Rev. Biomed. Eng., 2000, vol. 2, 339-76 [0044]
- **HUDSON**Curr. Opin. Biotechnol., 1998, vol. 9, 395-402 [0044]
- **S-Z HU et al.**Cancer Research, 1996, vol. 56, 3055-3061 [0044]
- **SMITHWATERMAN**Adv. Appl. Math., 1970, vol. 2, 482c- [0063]
- **NEEDLEMANWUNSCH**J. Mol. Biol., 1970, vol. 48, 443- [0063]
- **PEARSONLIPMAN**Proc. Nat'l. Acad. Sci. USA, 1988, vol. 85, 2444- [0063]
- **AUSUBEL et al.**Current Protocols in Molecular Biology, 1995, [0063]
- **ALTSCHUL et al.**Nuc. Acids Res., 1997, vol. 25, 3389-3402 [0064]
- **ALTSCHUL1990 et al.**J. Mol. Biol., vol. 215, 403-410 [0064]
- **HENIKOFFHENIKOFF**Proc. Natl. Acad. Sci. USA, 1992, vol. 89, 10915- [0064]
- **KARLINALTSCHUL**Proc. Natl. Acad. Sci. USA, 1993, vol. 90, 5873-5787 [0065]
- **CREIGHTON**Proteins:Structures and Molecular PropertiesW H Freeman & Co.19931200 [0068]
- **BATZER et al.**Nucleic Acid Res., 1991, vol. 19, 5081- [0112]
- **OHTSUKA et al.**J. Biol. Chem., 1985, vol. 260, 2605-2608 [0112]
- **ROSSOLINI et al.**Mol. Cell. Probes, 1994, vol. 8, 91-98 [0112]
- **GREENEWUTS**Protective Groups in Organic SynthesisJohn Wiley & Sons19990000 [0128]
- **TIJSSEN**Overview of principles of hybridization and the strategy of nucleic acid assaysLaboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes, 1993, [0142]
- **L. WANG et al.**Science, 2001, vol. 292, 498-500 [0167]
- **J. CHIN et al.**Science, 2003, vol. 301, 964-7 [0167]
- **J. W. CHIN et al.**Journal of the American Chemical Society, 2002, vol. 124, 9026-9027 [0167]
- **J. W. CHINP. G. SCHULTZ**ChemBioChem, 2002, vol. 3, 111135-1137 [0167]
- **J. W. CHIN et al.**PNAS United States of America, 2002, vol. 99, 1711020-11024 [0167]
- **L. WANGP. G. SCHULTZ**Chem. Comm., 2002, 1-11 [0167]
- **BAZAN, F.**Immunology Today, 1990, vol. 11, 350-354 [0190]
- **BAZAN, J. F.**Science, 1992, vol. 257, 410-411 [0190]
- **MOTT, H. R.CAMPBELL, I. D.**Current Opinion in Structural Biology, 1995, vol. 5, 114-121 [0190]

- **SILVENNOINEN, O.LHLE, J. N.**Signalling by the Hematopoietic Cytokine Receptors, 1996, [0190]
- **ZINK et al.**FEBS Lett., 1992, vol. 314, 435- [0191]
- **ZINK et al.**Biochemistry, 1994, vol. 33, 8453- [0191]
- **HILL et al.**Proc. Natl. Acad. Sci.USA, 1993, vol. 90, 5167- [0191]
- **DIEDERICHS, K. et al.**Science, 1991, vol. 154, 1779-1782 [0191]
- **WALTER et al.**J. Mol. Biol., 1992, vol. 224, 1075-1085 [0191]
- **BAZAN, J: F.MCKAY, D. B.**Science, 1992, vol. 257, 410-413 [0191]
- **REDFIELD et al.**Biochemistry, 1991, vol. 30, 11029-11035 [0191]
- **POWERS et al.**Science, 1992, vol. 256, 1673-1677 [0191]
- **MILBURN et al.**Nature, 1993, vol. 363, 172-176 [0191]
- **LEE et al.**J. Interferon Cytokine Res., 1995, vol. 15, 341- [0191]
- **MURGOLO et al.**Proteins, 1993, vol. 17, 62- [0191]
- **RADHAKRISHNAN et al.**Structure, 1996, vol. 4, 1453- [0191]
- **KLAUS et al.**J. Mol. Biol., 1997, vol. 274, 661- [0191]
- **MOTTCAMPBELL**Current Opinion in Structural Biology, 1995, vol. 5, 114-121 [0191] [0192] [0192]
- **DAVIS et al.**Science, 1993, vol. 260, 1805-1808 [0192]
- **PAONESSA et al.**EMBO J., 1995, vol. 14, 1942-1951 [0192]
- **MATTHEWS et al.**Proc. Natl. Acad. Sci. USA, 1996, vol. 93, 9471-9476 [0192]
- **SHAO, J.TAM, J.**J. Am. Chem. Soc., 1995, vol. 117, 3893-3899 [0201]
- **H. HANGC. BERTOZZI**Acc. Chem. Res., 2001, vol. 34, 9727-736 [0201]
- **LIU, D.R.SCHULTZ, P.G.**Progress toward the evolution of an organism with an expanded genetic codePNAS United States, 1999, vol. 96, 4780-4785 [0206]
- **STEMMER**Rapid evolution of a protein in vitro by DNA shufflingNature, 1994, vol. 370, 4389-391 [0209]
- **STEMMER**DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolutionProc. Natl. Acad. Sci. USA., 1994, vol. 91, 10747-10751 [0209]
- **DOUGHERTY**Unnatural Amino Acids as Probes of Protein Structure and FunctionCurrent Opinion in Chemical Biology, 2000, vol. 4, 645-652 [0216] [0368]
- **BARANY et al.**Proc. Natl. Acad. Sci., 1991, vol. 88, 189-193 [0224]
- **SAMBROOK et al.**Molecular Cloning, A Laboratory Manual20010000 [0225]
- **KRIEGLER**Gene Transfer and Expression: A Laboratory Manual, 1990, [0225]
- Current Protocols in Molecular Biology19940000 [0225]
- **BERGERKIMMEL**Guide to Molecular Cloning Techniques, Methods in EnzymologyAcademic Press, Inc.vol. 152, [0226]
- **SAMBROOK et al.**Molecular Cloning - A Laboratory ManualCold Spring Harbor Laboratory19890000vol. 1-3, [0226]
- Current Protocols in Molecular BiologyGreene Publishing Associates, Inc. and John Wiley & Sons, Inc.19990000 [0226]
- **LING et al.**Approaches to DNA mutagenesis: an overviewAnal Biochem., 1997, vol. 254, 2157-178 [0228]
- **DALE et al.**Oligonucleotide-directed random mutagenesis using the phosphorothioate methodMethods Mol. Biol., 1996, vol. 57, 369-374 [0228]
- **SMITH**In vitro mutagenesisAnn. Rev. Genet., 1985, vol. 19, 423-462 [0228]
- **BOTSTEINSHORTLE**Strategies and applications of in vitro mutagenesisScience, 1985, vol. 229, 1193-1201 [0228]
- **CARTER**Site-directed mutagenesisBiochem. J., 1986, vol. 237, 1-7 [0228]
- The efficiency of oligonucleotide directed mutagenesis**KUNKEL**Nucleic Acids & Molecular BiologySpringer Verlag19870000 [0228]
- **KUNKEL**Rapid and efficient site-specific mutagenesis without phenotypic selectionProc. Natl. Acad. Sci. USA, 1985, vol. 82, 488-492 [0228]
- **KUNKEL et al.**Rapid and efficient site-specific mutagenesis without phenotypic selectionsMethods in Enzymol., 1987, vol. 154, 367-382 [0228]
- **BASS et al.**Mutant Trp repressors with new DNA-binding specificitiesScience, 1988, vol. 242, 240-245 [0228]
- Methods in Enzymol., 1983, vol. 100, 468-500 [0228]
- Methods in Enzymol., 1987, vol. 154, 329-350 [0228]
- **ZOLLERSMITH**Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragmentNucleic Acids Res., 1982, vol. 10, 6487-6500 [0228]
- **ZOLLERSMITH**Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectorsMethods in Enzymol., 1983, vol. 100, 468-500 [0228]
- **ZOLLERSMITH**Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA templateMethods in Enzymol., 1987, vol. 154, 329-350 [0228]
- **TAYLOR et al.**The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNANucl. Acids Res., 1985, vol. 13, 8749-8764 [0228]
- **TAYLOR et al.**The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-

modified DNANucl. Acids Res., 1985, vol. 13, 8765-8785 [0228]

- **NAKAMAYEECKSTEIN**Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesisNucl. Acids Res., 1986, vol. 14, 9679-9698 [0228]
- **SAYERS et al.**5'-3' Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesisNucl. Acids Res., 1988, vol. 16, 791-802 [0228]
- **SAYERS et al.**Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromideNucl. Acids Res., 1988, vol. 16, 803-814 [0228]
- **KRAMER et al.**The gapped duplex DNA approach to oligonucleotide-directed mutation constructionNucl. Acids Res., 1984, vol. 12, 9441-9456 [0228]
- **KRAMERFRITZ**Oligonucleotide-directed construction of mutations via gapped duplex DNAMethods in Enzymol., 1987, vol. 154, 350-367 [0228]
- **KRAMER et al.**Nucl. Acids Res., 1988, vol. 16, 7207- [0228]
- **FRITZ et al.**Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitroNucl. Acids Res., 1988, vol. 16, 6987-6999 [0228]
- **KRAMER et al.**Point Mismatch RepairCell, 1984, vol. 38, 879-887 [0228]
- **CARTER et al.**Nucl. Acids Res., 1985, vol. 13, 4431-4443 [0228]
- **CARTER**Improved oligonucleotide-directed mutagenesis using M13 vectorsMethods in Enzymol., 1987, vol. 154, 382-403 [0228]
- **EGHTEDARZADEHHENIKOFF**Use of oligonucleotides to generate large deletionsNucl. Acids Res., 1986, vol. 14, 5115- [0228]
- **WELLS et al.**Importance of hydrogen-bond formation in stabilizing the transition state of subtilisinPhil. Trans. R. Soc. Lond. A, 1986, vol. 317, 415-423 [0228]
- **NAMBIAR et al.**Total synthesis and cloning of a gene coding for the ribonuclease S proteinScience, 1984, vol. 223, 1299-1301 [0228]
- **SAKMARKHORANA**Total synthesis and expression of a gene for the a-subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin)Nucl. Acids Res., 1988, vol. 14, 6361-6372 [0228]
- **WELLS et al.**Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sitesGene, 1985, vol. 34, 315-323 [0228]
- **GRUNDSTRÖM et al.**Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesisNucl. Acids Res., 1985, vol. 13, 3305-3316 [0228]
- **MANDECKI**Oligonucleotide-directed double-strand break repair in plasmids of Escherichia coli: a method for site-specific mutagenesisProc. Natl. Acad. Sci. USA, 1986, vol. 83, 7177-7181 [0228]
- **ARNOLD**Protein engineering for unusual environmentsCurrent Opinion in Biotechnology, 1993, vol. 4, 450-455 [0228]
- **SIEBER et al.**Nature Biotechnology, 2001, vol. 19, 456-460 [0228]
- **W. P. C. STEMMER**Nature, 1994, vol. 370, 389-91 [0228]
- **I. A. LORIMERI. PASTAN**Nucleic Acids Res., 1995, vol. 23, 3067-8 [0228]
- Methods in Enzymology, vol. 154, [0228]
- **FROMM et al.**Proc. Natl. Acad. Sci. USA, 1985, vol. 82, 5824- [0229]
- **KLEIN et al.**Nature, 1987, vol. 327, 70-73 [0229]
- **FRESHNEY**Culture of Animal Cells, a Manual of Basic TechniqueWiley-Liss19940000 [0230]
- **PAYNE et al.**Plant Cell and Tissue Culture in Liquid SystemsJohn Wiley & Sons, Inc.19920000 [0230]
- Plant Cell, Tissue and Organ CultureFundamental Methods Springer Lab ManualSpringer-Verlag19950000 [0230]
- The Handbook of Microbiological MediaCRC Press19930000 [0230]
- **GILLAMSMITH**Gene, 1979, vol. 8, 81- [0231]
- **ROBERTS et al.**Nature, 1987, vol. 328, 731- [0231]
- **SCHNEIDER, E. et al.**Protein Expr. Purif., 1995, vol. 6, 110-14 [0231]
- The ATCC Catalogue of bacteria and bacteriophageATCC19920000 [0231]
- **WATSON et al.**Recombinant DNAScientific American Books19920000 [0231]
- **SAYERS, J.R. et al.**5',3' Exonuclease in phosphorothioate-based oligonucleotide-directed mutagenesisNucleic Acids Res, 1988, vol. 16, 3791-802 [0233]
- **ANDERSON et al.**Exploring the Limits of Codon and Anticodon SizeChemistry and Biology, 2002, vol. 9, 237-244 [0235]
- **MAGLIERY**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 coliJ. Mol. Biol., 2001, vol. 307, 755-769 [0235]
- **MA et al.**Biochemistry, 1993, vol. 32, 7939-7945 [0236]
- **HOHSAKA et al.**J. Am. Chem. Soc., 1999, vol. 121, 34-40 [0236]
- **HOHSAKA et al.**J. Am. Chem. Soc., 1999, vol. 121, 12194-12195 [0236]
- **MOORE et al.**J. Mol. Biol., 2000, vol. 298, 195-205 [0236]

- **HIRAO et al.**An unnatural base pair for incorporating amino acid analogues into proteinNature Biotechnology, 2002, vol. 20, 177-182 [0238]
- **WU, Y.**J. Am. Chem. Soc., 2002, vol. 124, 14626-14630 [0238]
- **SWITZER et al.**J. Am. Chem. Soc., 1989, vol. 111, 8322-8322 [0239]
- **PICCIRILLI et al.**Nature, 1990, vol. 343, 33-37 [0239]
- **KOOL**Curr. Opin. Chem. Biol., 2000, vol. 4, 602-608 [0239] [0239]
- **GUCKIANKOOL**Angew. Chem. Int. Ed. Engl., 1998, vol. 36, 242825-2828 [0239]
- **MCMINN et al.**J. Am. Chem. Soc., 1999, vol. 121, 11585-11586 [0239]
- **OGAWA et al.**J. Am. Chem. Soc., 2000, vol. 122, 3274-3278 [0239]
- **OGAWA et al.**J. Am. Chem. Soc., 2000, vol. 122, 8803-8804 [0239]
- **TAE et al.**J. Am. Chem. Soc., 2001, vol. 123, 7439-7440 [0239]
- **MEGGERS et al.**J. Am. Chem. Soc., 2000, vol. 122, 10714-10715 [0239]
- **WANG, L. et al.**Proc. Natl. Acad. Sci. USA, 2003, vol. 100, 156-61 [0247]
- **ZHANG, Z. et al.**Biochem., 2003, vol. 42, 226735-6746 [0247]
- **MEHL et al.**J. Am. Chem. Soc., 2003, vol. 125, 935-939 [0247]
- **SANTORO et al.**Nature Biotechnology, 2002, vol. 20, 1044-1048 [0247]
- **CHIN, J. W. et al.**Science, 2003, vol. 301, 964-967 [0248]
- **WANG, L. et al.**Science, 2001, vol. 292, 498-500 [0251]
- **CHIN, J. W. et al.**J. Am. Chem. Soc., 2002, vol. 124, 9026-9027 [0251]
- **ZHANG, Z. et al.**Biochemistry, 2003, vol. 42, 6735-6746 [0251]
- **PALVA et al.**Gene, 1983, vol. 22, 229-235 [0270]
- **MOSBACH et al.**Nature, 1983, vol. 302, 543-545 [0270]
- **SIKORSKI et al.**GENETICS, 1998, vol. 112, 19- [0278]
- **ITO et al.**J. BACTERIOL., 1983, vol. 153, 163- [0278]
- **HINNEN et al.**PROC. NATL. ACAD. SCI. USA, 1978, vol. 75, 1929- [0278]
- **KURTZ et al.**MOL. CELL. BIOL., 1986, vol. 6, 142- [0278]
- **KUNZE et al.**J. BASIC MICROBIOL., 1985, vol. 25, 141- [0278] [0278]
- **GLEESON et al.**J. GEN. MICROBIOL., 1986, vol. 132, 3459- [0278]
- **ROGGENKAMP et al.**MOL. GEN. GENET., 1986, vol. 202, 302- [0278]
- **DAS et al.**J. BACTERIOL., 1984, vol. 158, 1165- [0278]
- **DE LOUVENCOURT et al.**J. BACTERIOL., 1983, vol. 154, 737- [0278]
- **VAN DEN BERG et al.**BIO/TECHNOLOGY, 1990, vol. 8, 135- [0278]
- **CREGG et al.**MOL. CELL. BIOL., 1985, vol. 5, 3376- [0278]
- **BEACHNURSE**NATURE, 1981, vol. 300, 706- [0278]
- **DAVIDOW et al.**CURR. GENET., 1985, vol. 10, 380- [0278]
- **GAILLARDIN et al.**CURR. GENET., 1985, vol. 10, 49- [0278]
- **BALLANCE et al.**BIOCHEM. BIOPHYS. RES. COMMUN., 1983, vol. 112, 284-89 [0278]
- **TILBURN et al.**GENE, 1983, vol. 26, 205-221 [0278]
- **YELTON et al.**PROC. NATL. ACAD. SCI. USA, 1984, vol. 81, 1470-74 [0278]
- **KELLYHYNES**EMBO J., 1985, vol. 4, 475-479 [0278]
- **MIYANOHARA et al.**PROC. NATL. ACAD. SCI. USA, 1983, vol. 80, 1- [0279]
- **HITZEMAN et al.**J. BIOL. CHEM., 1980, vol. 255, 412073-12080 [0279]
- **HOLLAND et al.**BIOCHEMISTRY, 1978, vol. 17, 234900-4907 [0279]
- **HESS et al.**J. ADV. ENZYME REG., 1969, vol. 7, 149-167 [0279]
- **HOLLAND et al.**J. BIOL. CHEM., 1981, vol. 256, 1385- [0281]
- **TSCHUMPER et al.**GENE, 1980, vol. 10, 157- [0281]
- **KINGSMAN et al.**GENE, 1979, vol. 7, 141- [0281]
- **HSIAO et al.**PROC. NATL. ACAD. SCI. USA, 1979, vol. 76, 3829- [0282]
- **VAN SOLINGEN et al.**J. BACT., 1977, vol. 130, 946- [0282]
- **SAMBROOK et al.**MOLECULAR CLONING: A LAB. MANUAL20010000 [0282]
- **GELLISSEN et al.**ANTONIE VAN LEEUWENHOEK, 1992, vol. 62, 1-279-93 [0283]
- **ROMANOS et al.**YEAST, 1992, vol. 8, 6423-488 [0283]
- **GOEDDEL**METHODS IN ENZYMOLOGY, 1990, vol. 185, 3-7 [0283]
- **ELLIOTT et al.**J. PROTEIN CHEM., 1990, vol. 9, 95- [0284]
- **FIESCHKO et al.**BIOTECH. BIOENG., 1987, vol. 29, 1113- [0284]
- Department of Biophysics and Medical PhysicsUniversity of California [0287]
- **SUMMERSSMITH**TEXAS AGRICULTURAL EXPERIMENT STATION BULLETIN No. 1555, 1987, [0289]
- **RICHARDSON**39 METHODS IN MOLECULAR BIOLOGY: BACULOVIRUS EXPRESSION PROTOCOLS, 1995, [0289]

- **AUSUBEL et al.**CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, 1994, vol. 16, 9-1611- [0289]
- **KINGPOSSEE**THE BACULOVIRUS SYSTEM: A LABORATORY GUIDE, 1992, [0289]
- **O'REILLY et al.**BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL, 1992, [0289]
- **REILLY et al.**EXPRESSION VECTORS: A LABORATORY MANUAL, 1992, [0291]
- **MILLER et al.**ANN. REV. MICROBIOL., 1988, vol. 42, 177- [0292]
- **LUCKOWSUMMERS**VIROLOGY, 1989, vol. 170, 31-39 [0293] [0294]
- **SUMMERSSMITH**TEXAS AGRICULTURAL EXPERIMENT STATION BULLETIN, 1987, [0294] [0298]
- **SMITH et al.**MOL. CELL. BIOL., 1983, vol. 3, 2156- [0294] [0299]
- **MILLER et al.**BIOESSAYS, 1989, vol. 4, 91- [0294] [0298]
- **TROTTERWOOD**METHODS IN MOLECULAR BIOLOGY, 1995, vol. 39, [0295] [0295]
- **MANNKING**J. GEN. VIROL., 1989, vol. 70, 3501- [0295] [0295]
- **LIEBMAN et al.**BIOTECHNIQUES, 1999, vol. 26, 136- [0295]
- **GRAVES et al.**BIOCHEMISTRY, 1998, vol. 37, 6050- [0295]
- **NOMURA et al.**J. BIOL. CHEM., 1998, vol. 273, 2213570- [0295]
- **SCHMIDT et al.**PROTEIN EXPRESSION AND PURIFICATION, 1998, vol. 12, 323- [0295]
- **SIFFERT et al.**NATURE GENETICS, 1998, vol. 18, 45- [0295]
- **TILKINS et al.**CELL BIOLOGY: A LABORATORY HANDBOOK, 1998, 145-154 [0295]
- **CAI et al.**PROTEIN EXPRESSION AND PURIFICATION, 1997, vol. 10, 263- [0295]
- **DOLPHIN et al.**NATURE GENETICS, 1997, vol. 17, 491- [0295]
- **KOST et al.**GENE, 1997, vol. 190, 139- [0295]
- **JAKOBSSON et al.**J. BIOL. CHEM., 1996, vol. 271, 22203- [0295]
- **ROWLES et al.**J. BIOL. CHEM., 1996, vol. 271, 3722376- [0295]
- **REVERSEY et al.**J. BIOL. CHEM., 1996, vol. 271, 3923607-10 [0295]
- **STANLEY et al.**J. BIOL. CHEM., 1995, vol. 270, 4121- [0295]
- **SISK et al.**J. VIROL., 1994, vol. 68, 2766- [0295]
- **PENG et al.**BIOTECHNIQUES, 1993, vol. 14, 2274- [0295]
- **KITTS**NAR, 1990, vol. 18, 195667- [0295]
- **FRIESEN et al.**The Regulation of Baculovirus Gene ExpressionTHE MOLECULAR BIOLOGY OF BACULOVIRUSES, 1986, [0297]
- **VLAK et al.**J. GEN. VIROL., 1988, vol. 69, 765- [0297]
- **WRIGHT**NATURE, 1986, vol. 321, 718- [0299]
- **CARBONELL et al.**J. VIROL., 1985, vol. 56, 153- [0299]
- **FRASER et al.**IN VITRO CELL. DEV. BIOL., 1989, vol. 25, 225- [0299]
- **RAIBAUD et al.**ANNU. REV. GENET., 1984, vol. 18, 173- [0302]
- **CHANG et al.**NATURE, 1977, vol. 198, 1056- [0303]
- **GOEDDEL et al.**Nuc. ACIDS RES., 1980, vol. 8, 4057- [0303]
- **YELVERTON et al.**NUCL. ACIDS RES., 1981, vol. 9, 731- [0303]
- The cloning of interferon and other mistakes**WEISSMANN**In Interferon19810000vol. 3, [0303]
- **SHIMATAKE et al.**NATURE, 1981, vol. 292, 128- [0303]
- **AMANN et al.**GENE, 1983, vol. 25, 167- [0304]
- **DE BOER et al.**PROC. NATL. ACAD. SCI., 1983, vol. 80, 21- [0304]
- **STUDIER et al.**J. MOL. BIOL., 1986, vol. 189, 113- [0304]
- **TABOR et al.**Proc Natl. Acad. Sci., 1985, vol. 82, 1074- [0304]
- **SHINE et al.**NATURE, 1975, vol. 254, 34- [0305]
- Genetic signals and nucleotide sequences in messenger RNA**STEITZ et al.**In Biological Regulation and Development: Gene Expression19790000 [0305]
- **SAMBROOK et al.**Expression of cloned genes in Escherichia coliMolecular Cloning: A Laboratory Manual, 1989, [0305]
- Bacterial Genetic Stock CenterDepartment of Biophysics and Medical Physics, University of California [0307]
- **R. SCOPES**Protein PurificationSpringer-Verlag19820000 [0319]
- **DEUTSCHER**Methods in Enzymology Vol. 182: Guide to Protein PurificationAcademic Press19900000 [0319]
- **SANDANA**Bioseparation of ProteinsAcademic Press, Inc.19970000 [0319]
- **BOLLAG et al.**Protein MethodsWiley-Liss19960000 [0319]
- **WALKER**The Protein Protocols HandbookHumana Press19960000 [0319]
- **HARRISANGAL**Protein Purification Applications: A Practical ApproachIRL Press19900000 [0319]
- **HARRISANGAL**Protein Purification Methods: A Practical ApproachIRL Press at Oxford [0319]
- **SCOPES**Protein Purification: Principles and Practice 3rd EditionSpringer Verlag19930000 [0319]
- **JANSONRYDEN**Protein Purification: Principles, High Resolution Methods and Applications, Second EditionWiley-

VCH19980000 [0319]

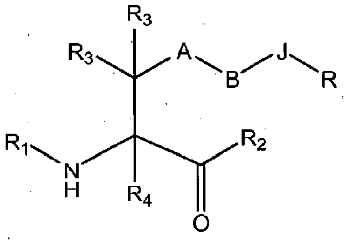
- **WALKER**Protein Protocols on CD-ROMHumana Press19980000 [0319]
- **DEBINSKI et al.**J. Biol. Chem., 1993, vol. 268, 14065-14070 [0321]
- **KREITMANPASTAN**Bioconjug. Chem., 1993, vol. 4, 581-585 [0321]
- **BUCHNER et al.**Anal. Biochem., 1992, vol. 205, 263-270 [0321]
- **PEARSON et al.**ANAL BIOCHEM., 1982, vol. 124, 217-230 [0337]
- **RIVIER et al.**J. CHROM., 1983, vol. 268, 112-119 [0337]
- **KUNITANI et al.**J. CHROM., 1986, vol. 359, 391-402 [0337]
- **CORNISH et al.**Am. Chem. Soc., 1996, vol. 118, 8150-8151 [0352]
- **MAHAL et al.**Science, 1997, vol. 276, 1125-1128 [0352]
- **WANG et al.**Science, 2001, vol. 292, 498-500 [0352]
- **CHIN et al.**Am. Chem. Soc., 2002, vol. 124, 9026-9027 [0352]
- **CHIN et al.**Proc. Natl. Acad. Sci., 2002, vol. 99, 11020-11024 [0352]
- **WANG et al.**Proc. Natl. Acad. Sci., 2003, vol. 100, 56-61 [0352]
- **ZHANG et al.**Biochemistry, 2003, vol. 42, 6735-6746 [0352]
- **CHIN et al.**Science, 2003, vol. 300, 964-967 [0352]
- **KIICK et al.**Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligtationPNAS, 2002, vol. 99, 119-24 [0352]
- **P. E. DAWSONS. B. H. KENT**Annu. Rev. Biochem., 2000, vol. 69, 923- [0353]
- **V. W. CORNISHD. MENDELP. G. SCHULTZ**Angew. Chem. Int. Ed. Engl., 1995, vol. 34, 621-633 [0353]
- **C.J. NORENS.J. ANTHONY-CAHILLM.C. GRIFFITHP.G. SCHULTZ**A general method for site-specific incorporation of unnatural amino acids into proteinsScience, 1989, vol. 244, 182-188 [0353]
- **J.D. BAINC.G. GLABET.A DIXA.R. CHAMBERLINE.S. DIALA**Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptideJ. Am. Chem. Soc., 1989, vol. 111, 8013-8014 [0353]
- **N. BUDISAC. MINKSS. ALEFELDERW. WENGERF. M. DONGL. MORODERR. HUBER**FASEB J., 1999, vol. 13, 41-51 [0354]
- **C. MINKSR. HUBERL. MORODERN. BUDISA**Anal. Biochem., 2000, vol. 284, 29-34 [0354]
- **H. DUEWELE. DAUBV. ROBINSONJ. F. HONEK**Biochemistry, 1997, vol. 36, 3404-3416 [0354]
- **Y. TANGG. GHIRLANDAW. A PETKAT. NAKAJIMAW. F. DEGRADOD. A TIRRELL**Angew. Chem. Int. Ed. Engl., 2001, vol. 40, 81494-1496 [0354]
- **W. A HENDRICKSONJ. R. HORTOND. M. LEMASTER**EMBO J., 1990, vol. 9, 51665-1672 [0354]
- **J. O. BOLESK. LEWINSKIM. KUNKLEJ. D. ODOMB. DUNLAPL. LEBIODAM. HATADA**Nat. Struct. Biol., 1994, vol. 1, 283-284 [0354]
- **N. BUDISAB. STEIPEP. DEMANGEC. ECKERSKORNJ. KELLERMANNR. HUBER**Eur. J. Biochem., 1995, vol. 230, 788-796 [0354]
- **N. BUDISAW. KARNBROCKS. STEINBACHERA. HUMML. PRADET. NEUEFEINDL. MORODERR. HUBER**J. Mol. Biol., 1997, vol. 270, 616-623 [0354]
- **J. C. M. VANHESTD. A TIRRELL**FEBS Lett., 1998, vol. 428, 68-70 [0354]
- **J. C. M. VAN HESTK. L. KIICKD. A TIRRELL**J. Am. Chem. Soc., 2000, vol. 122, 1282-1288 [0354]
- **K. L. KIICKD. A TIRRELL**Tetrahedron, 2000, vol. 56, 9487-9493 [0354]
- **M. IBBAP. KASTH. HENNECKE**Biochemistry, 1994, vol. 33, 7107-7112 [0355]
- **M. IBBAH. HENNECKE**FEBS Lett., 1995, vol. 364, 272-275 [0355]
- **N. SHARMAR. FURTERP. KASTD. A TIRRELL**FEBS Lett., 2000, vol. 467, 37-40 [0355]
- **F. HAMANO-TAKAKUT. IWAMAS. SAITO-YANOK. TAKAKUY. MONDENM. KITABATAKED. SOLLS. NISHIMURA**J. Biol. Chem., 2000, vol. 275, 5140324-40328 [0355]
- **V. DORINGH. D. MOOTZL. A NANGLET. L. HENDRICKSONV. DE CRECY-LAGARDP. SCHIMMELP. MARLIERE**Science, 2001, vol. 292, 501-504 [0356]
- **CRICK, F.J.C.BARRETT, L.BRENNER, S.WATTS-TOBIN, R.**General nature of the genetic code for proteinsNature, 1961, vol. 192, 48091227-1232 [0357]
- **HOFMANN, K.BOHN, H.**Studies on polypeptides. XXXVI. The effect of pyrazole-imidazole replacements on the S-protein activating potency of an 5-peptide fragmentJ. Am Chem, 1966, vol. 88, 245914-5919 [0357]
- **KAISER, E.T.**Synthetic approaches to biologically active peptides and proteins including enyzmesAcc Chem Res, 1989, vol. 22, 247-54 [0357]
- **NAKATSUKA, T. SASAKI, T. KAISER, E.T.**Peptide segment coupling catalyzed by the semisynthetic enzyme thiosubtilisinJ Am Chem Soc, 1987, vol. 109, 3808-3810 [0357]
- **SCHNOLZER, M.KENT, S B H**Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protedseScience, 1992, vol. 256, 221-225 [0357]
- **CHAIKEN, I.M**Semisynthetic peptides and proteinsCRC Crit Rev Biochem, 1981, 255-301 [0357]

- **OFFORD, R.E.**Protein engineering by chemical means?Protein Eng., 1987, vol. 1, 3151-157 [0357]
- **JACKSON, D.Y.BUMIER, J.QUAN, C.STANLEY, M.TOM, J.WELLS, J.A.**A Designed Peptide Ligase for Total Synthesis of Ribonuclease A with Unnatural Catalytic ResiduesScience, 1994, vol. 266, 243-247 [0357]
- **COREY, D.R.SCHULTZ, P.G.** Generation of hybrid sequence-specific single-stranded deoxyribonucleaseScience, 1987, vol. 238, 1401-1403 [0358]
- **KAISER, E.T.LAURENCE D.S.ROKITA, S.E.**The chemical modification of enzymatic specificityAnn. Rev Biochem, 1985, vol. 54, 565-595 [0358]
- **KAISER, E.T.LAWRENCE, D.S.**Chemical mutation of enyzme active sitesScience, 1984, vol. 226, 505-511 [0358]
- **NEET, K.E.NANCI AKOSHLAND, D.E.**Properties of thiol-subtilisinJ Biol. Chem, 1968, vol. 243, 246392-6401 [0358]
- **POLGAR, L.B.M.L. A**new enzyme containing a synthetically formed active site. Thiol-subtilisinJ. Am Chem Soc, 1966, vol. 88, 133153-3154 [0358]
- **POLLACK, S.J.NAKAYAMA, G.SCHULTZ, P.G.** Introduction of nucleophiles and spectroscopic probes into antibody combining sitesScience, 1988, vol. 1, 2421038-1040 [0358]
- **BRUNNER, J.**New Photolabeling and crosslinking methodsAnnu. Rev Biochem, 1993, 483-514 [0359]
- **KRIEG, U.C.WALTER, P. HOHNSON, A.E.**Photocrosslinking of the signal sequence of nascent preprolactin of the 54-kilodalton polypeptide of the signal recognition particleProc. Natl. Acad. Sci, 1986, vol. 83, 8604-8608 [0359]
- **NOREN, C.J.ANTHONY-CAHILLGRIFFITH, M.C.SCHULTZ, P.G.** A general method for site-specific incorporation of unnatural amino acids into proteinsScience, 1989, vol. 244, 182-188 [0360]
- **M.W. NOWAK et al.**Science, 1995, vol. 268, 439-42 [0360]
- **BAIN, J.D.GLABE, C.G.DIX, T.A.CHAMBERLIN, A.R.DIALA, E.S.**Biosynthetic site-specific Incorporation of a non-natural amino acid into a polypeptideJ. Am Chem Soc, 1989, vol. 111, 8013-8014 [0360]
- **N. BUDISA et al.**FASEB J., 1999, vol. 13, 41-51 [0360]
- **ELLMAN, J.A.MENDEL, D.ANTHONY-CAHILL, S.NOREN, C.J.SCHULTZ, P.G.** Biosynthetic method for introducing unnatural amino acids site-specifically into proteinsMethods in Enz., 1992, vol. 202, 301-336 [0360]
- **MENDEL, D.CORNISH, V.W.SCHULTZ, P.G.**Site-Directed Mutagenesis with an Expanded Genetic CodeAnnu Rev Biophys. Biomol Struct., 1995, vol. 24, 435-62 [0360]
- **SAYERS, J.R.SCHMIDT, W.ECKSTEIN, F.** 5', 3' Exonuclease in phosphorothioate-based oligonucleotide-direted mutagenesisNucleic Acids Res, 1988, vol. 16, 3791-802 [0361]
- **KOBAYASHI et al.**Nature Structural Biology, 2003, vol. 10, 6425-432 [0361]
- **ELLMAN, J.A.MENDEL, D.SCHULTZ, P.G.** Site-specific incorporation of novel backbone structures into proteinsScience, 1992, vol. 255, 197-200 [0361]
- **M. W. NOWAKP. C. KEARNEYJ. R. SAMPSONM. E. SAKSC. G. LABARCAS. K. SILVERMANW. G. ZHONGJ. THORSONJ. N. ABELSONN. DAVIDSON**Science, 1995, vol. 268, 439-442 [0362]
- **D. A. DOUGHERTY**Curr. Opin. Chem. Biol, 2000, vol. 4, 645- [0362]
- **G. TURCATTIK. NEMETHM. D. EDGERTONU. MESETHF. TALABOTM. PEITSCHJ. KNOWLESH. VOGELA. CHOLLET**J. Biol. Chem., 1996, vol. 271, 3319991-19998 [0362]
- **J. P. GALLIVANH. A. LESTERD. A. DOUGHERTY**Chem. Biol., 1997, vol. 4, 10739-749 [0362]
- **J. C. MILLERS. K. SILVERMANP. M. ENGLANDD. A. DOUGHERTYH. A. LESTER**Neuron, 1998, vol. 20, 619-624 [0362]
- **P. M. ENGLANDY. ZHANGD. A. DOUGHERTYH. A. LESTER**Cell, 1999, vol. 96, 89-98 [0362]
- **T. LUA. Y. TINGJ. MAINLANDL. Y. JANP. G. SCHULTZJ. YANG**Nat. Neurosci., 2001, vol. 4, 3239-246 [0362]
- **R. FURTER**Protein Sci., 1998, vol. 7, 419-426 [0364]
- **KIM, D.-M.J.R. SWARTZ**Biotechnology and Bioengineering, 2001, vol. 74, 4309-316 [0365]
- **KIM, D.-M.J.R. SWARTZ**Biotechnology Letters, 2000, vol. 22, 1537-1542 [0365]
- **KIM, D.-M.J.R. SWARTZ**Biotechnology Progress, 2000, vol. 16, 385-390 [0365]
- **KIM, D.-M.J.R. SWARTZ**Biotechnology and Bioengineering, 1999, vol. 66, 3180-188 [0365]
- **PATNAIK, R.J.R. SWARTZ**Biotechniques, 1998, vol. 24, 5862-868 [0365]
- **R. ROBERTSJ. SZOSTAK**Proc. Natl Acad. Sci., 1997, vol. 94, 12297-12302 [0365]
- **A. FRANKEL et al.**Chemistry & Biology, 2003, vol. 10, 1043-1050 [0365]
- **A. FORSTER et al.**Proc. Natl Acad. Sci., 2003, vol. 100, 116353-6357 [0365]
- **MARCH**ADVANCED ORGANIC CHEMISTRYWiley20010000 [0372]
- **CAREYSUNDBERG**ADVANCED ORGANIC CHEMISTRYPlenum20000000vol. A and B, [0372]
- **SANDLERKARO**Polymer SynthesisAcademic Pressvol. 3, 138-161 [0416]
- **BUCKMANN et al.**Makromol. Chem., 1981, vol. 182, 1379- [0418]
- **ZALIPSKY et al.**Eur. Polym. J., 1983, vol. 19, 1177- [0418]
- **ANDRESZ et al.**Makromol. Chem., 1978, vol. 179, 301- [0418]
- **OLSON et al.**Poly(ethylene glycol) Chemistry & Biological ApplicationsACS, Washington19970000170-181 [0418]
- **ABUCHOWSKI et al.**Cancer Biochem. Biophys., 1984, vol. 7, 175- [0418]

- **JOPPICH et al.**Makromol. Chem., 1979, vol. 180, 1381- [0418]
- **PITHA et al.**Eur. J Biochem., 1979, vol. 94, 11- [0418]
- **ELLING et al.**Biotech. Appl. Biochem., 1991, vol. 13, 354- [0418]
- **BEAUCHAMP et al.**Anal. Biochem., 1983, vol. 131, 25- [0418]
- **TONDELLI et al.**J. Controlled Release, 1985, vol. 1, 251- [0418]
- **VERONESE et al.**Appl. Biochem. Biotech., 1985, vol. 11, 141- [0418]
- **SARTORE et al.**Appl. Biochem. Biotech., 1991, vol. 27, 45- [0418]
- **HARRIS et al.**J. Polym. Sci. Chem. Ed., 1984, vol. 22, 341- [0418]
- **GOODSON et al.**Bio/Technology, 1990, vol. 8, 343- [0418]
- **ROMANI et al.**Chemistry of Peptides and Proteins, 1984, vol. 2, 29- [0418]
- **KOGAN**Synthetic Comm., 1992, vol. 22, 2417- [0418]
- **WOGHIREN et al.**Bioconj. Chem., 1993, vol. 4, 314- [0418]
- **SAWHNEY et al.**Macromolecules, 1993, vol. 26, 581- [0418]
- **HARRIS**Macromol. Chem. Phys., 1985, vol. C25, 325-373 [0441]
- **SCOUTEN**Methods in Enzymology, 1987, vol. 135, 30-65 [0441]
- **WONG et al.**Enzyme Microb. Technol., 1992, vol. 14, 866-874 [0441]
- **DELGADO et al.**Critical Reviews in Therapeutic Drug Carrier Systems, 1992, vol. 9, 249-304 [0441]
- **ZALIPSKY**Bioconjugate Chem., 1995, vol. 6, 150-165 [0441]
- **VERONESE**Biochem. Biotech., 1985, vol. 11, 141-152 [0442]
- **PRENETA AZ**PROTEIN PURIFICATION METHODS, A PRACTICAL APPROACHIRL Press19890000293-306 [0443]
- **PEPINSKY RB.**J. Pharmacol. & Exp. Ther., 2001, vol. 297, 31059-66 [0443]
- **MAKRIDES et al.**J. Pharmacol. Exp. Ther., 1996, vol. 277, 1534-542 [0446]
- **SJOLANDER et al.**J, Immunol. Methods, 1997, vol. 201, 115-123 [0446]
- **DENNIS et al.**J. Biol. Chem., 2002, vol. 277, 3835035-35043 [0446]
- **KURTZHALS et al.**Biochem. J., 1995, vol. 312, 725-731 [0447]
- **H. LIU et al.**J. Am. Chem. Soc., 2003, vol. 125, 1702-1703 [0450]
- **PIKAL, M.**Biopharm., 1990, vol. 3, 926-30 [0500]
- **ARAKAWA et al.**Pharm. Res., 1991, vol. 8, 3285-291 [0500]
- **BROADHEAD, J. et al.**The Spray Drying of PharmaceuticalsDrug Dev. Ind. Pharm, 1992, vol. 18, 11,121169-1206 [0501]
- **REMINGTON**The Science and Practice of PharmacyMack Publishing Company19950000 [0502]
- **HOOVER, JOHN E.**Remington's Pharmaceutical SciencesMack Publishing Co.19750000 [0502]
- Pharmaceutical Dosage FormsMarcel Decker19800000 [0502]
- Pharmaceutical Dosage Forms and Drug Delivery SystemsLippincott Williams & Wilkins19990000 [0502]
- **LANGER et al.**J. Biomed. Mater. Res., 1981, vol. 15, 167-277 [0503]
- **LANGER**Chem. Tech., 1982, vol. 12, 98-105 [0503]
- **U. SIDMAN et al.**Biopolymers, 1983, vol. 22, 547-556 [0503]
- **EPSTEIN et al.**Proc. Natl. Acad. Sci. U.S.A., 1985, vol. 82, 3688-3692 [0503] [0504]
- **HWANG et al.**Proc. Natl. Acad. Sci. U.S.A., 1980, vol. 77, 4030-4034 [0503] [0504]
- **PARK JW et al.**Proc. Natl. Acad. Sci. USA, 1995, vol. 92, 1327-1331 [0504]
- MEDICAL APPLICATIONS OF LIPOSOMES19980000 [0504]
- Liposomal drug delivery systems for cancer therapy**DRUMMOND DC et al.**CANCER DRUG DISCOVERY AND DEVELOPMENT20020000 [0504]
- **PARK JW et al.**Clin. Cancer Res., 2002, vol. 8, 1172-1181 [0504]
- **NIELSEN UB et al.**Biochim. Biophys. Acta, 2002, vol. 1591, 1-3109-118 [0504]
- **MAMOT C et al.**Cancer Res., 2003, vol. 63, 3154-3161 [0504]
- **BOISSEL et al.**J. Biol. Chem., 1993, vol. 268, 15983-93 [0541]

Patentkrav

1. Fremgangsmåde til derivatisering af et polypeptid omfattende en aminosyre med formel (I), hvilken fremgangsmåde omfatter at bringe polypeptidet i kontakt med et reagens med formel (XXVII), hvor formel (I) svarer til:



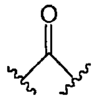
(I)

hvor:

A er eventuelt, og når til stede er lavere alkylen, substitueret lavere alkylen, lavere cycloalkylen, substitueret lavere cycloalkylen, lavere alkenylen, substitueret lavere alkenylen, alkynylen, lavere heteroalkylen, substitueret heteroalkylen, lavere heterocycloalkylen, substitueret lavere heterocycloalkylen, arylen, substitueret arylen, heteroarylen, substitueret heteroarylen, alkarylen, substitueret alkarylen, aralkylen, eller substitueret aralkylen;

B er eventuelt, og når til stede er en linker valgt fra gruppen bestående af lavere alkylen, substitueret lavere alkylen, lavere alkenylen, substitueret lavere alkenylen, lavere heteroalkylen, substitueret lavere heteroalkylen, -O-, -O-(alkylen eller substitueret alkylen)-, -S-, -S-(alkylen eller substitueret alkylen)-, $-S(O)_k-$ hvor k er 1, 2, eller 3, $-S(O)_k$(alkylen eller substitueret alkylen)-, -C(O)-, -C(O)-(alkylen eller substitueret alkylen)-, -C(S)-, -C(S)-(alkylen eller substitueret alkylen)-, -N(R')-, -NR'-(alkylen eller substitueret alkylen)-, -C(O)N(R')-, -CON(R')-(alkylen eller substitueret alkylen)-, -CSN(R')-, -CSN(R')-(alkylen eller substitueret alkylen)-, -N(R')CO-(alkylen eller substitueret alkylen)-, -N(R')C(O)O-, $-S(O)_kN(R')-$, -N(R')C(O)N(R')-, -N(R')C(S)N(R')-, $-N(R')S(O)_kN(R')-$, -N(R')-N=,-C(R')=N-, -C(R')=N-N(R')-, -C(R')=N-N=, $-C(R')_2-N=N-$, og $-C(R')_2-N(R')-N(R')-$, hvor hver R' uafhængigt er H, alkyl, eller substitueret alkyl

J er


,

R er H, alkyl, substitueret alkyl, cycloalkyl, eller substitueret cycloalkyl;

hver R' uafhængigt er H, alkyl, eller substitueret alkyl;

R_1 er H, en aminobeskyttelsesgruppe, resin, aminosyre, polypeptid, eller polynukleotid; og

R_2 er OH, en esterbeskyttelsesgruppe, resin, aminosyre, polypeptid, eller polynukleotid;

hver R_3 and R_4 uafhængigt er H, halogen, lavere alkyl, eller substitueret lavere alkyl;

hvor formel (XXVII) svarer til:

$$\boxed{\text{PEG}}\text{—L—O—NH}_2$$

(XXVII)

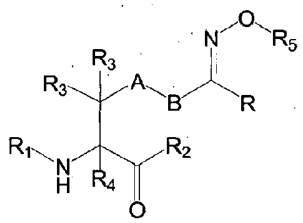
hvor:

hver L er en linker uafhængigt valgt fra gruppen bestående af alkylen, substitueret alkylen, alkenylen, substitueret alkenylen, -O-, -O-(alkylen eller substitueret alkylen)-, -S-, -S-(alkylen eller substitueret alkylen)-, $-S(O)_k-$ hvor k er 1, 2, eller 3, $-S(O)_k$(alkylen eller substitueret alkylen)-, -C(O)-, -C(O)-(alkylen eller substitueret alkylen)-, -C(S)-, -C(S)-(alkylen eller substitueret alkylen)-, -N(R')-, -NR'-(alkylen eller substitueret alkylen)-, -C(O)N(R')-, -CON(R')-(alkylen eller substitueret alkylen)-, -(alkylen eller substitueret alkylen)NR'C(O)O-(alkylen eller substitueret alkylen)-, -O-CON(R')-(alkylen eller substitueret alkylen)-, -CSN(R')-, -CSN(R')-(alkylen eller substitueret alkylen)-, -N(R')CO-(alkylen eller substitueret alkylen)-, -N(R')C(O)O-, -N(R')C(O)O-(alkylen eller substitueret alkylen)-, $-S(O)_kN(R')-$, -N(R')C(O)N(R')-, -N(R')C(O)N(R')-(alkylen eller substitueret alkylen)-, -N(R')C(S)N(R')-, $-N(R')S(O)_kN(R')-$, -N(R')-N=, -C(R')=N-,-C(R')=N-N(R')-, -C(R')=N-N=, $-C(R')_2-N=N-$, og

-C(R')$_2$-N(R')-N(R')-, hvor hver R' uafhængigt er H, alkyl, eller substitueret alkyl;

hvor "lavere alkyl" er en alkylkæde med otte eller færre carbonatomer.

2. Fremgangsmåden ifølge krav 1, hvor det derivatiserede polypeptid omfatter mindst en oxim-indeholdende aminosyre med strukturen med formel (XI):



(XI)

hvor:

A er eventuelt, og når til stede er lavere alkylen, substitueret lavere alkylen, lavere cycloalkylen, substitueret lavere cycloalkylen, lavere alkenylen, substitueret lavere alkenylen, alkynylen, lavere heteroalkylen, substitueret heteroalkylen, lavere heterocycloalkylen, substitueret lavere heterocycloalkylen, arylen, substitueret arylen, heteroarylen, substitueret heteroarylen, alkarylen, substitueret alkarylen, aralkylen, eller substitueret aralkylen;

B er eventuelt, og når til stede er en linker valgt fra gruppen bestående af lavere alkylen, substitueret lavere alkylen, lavere alkenylen, substitueret lavere alkenylen, lavere heteroalkylen, substitueret lavere heteroalkylen, -O-, -O-((alkylen eller substitueret alkylen)-, -S-, -S-(alkylen eller substitueret alkylen)-, -S(O)$_k$- hvor k er 1,2, eller 3,-S(O)$_k$(alkylen eller substitueret alkylen)-, -C(O)-, -C(O)-(alkylen eller substitueret alkylen)-, -C(S)-, -C(S)-(alkylen eller substitueret alkylen)-, -N(R')-, -NR'-(alkylen eller substitueret alkylen)-, -C(O)N(R')-, -CON(R')-(alkylen eller substitueret alkylen)-,-CSN(R')-, -CSN(R')-(alkylen eller substitueret alkylen)-, -N(R')CO-(alkylen eller substitueret alkylen)-, -N(R')C(O)O-, -S(O)$_k$N(R')-, -N(R')C(O)N(R')-, -N(R')C(S)N(R')-, -N(R')S(O)$_k$N(R')-, -N(R')-N=, -C(R')=N-, -C(R')=N-N(R')-,-C(R')=N-N=, -C(R')$_2$-N=N-, og

-C(R')$_2$-N(R')-N(R')-, hvor hver R' uafhængigt er H, alkyl, eller substitueret alkyl

R er H, alkyl, substitueret alkyl, cycloalkyl, eller substitueret cycloalkyl;

R_1 er H, en aminobeskyttelsesgruppe, resin, aminosyre, polypeptid, eller polynukleotid; og

R_2 er OH, en esterbeskyttelsesgruppe, resin, aminosyre, polypeptid, eller polynukleotid;

hver R_3 og R_4 uafhængigt er H, halogen, lavere alkyl, eller substitueret lavere alkyl;

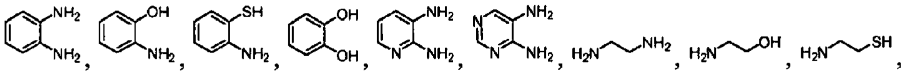
R_5 er -L-PEG og

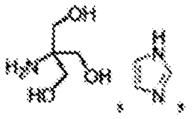
L er eventuelt, og når til stede er en linker valgt fra gruppen bestående af alkylen, substitueret alkylen, alkenylen, substitueret alkenylen, -O-, -O-(alkylen eller substitueret alkylen)-, -S-, -S-(alkylen eller substitueret alkylen)-,-S(O)$_k$- hvor k er 1, 2, eller 3, -S(O)$_k$(alkylen eller substitueret alkylen)-, -C(O)-, -C(O)-(alkylen eller substitueret alkylen)-, -C(S)-, -C(S)-(alkylen eller substitueret alkylen)-, -N(R')-, -NR'-(alkylen eller substitueret alkylen)-, -C(O)N(R')-, -CON(R')-(alkylen eller substitueret alkylen)-, -CSN(R')-, -CSN(R')-(alkylen eller substitueret alkylen)-, -N(R')CO-(alkylen eller substitueret alkylen)-, -N(R')C(O)O-, -S(O)$_k$N(R')-, -N(R')C(O)N(R')-, -N(R')C(S)N(R')-, -N(R')S(O)$_k$N(R')-, -N(R')-N=, -C(R')=N-, -C(R')=N-N(R')-, -C(R')=N-N=, -C(R')$_2$-N=N-, and -C(R')$_2$-N(R')-N(R')-, hvor hver R' uafhængigt er H, alkyl, eller substitueret alkyl.

3. Fremgangsmåden ifølge krav 1, hvor polypeptidet bringes i kontakt med reagenset med formel (XXVII) i vandig opløsning under mildt sure betingelser.

4. Fremgangsmåden ifølge krav 3, hvor betingelserne er pH 2 til 8.

5. Fremgangsmåden ifølge krav 1, hvor polypeptidet bringes i kontakt med reagenset med formel (XXVII) i tilstedeværelse af en accelerant valgt fra gruppen bestående af:





og

6. Fremgangsmåden ifølge krav 1, hvor hver R_1 er H.

7. Fremgangsmåden ifølge krav 1, hvor polyethylenglycolen afslutter på en ende med hydroxy eller methoxy.

8. Fremgangsmåden ifølge krav 1, hvor molekylvægten af polyethylenglycolen er mellem 100 Da og 100.000 Da.

9. Fremgangsmåden ifølge krav 8, hvor molekylvægten af polyethylenglycolen er mellem 1000 Da og 40.000 Da.

10. Fremgangsmåden ifølge krav 9, hvor molekylvægten af polyethylenglycolen er mellem 10.000 Da og 40.000 Da.

11. Fremgangsmåden ifølge krav 7, hvor polyethylenglycolen afslutter på en ende med methoxy.

DRAWINGS

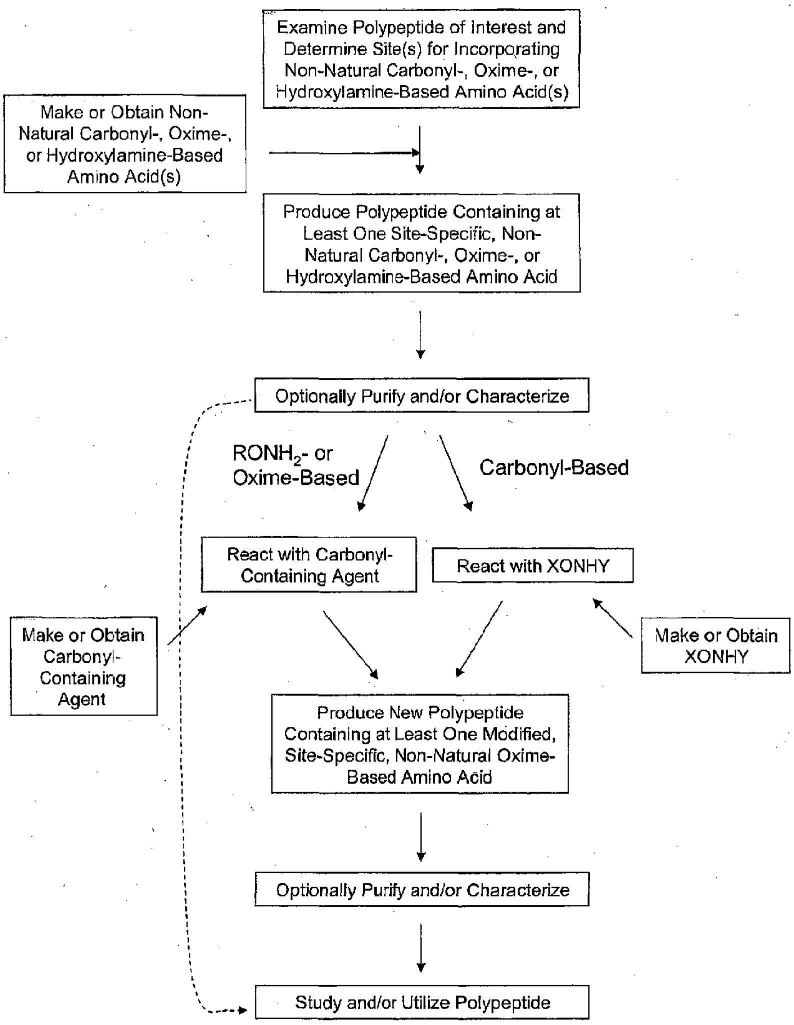
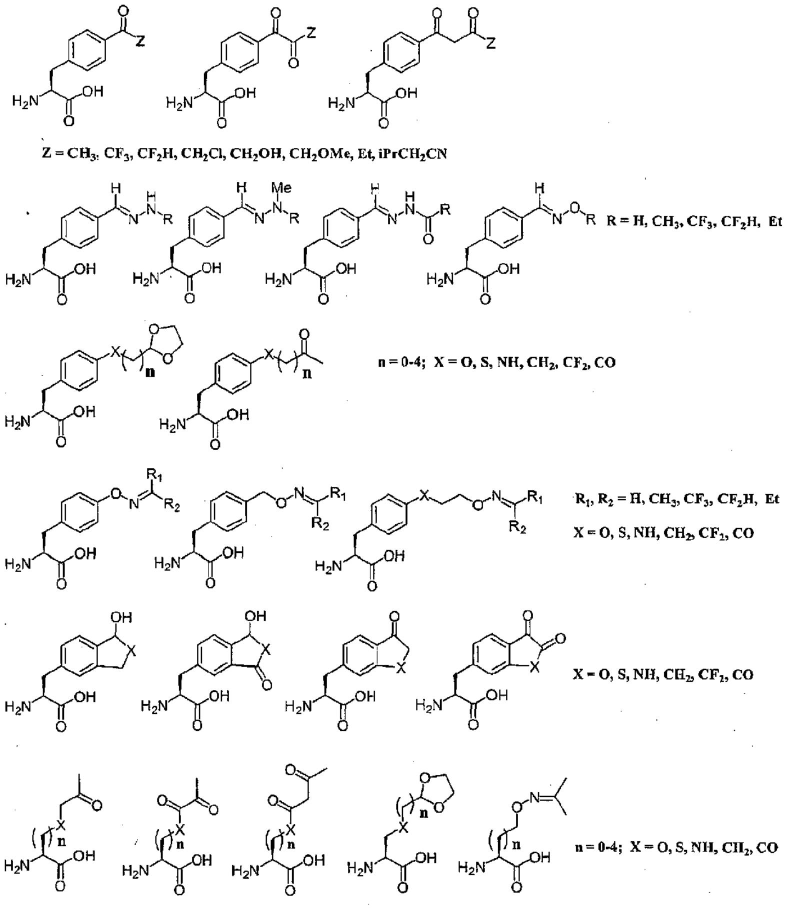


FIG. 1



$Z = CH_3, CF_3, CF_2H, CH_2Cl, CH_2OH, CH_2OMe, Et, iPrCH_2CN$

$R = H, CH_3, CF_3, CF_2H, Et$

$n = 0\text{-}4;\ X = O, S, NH, CH_2, CF_2, CO$

$R_1, R_2 = H, CH_3, CF_3, CF_2H, Et$

$X = O, S, NH, CH_2, CF_2, CO$

$X = O, S, NH, CH_2, CF_2, CO$

$n = 0\text{-}4;\ X = O, S, NH, CH_2, CO$

FIG. 2

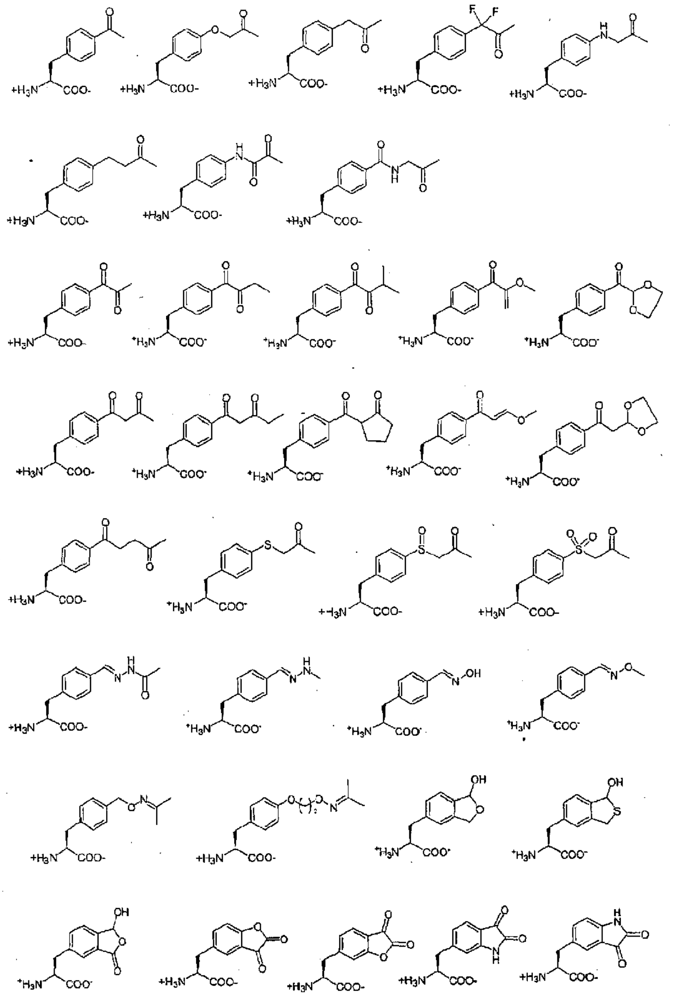
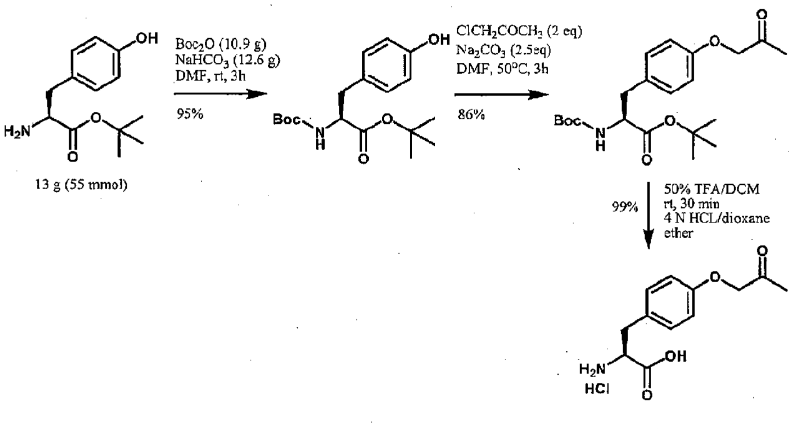


FIG. 3



13 g (55 mmol)
Boc2O (10.9 g)
NaHCO3 (12.6 g)
DMF, rt, 3h
95%
ClCH2COCH3 (2 eq)
Na2CO3 (2.5eq)
DMF, 50°C, 3h
86%
99%
50% TFA/DCM
rt, 30 min
4 N HCL/dioxane
ether
HCl

FIG. 4

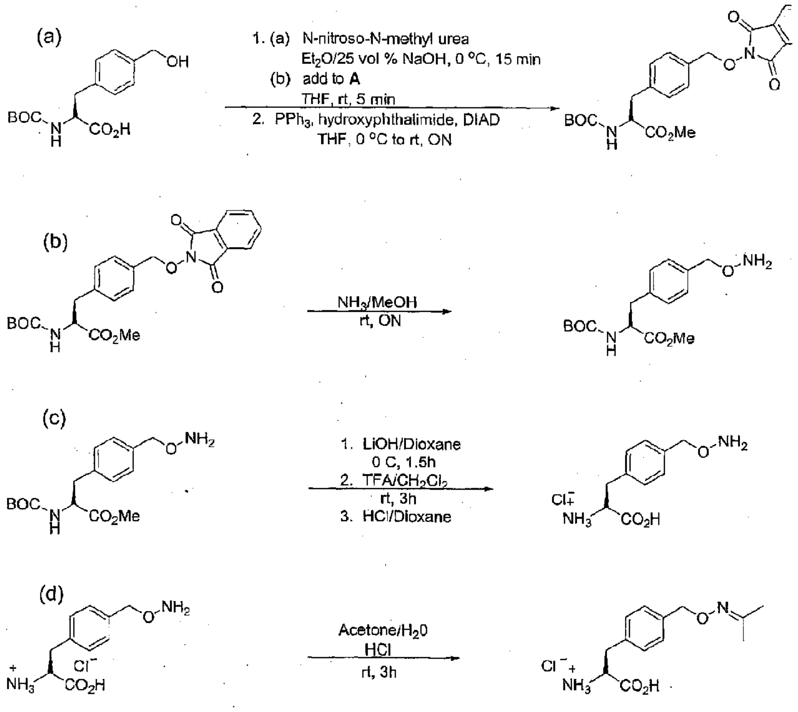

(a)
OH
BOC
N
H
CO2H
1. (a) N-nitroso-N-methyl urea
Et2O/25 vol % NaOH, 0 °C, 15 min
(b) add to A
THF, rt, 5 min
2. PPh3, hydroxyphthalimide, DIAD
THF, 0 °C to rt, ON
O
O
N
O
BOC
N
H
CO2Me
(b)
O
N
O
O
BOC
N
H
CO2Me
NH3/MeOH
rt, ON
O
NH2
BOC
N
H
CO2Me
(c)
O
NH2
BOC
N
H
CO2Me
1. LiOH/Dioxane
0 C, 1.5h
2. TFA/CH2Cl2
rt, 3h
3. HCl/Dioxane
O
NH2
Cl-
NH3
CO2H
(d)
O
NH2
+
Cl-
NH3
CO2H
Acetone/H20
HCl
rt, 3h
O
N
Cl- +
NH3
CO2H

FIG. 5

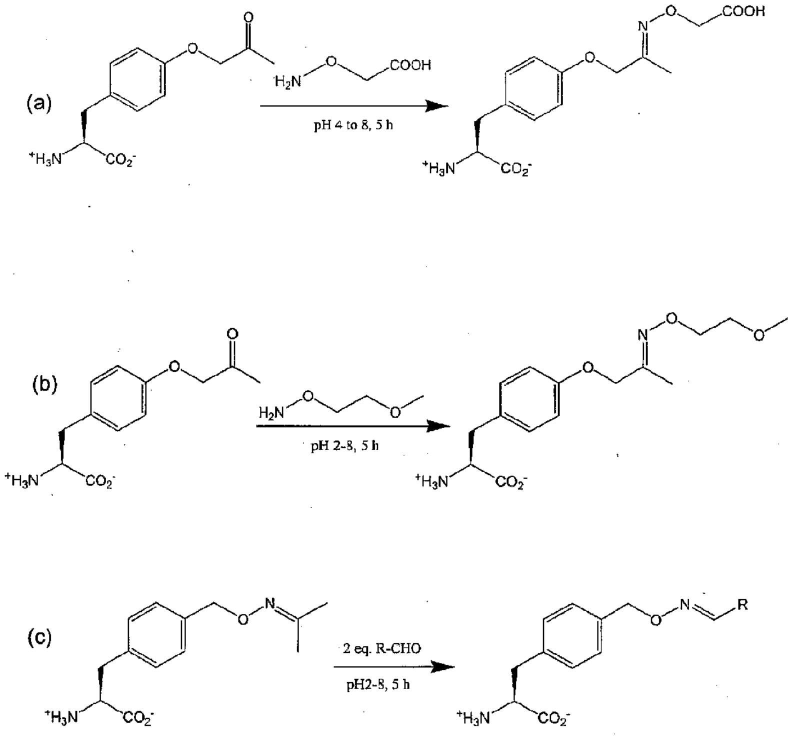
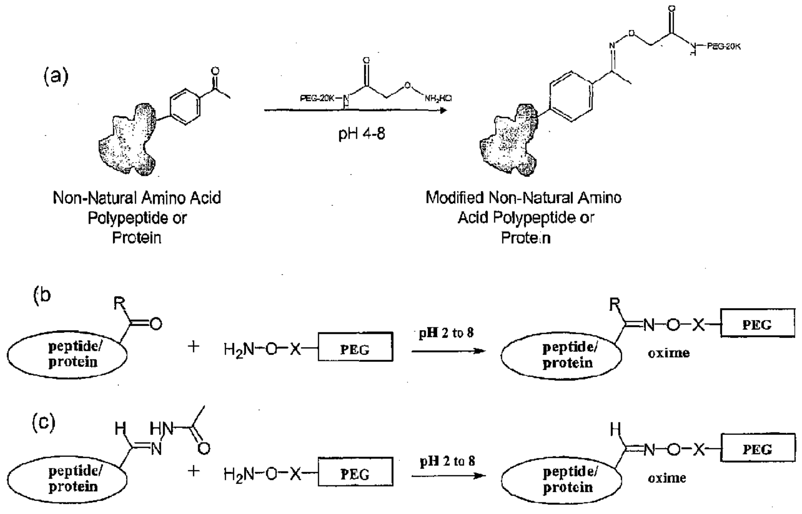

(a)
H2N–O–COOH
pH 4 to 8, 5 h
+H3N CO2-
COOH
(b)
H2N
pH 2-8, 5 h
+H3N CO2-
(c)
2 eq. R-CHO
pH2-8, 5 h
+H3N CO2-
R

FIG. 6



R = H, (substituted) alkyl, (substituted) aryl;

X = linker

FIG. 7

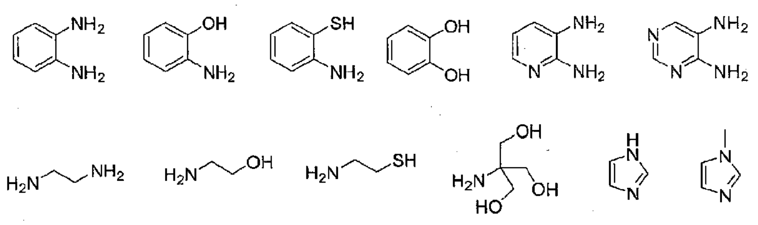
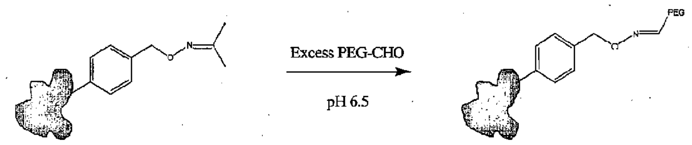


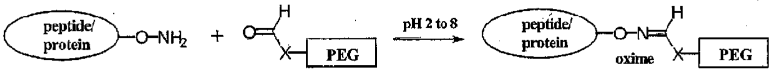
FIG. 8



Non-Natural Amino Acid
Polypeptide or
Protein

Modified Non-Natural Amino
Acid Polypeptide or
Protein

FIG. 9



R, R_1, R_2 = H, (substituted) alkyl, (substituted) aryl; X = linker

FIG. 10

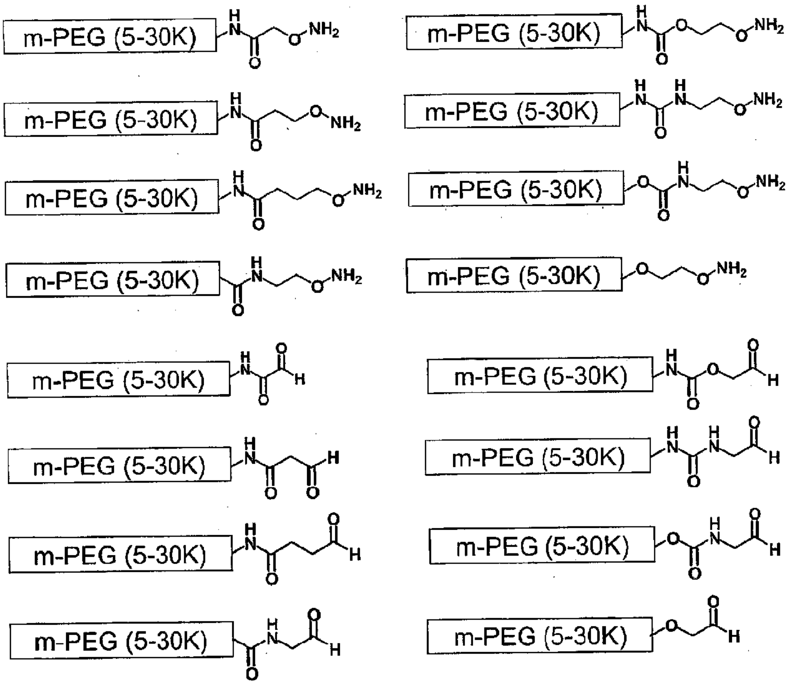

m-PEG (5-30K)
m-PEG (5-30K)
m-PEG (5-30K)
m-PEG (5-30K)
m-PEG (5-30K)
m-PEG (5-30K)
m-PEG (5-30K)
m-PEG (5-30K)
m-PEG (5-30K)
m-PEG (5-30K)
m-PEG (5-30K)
m-PEG (5-30K)
m-PEG (5-30K)
m-PEG (5-30K)
m-PEG (5-30K)
m-PEG (5-30K)

FIG. 11

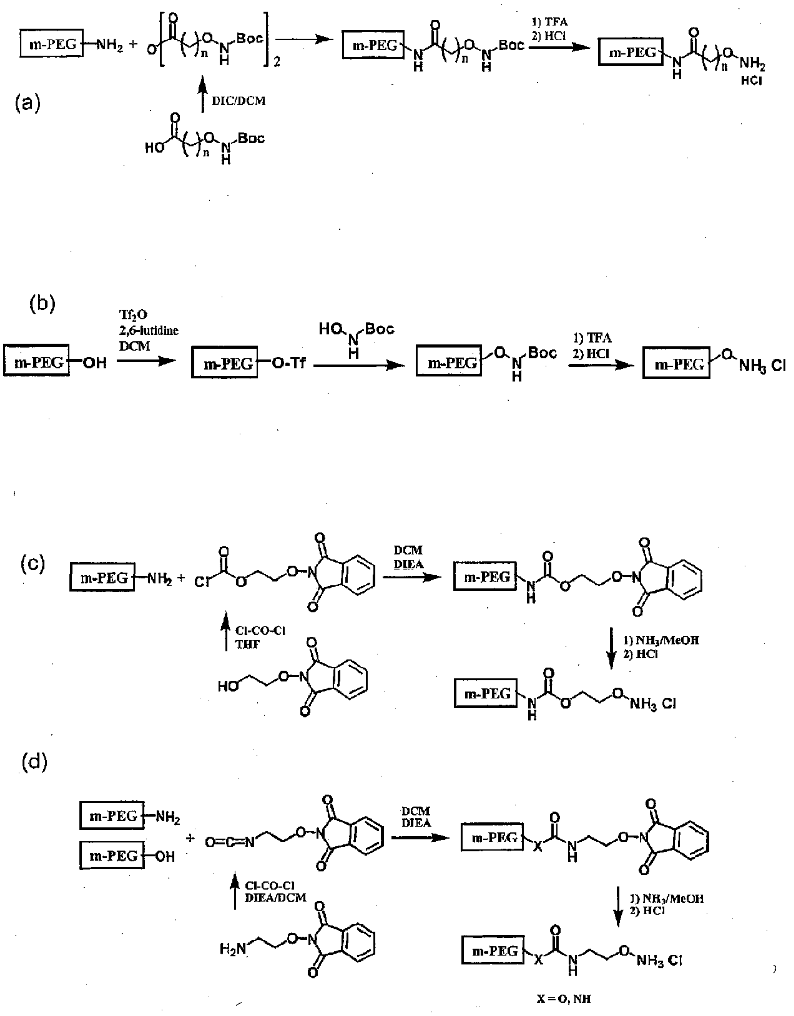

(a)
m-PEG
NH2
Boc
DIC/DCM
1) TFA
2) HCl
HCl
(b)
Tf2O
2,6-lutidine
DCM
O-Tf
HO N Boc
1) TFA
2) HCl
NH3 Cl
(c)
DCM
DIEA
Cl-CO-Cl
THF
1) NH3/MeOH
2) HCl
NH3 Cl
(d)
DCM
DIEA
Cl-CO-Cl
DIEA/DCM
O=C=N
1) NH3/MeOH
2) HCl
NH3 Cl
X = O, NH

FIG. 12

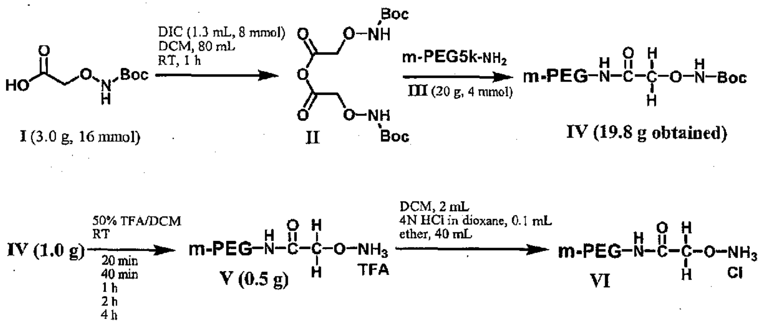
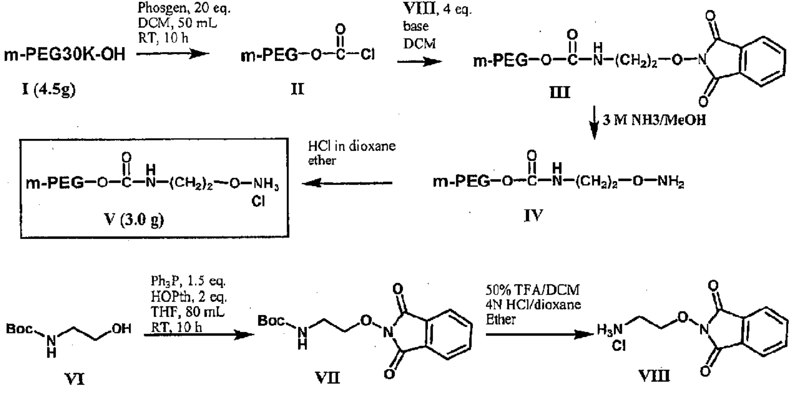

DIC (1.3 mL, 8 mmol)
DCM, 80 mL
RT, 1 h
I (3.0 g, 16 mmol)
II
m-PEG5k-NH2
III (20 g, 4 mmol)
IV (19.8 g obtained)
IV (1.0 g)
50% TFA/DCM
RT
20 min
40 min
1 h
2 h
4 h
V (0.5 g)
TFA
DCM, 2 mL
4N HCl in dioxane, 0.1 mL
ether, 40 mL
VI
Cl

FIG. 13

Route A:



Route B:

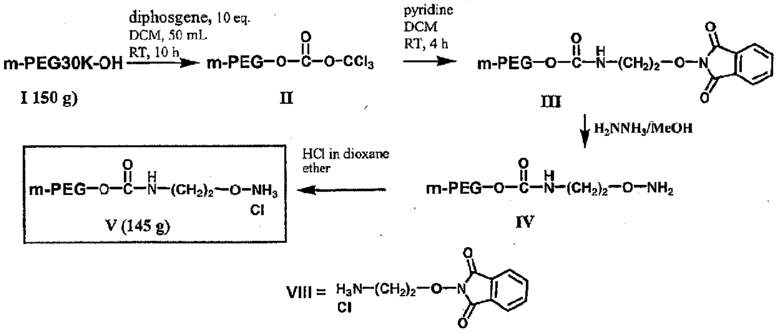
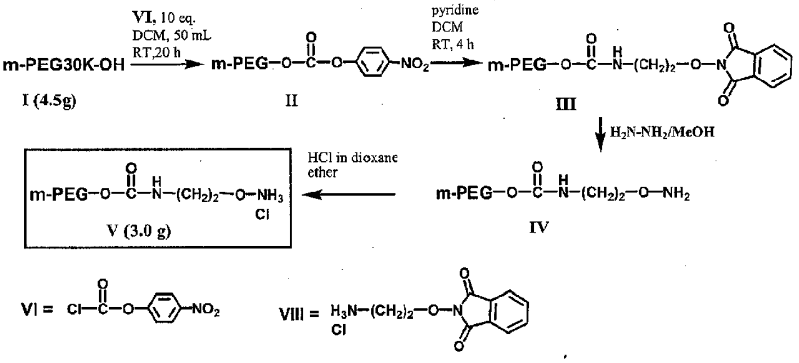


FIG. 14

Route C:



Route D:

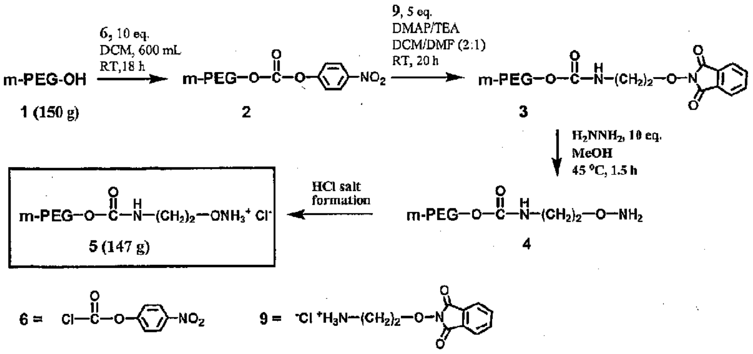


FIG. 15

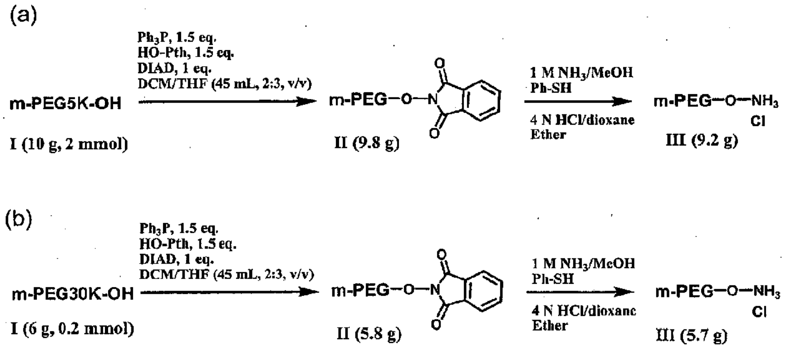

(a)
m-PEG5K-OH
I (10 g, 2 mmol)
Ph3P, 1.5 eq.
HO-Pth, 1.5 eq.
DIAD, 1 eq.
DCM/THF (45 mL, 2:3, v/v)
m-PEG—O—N
O
O
II (9.8 g)
1 M NH3/MeOH
Ph-SH
4 N HCl/dioxane
Ether
m-PEG—O—NH3
Cl
III (9.2 g)
(b)
m-PEG30K-OH
I (6 g, 0.2 mmol)
Ph3P, 1.5 eq.
HO-Pth, 1.5 eq.
DIAD, 1 eq.
DCM/THF (45 mL, 2:3, v/v)
m-PEG—O—N
O
O
II (5.8 g)
1 M NH3/MeOH
Ph-SH
4 N HCl/dioxane
Ether
m-PEG—O—NH3
Cl
III (5.7 g)

FIG. 16

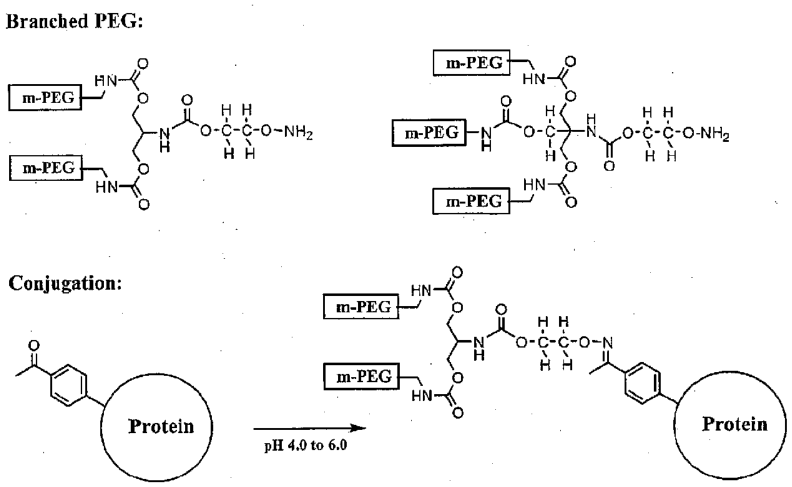

Branched PEG:
m-PEG
HN
O
N
H
O-C-C-O-NH₂
Conjugation:
Protein
pH 4.0 to 6.0
O-C-C-O-N

FIG. 17

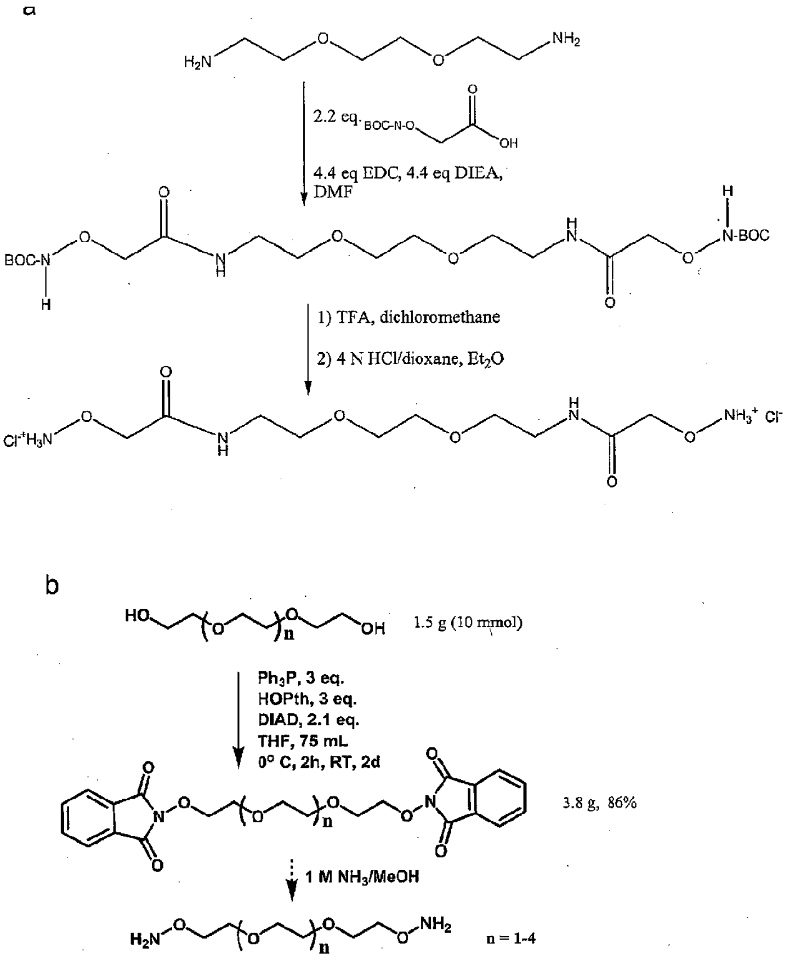

a
2.2 eq. BOC-N-O
4.4 eq EDC, 4.4 eq DIEA, DMF
1) TFA, dichloromethane
2) 4 N HCl/dioxane, Et2O
b
1.5 g (10 mmol)
Ph3P, 3 eq.
HOPth, 3 eq.
DIAD, 2.1 eq.
THF, 75 mL
0° C, 2h, RT, 2d
3.8 g, 86%
1 M NH3/MeOH
n = 1-4

FIG. 18

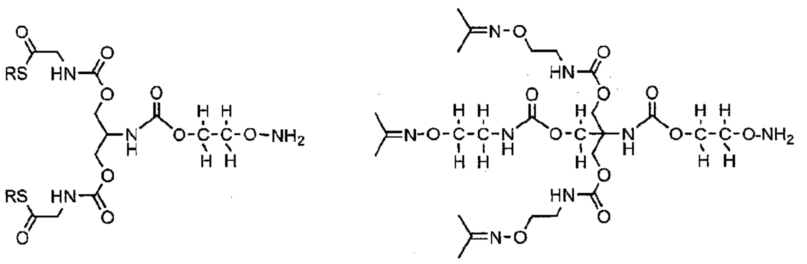


FIG. 19

Bifunctional linker:

Scheme:

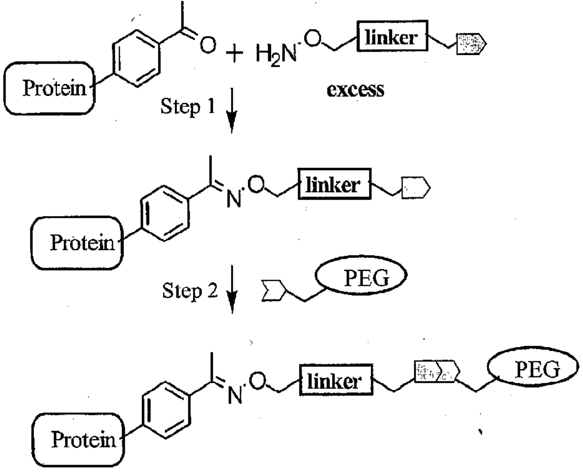


FIG. 20

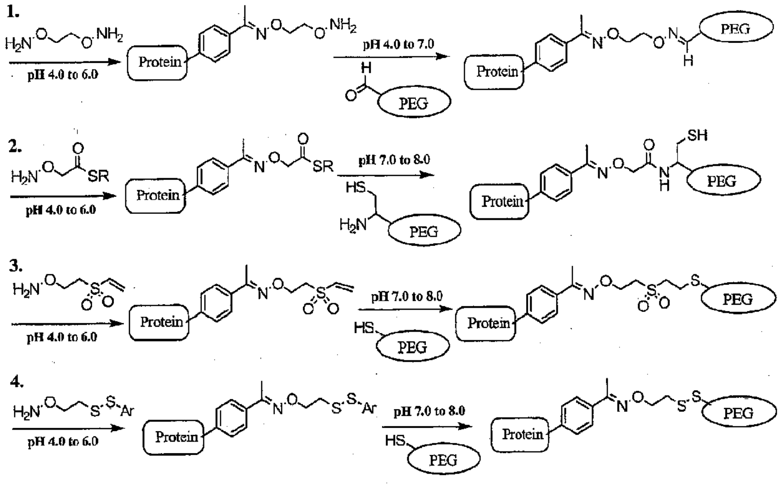

1.
2.
3.
4.
Protein
PEG
pH 4.0 to 6.0
pH 4.0 to 7.0
pH 7.0 to 8.0
H₂N
NH₂
SR
SH
HS
Ar

FIG. 21

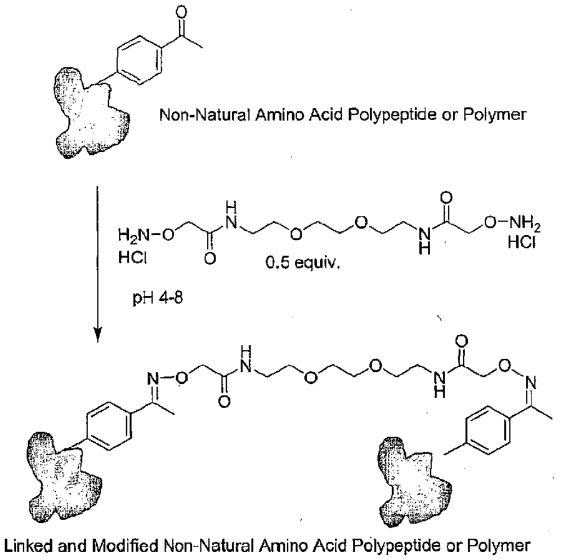

Non-Natural Amino Acid Polypeptide or Polymer
H2N-O
HCl
O-NH2
HCl
0.5 equiv.
pH 4-8
Linked and Modified Non-Natural Amino Acid Polypeptide or Polymer

FIG. 22

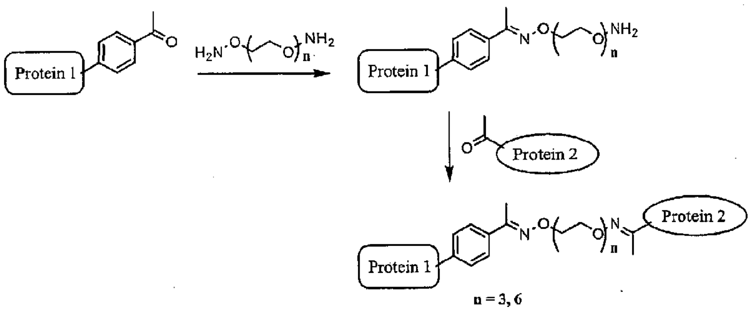

Protein 1
H_2N-O-(~O)n-NH_2
Protein 1
Protein 2
Protein 1
Protein 2
n = 3, 6

FIG. 23

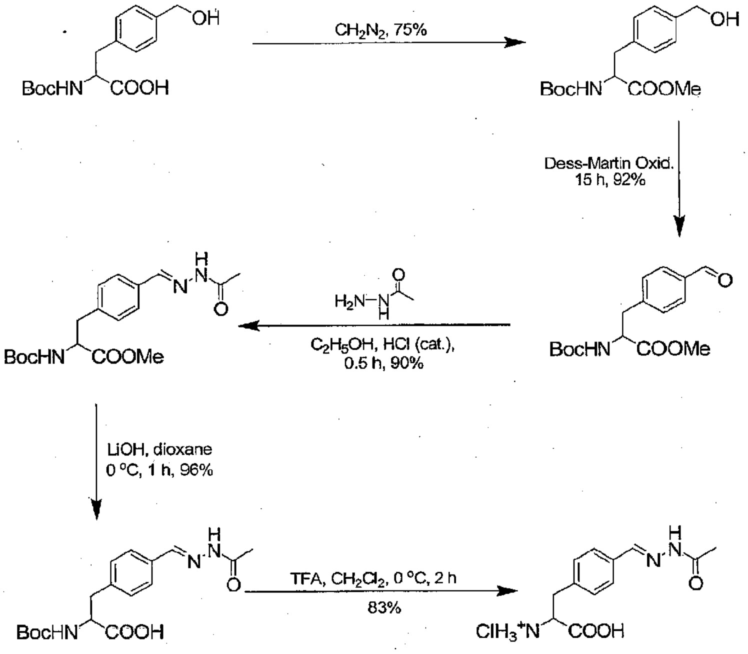

BocHN COOH
OH
CH2N2, 75%
BocHN COOMe
OH
Dess-Martin Oxid.
15 h, 92%
BocHN COOMe
O
H2N-N(H)-C(O)CH3
C2H5OH, HCl (cat.),
0.5 h, 90%
BocHN COOMe
N
H
N
O
LiOH, dioxane
0 °C, 1 h, 96%
BocHN COOH
TFA, CH2Cl2, 0 °C, 2 h
83%
ClH3+N COOH

FIG. 24

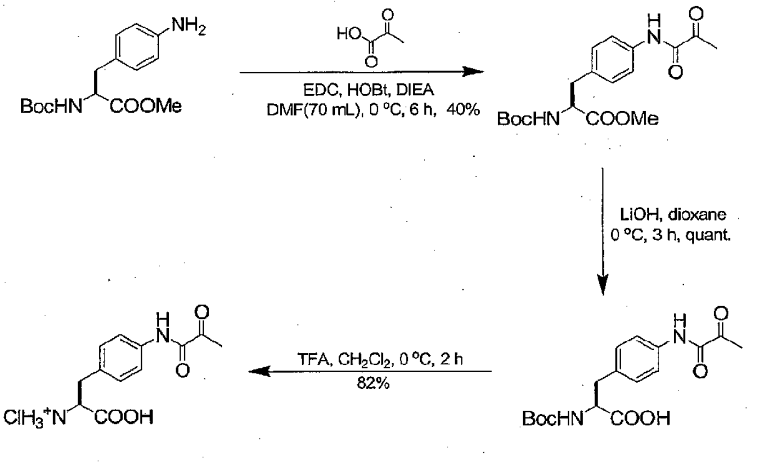

NH$_2$
BocHN COOMe
HO
EDC, HOBt, DIEA
DMF(70 mL), 0 °C, 6 h, 40%
H N
BocHN COOMe
LiOH, dioxane
0 °C, 3 h, quant.
BocHN COOH
TFA, CH$_2$Cl$_2$, 0 °C, 2 h
82%
ClH$_3$$^+$N COOH

FIG. 25

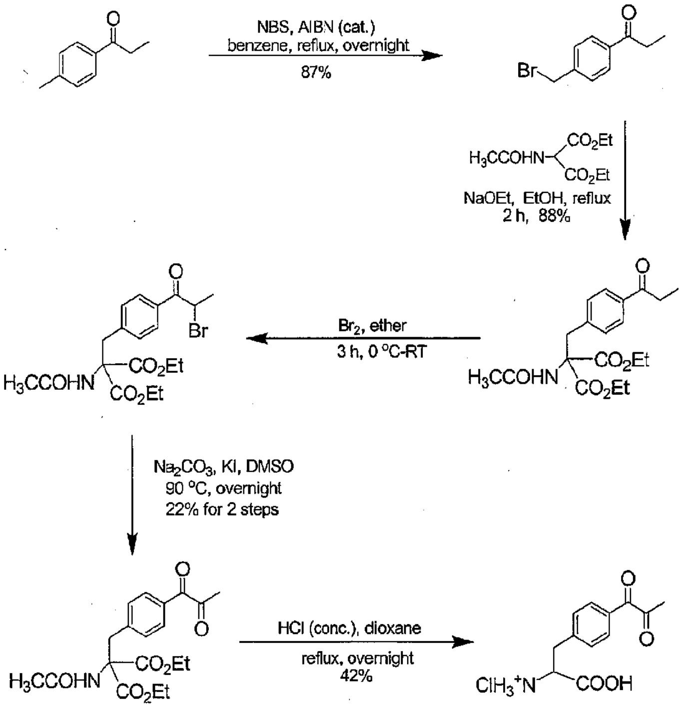

NBS, AIBN (cat.)
benzene, reflux, overnight
87%
Br
H3CCOHN
CO2Et
CO2Et
NaOEt, EtOH, reflux
2 h, 88%
O
Br2, ether
3 h, 0 °C-RT
Br
Na2CO3, KI, DMSO
90 °C, overnight
22% for 2 steps
HCl (conc.), dioxane
reflux, overnight
42%
ClH3+N
COOH

FIG. 26

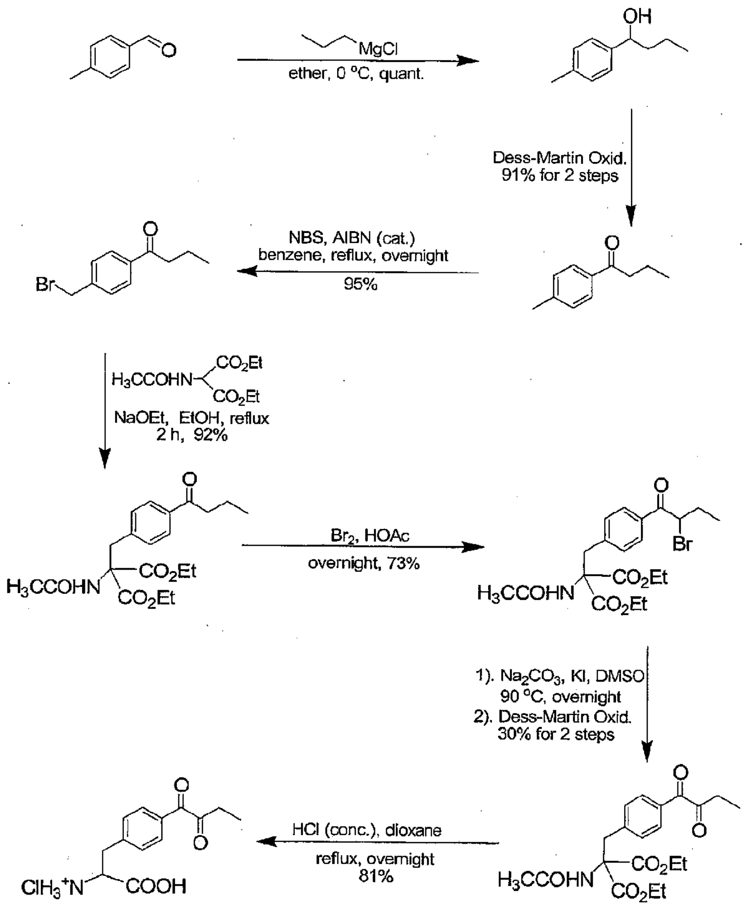

O
MgCl
ether, 0 °C, quant.
OH
Dess-Martin Oxid.
91% for 2 steps
O
NBS, AIBN (cat.)
benzene, reflux, overnight
95%
O
Br
H3CCOHN
CO2Et
CO2Et
NaOEt, EtOH, reflux
2 h, 92%
O
H3CCOHN
CO2Et
CO2Et
Br2, HOAc
overnight, 73%
O
Br
H3CCOHN
CO2Et
CO2Et
1). Na2CO3, KI, DMSO
90 °C, overnight
2). Dess-Martin Oxid.
30% for 2 steps
O
O
H3CCOHN
CO2Et
CO2Et
HCl (conc.), dioxane
reflux, overnight
81%
O
O
ClH3+N
COOH

FIG. 27

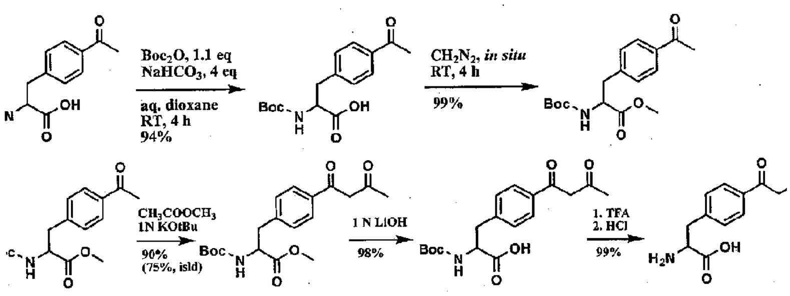

Boc$_2$O, 1.1 eq
NaHCO$_3$, 4 eq
aq. dioxane
RT, 4 h
94%
CH$_2$N$_2$, in situ
RT, 4 h
99%
CH$_3$COOCH$_3$
1N KOtBu
90%
(75%, isld)
1 N LiOH
98%
1. TFA
2. HCl
99%
Boc
OH
O
H
N
H$_2$N

FIG. 28

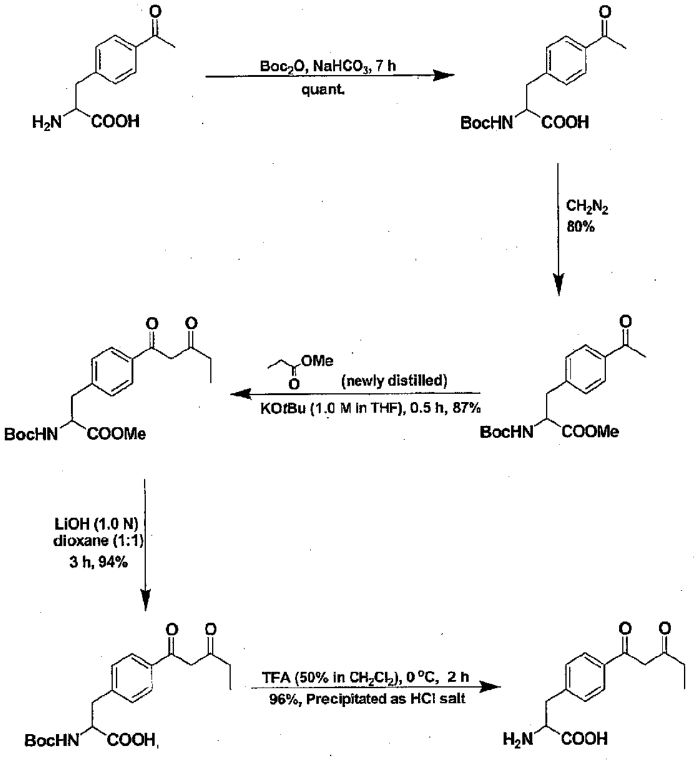
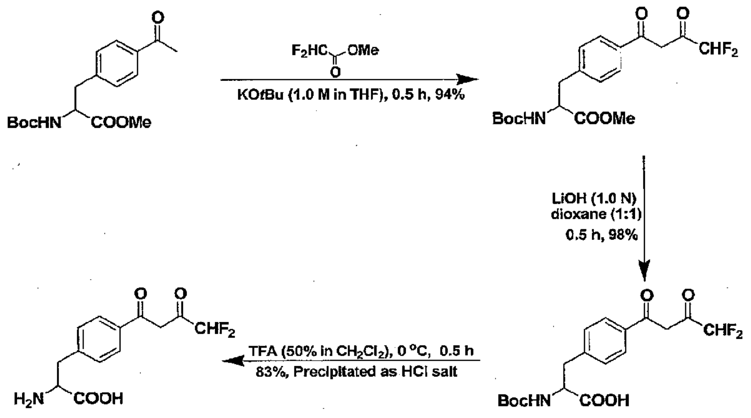

O
H₂N COOH
Boc₂O, NaHCO₃, 7 h
quant.
O
BocHN COOH
CH₂N₂
80%
O O
OMe
O
(newly distilled)
KOtBu (1.0 M in THF), 0.5 h, 87%
BocHN COOMe
O
BocHN COOMe
LiOH (1.0 N)
dioxane (1:1)
3 h, 94%
O O
BocHN COOH,
TFA (50% in CH₂Cl₂), 0 °C, 2 h
96%, Precipitated as HCl salt
O O
H₂N COOH

FIG. 29



BocHN
COOMe
F2HC OMe
O
KOtBu (1.0 M in THF), 0.5 h, 94%
O
O
CHF2
BocHN
COOMe
LiOH (1.0 N)
dioxane (1:1)
0.5 h, 98%
O
O
CHF2
BocHN
COOH
TFA (50% in CH2Cl2), 0 °C, 0.5 h
83%, Precipitated as HCl salt
O
O
CHF2
H2N
COOH

FIG. 30

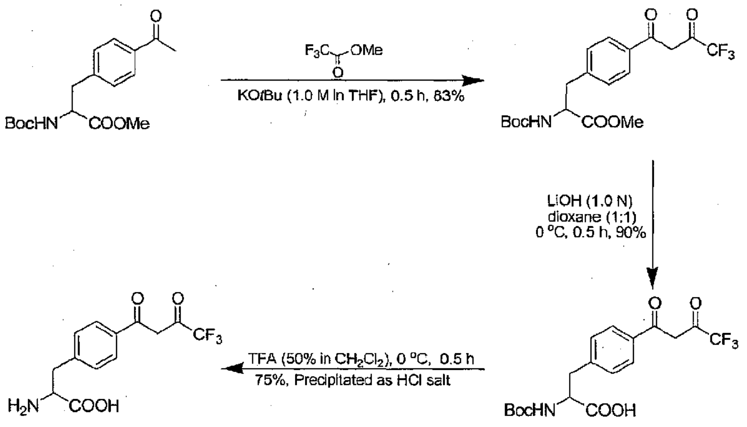

BocHN
COOMe
F3C OMe
O
KOtBu (1.0 M in THF), 0.5 h, 83%
O
O
CF3
BocHN
COOMe
LiOH (1.0 N)
dioxane (1:1)
0 °C, 0.5 h, 90%
O
O
CF3
BocHN
COOH
TFA (50% in CH2Cl2), 0 °C, 0.5 h
75%, Precipitated as HCl salt
O
O
CF3
H2N
COOH

FIG. 31

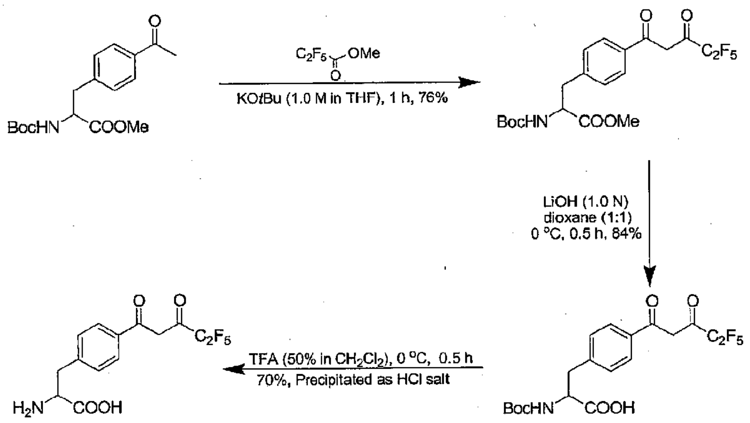

BocHN
COOMe
C2F5
OMe
O
KOtBu (1.0 M in THF), 1 h, 76%
O
LiOH (1.0 N)
dioxane (1:1)
0 °C, 0.5 h, 84%
COOH
TFA (50% in CH2Cl2), 0 °C, 0.5 h
70%, Precipitated as HCl salt
H2N

FIG. 32

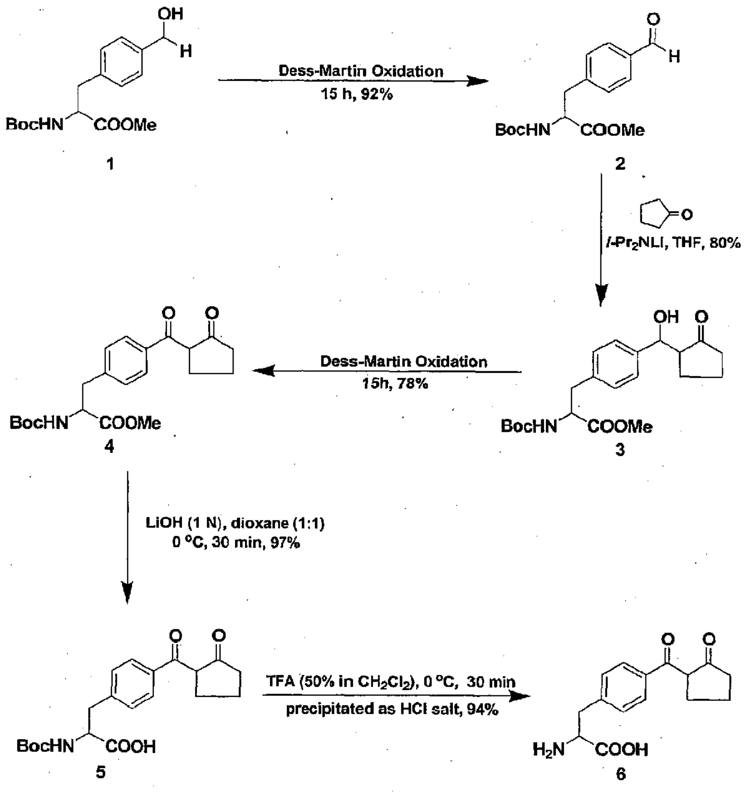

OH
H
BocHN
COOMe
1
Dess-Martin Oxidation
15 h, 92%
O
H
BocHN
COOMe
2
O
i-Pr$_2$NLi, THF, 80%
OH
O
BocHN
COOMe
3
Dess-Martin Oxidation
15h, 78%
O
O
BocHN
COOMe
4
LiOH (1 N), dioxane (1:1)
0 °C, 30 min, 97%
O
O
BocHN
COOH
5
TFA (50% in CH$_2$Cl$_2$), 0 °C, 30 min
precipitated as HCl salt, 94%
O
O
H$_2$N
COOH
6

FIG. 33

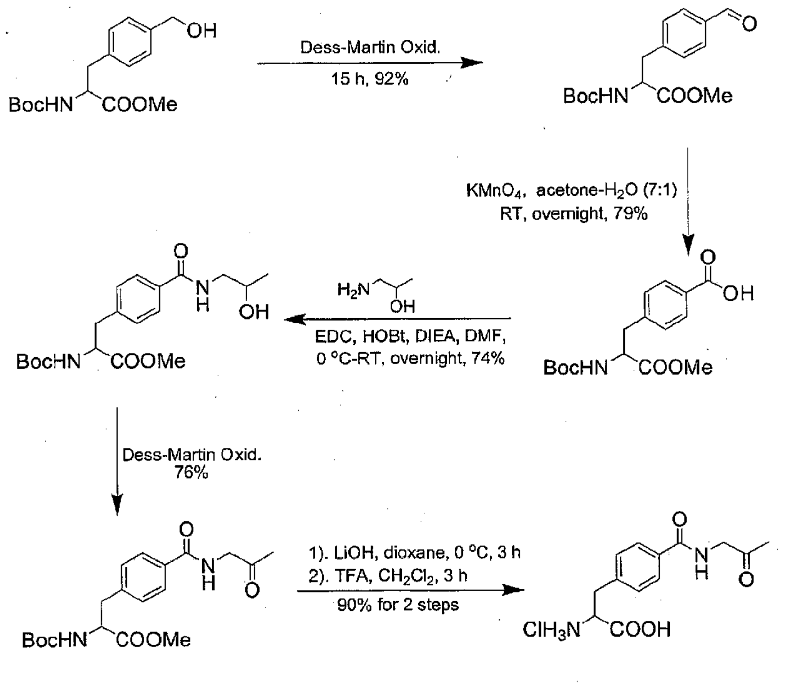

OH
BocHN
COOMe
Dess-Martin Oxid.
15 h, 92%
O
BocHN
COOMe
KMnO4, acetone-H2O (7:1)
RT, overnight, 79%
O
OH
BocHN
COOMe
H2N
OH
EDC, HOBt, DIEA, DMF,
0 °C-RT, overnight, 74%
O
N
H
OH
BocHN
COOMe
Dess-Martin Oxid.
76%
O
N
H
O
BocHN
COOMe
1). LiOH, dioxane, 0 °C, 3 h
2). TFA, CH2Cl2, 3 h
90% for 2 steps
O
N
H
O
ClH3N
COOH

FIG. 34

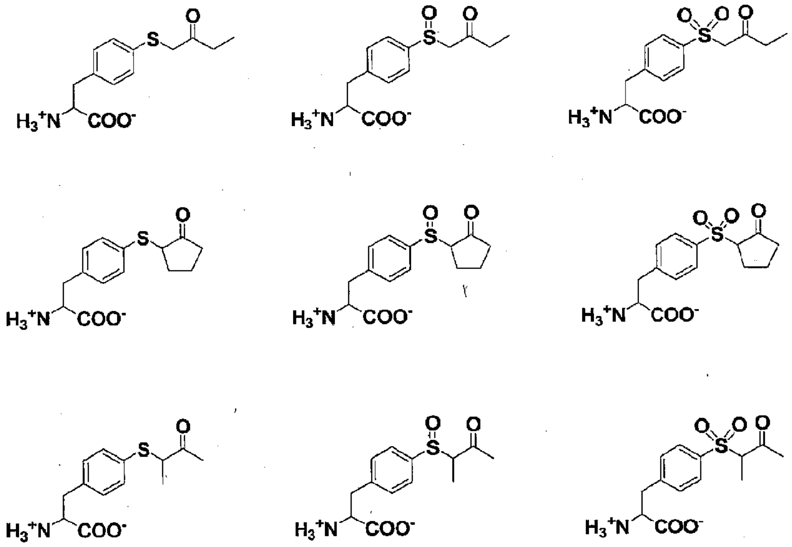


FIG. 35

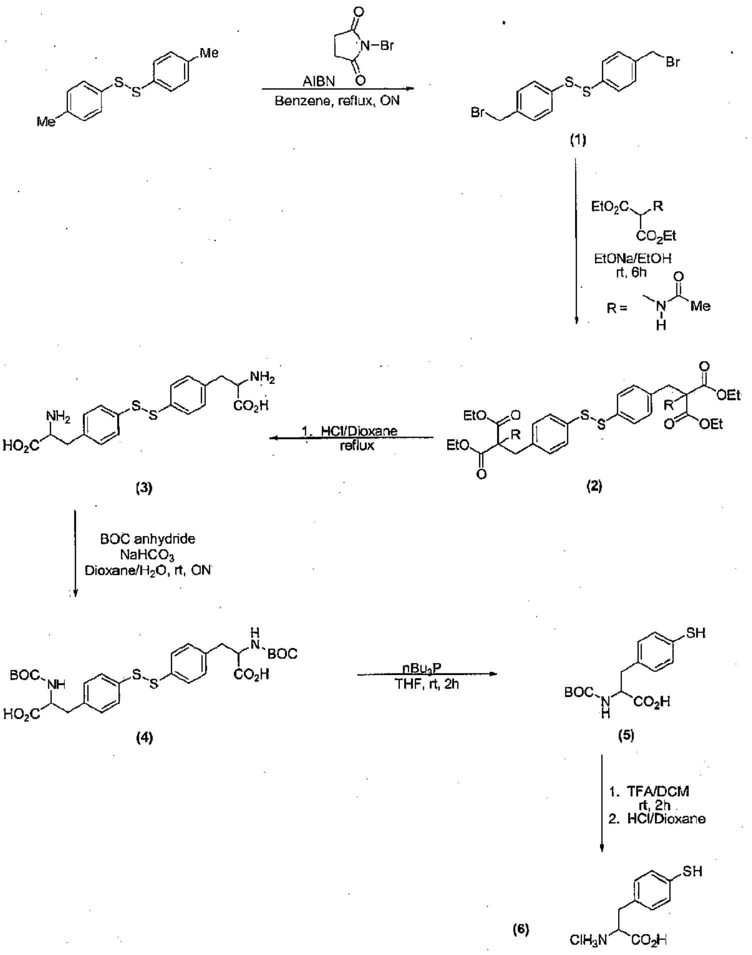

Me
Me
S
S
O
N−Br
O
AIBN
Benzene, reflux, ON
Br
Br
(1)
EtO_2C
R
CO_2Et
EtONa/EtOH
rt, 6h
R =
O
N
H
Me
EtO
O
R
EtO
O
O
OEt
R
O
OEt
(2)
1. HCl/Dioxane
reflux
NH_2
CO_2H
NH_2
HO_2C
(3)
BOC anhydride
$NaHCO_3$
Dioxane/H_2O, rt, ON
H
N
BOC
CO_2H
BOC
NH
HO_2C
(4)
nBu_3P
THF, rt, 2h
SH
BOC
N
H
CO_2H
(5)
1. TFA/DCM
rt, 2h
2. HCl/Dioxane
SH
(6)
ClH_3N
CO_2H

FIG. 36

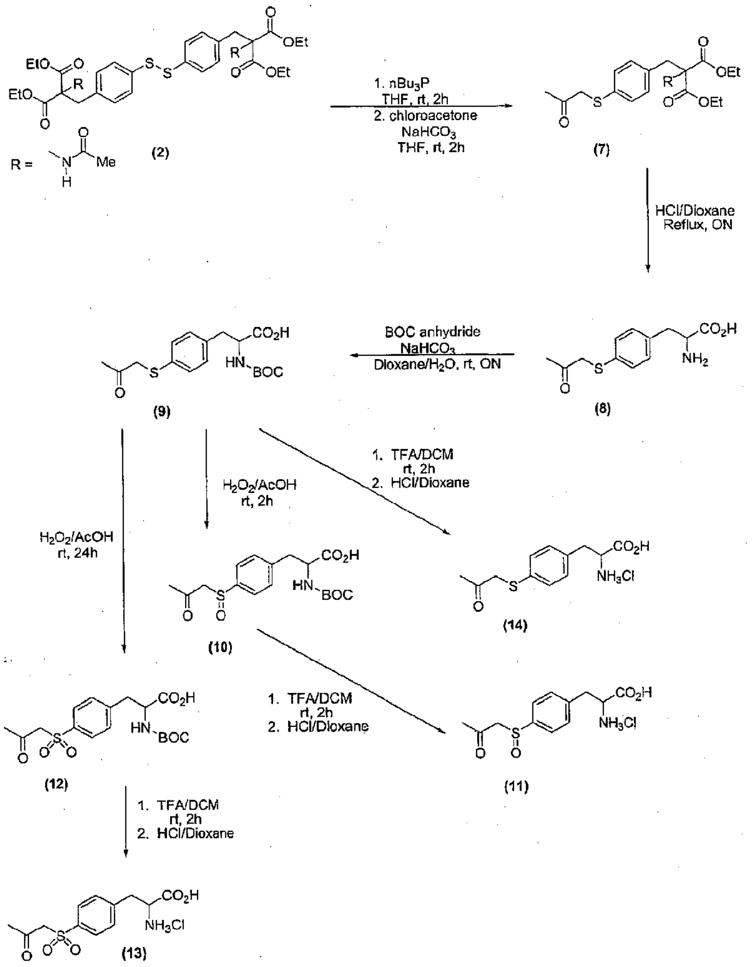

EtO
O
R
EtO
O
S
S
R
O
OEt
O
OEt
R =
N
H
O
Me
(2)
1. nBu$_3$P
THF, rt, 2h
2. chloroacetone
NaHCO$_3$
THF, rt, 2h
O
S
R
O
OEt
O
OEt
(7)
HCl/Dioxane
Reflux, ON
S
O
CO$_2$H
NH$_2$
(8)
BOC anhydride
NaHCO$_3$
Dioxane/H$_2$O, rt, ON
S
O
CO$_2$H
HN
BOC
(9)
H$_2$O$_2$/AcOH
rt, 24h
H$_2$O$_2$/AcOH
rt, 2h
1. TFA/DCM
rt, 2h
2. HCl/Dioxane
S
O
CO$_2$H
NH$_3$Cl
(14)
S
O
O
CO$_2$H
HN
BOC
(10)
1. TFA/DCM
rt, 2h
2. HCl/Dioxane
S
O
O
CO$_2$H
NH$_3$Cl
(11)
S
O
O
O
CO$_2$H
HN
BOC
(12)
1. TFA/DCM
rt, 2h
2. HCl/Dioxane
S
O
O
O
CO$_2$H
NH$_3$Cl
(13)

FIG. 37

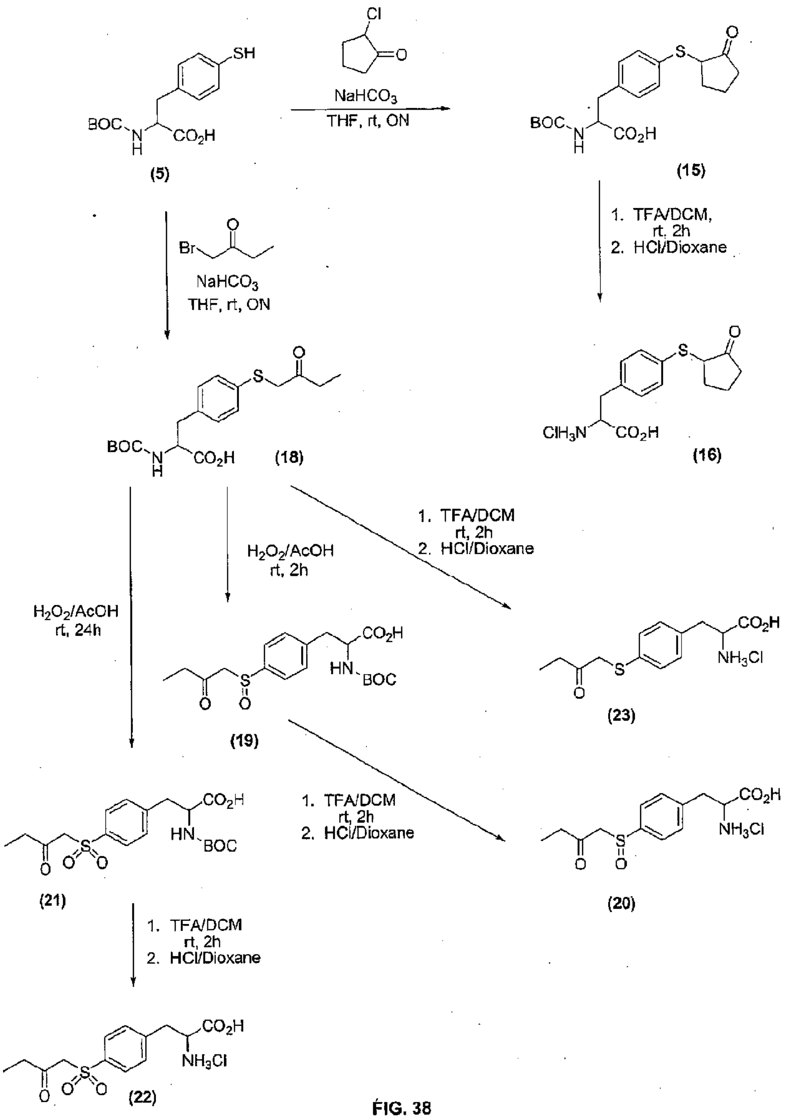

SH
BOC
N
H
CO2H
(5)
Cl
O
NaHCO3
THF, rt, ON
S
O
(15)
1. TFA/DCM, rt, 2h
2. HCl/Dioxane
ClH3N
(16)
Br
NaHCO3
THF, rt, ON
(18)
1. TFA/DCM rt, 2h
2. HCl/Dioxane
H2O2/AcOH rt, 2h
H2O2/AcOH rt, 24h
HN
(19)
NH3Cl
(23)
(21)
1. TFA/DCM rt, 2h
2. HCl/Dioxane
(20)
1. TFA/DCM rt, 2h
2. HCl/Dioxane
(22)

FIG. 38

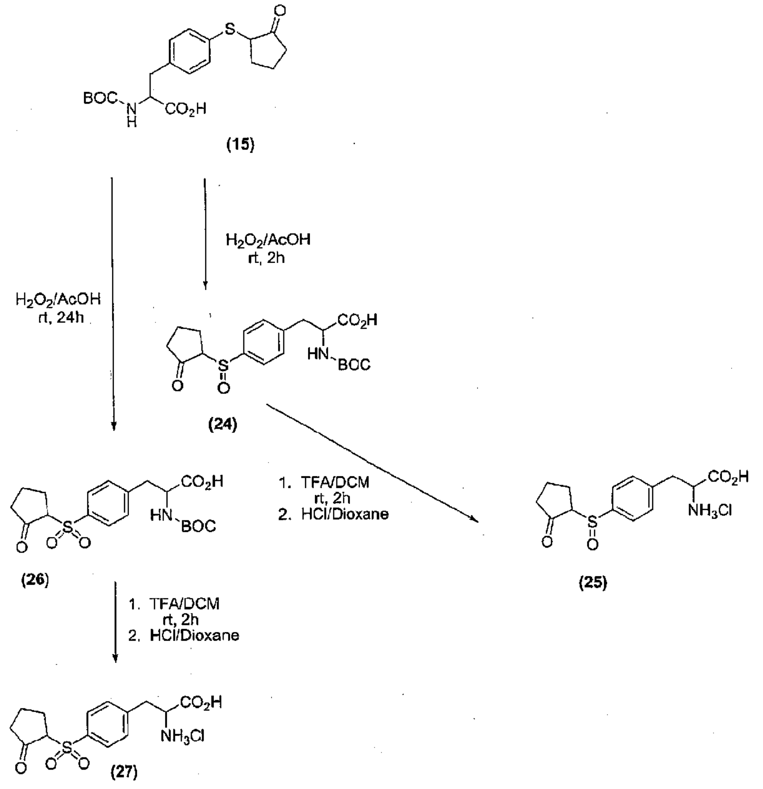

(15)
BOC
N
H
CO2H
S
O
H2O2/AcOH
rt, 2h
H2O2/AcOH
rt, 24h
(24)
CO2H
HN
BOC
S
O
(26)
1. TFA/DCM
rt, 2h
2. HCl/Dioxane
(25)
NH3Cl
1. TFA/DCM
rt, 2h
2. HCl/Dioxane
(27)

FIG. 39

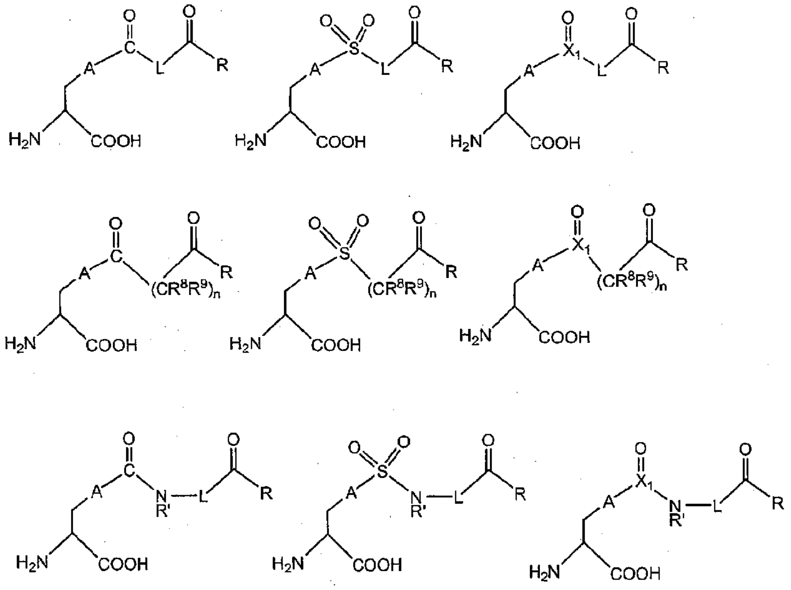


FIG. 40

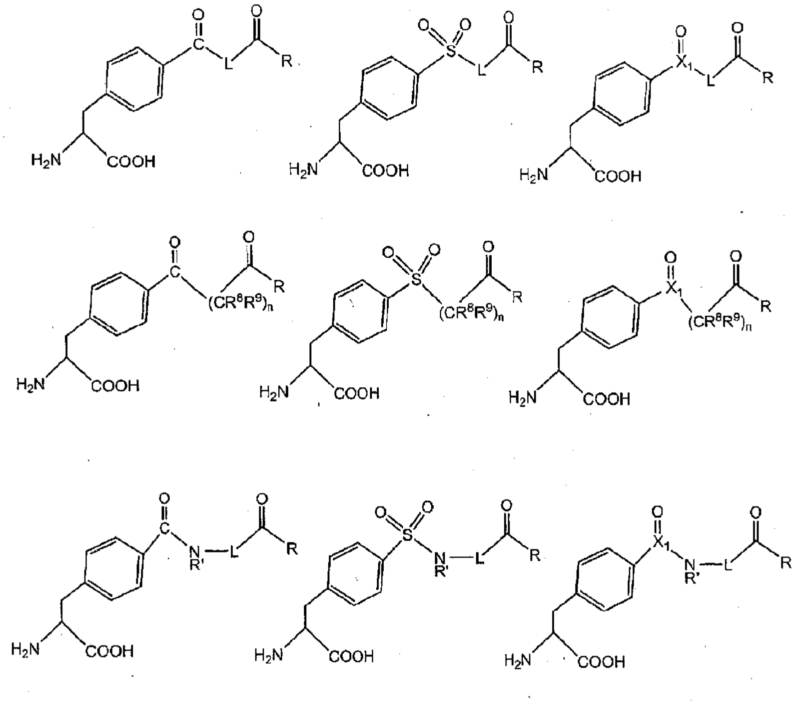


FIG. 41

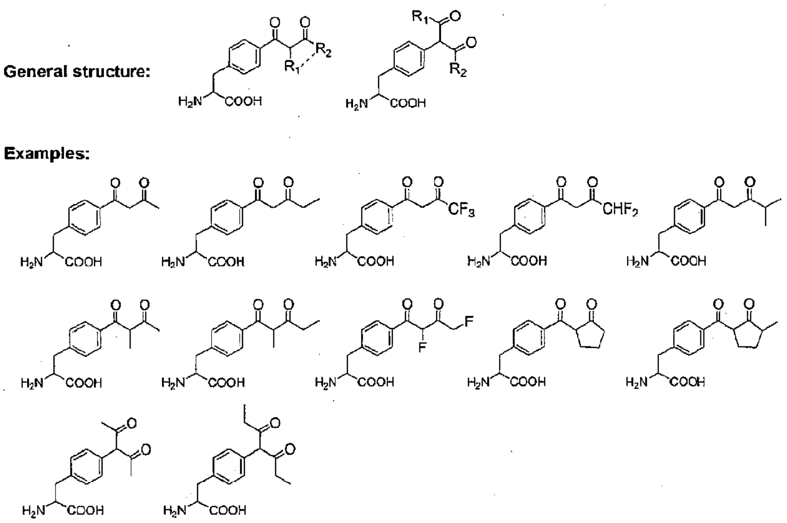
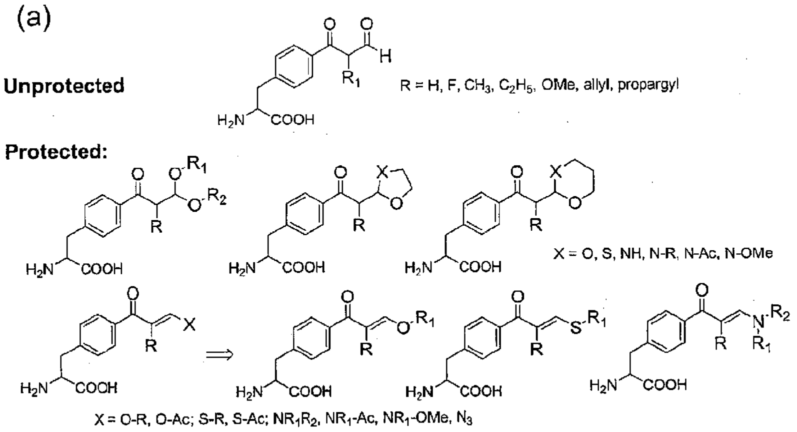


FIG. 42


(a)
Unprotected
R = H, F, CH3, C2H5, OMe, allyl, propargyl
Protected:
X = O, S, NH, N-R, N-Ac, N-OMe
X = O-R, O-Ac; S-R, S-Ac; NR1R2, NR1-Ac, NR1-OMe, N3

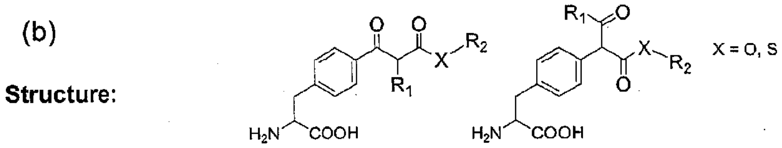

(b)
Structure:
X = O, S

FIG. 43

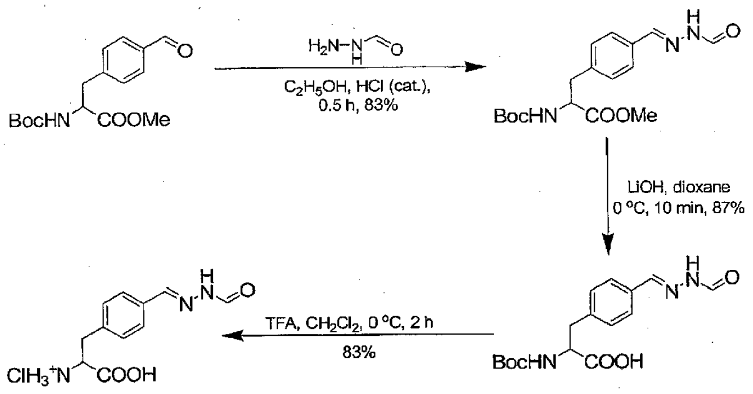

BocHN COOMe
H2N-NH CHO
C2H5OH, HCl (cat.), 0.5 h, 83%
BocHN COOMe
LiOH, dioxane
0 °C, 10 min, 87%
BocHN COOH
TFA, CH2Cl2, 0 °C, 2 h
83%
ClH3+N COOH

FIG. 44

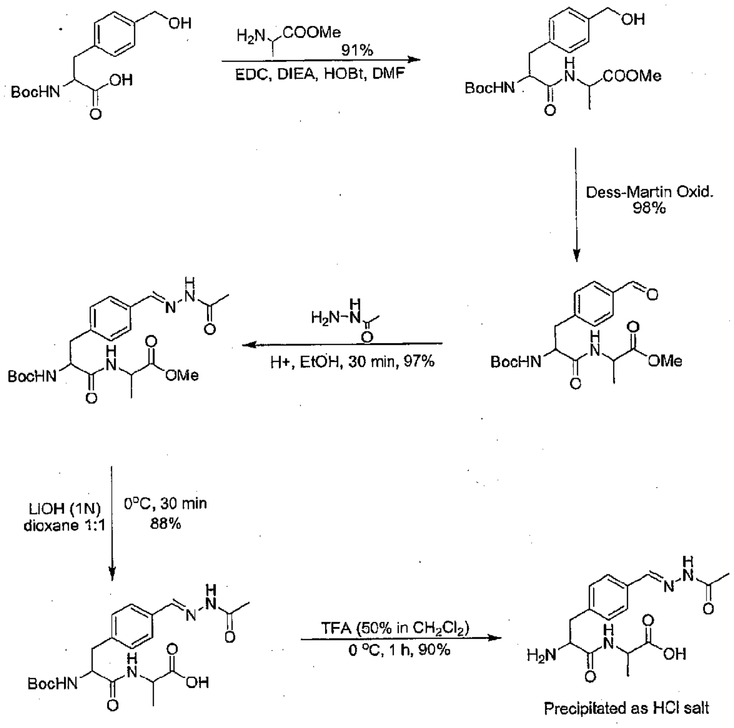

OH
BocHN
OH
O
H_2N COOMe
91%
EDC, DIEA, HOBt, DMF
OH
BocHN
H
N
COOMe
O
Dess-Martin Oxid.
98%
O
BocHN
H
N
O
OMe
O
H_2N-N H
O
H+, EtOH, 30 min, 97%
N H N
O
BocHN
H
N
O
OMe
O
LiOH (1N)
dioxane 1:1
0°C, 30 min
88%
N H N
O
BocHN
H
N
O
OH
O
TFA (50% in CH_2Cl_2)
0 °C, 1 h, 90%
N H N
O
H_2N
H
N
O
OH
O
Precipitated as HCl salt

FIG. 45

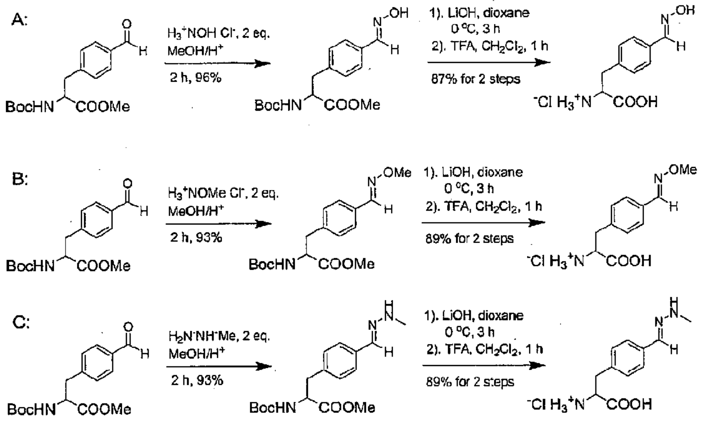
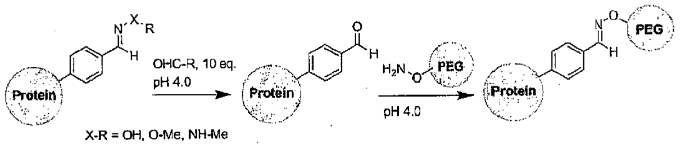

A:
H3+NOH Cl-, 2 eq.
MeOH/H+
2 h, 96%
1). LiOH, dioxane
0 °C, 3 h
2). TFA, CH2Cl2, 1 h
87% for 2 steps
B:
H3+NOMe Cl-, 2 eq.
MeOH/H+
2 h, 93%
1). LiOH, dioxane
0 °C, 3 h
2). TFA, CH2Cl2, 1 h
89% for 2 steps
C:
H2NNHMe, 2 eq.
MeOH/H+
2 h, 93%
1). LiOH, dioxane
0 °C, 3 h
2). TFA, CH2Cl2, 1 h
89% for 2 steps
BocHN
COOMe
-Cl H3+N
COOH
OH
OMe
H

FIG. 46

Two-step conjugation:



One-step conjugation:

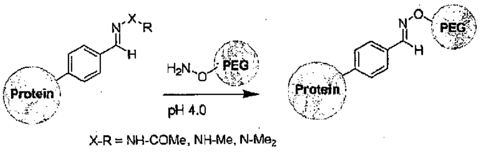


FIG. 47

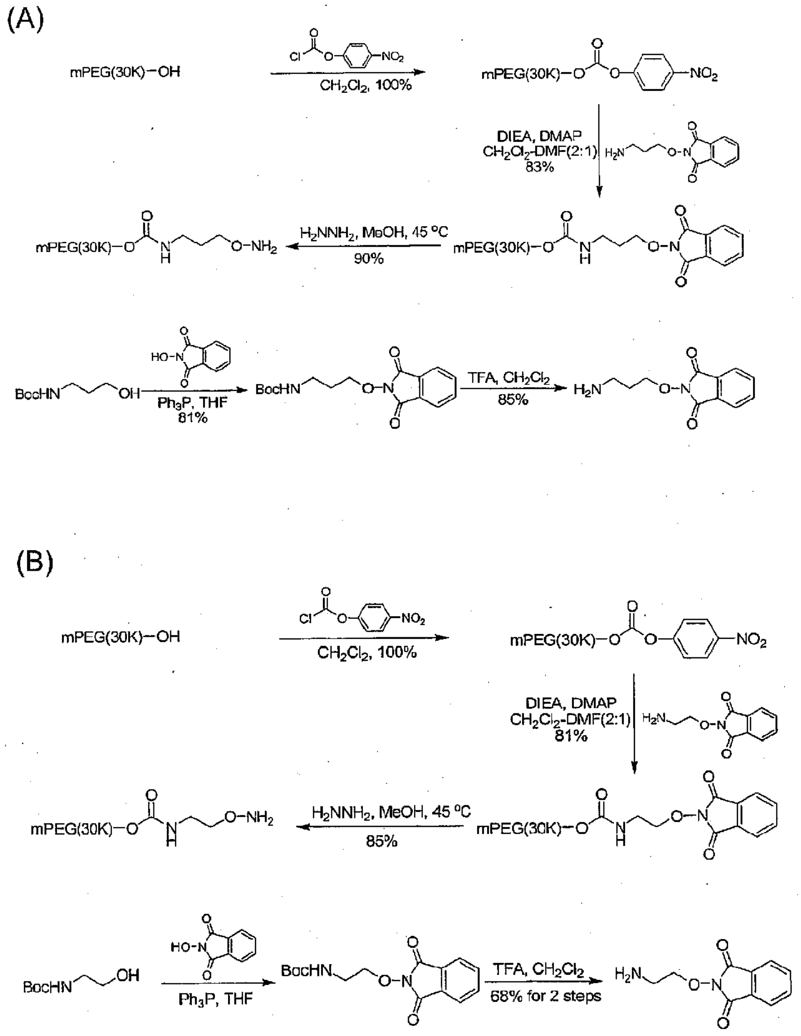

(A)
mPEG(30K)−OH
CH2Cl2, 100%
mPEG(30K)−O
NO2
DIEA, DMAP
CH2Cl2-DMF(2:1)
83%
H2N
O−N
H2NNH2, MeOH, 45 °C
90%
mPEG(30K)−O
O−NH2
BocHN
OH
HO−N
Ph3P, THF
81%
TFA, CH2Cl2
85%
(B)
mPEG(30K)−OH
CH2Cl2, 100%
DIEA, DMAP
CH2Cl2-DMF(2:1)
81%
H2NNH2, MeOH, 45 °C
85%
BocHN
Ph3P, THF
TFA, CH2Cl2
68% for 2 steps
H2N

FIG. 48

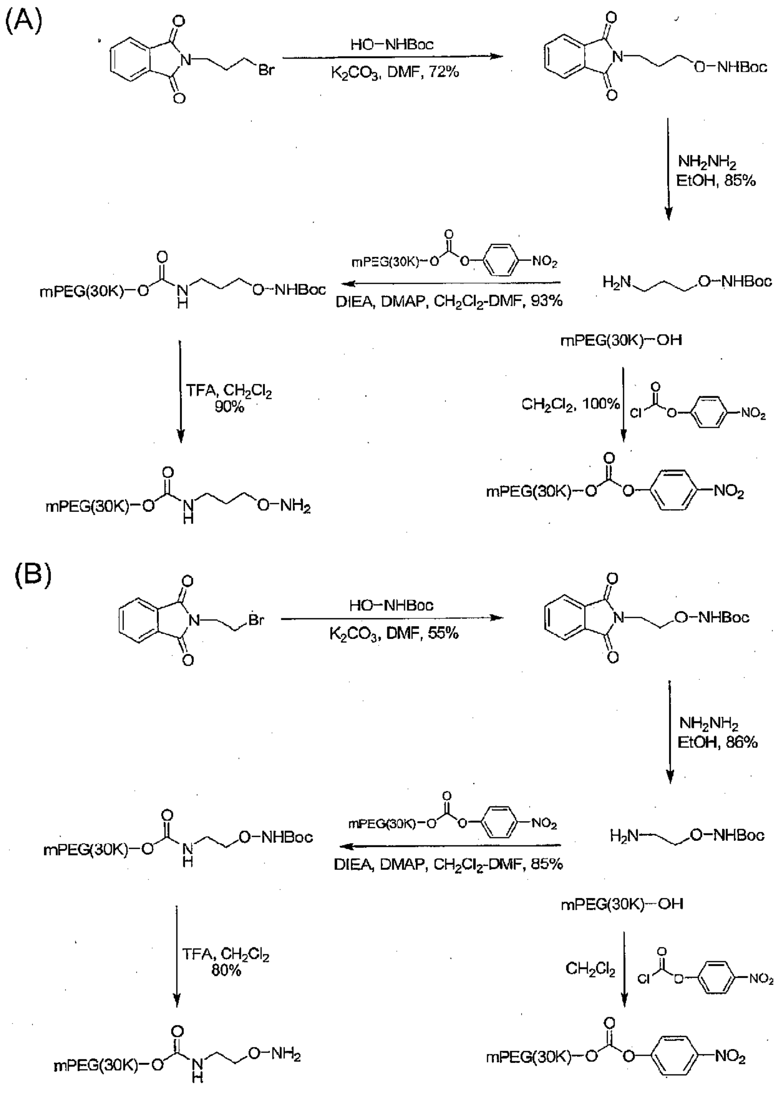

(A)
HO−NHBoc
K2CO3, DMF, 72%
O−NHBoc
NH2NH2
EtOH, 85%
mPEG(30K)−O
NO2
DIEA, DMAP, CH2Cl2-DMF, 93%
H2N
O−NHBoc
mPEG(30K)−O
N
H
O−NHBoc
mPEG(30K)−OH
TFA, CH2Cl2
90%
CH2Cl2, 100%
Cl
mPEG(30K)−O
O−NH2
(B)
Br
HO−NHBoc
K2CO3, DMF, 55%
O−NHBoc
NH2NH2
EtOH, 86%
DIEA, DMAP, CH2Cl2-DMF, 85%
mPEG(30K)−OH
TFA, CH2Cl2
80%
CH2Cl2
O−NH2

FIG. 49

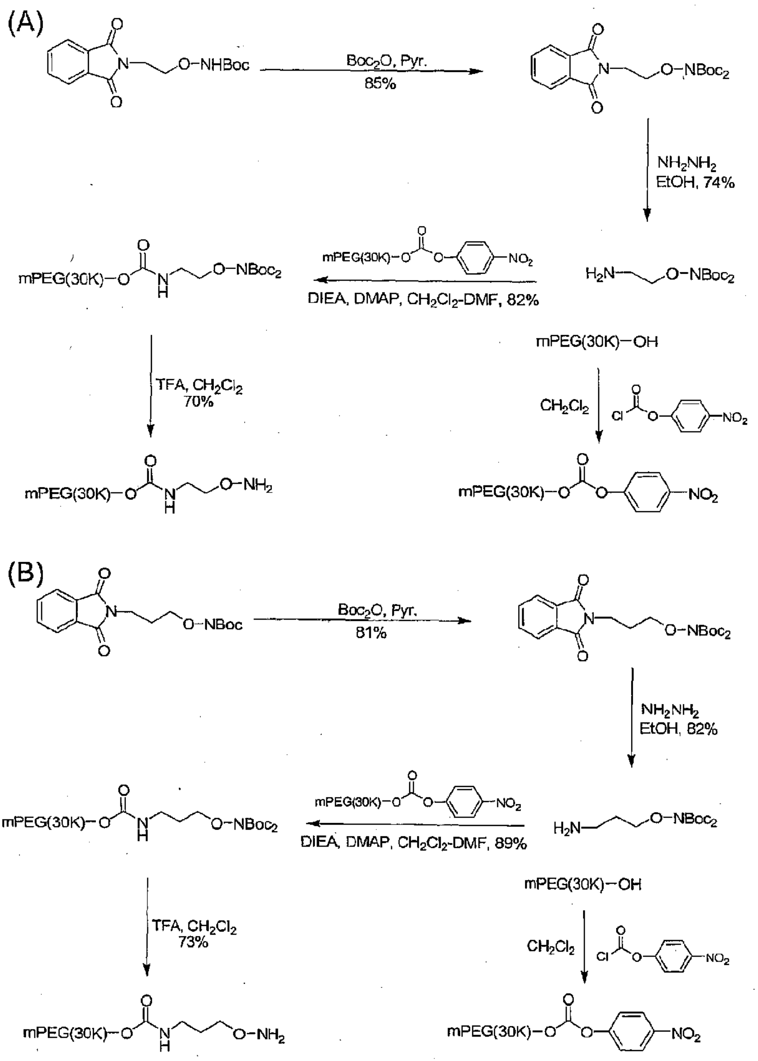

(A)
O-NHBoc
Boc₂O, Pyr.
85%
O-NBoc₂
NH₂NH₂
EtOH, 74%
mPEG(30K)-O
O-NBoc₂
mPEG(30K)-O
NO₂
DIEA, DMAP, CH₂Cl₂-DMF, 82%
H₂N
O-NBoc₂
mPEG(30K)-OH
TFA, CH₂Cl₂
70%
CH₂Cl₂
Cl
NO₂
mPEG(30K)-O
O-NH₂
mPEG(30K)-O
NO₂
(B)
O-NBoc
Boc₂O, Pyr.
81%
O-NBoc₂
NH₂NH₂
EtOH, 82%
mPEG(30K)-O
O-NBoc₂
mPEG(30K)-O
NO₂
DIEA, DMAP, CH₂Cl₂-DMF, 89%
H₂N
O-NBoc₂
mPEG(30K)-OH
TFA, CH₂Cl₂
73%
CH₂Cl₂
Cl
NO₂
mPEG(30K)-O
O-NH₂
mPEG(30K)-O
NO₂

FIG. 50

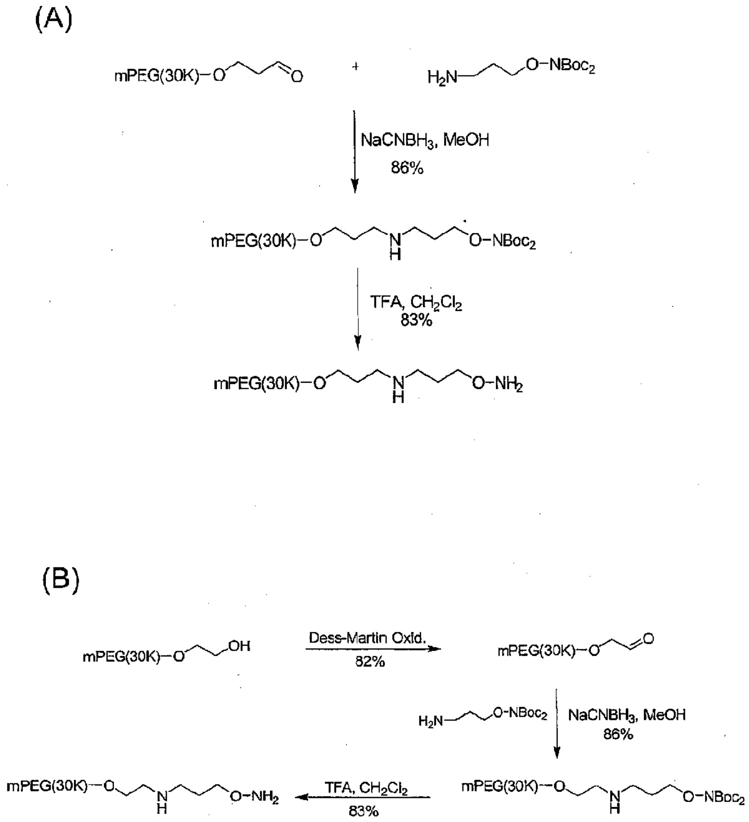

(A)
mPEG(30K)–O
O
+
H2N
O–NBoc2
NaCNBH3, MeOH
86%
mPEG(30K)–O
N
H
O–NBoc2
TFA, CH2Cl2
83%
mPEG(30K)–O
N
H
O–NH2
(B)
mPEG(30K)–O
OH
Dess-Martin Oxid.
82%
mPEG(30K)–O
O
H2N
O-NBoc2
NaCNBH3, MeOH
86%
mPEG(30K)–O
N
H
O–NH2
TFA, CH2Cl2
83%
mPEG(30K)–O
N
H
O–NBoc2

FIG. 51

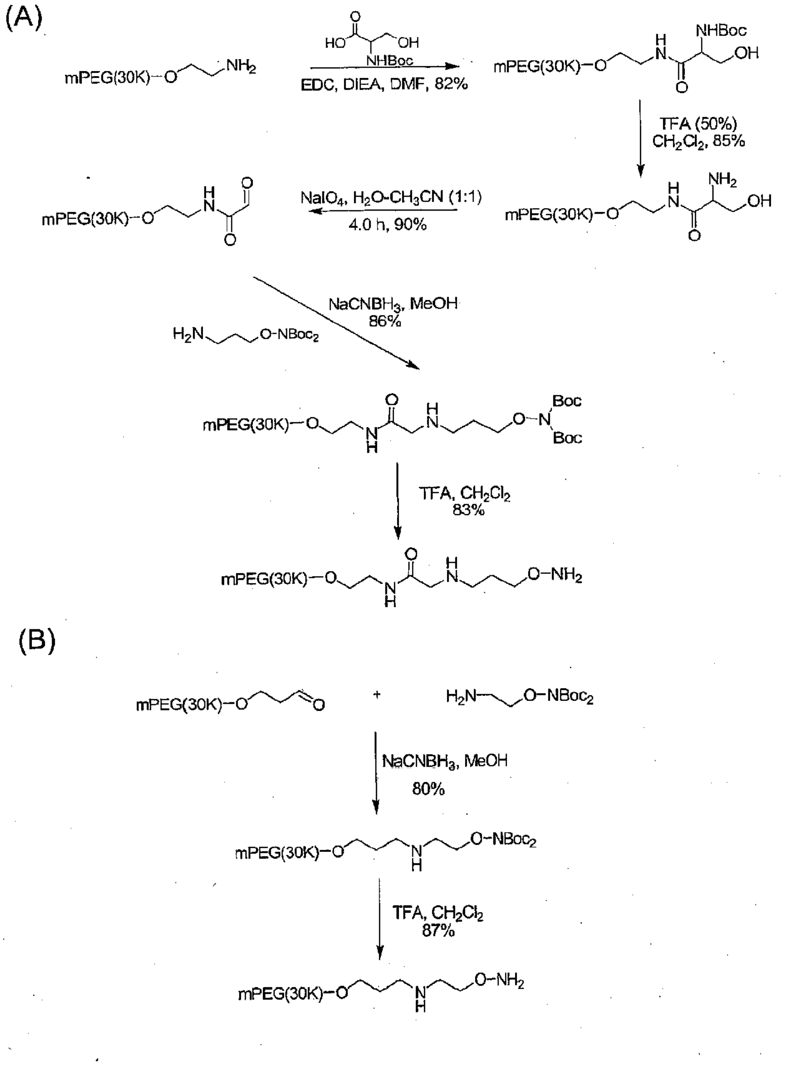

(A)
mPEG(30K)–O–NH2
HO–C(O)–CH(NHBoc)–CH2OH
EDC, DIEA, DMF, 82%
NHBoc
OH
TFA (50%)
CH2Cl2, 85%
NH2
NaIO4, H2O-CH3CN (1:1)
4.0 h, 90%
NaCNBH3, MeOH
86%
H2N–O-NBoc2
Boc
O–N
TFA, CH2Cl2
83%
O–NH2
(B)
mPEG(30K)–O
+
H2N–O–NBoc2
NaCNBH3, MeOH
80%
O–NBoc2
TFA, CH2Cl2
87%
O–NH2
mPEG(30K)–O
H
N
O

FIG. 52

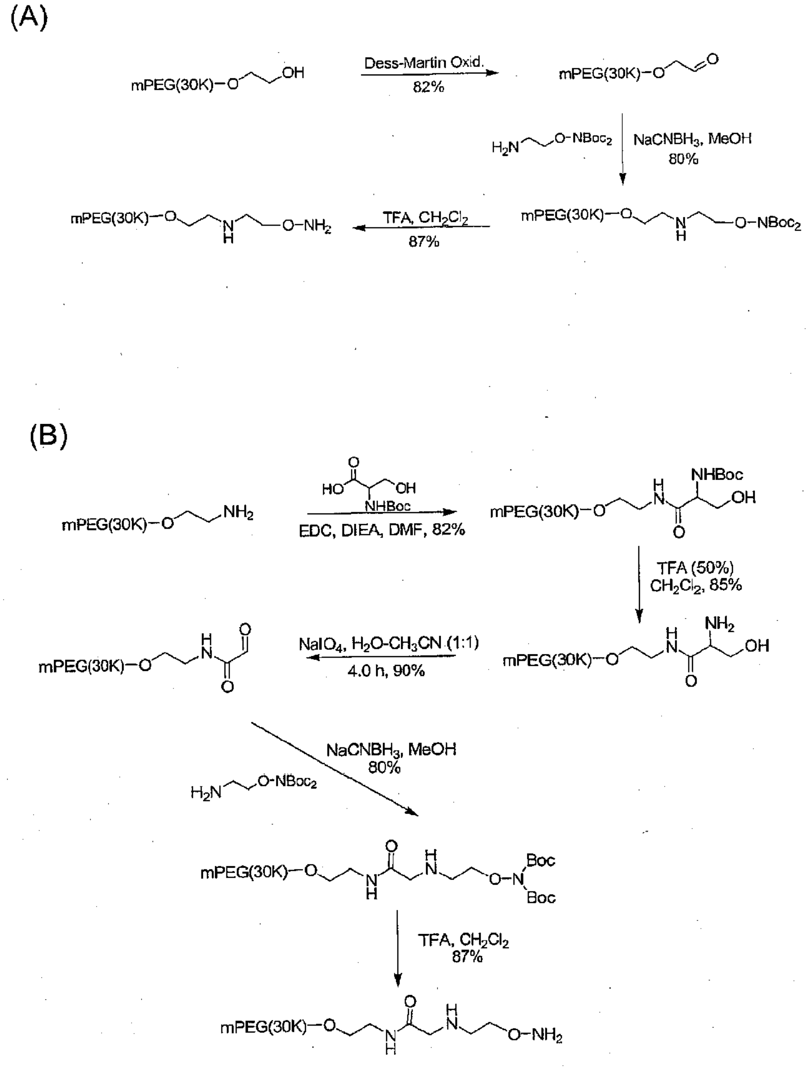

(A)
mPEG(30K)–O–OH
Dess-Martin Oxid.
82%
mPEG(30K)–O–=O
H2N–O-NBoc2
NaCNBH3, MeOH
80%
mPEG(30K)–O–N(H)–O–NBoc2
TFA, CH2Cl2
87%
mPEG(30K)–O–N(H)–O–NH2
(B)
mPEG(30K)–O–NH2
HO
O
OH
NHBoc
EDC, DIEA, DMF, 82%
mPEG(30K)–O
H
N
NHBoc
OH
O
TFA (50%)
CH2Cl2, 85%
NH2
NaIO4, H2O-CH3CN (1:1)
4.0 h, 90%
NaCNBH3, MeOH
80%
H2N–O-NBoc2
Boc
O–N
Boc
TFA, CH2Cl2
87%
O–NH2

FIG. 53

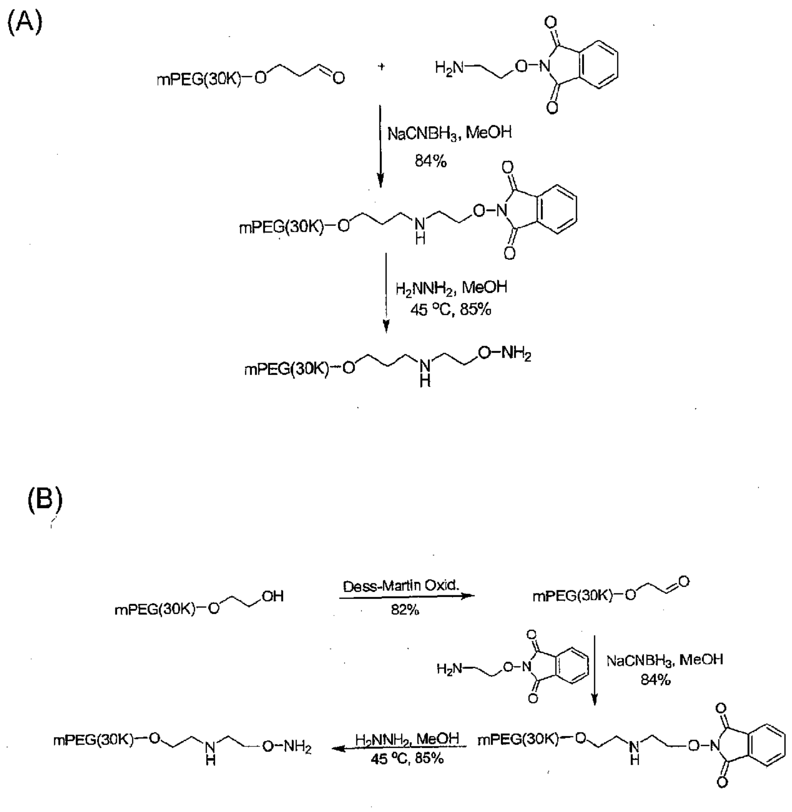

(A)
mPEG(30K)–O
+
H2N–O–N
NaCNBH3, MeOH
84%
H2NNH2, MeOH
45 °C, 85%
O–NH2
(B)
mPEG(30K)–O
OH
Dess-Martin Oxid.
82%
NaCNBH3, MeOH
84%
H2NNH2, MeOH
45 °C, 85%
O–NH2

FIG. 54

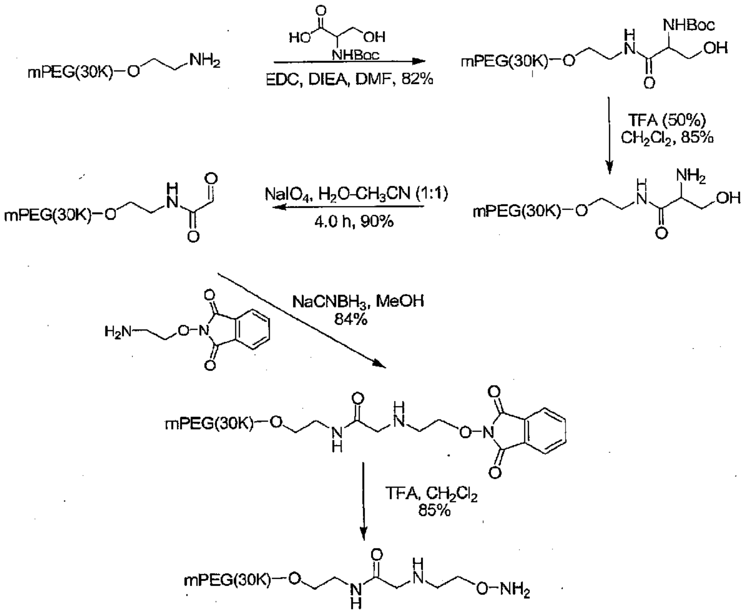

mPEG(30K)–O–NH2
EDC, DIEA, DMF, 82%
NHBoc
TFA (50%)
CH2Cl2, 85%
NaIO4, H2O-CH3CN (1:1)
4.0 h, 90%
NaCNBH3, MeOH
84%
TFA, CH2Cl2
85%
mPEG(30K)

FIG. 55

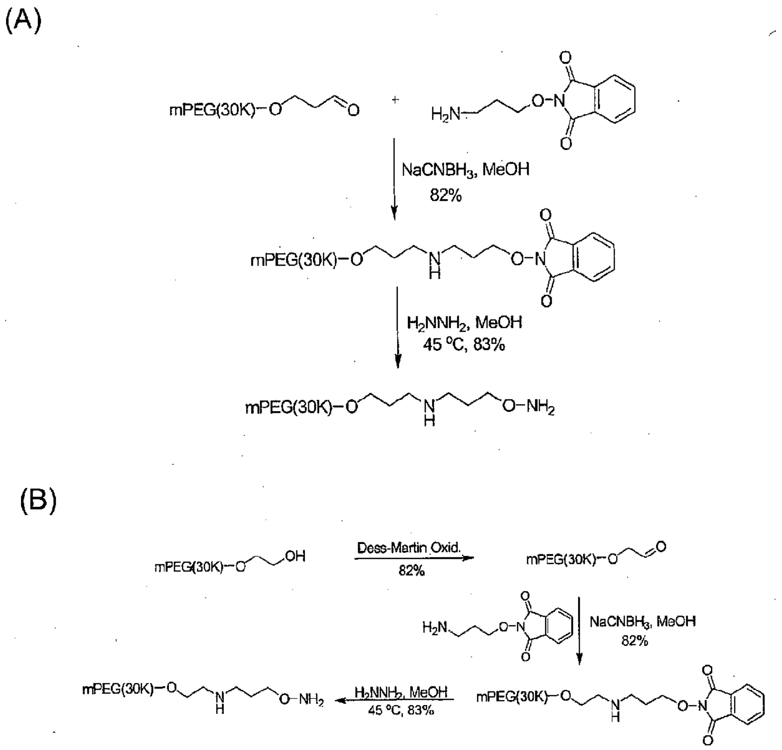

(A)
mPEG(30K)–O
O
+
H$_2$N
O–N
O
O
NaCNBH$_3$, MeOH
82%
mPEG(30K)–O
N
H
O–N
H$_2$NNH$_2$, MeOH
45 °C, 83%
mPEG(30K)–O
O–NH$_2$
(B)
mPEG(30K)–O
OH
Dess-Martin Oxid.
82%
mPEG(30K)–O
O
H$_2$N
O–N
NaCNBH$_3$, MeOH
82%
mPEG(30K)–O
N
H
O–N
H$_2$NNH$_2$, MeOH
45 °C, 83%
mPEG(30K)–O
O–NH$_2$

FIG. 56

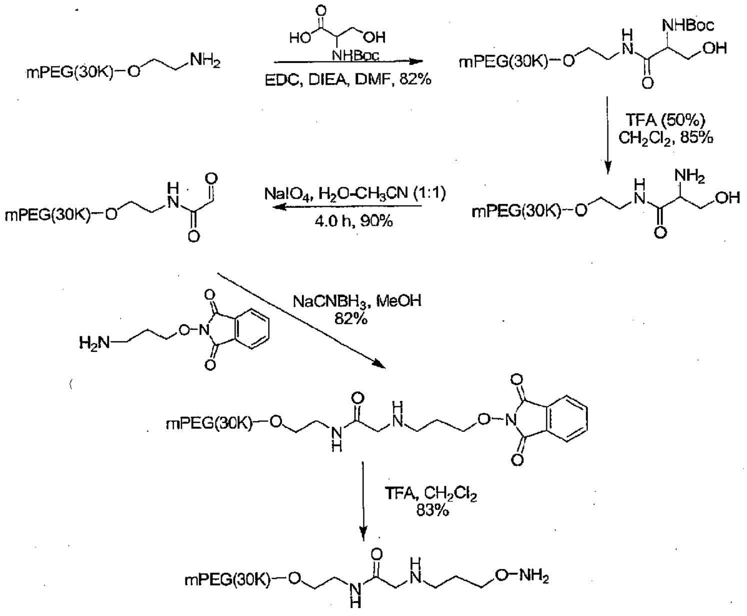

mPEG(30K)–O
NH$_2$
HO
OH
NHBoc
EDC, DIEA, DMF, 82%
TFA (50%)
CH$_2$Cl$_2$, 85%
NaIO$_4$, H$_2$O-CH$_3$CN (1:1)
4.0 h, 90%
NaCNBH$_3$, MeOH
82%
H$_2$N
TFA, CH$_2$Cl$_2$
83%
O–NH$_2$

FIG. 57

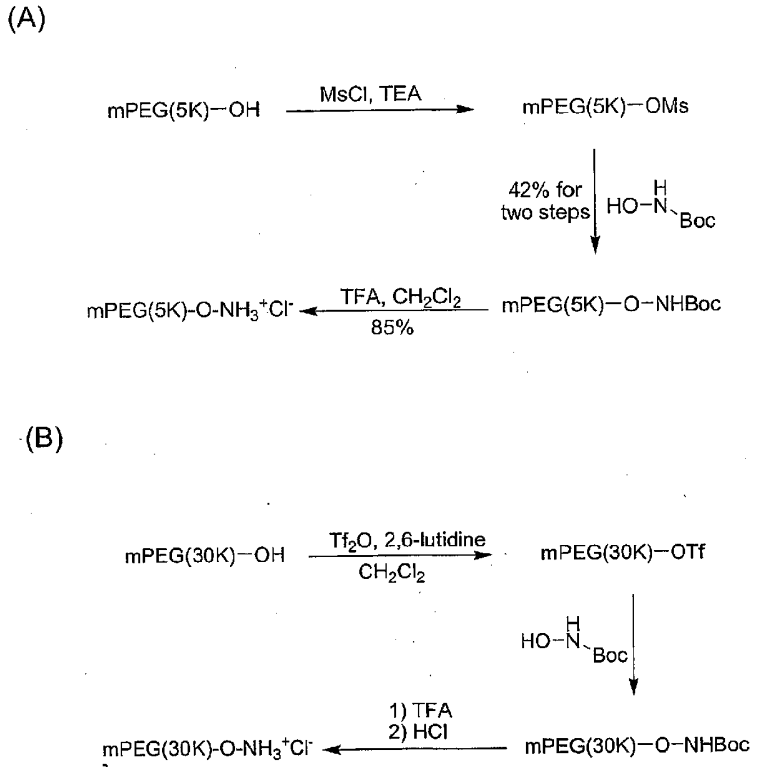

(A)
mPEG(5K)—OH
MsCl, TEA
mPEG(5K)—OMs
42% for two steps
HO–N(H)–Boc
mPEG(5K)—O–NHBoc
TFA, CH$_2$Cl$_2$
85%
mPEG(5K)-O-NH$_3^+$Cl$^-$
(B)
mPEG(30K)—OH
Tf$_2$O, 2,6-lutidine
CH$_2$Cl$_2$
mPEG(30K)—OTf
HO–N(H)–Boc
mPEG(30K)—O–NHBoc
1) TFA
2) HCl
mPEG(30K)-O-NH$_3^+$Cl$^-$

FIG. 58

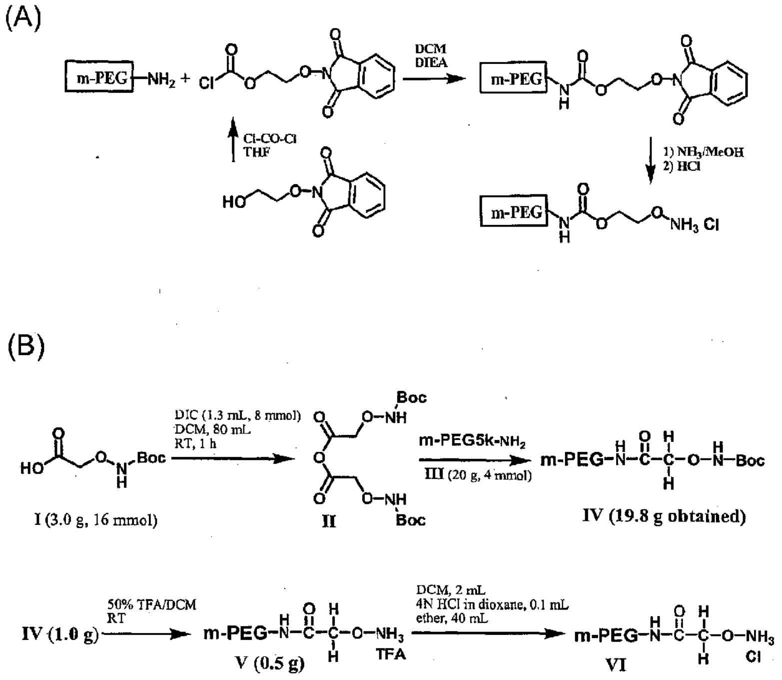

(A)
m-PEG
NH2 + Cl
DCM
DIEA
Cl-CO-Cl
THF
HO
1) NH3/MeOH
2) HCl
NH3 Cl
(B)
DIC (1.3 mL, 8 mmol)
DCM, 80 mL
RT, 1 h
Boc
I (3.0 g, 16 mmol)
II
m-PEG5k-NH2
III (20 g, 4 mmol)
m-PEG-N-C-C-O-N-Boc
IV (19.8 g obtained)
50% TFA/DCM
RT
IV (1.0 g)
m-PEG-N-C-C-O-NH3
TFA
V (0.5 g)
DCM, 2 mL
4N HCl in dioxane, 0.1 mL
ether, 40 mL
m-PEG-N-C-C-O-NH3
Cl
VI

FIG. 59

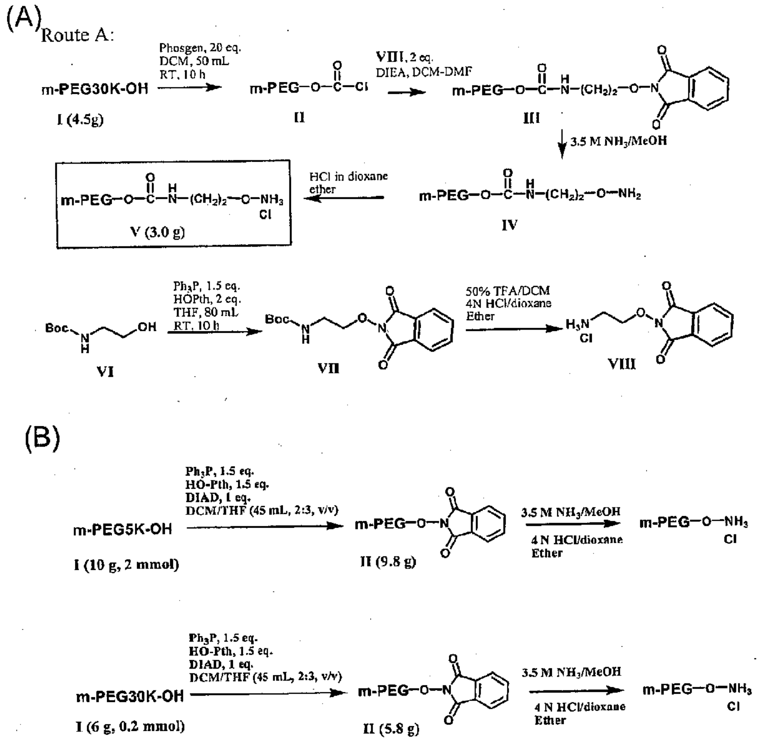

(A)
Route A:
Phosgen, 20 eq.
DCM, 50 mL
RT, 10 h
m-PEG30K-OH
I (4.5g)
m-PEG—O—C—Cl
II
VIII, 2 eq.
DIEA, DCM-DMF
m-PEG—O—C—N—(CH2)2—O—N
III
3.5 M NH3/MeOH
m-PEG—O—C—N—(CH2)2—O—NH2
IV
HCl in dioxane
ether
m-PEG—O—C—N—(CH2)2—O—NH3 Cl
V (3.0 g)
Ph3P, 1.5 eq.
HOPth, 2 eq.
THF, 80 mL
RT, 10 h
Boc
OH
VI
VII
50% TFA/DCM
4N HCl/dioxane
Ether
H3N Cl
VIII
(B)
Ph3P, 1.5 eq.
HO-Pth, 1.5 eq.
DIAD, 1 eq.
DCM/THF (45 mL, 2:3, v/v)
m-PEG5K-OH
I (10 g, 2 mmol)
m-PEG—O—N
II (9.8 g)
3.5 M NH3/MeOH
4 N HCl/dioxane
Ether
m-PEG—O—NH3 Cl
Ph3P, 1.5 eq.
HO-Pth, 1.5 eq.
DIAD, 1 eq.
DCM/THF (45 mL, 2:3, v/v)
m-PEG30K-OH
I (6 g, 0.2 mmol)
m-PEG—O—N
II (5.8 g)
3.5 M NH3/MeOH
4 N HCl/dioxane
Ether
m-PEG—O—NH3 Cl

FIG. 60

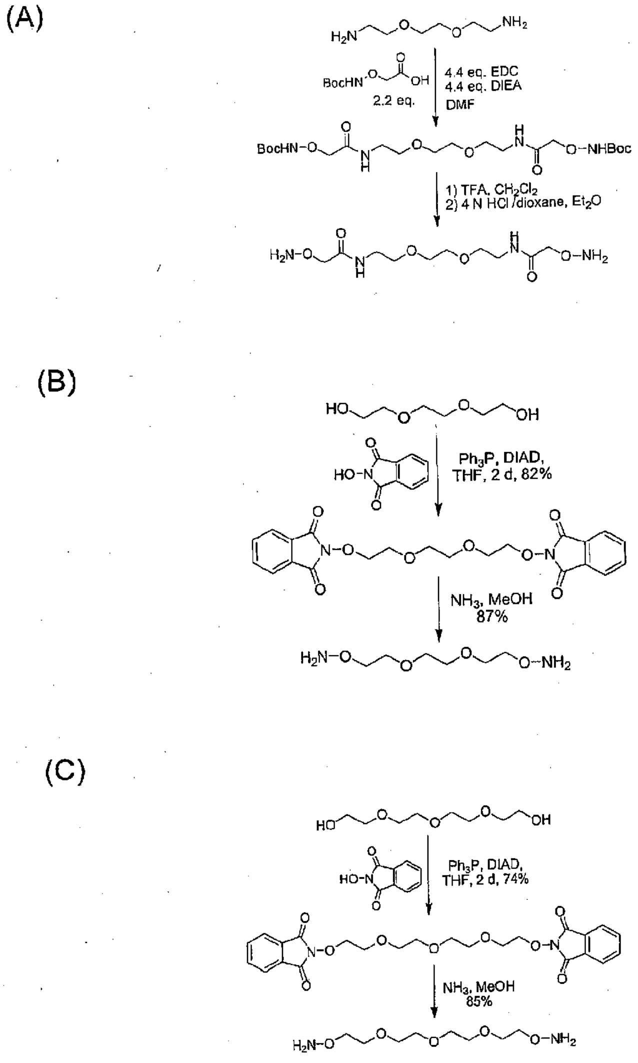

(A)
H2N
NH2
BocHN
OH
2.2 eq.
4.4 eq. EDC
4.4 eq. DIEA
DMF
BocHN-O
O-NHBoc
1) TFA, CH2Cl2
2) 4 N HCl /dioxane, Et2O
H2N-O
O-NH2
(B)
HO
OH
HO-N
Ph3P, DIAD,
THF, 2 d, 82%
NH3, MeOH
87%
H2N-O
O-NH2
(C)
HO
OH
HO-N
Ph3P, DIAD,
THF, 2 d, 74%
NH3, MeOH
85%
H2N-O
O-NH2

FIG. 61

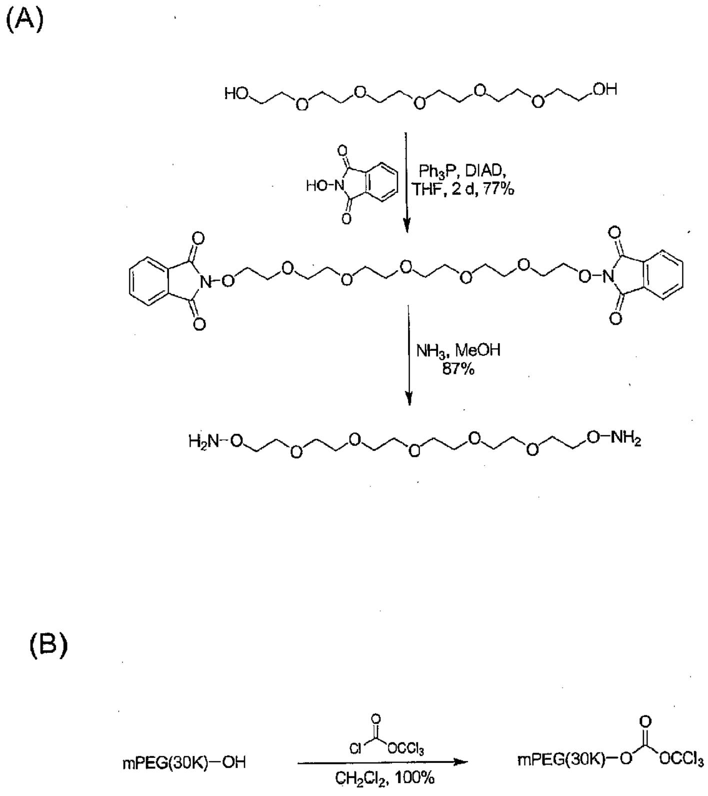

(A)
HO
OH
O
HO-N
Ph3P, DIAD,
THF, 2 d, 77%
N-O
O-N
NH3, MeOH
87%
H2N-O
O-NH2
(B)
mPEG(30K)—OH
Cl
OCCl3
CH2Cl2, 100%
mPEG(30K)—O
OCCl3

FIG. 62

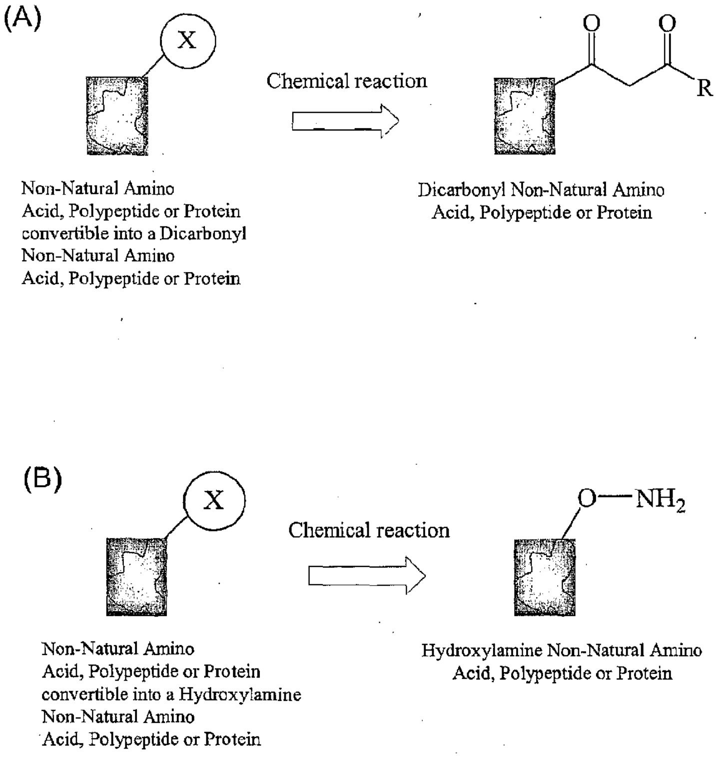

(A)
X
Chemical reaction
O
O
R
Non-Natural Amino Acid, Polypeptide or Protein convertible into a Dicarbonyl Non-Natural Amino Acid, Polypeptide or Protein
Dicarbonyl Non-Natural Amino Acid, Polypeptide or Protein
(B)
X
Chemical reaction
O—NH2
Non-Natural Amino Acid, Polypeptide or Protein convertible into a Hydroxylamine Non-Natural Amino Acid, Polypeptide or Protein
Hydroxylamine Non-Natural Amino Acid, Polypeptide or Protein

FIG. 63

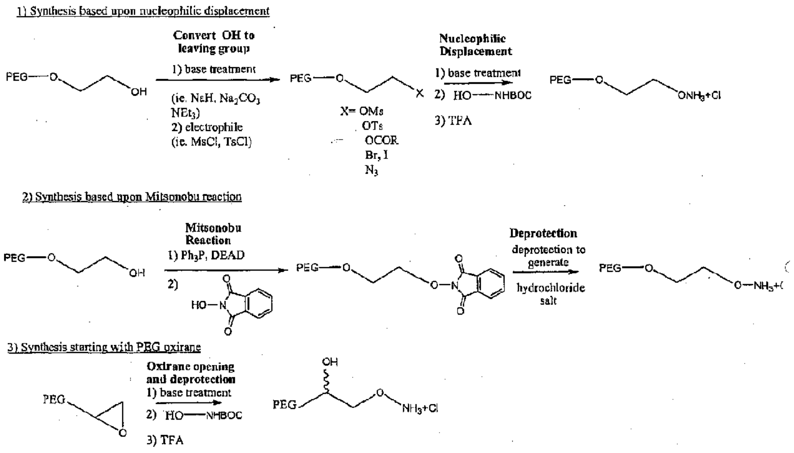
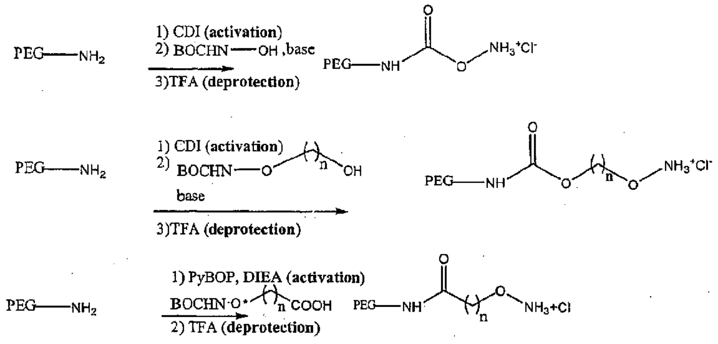

1) Synthesis based upon nucleophilic displacement
Convert OH to leaving group
1) base treatment
(ie. NaH, Na2CO3 NEt3)
2) electrophile
(ie. MsCl, TsCl)
PEG—O
OH
X
X= OMs
OTs
OCOR
Br, I
N3
Nucleophilic Displacement
1) base treatment
2) HO—NHBOC
3) TFA
ONH3+Cl
2) Synthesis based upon Mitsonobu reaction
Mitsonobu Reaction
1) Ph3P, DEAD
2)
HO—N
Deprotection
deprotection to generate hydrochloride salt
O—NH3+(
3) Synthesis starting with PEG oxirane
Oxirane opening and deprotection
1) base treatment
2) HO—NHBOC
3) TFA
PEG
OH
NH3+Cl

FIG. 64

4) Synthesis from PEG amine



5)Synthesis from PEG aldehyde

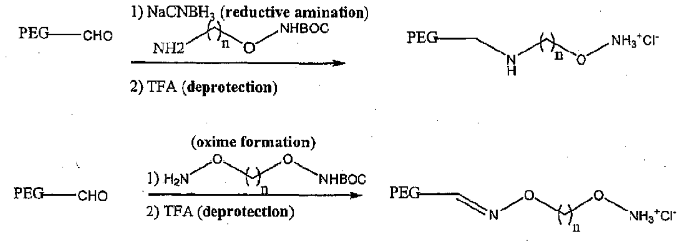


FIG. 65