

US 20140179560A1

(19) **United States**

(12) **Patent Application Publication**
Olson et al.

(10) **Pub. No.: US 2014/0179560 A1**

(43) **Pub. Date: Jun. 26, 2014**

(54) **DRUG DISCOVERY METHODS AND PLATFORMS**

Publication Classification

(71) Applicant: **Fred Hutchinson Cancer Research Center**, Seattle, WA (US)

(51) **Int. Cl.**
G01N 33/53 (2006.01)

(72) Inventors: **James Olson**, Seattle, WA (US); **Christopher Mehlin**, Seattle, WA (US); **Mark Stroud**, Seattle, WA (US); **Julian Simon**, Seattle, WA (US); **Colin Correnti**, Seattle, WA (US); **Patrick Paddison**, Seattle, WA (US); **Roland Strong**, Seattle, WA (US); **Damon May**, Seattle, WA (US)

(52) **U.S. Cl.**
CPC **G01N 33/5308** (2013.01)
USPC **506/12**

(73) Assignee: **Fred Hutchinson Cancer Research Center**, Seattle, WA (US)

(21) Appl. No.: **14/102,396**

(22) Filed: **Dec. 10, 2013**

Related U.S. Application Data

(60) Provisional application No. 61/735,516, filed on Dec. 10, 2012, provisional application No. 61/794,685, filed on Mar. 15, 2013.

(57) **ABSTRACT**

Methods and systems for discovering drug candidates are disclosed. Methods and systems can include generating libraries of potential drug candidates (e.g., libraries of peptides) that can be screened to identify sub-libraries of potential drug candidates (e.g., sub-libraries of peptides) having selected pharmacological properties. Methods of making and using peptide libraries are also provided. D-amino acid chlorotoxins and D-amino acid chlorotoxin variants are also provided.

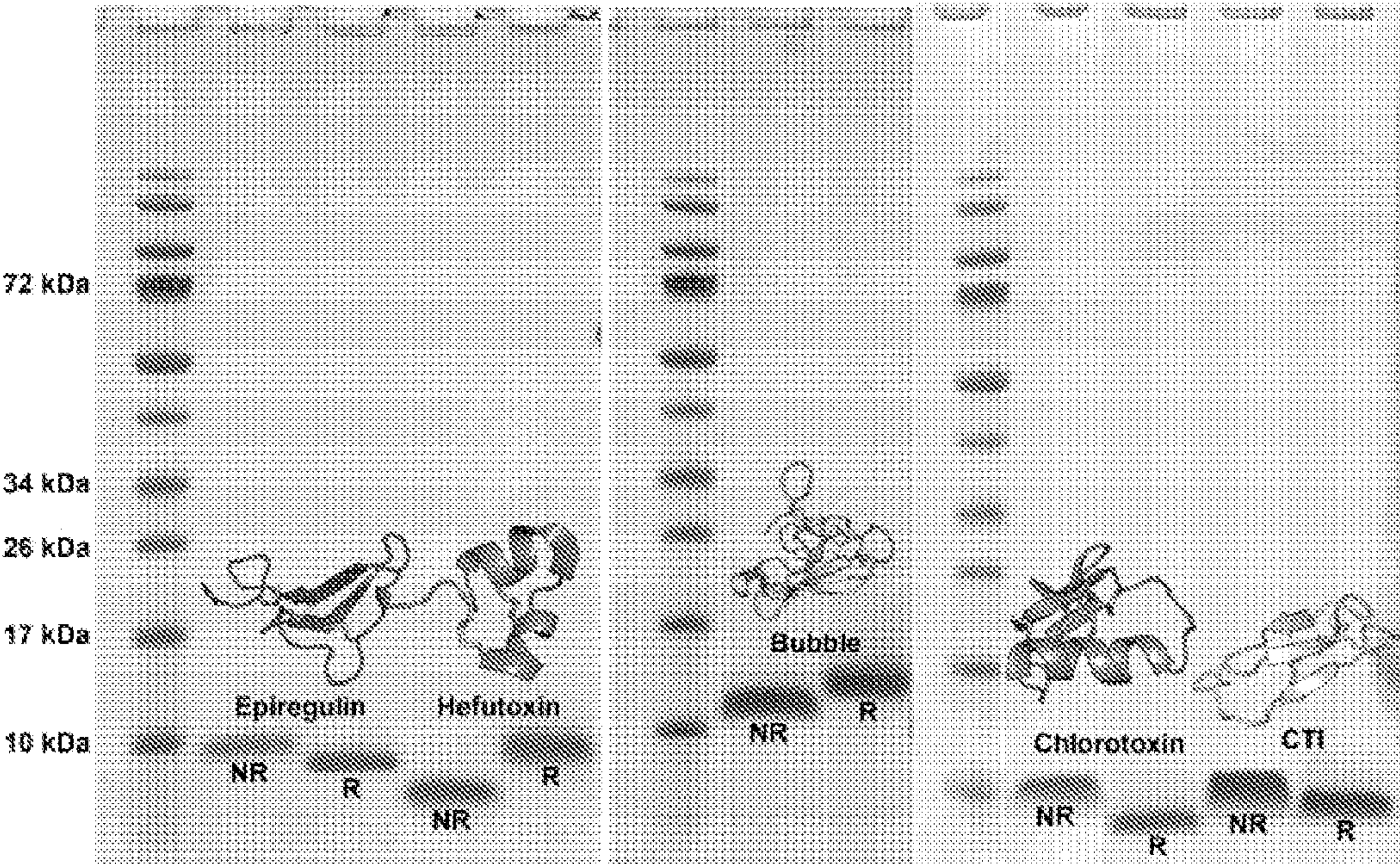


FIG. 1

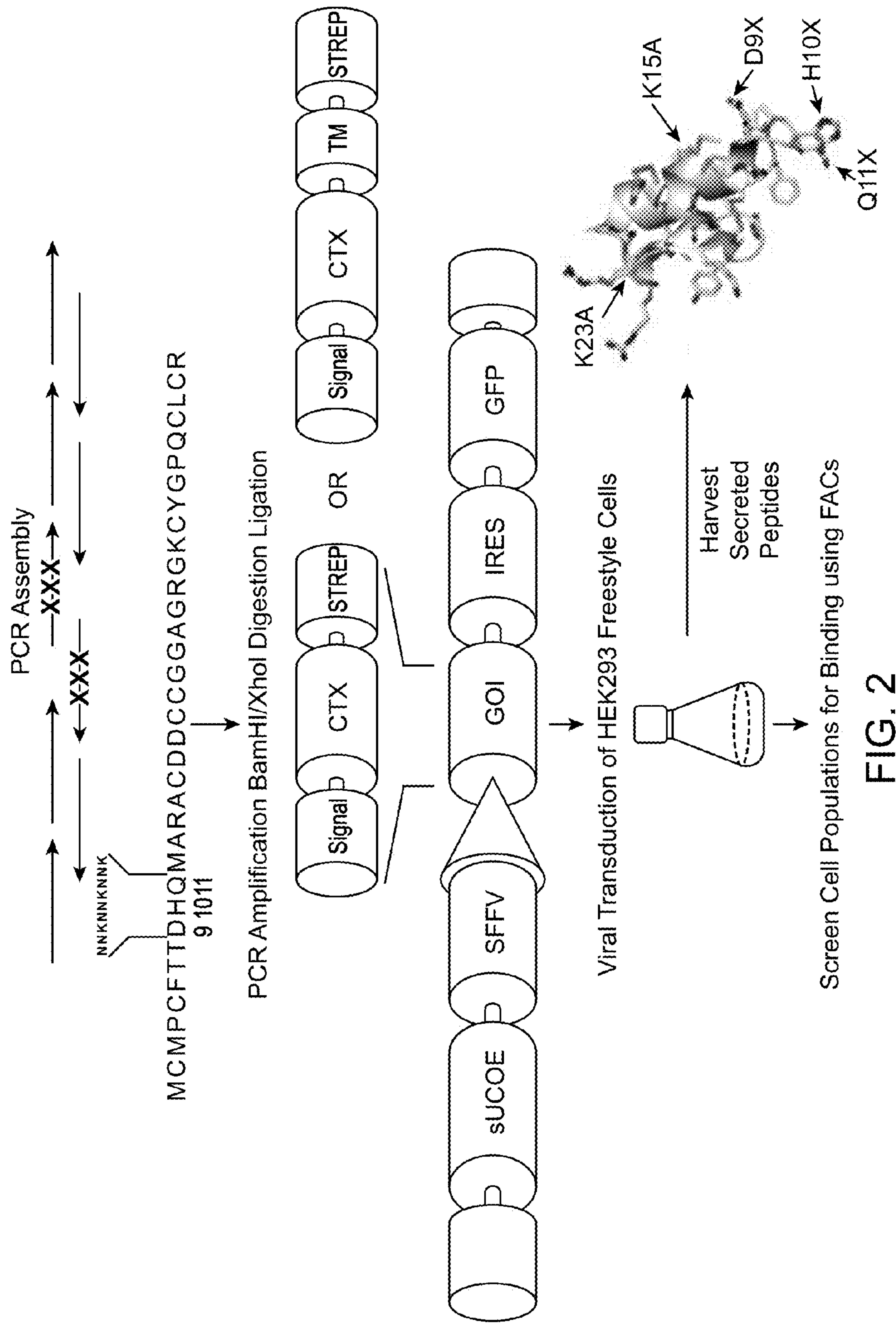


FIG. 2

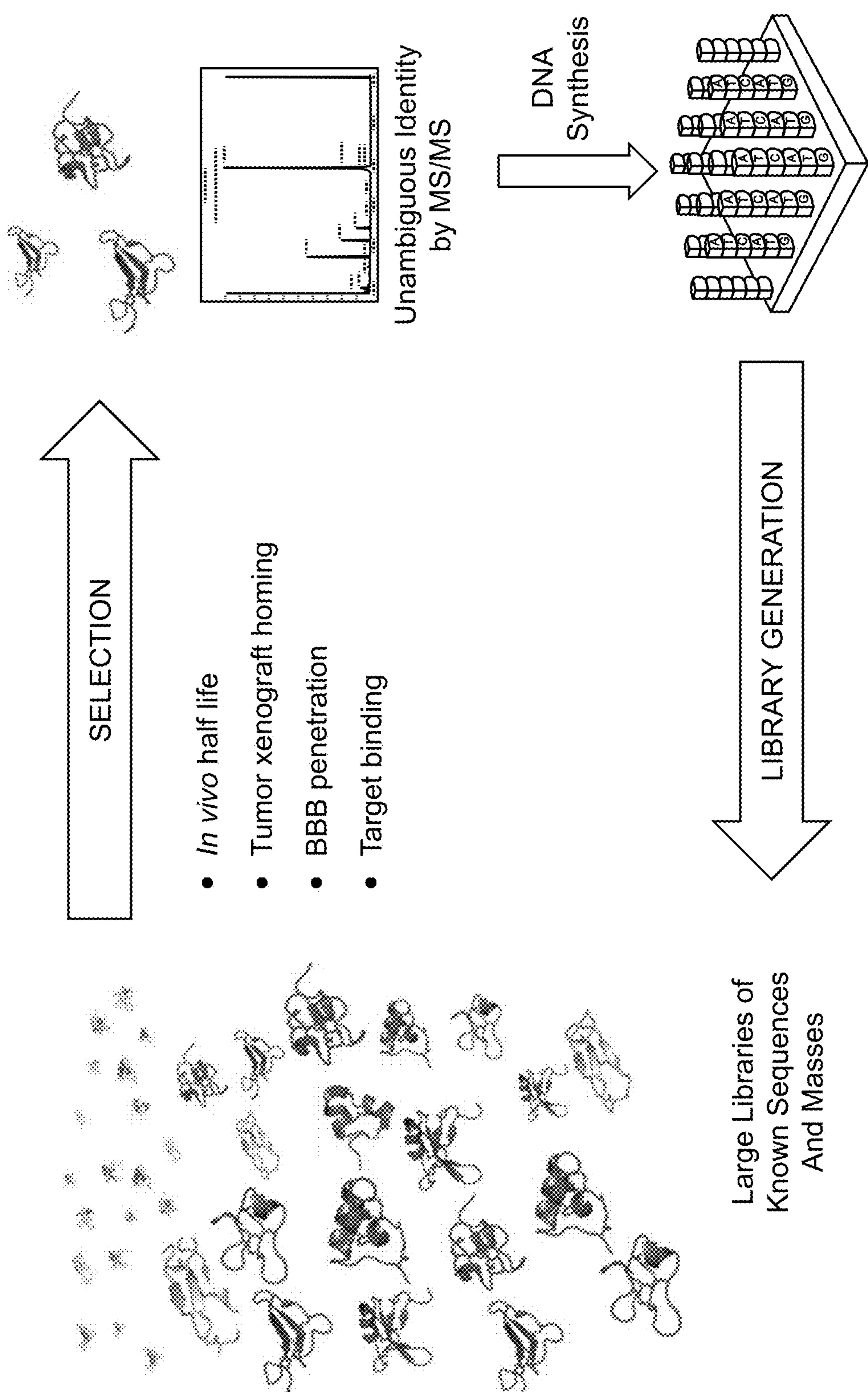


FIG. 3

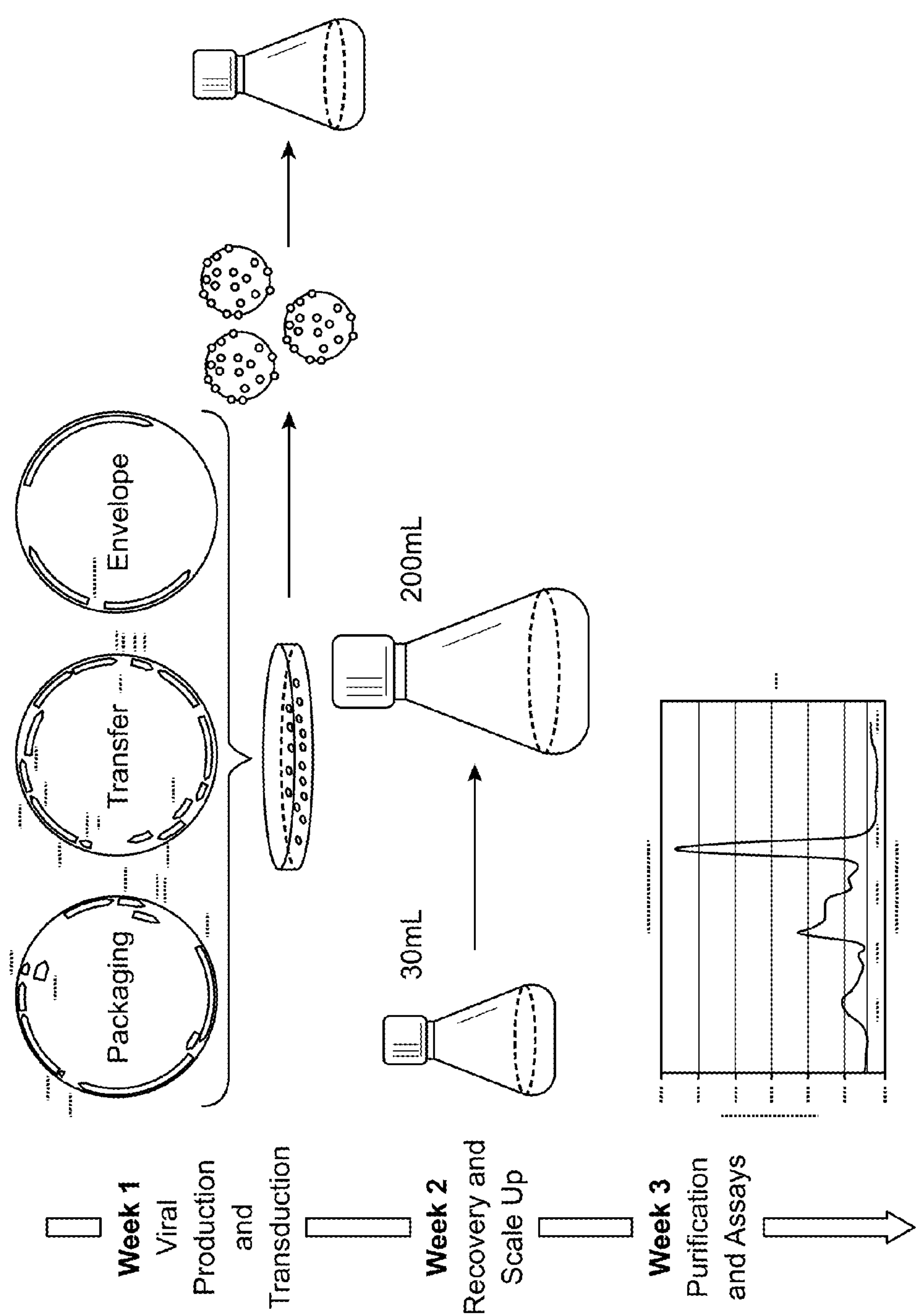


FIG. 4

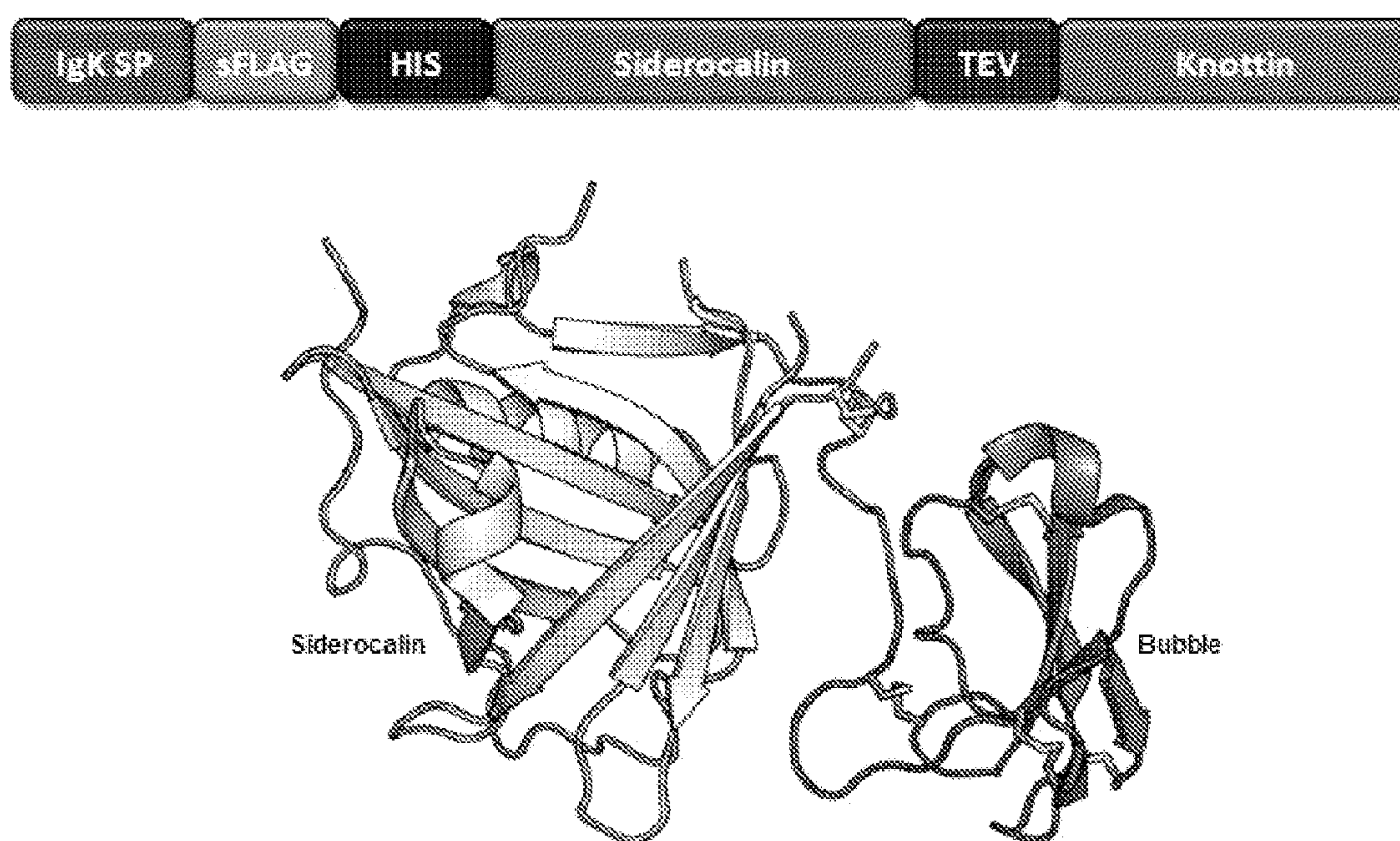


FIG. 5

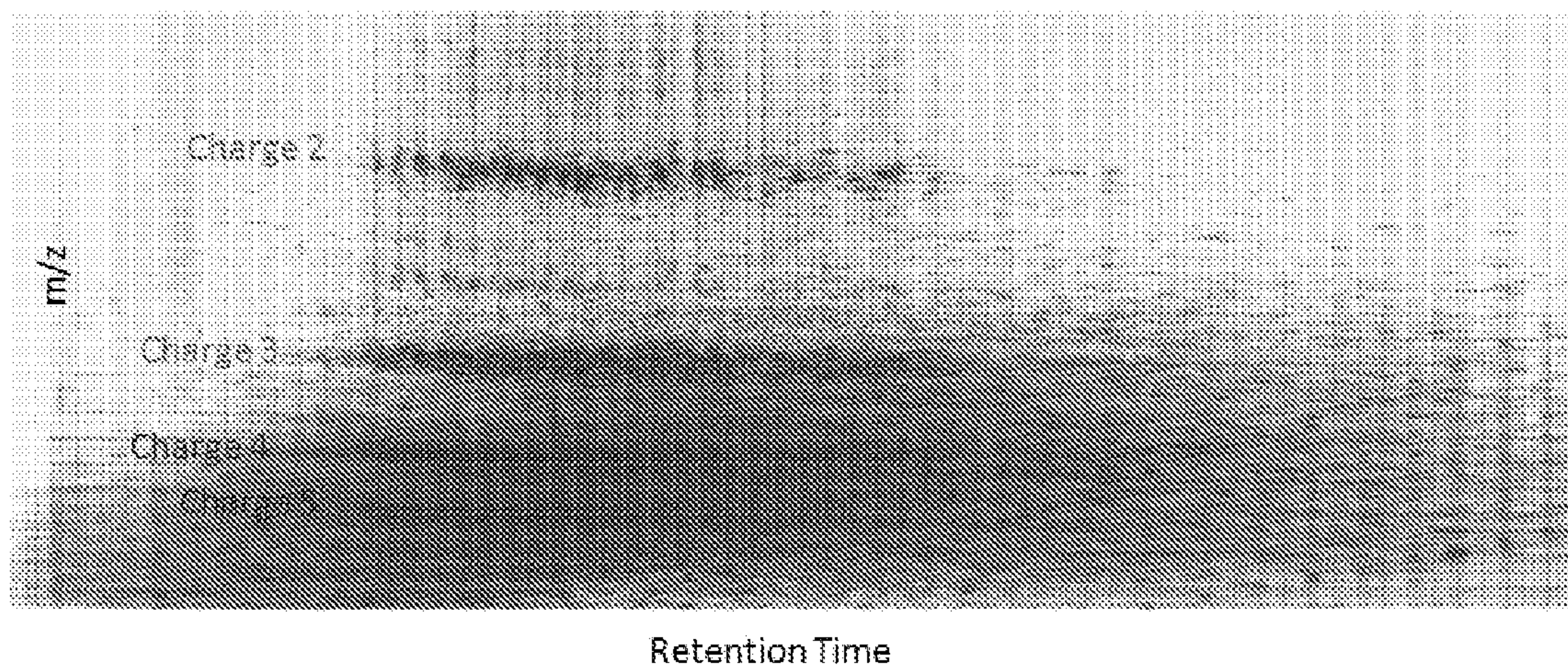


FIG. 6

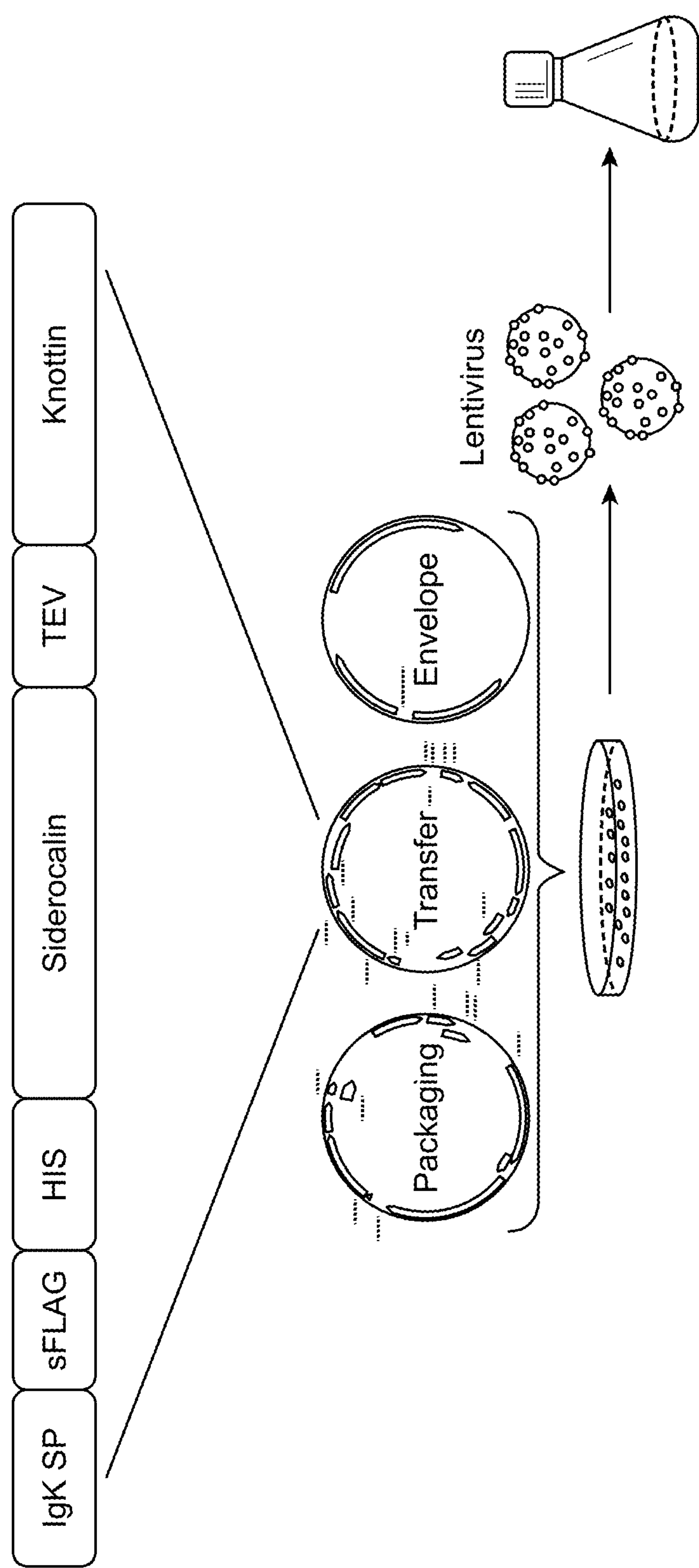


FIG. 7

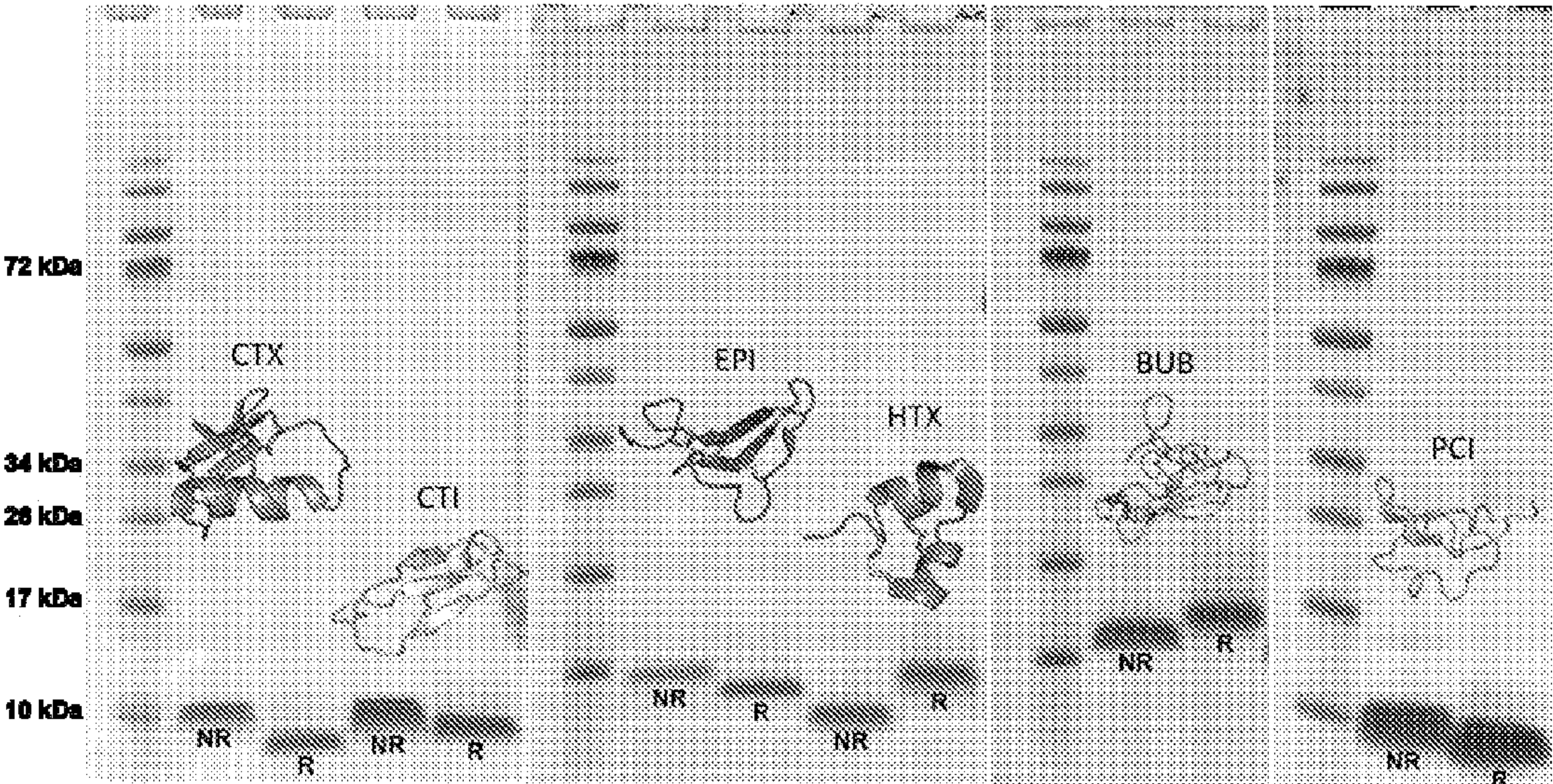
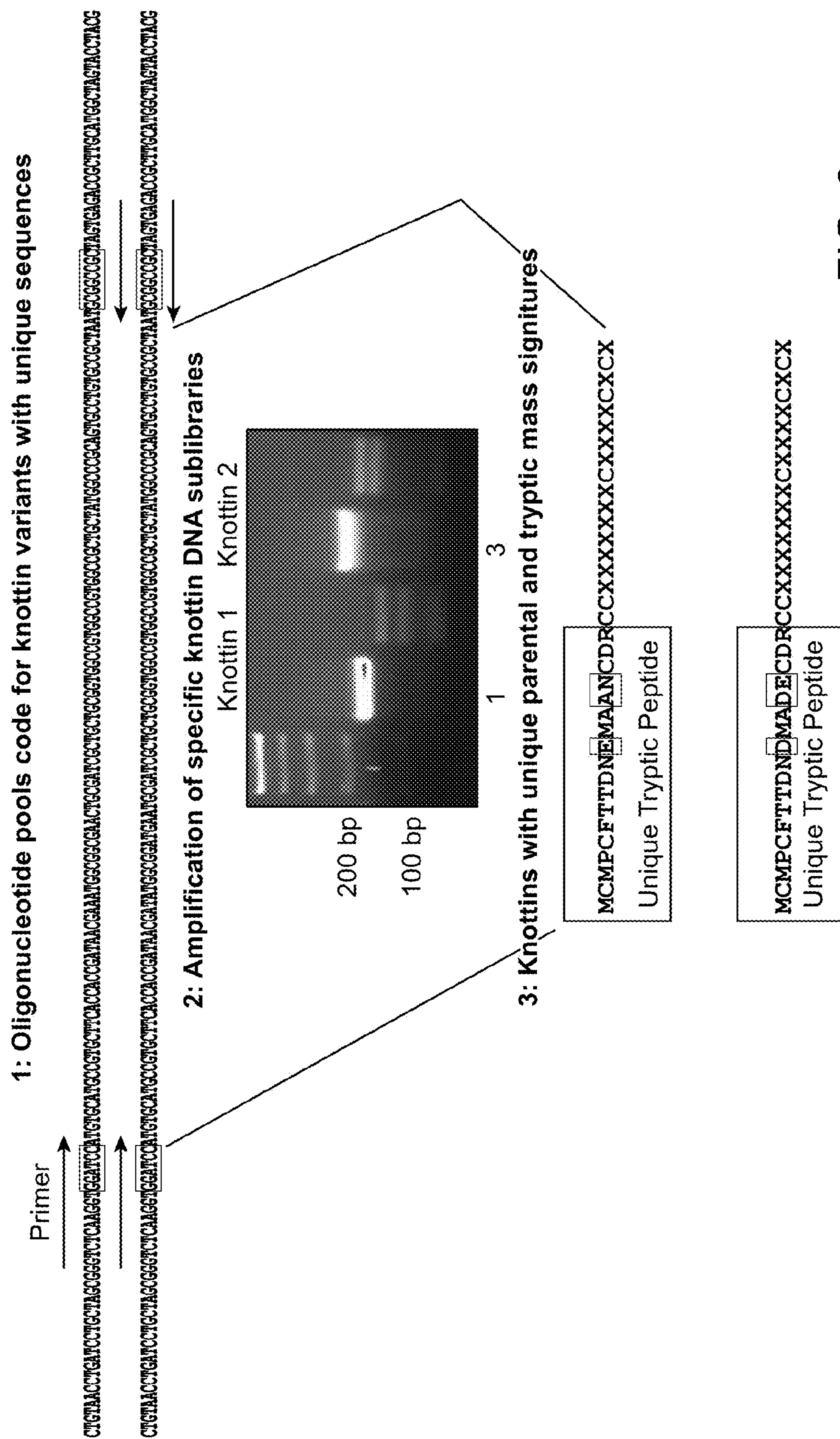


FIG. 8



9
G^x
L

B02.seq GSXCMPCFITDQXMANICDRCCGGGRCYGPQCLCR
C02.seq GSXCMPCFITFXXMARRCDDCCGGGAAAMARSACAANAAHHH
D02.seq GSXCMPCFITXXXMEYRCDDCCGGGRCYGPQCLCR
A03.seq GSECMPCFITDXXMARRCDECCGGGRCYGPQCLCR
B03.seq GSMCMPCFITXFXMAQRCDCCGGGRCYGPQCLCR
A04.seq GSMCMPCFITXXXMINRCDDCCGGGLICYGPQCLCR
B04.seq GSXCMPCFITAXXMAAICDDCCGGGRCYGPQCLCR
D04.seq GSXCMPCFITLXXMAARCDCCGGGRCYGPQCLCR
B01.seq GSRCMPCFITITXMARRCDDCCGGGAAAMARSACAANAAHHH
A05.seq GSXCMPCFITVXXMANRCDDCCGGGRCYGPQCLCR
B05.seq GXXSMPCFITXXXMAYFCDDCCGGGSXGICVGPQCLCR
C05.seq GSXCMPCFITDXXMARRCDDCCGGGDBCGPQCLCR
D05.seq GSCACRASPPIIKWRRAAMIAAAAAAMVSAMARIAACAA
C01.seq GSXCMPCFITXXXMARRCDDCCGGGLHGECLGPQCLCR
A06.seq GSXCMPCFITXXXMQQSCDRCCGGGRAVAAMARSVCAANAAHHE
B06.seq GSXCMPCFITXXXMARRCDDCCGGGDGNCRCGPQCLCR
C06.seq GXXCMPCFITNXXIALACDRCCGGGRCYGPQCLCR
D01.seq GSXCMPCFITDXXMAIICDRCCGGGRCYGPQXXCR
B07.seq GSXCMPSSPPXXX
C07.seq GSXCMPCFITXXXMASHCDDCCGGGRLGGCWGRSACAANAAHHH
D07.seq GSXCMPCFITDQMAAECDRCCGGGRCYGPQCLCR
B08.seq GSXCMPCFITXXXMANRCDRCCGGGRCYGPQYLCR
D08.seq GSXXMPCFITDXXMEHYCDRCCGGGRCYGPQCLCR
A09.seq GSXCMPCFITXXXMARRSMIAAALAMAAAMARSACAVMRPLITINHHH
B09.seq GXXXMPCFITXXXKARACDFCSVAAKXVAVLNPAKPVPLMRPLITINHHH
D09.seq GSXCMPCFITDHIMTWCDECCGPRQWPLLNPAVPVPLMRPLITINHHH
A10.seq GSXCMPCFITXXVMIAICDICCGRGEGACYGPLCLAANAAHHH
B10.seq GSXCMPCFITDXGMAGSCDRCCGGIRAVAAAMARAACAANAAHHH
C10.seq GSXCMPCFITFXMYIRCDDCCGGGRCYGPQCLCR
D10.seq GSXCMPCFITRSYDGVLLRLVLRWPNRVLRPVVPVPLMRPLITINHHH
A11.seq GSXCMPCFITYXXMWIRCDDCCGGGRCYGPQCLCR
B11.seq GSXCMPCFITDXXMARRCDHCCGGGRCYGPQLPVPLMRPLITINHHH
C11.seq GSMCMSCFITXXXMARRCDDCCGGGIGFICQGPQCLCR
D11.seq GSXCMPCFITDXXMARRCDDCCGGDGYGICHGPQCLCR
B12.seq GSXCMPCFITXXXMAETCDDCCGGGRCYGPQCLCR
C12.seq GSXCMPCFITXXXMAIQCDRCCSGDGRGRCYGPQCLCR
A02.seq GSMCMPCFITXXXMARRCDDCCGGYGHGHCWGPQCLCR

FIG. 10

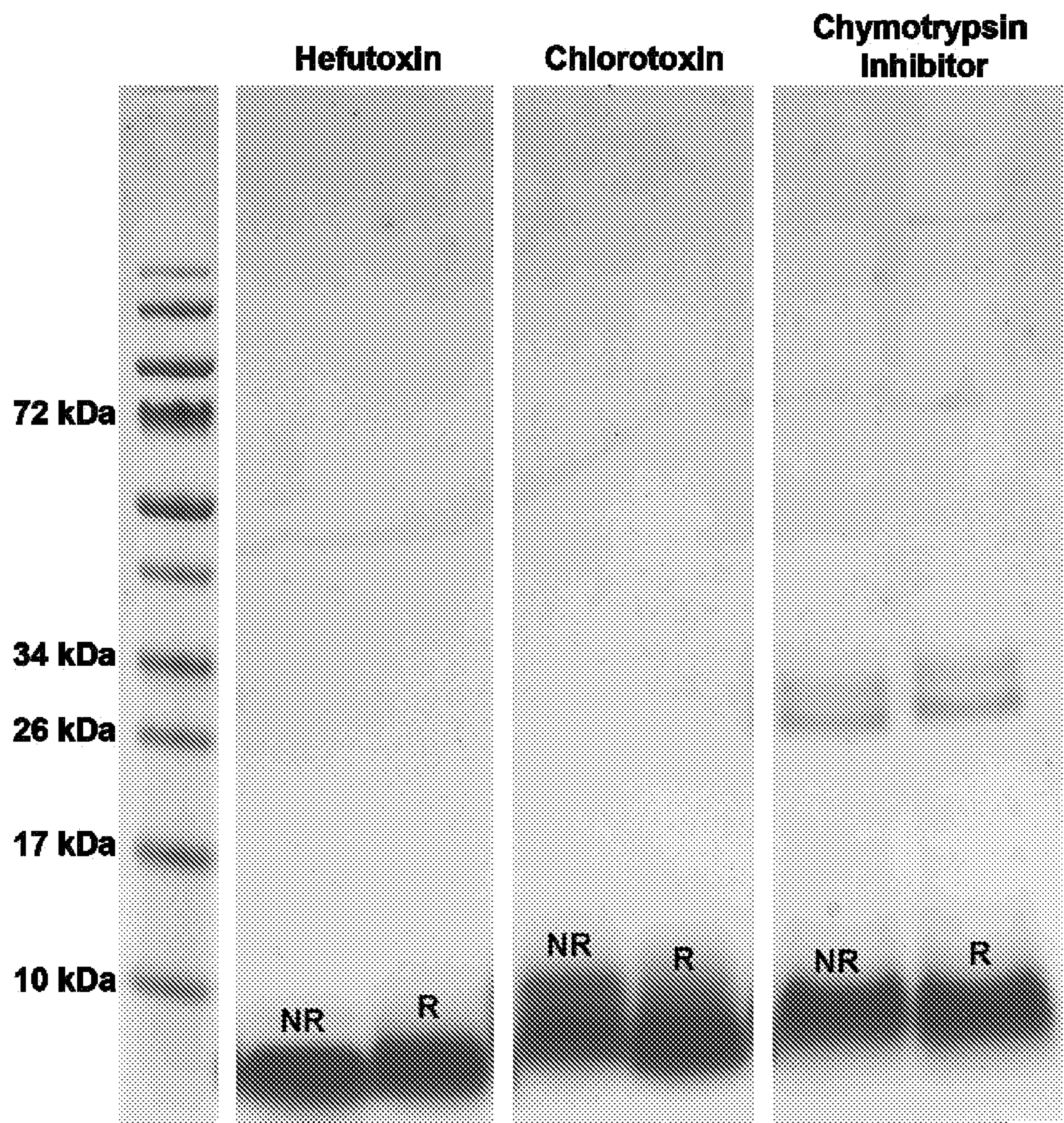


FIG. 11

DRUG DISCOVERY METHODS AND PLATFORMS

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 61/735,516, filed Dec. 10, 2012, and U.S. Provisional Application No. 61/794,685, filed Mar. 15, 2013, the contents of each of which is hereby incorporated by reference in their entirety.

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with the support of the United States government under Contract number R01CA135491 by the National Institutes of Health.

BACKGROUND OF THE INVENTION

[0003] Efforts towards drug discovery continue to use vast technical and financial resources to identify and develop new and useful drugs. Unfortunately, finding new drugs has continued to be difficult. For example, development of less damaging, more precisely targeted cancer therapies is essential in order to relieve patients' suffering and improve treatment success. But even after decades of research, scientists still struggle to identify therapeutic compounds with the right mix of medicinal and cancer-targeting properties.

[0004] A wide variety of types of compounds have been studied and pursued for different therapeutic purposes. For example, small chemical molecules and larger biologics (e.g., antibodies) have been used for a plethora of therapeutic applications with varied success. Some smaller peptides have also been shown to be useful as drugs, e.g., by virtue of their natural potency.

[0005] Despite the numerous examples of drugs that have been discovered, there is still a need for drug discovery platforms, further identification of useful drugs that have, e.g., improved pharmacological properties, and improved methods for making and analyzing potential drug candidates.

SUMMARY OF THE INVENTION

[0006] The present invention relates to drug discovery platforms and methods. In some aspects, the present invention relates to methods of generating libraries of potential drug candidates (e.g., libraries of peptides) that can be screened to identify sub-libraries of potential drug candidates (e.g., sub-libraries of peptides) having selected pharmacological properties. In certain aspects, the present invention relates to methods of identifying drug candidates (e.g., peptides) that can be lead compounds for further drug development by additional modification of the pharmacological properties of the identified candidates. The present invention also includes methods of making libraries of potential drug candidates (e.g., libraries of peptides).

INCORPORATION BY REFERENCE

[0007] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0009] FIG. 1 provides a gel image showing a variety of scaffolds, in accordance with an embodiment of the present invention.

[0010] FIG. 2 depicts a schematic of initial library creation from PCR assembly to protein harvesting. One potential patch of surface residues (DHQ loop) is shown to illustrate the random mutagenesis approach for the initial peptide libraries. Residues K15 and K23 have been mutated to alanine to facilitate chemical conjugations.

[0011] FIG. 3 illustrates an example method of generating peptide libraries, in accordance with an embodiment of the present invention.

[0012] FIG. 4 shows an example method of producing peptide libraries, in accordance with an embodiment of the present invention.

[0013] FIG. 5 depicts an example fusion system that can be used to make knottins (e.g., bubble protein), in accordance with an embodiment of the present invention.

[0014] FIG. 6 shows analysis of a large library using mass spectrometry, in accordance with an embodiment of the present invention.

[0015] FIG. 7 shows an example method of using siderocalin fusions to express knottin variants, in accordance with an embodiment of the present invention.

[0016] FIG. 8 depicts SDS-PAGE analysis of expressing knottin scaffolds, in accordance with embodiments of the present invention.

[0017] FIG. 9 provides a schematic of pooled library production, in accordance with an embodiment of the present invention.

[0018] FIG. 10 describes representative sequencing data from a cloned knottin library, in accordance with an embodiment of the present invention.

[0019] FIG. 11 shows SDS-PAGE analysis of 3000 member knottin libraries, in accordance with embodiments of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

Generating Selected Libraries

[0020] In one aspect, the present invention relates to methods and systems for generating libraries of potential drug candidates (e.g., libraries of peptides) that can be screened to identify sub-libraries of potential drug candidates (e.g., sub-libraries of peptides) having selected pharmacological properties. In contrast to conventional drug screening methods that screen one potential drug candidate at a time, the present invention first screens a library of potential drug candidates (e.g., greater than 100 candidates) to search for certain drug-like qualities for some or all of the candidates in the library. Screening of the library can include administering some or all of the potential drug candidates in the library to a subject and determining which of the drug candidates possess a pharmacological property of interest. For example, large libraries of potential drug candidates can be screened for certain pharma-

cological properties, such as, but not limited to, oral bioavailability, pharmacokinetics, distribution, targeting capability, serum half-life, tissue penetration, tumor penetration, and/or blood-brain barrier penetration. The initial screen of the library can be used to identify sub-libraries of potential drug candidates that share at least one pharmacological property, thereby generating panels of optimized drug candidates for subsequent target-based or phenotypic screens. In some embodiments, the screening of the libraries can be used to identify potential drug candidates that share more than one pharmacological property.

[0021] In some embodiments, the screening of the libraries can be used to identify candidates that share 2 pharmacological properties, 3 pharmacological properties, 4 pharmacological properties, 5 pharmacological properties, 6 pharmacological properties, 7 pharmacological properties, 8 pharmacological properties, 9 pharmacological properties or 10 pharmacological properties. In some embodiments, the screening of the libraries can be used to identify potential drug candidates sharing more than 10 pharmacological properties.

[0022] Large number of drug candidates can be administered to a subject and screened to, e.g., which drug candidates have pharmacological properties of interest. For example, the methods can include administering more than five drug candidates, more than ten drug candidates, more than 50 drug candidates, more than 100 drug candidates, more than 500 drug candidates, more than 1000 drug candidates, more than 1500 drug candidates, more than 2000 drug candidates, more than 3000 drug candidates, more than 5000 drug candidates, more than 8000 drug candidates, or more than 10000 drug candidates.

[0023] In some, embodiments the library screens can be used to identify candidates with an ability to cross the blood brain barrier. In certain embodiments, the screens can be used to identify candidates based on serum half-life, oral bioavailability, proteolytic degradation, toxicity, clearance, capability to penetrate cells or capability to enter sub cellular organelles. In some examples, the screening can be used to identify members of the library with more than one desired pharmacological properties. For example, the screening can be used to identify candidates that exhibit two or more of the properties, such as the ability to cross the blood brain barrier, exclusion by the blood-brain barrier, optimum serum half-life, oral bioavailability, reduced proteolytic degradation, capability to penetrate cells or a capability to enter sub cellular organelles, capability to target organs or tissue, and/or capability to target cancerous tissue. In some embodiments, screening of the peptides can be performed to identify which peptides exhibit an activity for inhibiting a protein:protein interaction, inhibiting antagonism of a receptor, inhibiting binding of an agonist to a receptor, modulating an ion channel, inhibiting a signaling pathway, activating a signaling pathway, and/or a inhibiting a protein:small molecule interaction. The activity of the peptides (e.g., the protein:protein interaction) can be associated with a disease or disorder, such as, e.g., a cancer, an infectious disease, an inflammatory disease, an immune disease, a metabolic disease, a cardiac disease, an aging-related disease, and a neurologic disease.

[0024] In some aspects, the methods of generating libraries of the present invention can include iterative processes that progressively enrich for sub-libraries of drug candidates (e.g., peptides) having one or more drug-like properties, such as those described above. For example, in a first screen, thou-

sands of candidates (e.g., peptides) can be tested for one or more of the pharmacological properties, such as cell penetration, oral bioavailability, serum half-life, blood-brain barrier penetration, exclusion by the blood-brain barrier, capability to target organs or tissue, and/or capability to target cancerous tissue. The identified drug candidates that possess certain properties can then be screened again to generate a subsequent library of candidates possessing other pharmacological properties. The subsequent screen can include any desired number of the identified drug candidates from the first screen. Another sublibrary of candidates can be generated after the iteration from the first screen and the second screen. Any number of iterations can be used. In some embodiments, the screening methods of this invention include 1 iteration, 2 iterations, 3 iterations, 4, 5 iterations, 6 iterations, 7 iterations, 8 iterations, 9 iterations, 10 iterations, 11 iterations, 12 iterations, 13 iterations, 14 iterations, 15 iterations, 16 iterations, 17 iterations, 18 iterations, 19 iterations, or 20 iterations. The drug-like properties tested in the various iterations can be the same or different for every iteration. For example, in one round the library of the candidates (e.g., peptides) could be tested for capability to cross the brain blood barrier. In the second iteration, the library could be tested for one or more pharmacological properties including serum half-life.

[0025] Furthermore, through, e.g., iterative modeling and structural analysis, rules can be identified to allow for better selection of drugs for lead compound identification. With peptide libraries, e.g., the peptides can have a variety of common characteristics that cause the peptides to have certain drug-like properties. Rules can be identified to help define which of the peptides may be more likely to have the drug-like properties. For example, rules can include acidic/basic amino acid ratios, hydrophobic/hydrophilic ratios or absolute percentage, neighboring amino acid substitutions, and/or lengths of structural loops in the peptides.

[0026] In some embodiments, a rule could include having the first residue of the members of the library to be glycine in order to facilitate cleavage of fusion proteins. The rules could also include that no proline and cysteine residues be mutated or that glycine mutation in alpha helices be avoided. The rule could include avoiding mutation schemes that change all the amino acid residues to highly hydrophobic amino acids. Examples of hydrophobic acids include phenylalanine, tryptophan, isoleucine, leucine, valine, and methionine. In other cases, rules could indicate bias towards mutation of common protein-protein hotspot residues. Such residues can include tyrosine, tryptophan, phenylalanine, isoleucine, histidine, aspartic acid and arginine moieties. In some embodiments, rules could also favor mutation of structurally adjacent residues. A rule, e.g., could require that mutations in the alpha helices be biased towards the amino acids that have high alpha helix-forming propensities to avoid mutations. Examples for such amino acids include methionine, alanine, leucine, glutamic acid, lysine, alanine, and isoleucine). Rules could also avoid mutations to glycine and proline in alpha helices. The rules could include mutations at structurally adjacent residues. The rules can also include retaining the cysteine residues are not mutated and/or moved in order to preserve any cysteine knots that might be present in the peptides.

[0027] In some embodiments, the rules could also include identifying the number of mutations per peptide construct. In some cases, a rule might include a number of mutations in the range of 4-8 amino acid mutations per peptide construct. In some cases, a rule could include about 4 amino acid muta-

tions, about 5 amino acid mutations, about 6 amino acid mutations, about 7 amino acid mutations or about 8 amino acid mutations per peptide construct. In some cases, the rules could include that no lysine be present in the peptide sequence in order to facilitate conjugation of the peptide libraries at the N-terminus. The rules could also be designed to facilitate easier identification and analysis of the peptide libraries. For example a rule could include having every peptide be identifiable by a unique mass signature or unique fragment mass signature in order to aid identification by MS.

[0028] In addition to generating selected libraries, the present invention provides further advancements to discovering drugs based on knottins (or knotted peptides). For example, the present invention can provide knotted peptides having longer serum half life (e.g., by conjugation), knotted peptides having better solubility, time and cost efficient ways to produce libraries diversities of variants ranging from the tens to hundreds to thousands and more peptides, methods for generating properly folded knotted peptides, and methods for producing high yields of knotted peptides (e.g., mg quantities)

[0029] In some aspects, the methods of generating libraries provide for more efficient and focused selection of potential drug candidates that may have a higher probability of proving to be a lead compound for further drug development. Conventional drug discovery, for example, typically focuses initially on a target and develops drug leads with potency in vitro. Unfortunately, the drug leads identified in vitro can have problems with, e.g., pharmacological properties or dosing when administered in vivo. For example, conventional approaches identify a biological function of a particular drug candidate and then issues of blood-brain barrier penetration, serum half-life, etc. are left to fix later. In most cases, any attempts to fix an active molecule destroy the activity. The approaches of the present invention in a sense turn drug discovery on its head by, e.g., screening for “pre-selected” libraries in vivo, selected for biologically relevant properties, directing the modification/evolution of desired drug-like and pharmacological properties, and enabling efficient and rapid synthesis of thousands of variants that can be further tested and screened to more efficiently and selectively identify lead compounds of interest. Limiting additional screening, e.g., molecules that already possess the drug-like qualities required by the final molecule can assist with avoiding selection of drug candidates lacking a necessary pharmacological property that might not be identified until later in the conventional approaches.

[0030] The methods of generating libraries of the present invention can be applied to any type of potential drug candidate. Examples of potential drug candidates can include, e.g., any biologically, physiologically, or pharmacologically active substances that act locally or systemically in a subject. For example, libraries of small chemical molecules, peptides, and/or larger biologics (e.g., antibodies) can be screened (e.g., in vitro or in vivo). For example, existing libraries of small molecule drugs can be selected and administered to a subject to identify sub-libraries of the small molecule drugs that have pharmacological properties of interest. Similarly, existing libraries of biologics (e.g., antibodies) can be selected and administered to a subject to identify sub-libraries of the small molecule drugs that have certain pharmacological properties. As will be further described herein, the present invention further includes screening of peptides that can provide unique properties as compared to small chemical mol-

ecules and/or large biologics (e.g., antibodies). In one example, a particular target of interest could be screened against libraries (e.g., peptide libraries) that were already pre-selected to have drug-like properties aligned with those of the target. For example, if the target was in the central nervous system (CNS), a library of drug candidates that passed the blood-brain barrier could be screened for the CNS target. Alternatively, a drug screen for osteoarthritis could use a library designed to target cartilage. Once a lead was identified, then variants (e.g., thousands of variants) could be synthesized and screened in the same way to narrow down the number of potential lead compounds.

[0031] In some embodiments, the methods of the present invention can be used for generating small molecule libraries of potential drug candidates. For example, drug candidates can include known drugs such as those described in well-known literature references such as the Merck Index, the Physicians Desk Reference, and The Pharmacological Basis of Therapeutics, and they include, without limitation, medicaments; vitamins; mineral supplements; substances used for the treatment, prevention, diagnosis, cure or mitigation of a disease or illness; substances which affect the structure or function of the body; or prodrugs, which become biologically active or more active after they have been placed in a physiological environment. Candidates can also include, for example, small molecules, antibiotics, antivirals, antifungals, enediynes, heavy metal complexes, hormone antagonists, non-specific (non-antibody) proteins, sugar oligomers, aptamers, oligonucleotides (e.g., antisense oligonucleotides that bind to a target nucleic acid sequence (e.g., mRNA sequence)), siRNA, shRNA, peptides, proteins, radionuclides, and transcription-based pharmaceuticals. In some embodiments, potential drug candidates can include nucleic acids, peptides, small molecule compounds (e.g., pharmaceutical compounds), and peptidomimetics.

[0032] In certain aspects, drug candidates can be provided from large libraries of synthetic or natural compounds (e.g. pharmaceutical small molecule compounds and/or peptides). One example is an FDA approved library of compounds that can be used by humans. In addition, synthetic compound libraries are commercially available from a number of companies including Maybridge Chemical Co. (Trevillet, Cornwall, UK), Comgenex (Princeton, N.J.), Brandon Associates (Merrimack, N.H.), and Microsource (New Milford, Conn.), and a rare chemical library is available from Aldrich (Milwaukee, Wis.). Combinatorial libraries are available and can be prepared. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are also available, for example, Pan Laboratories or MycoSearch, or can be readily prepared by methods well known in the art. Compounds isolated from natural sources, such as animals, bacteria, fungi, plant sources, including leaves and bark, and marine samples may be assayed as candidates for the presence of potentially useful pharmaceutical agents. Several commercial libraries can immediately be used in the screens. Such libraries can include of analogues of naturally occurring or synthetic small molecules. Non limiting examples of naturally occurring small molecules include alkaloids, glycoside, lipids, phenazines, phenols, polyketide, terpenes, tetrapyrroles or nonribosomal peptides.

[0033] The small molecules to be screened as potential drug candidates can have molecular weights in the range of 200-1000 Da. In some embodiments, the potential drug candidates can have a molecular weight in the range of about

200-900 Da, about 200-800 Da, about 200-700 Da, about 200-600 Da, about 200-500 Da, about 200-400 Da, about 200-300 Da, about 300-1000 Da, about 300-900 Da, about 300-800 Da, about 300-700 Da, about 300-600 Da, about 300-500 Da, about 300-400 Da, about 400-1000 Da, about 400-900 Da, about 400-800 Da, about 400-700 Da, about 400-600 Da, about 400-500 Da, about 500-1000 Da, about 500-900 Da, about 500-800 Da, about 500-700 Da, about 500-600 Da, about 600-1000 Da, about 600-900 Da, about 600-800 Da, about 600-700 Da, about 700-1000 Da, about 700-900 Da, about 700-800 Da, about 800-1000 Da, about 800-900 Da, about 900-1000 Da, about 1000-1500 Da, about 1500-2000 Da, about 2000-2500 Da, about 2500-3000 Da, about 3000-3500 Da, about 3500-4000 Da, about 4000-4500 Da, or about 4500-5000 Da.

[0034] In some aspects, the present invention includes methods and systems for generating large libraries of knotted peptides (or knottins). These kinds of peptides are often found in the venom of poisonous animals and have evolved over millions of years to have remarkable serum stability and to bind specific targets. In an example embodiment, the present invention relates to knotted peptides that can include 15-40, or in some embodiments 11-57, amino acid, toxin-like, disulfide-linked peptides as potential drug scaffolds. Libraries of these knotted peptides can be generated and then further categorized by in vivo selection from the library of those peptide candidates with appropriate “drug like properties,” such as pharmacokinetics, distribution, oral availability, specific tissue distribution or targeting capability (e.g. tumor). Some selected properties of the generated peptide libraries and/or individual drug peptides can include 1) GI stability/oral bioavailability, 2) controllable/programmable pharmacokinetics, 3) ability to disrupt protein/protein interactions, 4) ability to cross the blood-brain barrier, 5) ability to penetrate cells, ability to penetrate specific organelles and/or cellular compartments, 6) potential for selectable tissue/organ targeting, 7) ability to be conjugated specifically to imaging agents or drugs, 8) “programmable diversity” (e.g., DNA encodes specific positions that can be fixed, variable, or selectively variable, 9) readily engineered (because the peptides can be “programmed” by DNA coding sequences), and 10) potential for programmable intracellular targeting (e.g., incorporation of nuclear translocation sequences).

[0035] The present invention further includes peptide scaffolds that, e.g., can be used as a starting point for generating peptide libraries. In some embodiments, these scaffolds can be derived from a variety of knotted peptides (or knottins). As referred to herein, “knotted peptides” or “knottins” include, for example, small disulfide-rich proteins characterized by a disulfide through disulfide knot. This knot can be, e.g., obtained when one disulfide bridge crosses the macrocycle formed by two other disulfides and the interconnecting backbone. In some aspects, the knotted peptides can include, e.g., growth factor cystine knots or inhibitor cystine knots. Other possible peptide structures include peptide having two parallel helices linked by two disulfide bridges without O—sheets (e.g., hefutoxin). Some suitable peptides for scaffolds can include, but are not limited to, chlorotoxin, brazzein, circulin, stecrisp, hanatoxin, midkine, hefutoxin, potato carboxypeptidase inhibitor, bubble protein, attractin, α -GI, α -GID, μ -PIIIA, ω -MVIIA, ω -CVID, χ -MrIA, ρ -TIA, conantokin G, contulakin G, GsMTx4, margatoxin, shK, toxin K, chymotrypsin inhibitor (CTI), and EGF epiregulin core. As shown in FIG. 1, several of these scaffolds, e.g., epiregulin, hefutoxin,

bubble, chlorotoxin and CTI, can be produced using the methods of the present invention and identified on a gel.

[0036] Based on a particular scaffold, larger libraries of scaffold variants can be generated (e.g., with diversity of tens, hundreds, thousands, tens of thousands, hundreds of thousands, or more different scaffold variants). Random or pre-defined mutagenesis can be used, e.g., to design the libraries of scaffold variants. In some instances, the methods can include identification of conserved amino acid residues in the peptide scaffold, and residues involved in forming tertiary structures, then biosynthesis of peptides with controlled variability, by encoding the peptides with nucleotides designed to conserve necessary amino acids, while varying other positions. For example, scans of the amino acids through the loop regions of a scaffold can be conducted systematically, applying mutagenesis to three residue patches and judging the outcome by tandem mass spectrometry. Once the mutation-tolerant regions are defined, this information can be used to make larger libraries through the mutation of larger numbers of residues.

[0037] In some embodiments, the methods described herein can use more than 10 peptide variants, 20 peptide variants, 30 peptide variants, 40 peptide variants, 50 peptide variants, 60 peptide variants, 70 peptide variants, 80 peptide variants, 90 peptide variants, 100 peptide variants, 110 peptide variants, 120 peptide variants, 130 peptide variants, 140 peptide variants, 150 peptide variants, 160 peptide variants, 170 peptide variants, 180 peptide variants, 190 peptide variants, 200 peptide variants, 210 peptide variants, 220 peptide variants, 230 peptide variants, 240 peptide variants, 250 peptide variants, 260 peptide variants, 270 peptide variants, 280 peptide variants, 290 peptide variants, 300 peptide variants, 350 peptide variants, 400 peptide variants, 450 peptide variants, 500 peptide variants, 550 peptide variants, 600 peptide variants, 650 peptide variants, 700 peptide variants, 750 peptide variants, 800 peptide variants, 850 peptide variants, 900 peptide variants, 950 peptide variants, 1000 peptide variants, 1100 peptide variants, 1200 peptide variants, 1300 peptide variants, 1400 peptide variants, 1500 peptide variants, 1600 peptide variants, 1700 peptide variants, 1800 peptide variants, 1900 peptide variants, 2000 peptide variants, 2500 peptide variants, 3000 peptide variants, 3500 peptide variants, 4000 peptide variants, 4500 peptide variants, 5000 peptide variants, 6000 peptide variants, 7000 peptide variants, 8000 peptide variants, 9000 peptide variants, 10,000 peptide variants, 11,000 peptide variants, 12,000 peptide variants, 13,000 peptide variants, 14,000 peptide variants, 15,000 peptide variants, 16,000 peptide variants, 17,000 peptide variants, 18,000 peptide variants, 19,000 peptide variants, 20,000 peptide variants, 21,000 peptide variants, 22,000 peptide variants, 23,000 peptide variants, 24,000 peptide variants, 25,000 peptide variants, 26,000 peptide variants, 27,000 peptide variants, 28,000 peptide variants, 29,000 peptide variants, 30,000 peptide variants, 34,000 peptide variants, 35,000 peptide variants, 36,000 peptide variants, 37,000 peptide variants, 38,000 peptide variants, 39,000 peptide variants, 40,000 peptide variants, 41,000 peptide variants, 42,000 peptide variants, 43,000 peptide variants, 44,000 peptide variants, 45,000 peptide variants, 46,000 peptide variants, 47,000 peptide variants, 48,000 peptide variants, 49,000 peptide variants, 50,000 peptide variants, 5,5000 peptide variants, 60,000 peptide variants, 65,000 peptide variants, 70,000 peptide variants, 75,000 peptide variants, 80,000 peptide variants, 85,000 peptide variants, 90,000 peptide variants, 95,000 pep-

peptide variants, 100,000 peptide variants, 110,000 peptide variants, 120,000 peptide variants, 130,000 peptide variants, 140,000 peptide variants, 150,000 peptide variants, 160,000 peptide variants, 170,000 peptide variants, 180,000 peptide variants, 190,000 peptide variants, 200,000 peptide variants, 210,000 peptide variants, 220,000 peptide variants, 230,000 peptide variants, 240,000 peptide variants, 250,000 peptide variants, 260,000 peptide variants, 270,000 peptide variants, 280,000 peptide variants, 290,000 peptide variants, or 300,000 peptide variants. In some embodiments more than about 300,000 peptide variants can be produced and/or used according to the methods of the present invention.

[0038] In some embodiments, about 100-15,000 peptides variants, 100-14,000 peptides variants, 100-13,000 peptides variants, 100-12,000 peptides variants, 100-11,000 peptides variants, 100-10,000 peptides variants, 100-9000 peptides variants, 100-8000 peptides variants, 100-7000 peptides variants, 100-6000 peptides variants, 100-5000 peptides variants, 100-4000 peptides variants, 100-3000 peptides variants, 100-2000 peptides variants, 100-1000 peptides variants, 100-900 peptides variants, 100-800 peptides variants, 100-700 peptides variants, 100-600 peptides variants, 100-500 peptides variants, 100-400 peptides variants, 100-300 peptides variants, 100-200 peptides variants can be produced and/or used.

[0039] In some embodiments, less than 100 peptide variants can be used. In some embodiments about 10-100 peptide variants, about 10-90 peptide variants, about 10-80 peptide variants, about 10-70 peptide variants, about 10-60 peptide variants, about 10-50 peptide variants, about 10-40 peptide variants, about 10-30 peptide variants, about 10-20 peptide variants, about 20-100 peptide variants, about 20-90 peptide variants, about 20-80 peptide variants, about 20-70 peptide variants, about 20-60 peptide variants, about 20-50 peptide variants, about 20-40 peptide variants, about 20-30 peptide variants, about 30-100 peptide variants, about 30-90 peptide variants, about 30-80 peptide variants, about 30-70 peptide variants, about 30-60 peptide variants, about 30-50 peptide variants, about 30-40 peptide variants, about 40-100 peptide variants, about 40-90 peptide variants, about 40-80 peptide variants, about 40-70 peptide variants, about 40-60 peptide variants, about 40-50 peptide variants, about 50-100 peptide variants, about 50-90 peptide variants, about 50-80 peptide variants, about 50-70 peptide variants, about 50-60 peptide variants, about 60-100 peptide variants, about 60-90 peptide variants, about 60-80 peptide variants, about 60-70 peptide variants, about 70-100 peptide variants, about 70-90 peptide variants, about 70-80 peptide variants, about 80-100 peptide variants, about 80-90 peptide variants, or about 90-100 peptide variants can be used in the methods of the present invention. In some embodiments less than about 90 peptide variants, about 80 peptide variants, about 70 peptide variants, about 60 peptide variants, about 50 peptide variants, about 40 peptide variants, about 30 peptide variants, about 20 peptide variants or about 10 peptide variants can be used in the methods of the present invention.

[0040] In some embodiments, the peptides used in the methods of the present invention can include about 2-100 amino acids. In some embodiments, the peptides of the present invention can include about 10-100 amino acids, about 10-90 amino acids, about 10-80 amino acids, about 10-70 amino acids, about 10-60 amino acids, about 10-50 amino acids, about 10-40 amino acids, about 10-30 amino acids, about 10-20 amino acids, about 20-30 amino acids, about 20-40 amino acids, about 20-50 amino acids, about

20-60 amino acids, about 20-70 amino acids, about 20-80 amino acids, about 20-90 amino acids, about 20-100 amino acids, about 30-40 amino acids, about 30-50 amino acids, about 30-60 amino acids, about 30-70 amino acids, about 30-80 amino acids, about 30-90 amino acids, about 30-100 amino acids, about 40-50 amino acids, about 40-60 amino acids, about 40-70 amino acids, about 40-80 amino acids, about 40-90 amino acids, about 40-100 amino acids, about 50-60 amino acids, about 50-70 amino acids, about 50-80 amino acids, about 50-90 amino acids, about 50-100 amino acids, about 50-100 amino acids or about 60-70 amino acids, about 60-80 amino acids, about 60-90 amino acids, about 60-100 amino acids, about 70-80 amino acids, about 70-90 amino acids, about 70-100 amino acids, about 80-90 amino acids, about 80-100 amino acids, or about 90-100 amino acids. In some embodiments, the peptides of the present invention can include about 10 amino acids, about 11 amino acids, about 12 amino acids, about 13 amino acids, about 14 amino acids, about 15 amino acids, about 16 amino acids, about 17 amino acids, about 18 amino acids, about 19 amino acids, about 20 amino acids, about 21 amino acids, about 22 amino acids, about 23 amino acids, about 24 amino acids, about 25 amino acids, about 26 amino acids, about 27 amino acids, about 28 amino acids, about 29 amino acids, about 30 amino acids, about 31 amino acids, about 32 amino acids, about 33 amino acids, about 34 amino acids, about 35 amino acids, about 36 amino acids, about 37 amino acids, about 38 amino acids, about 39 amino acids, about 40 amino acids, about 41 amino acids, about 42 amino acids, about 43 amino acids, about 44 amino acids, about 45 amino acids, about 46 amino acids, about 47 amino acids, about 48 amino acids, about 49 amino acids or about 50 amino acids.

[0041] In some embodiments, the methods of the present invention are used to generate libraries of knotted peptides. In certain embodiments, the knotted peptides of the present invention can include about 10-60 amino acids (e.g., 11-57 amino acids). In some embodiments, the knotted peptides of the present invention can include about 10-45 amino acids, about 10-40 amino acids, about 10-35 amino acids, about 10-30 amino acids, about 10-25 amino acids, about 15-50 amino acids, about 15-45 amino acids, about 15-40 amino acids, about 15-35 amino acids, about 15-30 amino acids, about 15-25 amino acids, about 15-20 amino acids, about 20-50 amino acids, about 20-45 amino acids, about 20-40 amino acids, about 20-35 amino acids, about 20-30 amino acids, about 20-25 amino acids, about 25-50 amino acids, about 25-45 amino acids, about 25-40 amino acids, about 25-35 amino acids, about 25-30 amino acids, about 30-50 amino acids, about 30-45 amino acids, about 30-40 amino acids, about 30-35 amino acids, about 35-50 amino acids, about 35-45 amino acids, about 35-40 amino acids, about 40-50 amino acids, about 40-45 amino acids, about 45-50 amino acids, or about 50-60 amino acids. In some embodiments, the knotted peptides of the current invention comprise of about 10 amino acids, about 11 amino acids, about 12 amino acids, about 13 amino acids, about 14 amino acids, about 15 amino acids, about 16 amino acids, about 17 amino acids, about 18 amino acids, about 19 amino acids, about 20 amino acids, about 21 amino acids, about 22 amino acids, about 23 amino acids, about 24 amino acids, about 25 amino acids, about 26 amino acids, about 27 amino acids, about 28 amino acids, about 29 amino acids, about 30 amino acids, about 31 amino acids, about 32 amino acids, about 33 amino acids, about 34

amino acids, about 35 amino acids, about 36 amino acids, about 37 amino acids, about 38 amino acids, about 39 amino acids, about 40 amino acids, about 41 amino acids, about 42 amino acids, about 43 amino acids, about 44 amino acids, about 45 amino acids, about 46 amino acids, about 47 amino acids, about 48 amino acids, about 49 amino acids, about 50 amino acids, about 51 amino acids, about 52 amino acids, about 53 amino acids, about 54 amino acids, about 55 amino acids, about 56 amino acids, or about 57 amino acids.

[0042] In some embodiments, the peptides of the present invention can include disulfide linkages. The peptides provided herein can include 1-10 disulfide linkages. In some embodiments, the peptides of the provided herein can have 1-9 disulfide linkages, 1-8 disulfide linkages, 1-7 disulfide linkages, 1-6 disulfide linkages, 1-5 disulfide linkages, 1-4 disulfide linkages, 1-3 disulfide linkages, 1-2 disulfide linkages, 2-3 disulfide linkages, 2-4 disulfide linkages, 2-5 disulfide linkages, 2-6 disulfide linkages, 2-7 disulfide linkages, 2-8 disulfide linkages, 2-9 disulfide linkages, 2-10 disulfide linkages, 3-4 disulfide linkages, 3-5 disulfide linkages, 3-6 disulfide linkages, 3-7 disulfide linkages, 3-8 disulfide linkages, 3-9 disulfide linkages, 3-10 disulfide linkages, 4-5 disulfide linkages, 4-6 disulfide linkages, 4-7 disulfide linkages, 4-8 disulfide linkages, 4-9 disulfide linkages, 5-6 disulfide linkages, 5-7 disulfide linkages, 5-8 disulfide linkages, 5-9 disulfide linkages, 5-10 disulfide linkages, 6-7 disulfide linkages, 6-8 disulfide linkages, 6-9 disulfide linkages, 6-10 disulfide linkages, 7-8 disulfide linkages, 7-9 disulfide linkages, 7-10 disulfide linkages, 8-9 disulfide linkages, 8-10 disulfide linkages, or 9-10 disulfide linkages. In some embodiments of the invention, the knotted peptides have 1 disulfide linkage, 2 disulfide linkages, 3 disulfide linkages, 4 disulfide linkages, 5, 6 disulfide linkages, 7 disulfide linkages, 8 disulfide linkages, 9 disulfide linkages or 10 disulfide linkages.

[0043] In some embodiments, the molecular weight of the peptides (e.g., knotted peptides) can be in the range of about 0.5 kDa to about 20 kDa. In some embodiments, the molecular weight of the peptides (e.g., knotted peptides) described herein can have a molecular weight in the range of about 0.5-15 kDa, about 0.5-10 kDa, about 0.5-5 kDa, about 1-20 kDa, about 1-15 kDa, about 1-10 kDa, about 1-5 kDa, about 5-20 kDa, about 5-15 kDa, about 5-10 kDa, about 10-20 kDa, about 10-15 kDa, or about 15-20 kDa. In some embodiments, the peptides (e.g., knotted peptides) described herein can have a molecular weight in the range of about 0.5-10 kDa, about 0.5-9 kDa, about 0.5-8, about 0.5-7 kDa, about 0.5-6 kDa, about 0.5-5 kDa, about 0.5-4 kDa, about 0.5-3 kDa, about 0.5-2 kDa, about 0.5-1 kDa, about 1-2 kDa, about 1-3 kDa, about 1-4 kDa, about 1-5 kDa, about 1-6 kDa, about 1-7 kDa, about 1-8 kDa, about 1-9 kDa, about 1-10 kDa, about 2-3 kDa, about 2-4 kDa, about 2-5 kDa, about 2-6 kDa, about 2-7 kDa, about 2-8 kDa, about 2-9 kDa, about 2-10 kDa, about 3-4 kDa, about 3-5 kDa, about 3-6 kDa, about 3-7 kDa, about 3-8 kDa, about 3-9 kDa, about 3-10 kDa, about 4-5 kDa, about 4-6 kDa, about 4-7 kDa, about 4-8 kDa, about 4-9 kDa, about 4-10 kDa, about 5-6 kDa, about 5-7 kDa, about 5-8 kDa, about 5-9 kDa, about 5-10 kDa, about 6-7 kDa, about 6-8 kDa, about 6-9 kDa, about 6-10 kDa, about 7-10 kDa, about 7-8 kDa, about 7-9 kDa, about 7-10 kDa, about 8-9 kDa, about 8-10 kDa, or about 9-10 kDa.

[0044] The peptide (e.g., knotted peptide) variants can be random (via technologies such as degenerate codons or error-prone PCR) or completely defined. Advances in chip-based oligonucleotide synthesis has made the generation of DNA

libraries inexpensive and routine, allowing the composition of the peptide libraries made from the DNA to be completely defined. By designing the peptides to each have a unique mass or tryptic fragment, mass spectrometry can be used to screen for the presence of each of the peptides simultaneously.

[0045] In some aspects, the peptide variants can be generated using a wide range of techniques. In certain embodiments, peptide libraries can be prepared with controlled variability by, e.g., encoding the peptides with nucleotides designed to conserve necessary amino acids, while varying other positions. The amino acids to retain in the variants can be identified by, e.g., alanine scanning techniques in which alanine can be systematically substituted into each amino acid position. This strategy identifies the amino acids in the native sequences that are result in an active peptide. Substitution of an essential amino acid can result in a reduction in peptide activity, and the degree of reduction in activity can be, e.g., used a relative measure of the importance of the amino acid being substituted.

[0046] With retained amino acids identified, large libraries of peptides can be developed around those retained amino acid residues. For example, the peptide library can be generated by systematic truncation of the flanking residues around the retained amino acids. A library can be a random library generated by substituting selected positions on the original peptide randomly and simultaneously with all other natural amino acids in a shot gun approach. In some embodiments, the libraries generated can be positional peptide library where a selected position or positions in a peptide sequence are each systematically replaced with different amino acids. This approach can show the effect of amino acid substitution at a particular position. The peptide libraries can also be constructed as scramble library by carrying out permutation on the original peptide's sequence. Such libraries have the potential to give all possible alternatives and offer the highest degree of variability for peptide library.

[0047] A variety of screening technologies can be used to identify the various libraries of drug candidates. For example, in vitro and/or in vivo assays can be used separately or in combination to identify drug candidates having the various pharmacological properties described herein. In one example embodiment, oral bioavailability can be analyzed for a library. For example, a peptide library of hundreds or thousands of peptides can be fed (e.g., less than about 1 g/mL or 100 mg/mL) to a subject (e.g., an animal, a rat, a mouse, a human or a pig). The blood, for example, of the subject can be analyzed at a time point after the library of peptides have been processed in the stomach and possibly passed into the blood. In this example, peptides of the library that are found in the blood can be identified (e.g., by mass spectrometry) as capable of chemically withstanding the acidic environment of the stomach and capable of crossing into the bloodstream of the subject. After such an analysis, a sub-library of the larger library can then be used to identify drugs having oral bioavailability.

[0048] Other pharmacological properties can be identified for larger libraries and/or sub-libraries of drugs having another pharmacological property. For example, serum half-life can also be analyzed to determine a sub-library of drug candidates (e.g., knotted peptides) having a desired serum half-life. In some embodiments, sub-libraries of drug candidates (e.g., peptides) can have serum half-lives ranging from about 15 minutes to about 72 hours. Lower and longer half-lives can be identified, as well. In one example embodiment,

libraries of drug candidates (e.g., knotted peptides) can be injected (e.g., less than about 1 g/mL or 100 mg/mL) into the bloodstream of a subject and then, after a determined time point (e.g., every 15 minutes, every hour, every 4 hours, or other schedule), a sample can be taken from the subject (e.g., blood, urine, mucous, spinal fluid, or tissue) and analyzed (e.g., using mass spectrometry) to determine the drug candidates that are still present after injection. Tissue samples from subjects can include biopsy and/or necropsy tissue. The tissue samples can also include tissue from a brain, a lung, a kidney, a muscle, a liver, a heart, a stomach, a pancreas, or any other organ. Depending on the time points, a serum half-life can be determined and a sub-library of drug candidates (e.g., knotted peptides) having a specific serum half-life can be identified.

[0049] In some embodiments, the identification of drug candidates having certain pharmacological properties can be done in sequence or in parallel. For example, one sub-library can be identified as having a specific serum half-life. That sub-library can then be screened for oral bioavailability as described above. Alternatively, a library can be ingested by a subject and then later analyzed to determine both whether a sub-library is orally bioavailable and if the drug candidates in the sub-library have a desired serum half-life.

[0050] The screening processes are also not limited in the types of drug-like properties that can be identified for sub-libraries of the drug candidates. Other example screenings can be used to identify, e.g., sub-libraries of drug candidates (e.g., knotted peptides) that can cross the blood brain barrier, penetrate certain cell types, and/or penetrate subcellular organelles or cellular compartments. For example, libraries can be administered to a subject and then brain tissue can be analyzed (e.g., by mass spectrometry) to determine a sub-library of drug candidates (e.g., knotted peptides) that are capable of crossing the blood brain barrier. In some embodiments, sub-libraries can be generated that have candidates exhibiting other characteristics. For example, drug candidates (e.g., peptides) may be excluded from the libraries if they adversely target critical organs, thereby creating a liability. Alternatively, drug candidates may be selected or deselected depending on whether they bind serum albumin and/or IgG and/or FcRn (e.g., by pH sensitive binding).

[0051] In some embodiments, for example, the libraries of knottin peptide drug candidates can be expressed in a particular animal model to screen for a desired effect or therapeutically treat the animal. The expression, e.g., can be accomplished using a viral transduction system (e.g., lentivirus, adenovirus, or adeno-associated virus) to deliver the library of peptides to a particular organ, or area of the body that will incorporate the viral material and express the knottin peptide (s). In some aspects, these methods can be used for the determination of a therapeutically useful peptide in an experimental animal system. For example, a library of knottins can be expressed in brain cells of a mouse that is genetically engineered to develop a neurodegenerative disease (e.g., Huntington's disease) or muscular disease (e.g., sarcopenia). Regions where neurons remain healthy can be identified in comparison areas of degeneration, and then the peptide that is secreted locally in this region can be identified (e.g., via mass spectrometric analysis of tissue extracted from the healthy regions).

[0052] As noted above, the present invention includes a variety of methods for determining sub-libraries of drug candidates have certain pharmacological properties. In some aspects of the present invention, the sub-libraries of drug

candidates, having selected pharmacological properties and identified using in vitro and/or in vivo screening, can be then screened to identify one or more lead drug candidates that then can be modified to potentially improve certain characteristics for ultimately designing a useful drug for binding to a target of interest (e.g., a therapeutic target).

[0053] In an example embodiment, peptide libraries can be created using an oligonucleotide assembly approach (described further herein), where particular surface residue positions will be targeted for random mutagenesis using degenerate codons (FIG. 2). In one aspect, the methods can include parsing through the scaffold by mutating ten unique or overlapping three amino acid subdomains. Once we have determined which sites will tolerate mutation, larger randomized libraries will be built. These libraries can then be screened individually or in a pooled format. The assembly reactions can be amplified with primers, e.g., containing XhoI/BamHI sites and the purified PCR products can be cloned into a vector (e.g., the Daedalus vector pCVL-sUCOE-SFFV-IRES-GFP). The resulting lentiviral library can be used to transduce cells (e.g., HEK 293 Freestyle cells) to create a polyclonal population of cells secreting variants. Secreted constructs can be engineered with an N-terminal secretion signal and a C terminal tag (e.g., a STREPII tag (SAWSH-PQFEK)). The secreted peptides can be harvested from the culture supernatants using, e.g., size exclusion or affinity chromatography. Several libraries of limited diversity or wide diversity can be generated using this approach. The purified pool of peptides can be analyzed using tandem mass spectrometry in order to gauge the diversity of the secreted protein libraries and establish detection limits for downstream screening applications.

[0054] In some embodiments, the methods of generating libraries can be iterative and parallel. As shown in FIG. 3, for example, a single experiment can be conducted in parallel to test thousands of peptides for cell penetration, oral bioavailability, serum half-life, blood-brain barrier penetration and tissue homing. Those peptides that possess favorable properties can form the basis for making a subsequent library of peptides possessing the same and slightly altered sequences. This set may be thousands in size and can be tested in parallel for all these favorable ("drug-like") qualities. And the best of this group will form the basis for the set. This process can be used to produce general principles with predictive value for the generation of further drug-like peptides.

Peptide Conjugates

[0055] In some aspects, the present invention includes libraries of peptide conjugates. For example, some or all of a library of peptides can be conjugated to a moiety selected to modify a property of the peptides.

[0056] In an example embodiment, the present invention includes libraries of knottins or knotted peptides conjugated at the N-terminus to hydrophobic (e.g., lipophilic) moieties. All or some of the knottins can be lacking internal lysines, e.g., to avoid conjugation at the internal lysine positions, thereby allowing conjugation to the amino terminus of the peptide. In some embodiments, the attachment of a hydrophobic moiety to the N-terminus can be used to extend half-life of the knotted peptides. In some embodiments, simple carbon chains (e.g., by myristoylation and/or palmitoylation) can be conjugated to the peptides. The lipophilic moieties can extend half-life through reversible binding to serum albumin. In certain embodiments, attachment of a near infrared dye to

the N-terminus of the peptide can also be performed to allow for tracing of the conjugated peptide. In certain embodiments, attachment of a near infrared dye to a lysine of the peptide can also be performed to allow for tracing of the conjugated peptide. An antibody to the dye can further allow the dye to fill a dual role of both a tracking marker and a retrieval handle. The conjugated peptides can also be conjugated to other moieties that can serve other roles, such as providing an affinity handle (e.g. biotin) for retrieval of the peptides from tissues or fluids.

[0057] Other modifications can be used. For example, the knotted peptides can include post-translational modifications (e.g., methylation and/or amidation), or the knotted peptides can also be composed of some or all D-amino acids, which can affect, e.g., serum half-life. In some embodiments, the peptides in the libraries can be conjugated to other moieties that, e.g., can modify or effect changes to the properties of the peptides. The conjugated moieties can, e.g., be lipophilic moieties that extend half-life of the peptides through reversible binding to serum albumin. In some embodiments, the lipophilic moiety can be cholesterol or a cholesterol derivative including cholestenes, cholestanes, cholestadienes and oxysterols. In some embodiments, the peptides can be conjugated to myristic acid (tetradecanoic acid) or a derivative thereof.

[0058] In some embodiments, the peptides can be conjugated to detectable labels to enable tracking detecting or visualizing of the bio-distribution of a conjugated peptide. The detectable labels can be fluorescent labels (e.g., fluorescent dyes). In certain embodiments, the fluorescent label can have emission characteristics that are desired for a particular application. For example, the fluorescent label can be a fluorescent dye that has a emission wavelength maximum between a range of 500 nm to 1100 nm, between a range of 600 nm to 1000 nm, between a range of 600 to 800 nm, between a range of 650 nm to 850 nm, between a range of 700 nm to 800 nm, between a range of 720 to 780 nm, or between a range of 720 to 750 nm. For example, under certain conditions, cyanine 5.5 can have an emission maximum around 695 nm, IRdye 800 can have an emission maximum around 800 nm, and indocyanine green can have an emission maximum around 820 nm. One of ordinary skill in the art will appreciate the various dyes that can be used as detectable labels and that have the emission characteristics above.

[0059] Non limiting examples of fluorescent dyes that could be used as a conjugating molecule in the present disclosure include rhodamine, rhodol, fluorescein, thiofluorescein, aminofluorescein, carboxyfluorescein, chlorofluorescein, methylfluorescein, sulfofluorescein, a minorrhodol, carboxyrhodol, chlororhodol, methylrhodol, sulforhodol; a minorrhodamine, carboxyrhodamine, chlororhodamine, methylrhodamine, sulforhodamine, and thiorhodamine, cyanine, indocarbocyanine, oxacarbocyanine, thiocarbocyanine, merocyanine, a cyanine dye (e.g., cyanine 2, cyanine 3, cyanine 3.5, cyanine 5, cyanine 5.5, cyanine 7), oxadiazole derivatives, pyridyloxazole, nitrobenzoxadiazole, benzoxadiazole, pyrene derivatives, cascade blue, oxazine derivatives, Nile red, Nile blue, cresyl violet, oxazine 170, acridine derivatives, proflavin, acridine orange, acridine yellow, arylmethine derivatives, xanthene dyes, sulfonated xanthenes dyes, Alexa Fluors (e.g., Alexa Fluor 594, Alexa Fluor 633, Alexa Fluor 647, Alexa Fluor 700), auramine, crystal violet, malachite green, tetrapyrrole derivatives, porphyrin, phthalocyanine, and bilirubin. In some embodiments, the dyes can be

near-infrared dyes including, e.g., Cy5.5, IRdye 800, DyLight 750 or indocyanine green (ICG). In some embodiments, near infrared dyes can include cyanine dyes (e.g., cyanine 2, cyanine 3, cyanine 3.5, cyanine 5, cyanine 5.5, cyanine 7). In certain embodiments, the detectable label can include xanthene dyes or sulfonated xanthenes dyes, such as Alexa Fluors (e.g., Alexa Fluor 594, Alexa Fluor 633, Alexa Fluor 647, Alexa Fluor 700). If an antibody to the dye could be found the conjugated dyes could be used both as a tracking, detecting or visualizing marker and as a retrieval handle.

[0060] The peptides in the libraries of the present invention can also be conjugated to biotin. In addition of extension of half-life, biotin could also act as an affinity handle for retrieval of the peptides from tissues or other locations. In one embodiment, the peptides can be conjugated, e.g., to a biotinidase resistant biotin with a PEG linker (e.g., NHS-dPEG₄-Biotinidase resistant biotin). In some embodiments, fluorescent biotin conjugates that can act both as a detectable label and an affinity handle can be used. Non limiting examples of commercially available fluorescent biotin conjugates include Atto 425-Biotin, Atto 488-Biotin, Atto 520-Biotin, Atto-550 Biotin, Atto 565-Biotin, Atto 590-Biotin, Atto 610-Biotin, Atto 620-Biotin, Atto 655-Biotin, Atto 680-Biotin, Atto 700-Biotin, Atto 725-Biotin, Atto 740-Biotin, fluorescein biotin, biotin-4-fluorescein, biotin-(5-fluorescein) conjugate, and biotin-B-phycoerythrin, alexa fluor 488 biocytin, alexa fluor 546, alexa fluor 549, lucifer yellow cadaverine biotin-X, Lucifer yellow biocytin, Oregon green 488 biocytin, biotin-rhodamine and tetramethylrhodamine biocytin. In some other examples, the conjugates could include chemiluminescent compounds, colloidal metals, luminescent compounds, enzymes, radioisotopes, and paramagnetic labels.

[0061] The peptides in the libraries of the present invention can be conjugated to vitamins or other molecules typically found in foods that are absorbed into the bloodstream from the stomach, small intestine, or colon. Examples include, but are not limited to, vitamin A, vitamin C, vitamin B₂, vitamin B₃, vitamin B₆, vitamin B₁₂, vitamin D, vitamin E, vitamin K. The goal of these conjugations is to improve oral bioavailability or absorption of the peptide from the gastrointestinal system.

[0062] In some instances, selected series of amino acids that appear to help certain peptides cross biologic barriers such as the gastrointestinal tract, the blood brain barrier, the cell membrane, the nuclear membrane can be identified and genetically or physically grafted onto other peptides for the purpose of helping the new peptide cross the same biologic barriers. In other cases, the same approach might be used to graft sequences onto peptides that would prevent the new peptide from crossing certain biological barriers. For example, a drug could be modified in this manner to prevent BBB penetration and thus reduce the likelihood of central nervous system side effects.

Enantiomers of Peptide Libraries

[0063] The present invention further includes enantiomers of the peptide libraries. For example, the peptides can include all naturally occurring L-amino acids. In certain embodiments, some or all of the amino acids in the peptides can be D-amino acids.

[0064] In some aspects, the present invention includes peptides having D-amino acids and being based on a chlorotoxin scaffold, e.g., D-amino acid chlorotoxin and/or D-amino acid

chlorotoxin variants. The peptides based on chlorotoxin can include varying amounts of D-amino acids. The natural peptide of chlorotoxin has the following amino acid sequence: MCMPCFTTDHQMARCDDCCGGKGRGK-CYGPQCLCR. The peptide can be further crosslinked by four disulfide bonds formed among the cysteine residues present in the sequence. The natural peptide includes all L-amino acids. Chemically, the all-D-amino acid form of chlorotoxin can be indistinguishable from the all-L-amino acid form of chlorotoxin, except in chiral environments such as those that are present in vivo. As D-amino acids are quite rare in nature, peptides including D-amino acids can be resistant to proteolysis. Accordingly, D-amino acid chlorotoxin can be resistant to being cleaved, e.g., in biological fluids after administration to a subject. Thus, the D-amino acid peptide can, e.g., be administered orally and may have a lower immunogenic response. The D-amino acids can also assist with preventing antigen processing and display, thereby reducing antigenicity. The D-amino acids for use in the present invention can include, e.g., dArg, dHis, dLys, dAsp, dGlu, dSer, dThr, dAsn, dGln, dCys, dGly, dPro, dAla, dVal, dIle, dLeu, dMet, dPhe, dTyr, and dTrp.

[0065] In some embodiments, the peptides can include a peptide having the same sequence as the natural chlorotoxin sequence in which at least three of the amino acids in the sequence include a D-amino acid in place of the natural amino acid. In some embodiments, the peptides can include chlorotoxin variants, e.g., mutated chlorotoxin peptides, in which at least three of the amino acids in the sequence include a D-amino acid in place of the natural amino acid. In some embodiments, at least four, at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, or at least 35 of the amino acids in the sequence can include a D-amino acid in place of the natural amino acid. In certain embodiments, at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, or at least 90% of the amino acids in the peptides based on a chlorotoxin scaffold can be D-amino acids. In one aspect, the present invention includes peptides based on a chlorotoxin scaffold, in which all the amino acids in the peptides are D-amino acids.

[0066] In accordance, e.g., with the methods of making peptides of the present invention, a wide variety of variants of chlorotoxin can be generated. For example, using conventional mutagenesis and synthetic-based systems, thousands of chlorotoxin D-amino acid variants can be prepared based on the naturally occurring amino acid sequence of chlorotoxin. In some embodiments, the peptides of the present invention can include an amino acid sequence at least 60%, 65%, 70%, 75%, 80%, 83%, 85%, 86%, 89%, 90%, 92% or 95% identical to the following sequence of MCMPCFTTDHQMARCDDCCGGKGRGK-CYGPQCLCR, in which some or all of the amino acids are D-amino acids. In some embodiments, the chlorotoxin variants can have at least three of the amino acids in the sequence that are a D-amino acid in place of the natural amino acid. In some embodiments, at least four, at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, or at least 35 of the amino acids in the sequence of the chlorotoxin variants can include a D-amino acid in place of a natural amino acid. In certain embodiments, at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, or at least 90% of the amino acids in the sequence of the chlorotoxin variants can include a D-amino acid in place of a natural amino acid.

[0067] In one embodiment, the all D-amino acid peptide can have the following formula: H-dMet-dCys-dMet-dPro-dCys-dPhe-dThr-dThr-dAsp-dHis-dGln-dMet-dAla-dArg-dXaa-dCys-dAsp-dAsp-dCys-dCys-dGly-dGly-dXaa-dGly-dArg-dGly-dXaa-dCys-dTyr-dGly-dPro-dGln-dCys-dLeu-dCys-dArg-OH acetate salt (disulfide bonds, air oxidized), wherein dXaa is dArg, dAla, or dLys.

[0068] In another embodiment, the all D-amino acid peptide can have the following formula: H-dMet-dCys-dMet-dPro-dCys-dPhe-dThr-dThr-dAsp-dHis-dGln-dMet-dAla-dArg-dXaa-dCys-dAsp-dAsp-dCys-dCys-dGly-dGly-dXaa-dGly-dArg-dGly-dLys-dCys-dTyr-dGly-dPro-dGln-dCys-dLeu-dCys-dArg-OH acetate salt (disulfide bonds, air oxidized), wherein dXaa is dArg, dAla, or dLys.

[0069] In another embodiment, the all D-amino acid peptide can have the following formula: H-dMet-dCys-dMet-dPro-dCys-dPhe-dThr-dThr-dAsp-dHis-dGln-dMet-dAla-dArg-dArg-dCys-dAsp-dAsp-dCys-dCys-dGly-dGly-dArg-dGly-dArg-dGly-dLys-dCys-dTyr-dGly-dPro-dGln-dCys-dLeu-dCys-dArg-OH acetate salt (disulfide bonds, air oxidized); Molecular weight: 4065.6 Da.

[0070] In another embodiment, the all D-amino acid peptide can have the following formula: H-dMet-dCys-dMet-dPro-dCys-dPhe-dThr-dThr-dAsp-dHis-dGln-dMet-dAla-dArg-dArg-dCys-dAsp-dAsp-dCys-dCys-dGly-dGly-dAla-dGly-dArg-dGly-dLys-dCys-dTyr-dGly-dPro-dGln-dCys-dLeu-dCys-dArg-OH acetate salt (disulfide bonds, air oxidized).

[0071] In another embodiment, the all D-amino acid peptide can have the following formula: H-dMet-dCys-dMet-dPro-dCys-dPhe-dThr-dThr-dAsp-dHis-dGln-dMet-dAla-dArg-dAla-dCys-dAsp-dAsp-dCys-dCys-dGly-dGly-dAla-dGly-dArg-dGly-dLys-dCys-dTyr-dGly-dPro-dGln-dCys-dLeu-dCys-dArg-OH acetate salt (disulfide bonds, air oxidized).

[0072] In another embodiment, the all D-amino acid peptide can have the following formula: H-dMet-dCys-dMet-dPro-dCys-dPhe-dThr-dThr-dAsp-dHis-dGln-dMet-dAla-dArg-dAla-dCys-dAsp-dAsp-dCys-dCys-dGly-dGly-dArg-dGly-dArg-dGly-dLys-dCys-dTyr-dGly-dPro-dGln-dCys-dLeu-dCys-dArg-OH acetate salt (disulfide bonds, air oxidized).

[0073] In another embodiment, the all D-amino acid peptide can have the following formula: H-dMet-dCys-dMet-dPro-dCys-dPhe-dThr-dThr-dAsp-dHis-dGln-dMet-dAla-dArg-dXaa-dCys-dAsp-dAsp-dCys-dCys-dGly-dGly-dXaa-dGly-dArg-dGly-dLys(6-ICGhexanoyl)-dCys-dTyr-dGly-dPro-dGln-dCys-dLeu-dCys-dArg-OH acetate salt (disulfide bonds, air oxidized), wherein dXaa is dArg, dAla, or dLys.

[0074] In another embodiment, the all D-amino acid peptide can have the following formula: H-dMet-dCys-dMet-dPro-dCys-dPhe-dThr-dThr-dAsp-dHis-dGln-dMet-dAla-dArg-dArg-dCys-dAsp-dAsp-dCys-dCys-dGly-dGly-dArg-dGly-dArg-dGly-dLys(6-ICGhexanoyl)-dCys-dTyr-dGly-dPro-dGln-dCys-dLeu-dCys-dArg-OH acetate salt (disulfide bonds, air oxidized); Molecular weight: 4765 Da.

[0075] In another embodiment, the all D-amino acid peptide can have the following formula: H-dMet-dCys-dMet-dPro-dCys-dPhe-dThr-dThr-dAsp-dHis-dGln-dMet-dAla-dArg-dArg-dCys-dAsp-dAsp-dCys-dCys-dGly-dGly-dAla-dGly-dArg-dGly-dLys(6-ICGhexanoyl)-dCys-dTyr-dGly-dPro-dGln-dCys-dLeu-dCys-dArg-OH acetate salt (disulfide bonds, air oxidized).

[0076] In another embodiment, the all D-amino acid peptide can have the following formula: H-dMet-dCys-dMet-dPro-dCys-dPhe-dThr-dThr-dAsp-dHis-dGln-dMet-dAla-dArg-dAla-dCys-dAsp-dAsp-dCys-dCys-dGly-dGly-dAla-dGly-dArg-dGly-dLys(6-ICGhexanoyl)-dCys-dTyr-dGly-dPro-dGln-dCys-dLeu-dCys-dArg-OH acetate salt (disulfide bonds, air oxidized).

[0077] In another embodiment, the all D-amino acid peptide can have the following formula: H-dMet-dCys-dMet-dPro-dCys-dPhe-dThr-dThr-dAsp-dHis-dGln-dMet-dAla-dArg-dAla-dCys-dAsp-dAsp-dCys-dCys-dGly-dGly-dArg-dGly-dArg-dGly-dLys(6-ICGhexanoyl)-dCys-dTyr-dGly-dPro-dGln-dCys-dLeu-dCys-dArg-OH acetate salt (disulfide bonds, air oxidized).

[0078] The present invention further includes conjugates of D-amino acid chlorotoxin and D-amino acid chlorotoxin variants described herein. For example, the D-amino acid chlorotoxin and D-amino acid chlorotoxin variants can be conjugated with a variety of moieties that can, e.g., modify the pharmacological properties of the peptides. In some embodiments, some or all of the lysines in the amino acid sequence of the peptides can be replaced with alanine or arginine to facilitate N-terminal conjugation (e.g., by reducing competition from the amino terminus of the lysine).

[0079] For example, the present invention can include D-amino acid chlorotoxin and D-amino acid chlorotoxin variants that have an amino acid sequence at least 60%, 65%, 70%, 75%, 80%, 83%, 85%, 86%, 89%, 90%, 92% or 95% identical to the following sequence of MCMPCFTTDHQ-MARXCDDCCGGXGRGXCYGPQCLCR, in which some or all of the amino acids are D-amino acids and X can include K, A or R. In some embodiments, the chlorotoxin variants can have at least three of the amino acids in the sequence that are a D-amino acid in place of the natural amino acid. In some embodiments, at least four, at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, or at least 35 of the amino acids in the sequence of the chlorotoxin variants can include a D-amino acid in place of a natural amino acid. In certain embodiments, at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, or at least 90% of the amino acids in the sequence of the chlorotoxin variants can include a D-amino acid in place of a natural amino acid.

[0080] The conjugates of D-amino acid chlorotoxin and D-amino acid chlorotoxin variants can include moieties that are conjugated to various locations of the peptides. For example, moieties can be conjugated to any one of the lysine residues (e.g., the L- or D-amino acid form) in chlorotoxin sequence (e.g., Lys-15, Lys-23, and/or Lys-27). Alternatively, the peptides may be mutated to be free of lysine residues. In some embodiments, the moieties can be conjugated to the N-terminus of the D-amino acid chlorotoxin peptides and D-amino acid chlorotoxin peptide variants. In some embodiments, the moieties can be conjugated to at least one of the amino acids in the peptides.

[0081] In certain embodiments, the D-amino acid chlorotoxin and D-amino acid chlorotoxin variants can be conjugated to moieties, such as detectable labels (e.g., dyes) that can be detected (e.g., visualized) in a subject. In some embodiments, the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants can be conjugated to detectable labels to enable tracking of the bio-distribution of a conjugated peptide. The detectable labels can include fluorescent labels (e.g., fluorescent dyes).

[0082] In certain embodiments, the fluorescent label can have emission characteristics that are desired for a particular application. For example, the fluorescent label can be a fluorescent dye that has a emission wavelength maximum between a range of 500 nm to 1100 nm, between a range of 600 nm to 1000 nm, between a range of 600 to 800 nm, between a range of 650 nm to 850 nm, between a range of 700 nm to 800 nm, between a range of 720 to 780 nm, or between a range of 720 to 750 nm. One of ordinary skill in the art will appreciate the various dyes that can be used as detectable labels and that have the emission characteristics above. For example, under certain conditions, cyanine 5.5 can have an emission maximum around 695 nm, IRdye can have an emission maximum around 800 nm, and indocyanine green can have an emission maximum around 820 nm.

[0083] Non limiting examples of fluorescent dyes that could be used as a conjugating molecule in the present disclosure include rhodamine, rhodol, fluorescein, thiofluorescein, aminofluorescein, carboxyfluorescein, chlorofluorescein, methylfluorescein, sulfofluorescein, a minorhodol, carboxyrhodol, chlororhodol, methylrhodol, sulforhodol; a minorrhodamine, carboxyrhodamine, chlororhodamine, methylrhodamine, sulforhodamine, and thiorhodamine, cyanine, indocarbocyanine, oxacarbocyanine, thiocarbocyanine, merocyanine, oxadiazole derivatives, pyridyloxazole, nitrobenzoxadiazole, benzoxadiazole, pyrene derivatives, cascade blue, oxazine derivatives, Nile red, Nile blue, cresyl violet, oxazine 170, acridine derivatives, proflavin, acridine orange, acridine yellow, arylmethine derivatives, auramine, crystal violet, malachite green, tetrapyrrole derivatives, porphyrin, phthalocyanine, and bilirubin. In some embodiments, the detectable label can include near-infrared dyes, such as, but not limited to, Cy5.5, indocyanine green (ICG), DyLight 750 or IRdye 800. In some embodiments, near infrared dyes can include a cyanine dye (e.g., cyanine 2, cyanine 3, cyanine 3.5, cyanine 5, cyanine 5.5, cyanine 7). In certain embodiments, the detectable label can include xanthene dyes or sulfonated xanthenes dyes, such as Alexa Fluors (e.g., Alexa Fluor 594, Alexa Fluor 633, Alexa Fluor 647, Alexa Fluor 700). In addition, if an antibody to the dyes can be identified, then conjugated dyes could be used both as a tracking, detecting or visualizing marker and as a retrieval handle.

[0084] Other modifications to peptides can be used. For example, the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants can include post-translational modifications (e.g., methylation and/or amidation), which can affect, e.g., serum half-life. In some embodiments, the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants can be conjugated to other moieties that, e.g., can modify or effect changes to the properties of the peptides. The conjugated moieties can, e.g., be lipophilic moieties that extend half-life of the peptides through reversible binding to serum albumin. In some embodiments, simple carbon chains (e.g., by myristoylation) can be conjugated to the peptides. In some embodiments, the lipophilic moiety can be cholesterol or a cholesterol derivative including cholestenes, cholestanes, cholestadienes and oxysterols. In some embodiments, the peptides can be conjugated to myristic acid (tetradecanoic acid) or a derivative thereof.

[0085] The conjugated D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants can also be conjugated to other moieties that can serve other roles, such as providing an affinity handle (e.g. biotin) for retrieval of the peptides from tissues or fluids. For example, the peptides in the libraries of

the present invention can also be conjugated to biotin. In addition of extension of half-life, biotin could also act as an affinity handle for retrieval of the peptides from tissues or other locations. In some embodiments, fluorescent biotin conjugates that can act both as a detectable label and an affinity handle can be used. Non limiting examples of commercially available fluorescent biotin conjugates include Atto 425-Biotin, Atto 488-Biotin, Atto 520-Biotin, Atto-550 Biotin, Atto 565-Biotin, Atto 590-Biotin, Atto 610-Biotin, Atto 620-Biotin, Atto 655-Biotin, Atto 680-Biotin, Atto 700-Biotin, Atto 725-Biotin, Atto 740-Biotin, fluorescein biotin, biotin-4-fluorescein, biotin-(5-fluorescein) conjugate, and biotin-B-phycoerythrin, Alexa fluor 488 biocytin, Alexa fluor 546, Alexa Fluor 549, lucifer yellow cadaverine biotin-X, Lucifer yellow biocytin, Oregon green 488 biocytin, biotin-rhodamine and tetramethylrhodamine biocytin. In some other examples, the conjugates could include chemiluminescent compounds, colloidal metals, luminescent compounds, enzymes, radioisotopes, and paramagnetic labels.

[0086] In some embodiments, the present invention provides methods of detecting D-amino acid chlorotoxin peptides and/or D-amino acid chlorotoxin peptide variants. In one aspect, the present invention includes a method of detecting peptides in a subject. The method can include administering to the subject an effective amount of a composition including D-amino acid chlorotoxin peptides and/or D-amino acid chlorotoxin peptide variants. The peptides can include detectable labels, and the method can further include detecting signals from the detectable labels. In some embodiments, after the administering, the peptides can bind to tissue (e.g., cells) expressing a chlorotoxin target, and the method can include detecting signals from the peptides bound to the tissue in the subject to determine a level of binding of the peptides to the tissue. In certain embodiments, detecting an increased level of binding of the peptides to the tissue compared to normal tissue can indicate that the tissue includes tumor tissue expressing the chlorotoxin target. In addition to detecting binding of the peptides to tissue, the methods of detecting can include tracing the peptides in a patient, e.g., by detecting the peptides in a subject's blood and/or tissue.

[0087] The present invention further includes methods of detecting and/or imaging tissue (e.g., cancerous tissue) in a subject. The D-amino acid chlorotoxin peptides and/or D-amino acid chlorotoxin peptide variants can bind to a variety of chlorotoxin targets. For example, in some embodiments, the chlorotoxin targets can include Annexin A2 and Calpactin. With a binding affinity for the chlorotoxin targets, the compositions including D-amino acid chlorotoxin peptides and/or D-amino acid chlorotoxin peptide variants can be administered to a subject and the peptides can then specifically bind to the chlorotoxin targets. Detection and/or imaging of the binding to the chlorotoxin targets can be used, e.g., to identify tissue having chlorotoxin targets, such as Annexin A2 and Calpactin.

[0088] Detection and/or imaging of binding can also be used, e.g., for determining whether a subject has a disease, condition and/or disorder associated with cells having (e.g., expressing) the chlorotoxin target. In some embodiments, the methods can include obtaining an image of a region in the subject by detecting signals from the detectable labels. The region, for example, can include cancerous tissue and the peptides can be bound to the cancerous tissue, thereby providing detection and/or imaging of the cancerous tissue. The conjugated dyes can be used, e.g., in tumor imaging using the

D-amino acid chlorotoxin and D-amino acid chlorotoxin variants that can specifically bind to tumors (e.g., brain tumors). In some embodiments, the region can include a tumor and the peptides can be bound to the tumor, thereby providing imaging of the tumor. In some embodiments, the region can include a brain tumor and the peptides are bound to the brain tumor, thereby providing imaging of the brain tumor. In certain embodiments, the present invention can include detection and/or imaging of cancerous tissue associated with gliomas, astrocytomas, medulloblastomas, choroids plexus carcinomas, ependymomas, other brain tumors, neuroblastoma, head and neck cancer, lung cancer, breast cancer, intestinal cancer, pancreatic cancer, liver cancer, kidney cancer, sarcomas, osteosarcoma, rhabdomyosarcoma, Ewing's sarcoma, carcinomas, melanomas (including amelanotic melanoma), ovarian cancer, cervical cancer, lymphoma, thyroid cancer, anal cancer, colo-rectal cancer, endometrial cancer, germ cell tumors, laryngeal cancer, multiple myeloma, prostate cancer, retinoblastoma, gastric cancer, testicular cancer, and Wilm's tumor. In some embodiments, the cancerous tissue can be associated with a glioma, a skin cancer, a lung cancer, a lymphoma, a medulloblastoma, a prostate cancer, a pancreatic cancer, or a combination thereof. In some embodiments, the cancerous tissue can be associated with breast and mammary cancers, colon, skin, lung, lymphoma, glioma, medulloblastoma, prostate, pancreatic cancers, oral squamous cell carcinoma, and/or hemangiopericytoma.

[0089] In one aspect, the present invention further includes a peptide for imaging a tumor in a subject. The peptide can include an amino acid sequence having at least three D-amino acids and the peptide having a secondary structure configured to bind to the tumor, wherein the peptide further includes a detectable label. As described herein, for example, the peptide can have a secondary structure having a shape that binds to a chlorotoxin target, such as Annexin A2 and Calpactin. In some embodiments, the peptide includes a knotted peptide, such as the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants described herein. In certain embodiments, the peptide can include an amino acid sequence that is the same as the amino acid sequence of natural chlorotoxin. The detectable label can, e.g., be conjugated to the N-terminus of the peptide and/or to a lysine residue in the peptide.

[0090] The D-amino acid chlorotoxin peptides can be used for a variety of other applications, such as therapeutic and/or diagnostic applications. In some embodiments, the D-amino acid chlorotoxin and D-amino acid chlorotoxin variants can be used for methods of treating diseases. In some embodiments, the D-amino acid chlorotoxin peptides can be used to deliver drugs via a chlorotoxin-drug conjugate to, e.g., tumors in the brain of a subject.

[0091] The present invention also provides compositions for administering the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants, or conjugates thereof, described herein to a subject to facilitate diagnostic and/or therapeutic applications. In certain embodiments, the compositions can include a pharmaceutically acceptable excipient. Pharmaceutical excipients useful in the present invention include, but are not limited to, binders, fillers, disintegrants, lubricants, coatings, sweeteners, flavors and colors. One of skill in the art will recognize that other pharmaceutical excipients are useful in the present invention. The term "pharmaceutical composition" as used herein includes, e.g., solid and/or liquid dosage forms such as tablet, capsule, pill and the like.

[0092] The D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants, or conjugates thereof, of the present invention may be administered by any suitable technique available in the art, e.g., as compositions. The D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants, or conjugates thereof, of the present invention can be administered as frequently as necessary, including hourly, daily, weekly or monthly. The D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants, or conjugates thereof, utilized in the methods of the invention can be, e.g., administered at dosages that may be varied depending upon the requirements of the method being employed. The D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants, or conjugates thereof, described herein can be administered to the subject in a variety of ways, including parenterally, subcutaneously, intravenously, intratracheally, intranasally, intradermally, intramuscularly, colonically, rectally, urethrally or intraperitoneally. In some embodiments, the pharmaceutical compositions can be administered parenterally, intravenously, intramuscularly or orally. In some embodiments, the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants, or conjugates thereof, can be administered systemically. In some embodiments, the compositions can be administered intratumorally and/or intranodally, such as delivery to a subject's lymph node(s). In certain embodiments, administration can include enteral administration including oral administration, rectal administration, and administration by gastric feeding tube or duodenal feeding tube. Administration can also be including intravenous injection, intra-arterial injection, intra-muscular injection, intracerebral, intracerebroventricular or subcutaneous (under the skin) administration. In some embodiments, administration can be achieved by topical means including epicutaneous (application to skin) and inhalation.

[0093] The oral agents comprising a D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants, or conjugates thereof, described herein can be in any suitable form for oral administration, such as liquid, tablets, capsules, or the like. The oral formulations can be further coated or treated to prevent or reduce dissolution in stomach. The compositions of the present invention can be administered to a subject using any suitable methods known in the art. Suitable formulations for use in the present invention and methods of delivery are generally well known in the art. For example, the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants, or conjugates thereof, described herein can be formulated as pharmaceutical compositions with a pharmaceutically acceptable diluent, carrier or excipient. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions including pH adjusting and buffering agents, tonicity adjusting agents, wetting agents and the like, such as, for example, sodium acetate, sodium lactate, sodium chloride, potassium chloride, calcium chloride, sorbitan monolaurate, triethanolamine oleate, etc.

[0094] As used herein, a "subject" is a human or non-human animal. In some embodiments, a subject can include, but is not limited to, a mouse, a rat, a rabbit, a human, or other animal. In another embodiment, a subject is a human, such as a human having or at risk of having a cancer. In some embodiments, a subject or biological source may be suspected of having or being at risk for having a disease, disorder or condition, including a malignant disease, disorder or condition (e.g., cancer). In certain embodiments, a subject or bio-

logical source may be suspected of having or being at risk for having a hyperproliferative disease (e.g., carcinoma, sarcoma), and in certain other embodiments of this disclosure a subject or biological source may be known to be free of a risk or presence of such disease, disorder, or condition.

[0095] "Treatment," "treating" or "ameliorating" refers to either a therapeutic treatment or prophylactic/preventative treatment. A treatment is therapeutic if at least one symptom of disease (e.g., a hyperproliferative disorder, such as cancer) in an individual receiving treatment improves or a treatment may delay worsening of a progressive disease in an individual, or prevent onset of additional associated diseases (e.g., metastases from cancer).

[0096] A "therapeutically effective amount (or dose)" or "effective amount (or dose)" of a composition including D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants described herein, refers to that amount of compound sufficient to result in amelioration of one or more symptoms of the disease being in a statistically significant manner. When referring to an individual active ingredient, administered alone, a therapeutically effective dose refers to that ingredient alone (e.g., a D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variant described herein). When referring to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered serially or simultaneously (in the same formulation or in separate formulations).

[0097] The term "pharmaceutically acceptable" refers to molecular entities and compositions that do not produce allergic or other serious adverse reactions when administered to a subject using routes well known in the art.

[0098] A "patient in need" or "subject in need" refers to a patient or subject at risk of, or suffering from, a disease, disorder or condition (e.g., cancer) that is amenable to treatment or amelioration with a D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variant described herein.

[0099] In one aspect, the present invention includes methods of treating a disease associated with cells expressing a chlorotoxin target. The methods can include administering, to a subject in need thereof, a therapeutically effective amount of a pharmaceutical composition comprising a D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variant, thereby treating the disease. In some embodiments, the chlorotoxin target can include Annexin A2, Calpactin, or a combination thereof.

[0100] In some embodiments, the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variant can further include other agents to facilitate treatment. For example, a D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants can further include cytotoxic agents (e.g., mitotic inhibitors), toxins, antisense nucleotides, cancer treatment drugs (e.g., alkylating agents), nucleotide drugs, anti-metabolites, metabolic modulators, radiosensitizers, peptide therapeutics, peptide-drug conjugates, radionuclides, or a combination thereof.

[0101] Cytotoxic agents can include drugs that can be used to treat cancer, e.g., by inhibiting cell proliferation. Some example cytotoxic agents can include, e.g., the vinca alkaloids, mitomycins, bleomycins, cytotoxic nucleosides, taxanes, and epothilones. Members of those classes include, for example, doxorubicin, caminomycin, daunorubicin, aminopterin, methotrexate, methopterin, dichloromethotrexate, mitomycin C, porfiromycin, 5-fluorouracil, 6-mercaptopu-

rine, gemcitabine, cytosine arabinoside, podophyllotoxin or podo-phyllotoxin derivatives, such as etoposide, etoposide phosphate or teniposide, melphalan, vinblastine, vincristine, leurosine, vindesine, paclitaxel and therapeutically effective analogs and derivatives of the same. Other useful antineoplastic agents include estramustine, cisplatin, carboplatin, cyclophosphamide, bleomycin, gemcitabine, ifosamide, melphalan, hexamethyl melamine, thiopeta, cytarabine, idatrexate, trimetrexate, dacarbazine, L-asparaginase, camptothecin, CPT-11, topotecan, ara-C, bicalutamide, flutamide, leuprolide, pyridobenzoindole derivatives, interferons and interleukins.

[0102] Suitable metabolic modulators can include, but are not limited to, lonidamine, dichloroacetate, alpha-tocopheryl succinate, methyl jasmonate, betulinic acid, and resveratrol

[0103] Radiosensitizers are known to increase the sensitivity of cancerous cells to the toxic effects of electromagnetic radiation, e.g., x-rays. Examples of x-ray activated radiosensitizers include, but are not limited to, metronidazole, misonidazole, desmethylmisonidazole, pimonidazole, etanidazole, nimorazole, mitomycin C, RSU 1069, SR 4233, EO 9, RB 6145, nicotinamide, 5-bromodeoxyuridine (BUdR), 5-iododeoxyuridine (IUdR), bromodeoxycytidine, fluorodeoxyuridine (FudR), hydroxyurea, cisplatin, and therapeutically effective analogs and derivatives of the same.

[0104] In some embodiments, the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants can include radionuclides and/or complexed radionuclides. Suitable radionuclides can include, but are not limited to, Sc-47, Ga-67, Y-90, Ag-111, In-111, Sm-153, Tb-166, Lu-177, Bi-213, Ac-225, Cu-64, Cu-67, Pd-109, Ag-111, Re-186, Re-188, Pt-197, Bi-212, Bi-213, Pb-212 or Ra-223.

[0105] In certain embodiments, the present invention can include treating diseases, disorders, and/or conditions, such as gliomas, astrocytomas, medulloblastomas, choroid plexus carcinomas, ependymomas, other brain tumors, neuroblastoma, head and neck cancer, lung cancer, breast cancer, intestinal cancer, pancreatic cancer, liver cancer, kidney cancer, sarcomas, osteosarcoma, rhabdomyosarcoma, Ewing's sarcoma, carcinomas, melanomas, ovarian cancer, cervical cancer, lymphoma, thyroid cancer, anal cancer, colo-rectal cancer, endometrial cancer, germ cell tumors, laryngeal cancer, multiple myeloma, prostate cancer, retinoblastoma, gastric cancer, testicular cancer, and Wilm's tumor. In some embodiments, the methods can include treating a disease, disorder and/or condition including a glioma, a skin cancer, a lung cancer, a lymphoma, a medulloblastoma, a prostate cancer, a pancreatic cancer, or a combination thereof. In certain embodiments, the methods can be used to treat breast and mammary cancers, colon, skin, lung, lymphoma, glioma, medulloblastoma, prostate, pancreatic cancers, oral squamous cell carcinoma, and/or hemangiopericytoma.

[0106] The present invention further includes methods of administering a D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variant described herein. For example, in one aspect, the present invention includes a method comprising a step of administering an effective dose of a D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variant described herein or a composition including a D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variant described herein to a subject with a tumor such that the peptide selectively targets tumor tissue over normal tissue.

[0107] The methods can further include facilitating surgical removal of cancerous tissue (e.g., a tumor) in a subject.

For example, the present invention can include a method comprising administering an effective dose of a D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variant described herein or a composition including a D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variant described herein to a subject with cancerous tissue (e.g., a tumor) such that the peptide selectively targets cancerous tissue (e.g., tumor tissue) over normal tissue. The methods can include imaging the cancerous tissue by, e.g., detecting the tissue that shows elevated binding of the peptides, thereby indicating the location of the cancerous tissue. Identification of the location can provide a step of surgically removing the cancerous tissue from the subject. The surgically removing can include, e.g., intraoperative visualization of the cancerous tissue as identified by binding of the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants described herein.

[0108] The present invention also provides compositions for administering the D-amino acid chlorotoxin and D-amino acid chlorotoxin variants described herein to a subject to facilitate diagnostic and/or therapeutic applications. In certain embodiments, the compositions can include a pharmaceutically acceptable excipient. Pharmaceutical excipients useful in the present invention include, but are not limited to, binders, fillers, disintegrants, lubricants, coatings, sweeteners, flavors and colors. One of skill in the art will recognize that other pharmaceutical excipients are useful in the present invention. The term "pharmaceutical composition" as used herein includes, e.g., solid and/or liquid dosage forms such as tablet, capsule, pill and the like.

[0109] The D-amino acid chlorotoxin and D-amino acid chlorotoxin variants of the present invention can be administered as frequently as necessary, including hourly, daily, weekly or monthly. The D-amino acid chlorotoxin and D-amino acid chlorotoxin variants utilized in the methods of the invention can be, e.g., administered at dosages that may be varied depending upon the requirements of the method being employed. The D-amino acid chlorotoxin and D-amino acid chlorotoxin variants described herein can be administered to the subject in a variety of ways, including parenterally, subcutaneously, intravenously, intratracheally, intranasally, intradermally, intramuscularly, colonically, rectally, urethrally or intraperitoneally. In some embodiments, the pharmaceutical compositions can be administered parenterally, intravenously, intramuscularly or orally. In some embodiments, the compositions can be administered intratumorally and/or intranodally, such as delivery to a subject's lymph node(s). In certain embodiments, administration can include enteral administration including oral administration, rectal administration, and administration by gastric feeding tube or duodenal feeding tube. Administration can also be including intravenous injection, intra-arterial injection, intra-muscular injection, intracerebral, intracerebroventricular or subcutaneous (under the skin) administration.

[0110] The oral agents comprising a drug candidates (e.g., peptides) described herein can be in any suitable form for oral administration, such as liquid, tablets, capsules, or the like. The oral formulations can be further coated or treated to prevent or reduce dissolution in stomach. The compositions of the present invention can be administered to a subject using any suitable methods known in the art. Suitable formulations for use in the present invention and methods of delivery are generally well known in the art. For example, the drug candidates (e.g., peptides) described herein can be formulated as

pharmaceutical compositions with a pharmaceutically acceptable diluent, carrier or excipient. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions including pH adjusting and buffering agents, tonicity adjusting agents, wetting agents and the like, such as, for example, sodium acetate, sodium lactate, sodium chloride, potassium chloride, calcium chloride, sorbitan monolaurate, triethanolamine oleate, etc.

[0111] The present invention further includes functional assays of the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants. The capacity of D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants and conjugates thereof, to bind to tumor or cancerous tissue can be assayed by in vitro binding, ex vivo imaging, animal models, and other assays known in the art and as previously described. See, for example, US Patent Publication Number US20080279780 and WO 2011/142858, both of which are incorporated by reference herein for the description of functional assays to detect and measure binding to tumor cells and tumor tissue.

[0112] One skilled in the art will be knowledgeable about animal models that are useful for measuring the in vivo activity of the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants, and conjugates thereof. For example, the National Cancer Institute maintains a database of specific cancer models. See the "Cancer Models Database" at the National Cancer Institute website. All animals are handled in strict accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals. ND2:SmoA1 medulloblastoma mice, TRAMP prostate cancer mice and Apc^{1638N} intestinal adenoma and adenocarcinoma mice have been previously described. See, Fodde, R., et al., A targeted chain-termination mutation in the mouse Apc gene results in multiple intestinal tumors. *Proc. Natl. Acad. Sci. U.S.A.*, 1994. 91 (19): p. 8969-73; Greenberg, N. M., et al., Prostate cancer in a transgenic mouse. *Proc. Natl. Acad. Sci. U.S.A.*, 1995. 92 (8): p. 3439-43; Kaplan-Lefko, P. J., et al., Pathobiology of autochthonous prostate cancer in a pre-clinical transgenic mouse model. *Prostate*, 2003. 55 (3): p. 219-37; Hallahan, A. R., et al., The SmoA1 mouse model reveals that notch signaling is critical for the growth and survival of sonic hedgehog-induced medulloblastomas. *Cancer Res.*, 2004. 64 (21): p. 7794-800; each expressly incorporated herein by reference in its entirety.

[0113] One example of a mouse model of cancer is the autochthonous mouse model of medulloblastoma, ND2:SmoA1. See A. R. Hallahan, et al., "The SmoA1 Mouse Model Reveals That Notch Signaling is Critical for the Growth and Survival of Sonic Hedgehog-Induced Medulloblastomas," *Cancer Research* (54:7794-7800, 2004), and B. A. Hatton, et al. "The Smo/Smo Model: Hedgehog-Induced Medulloblastoma With 90% Incidence and Leptomeningeal Spread," *Cancer Research* (58: 1768-1776, 2008). A C57M/6 background is used and the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants, and conjugates thereof is or are administered to evaluate the preferential binding of the peptides or peptide conjugates to cancerous tissue over normal tissue. Hemizygous or homozygous (referred as ND2:SmoA1) mice with symptomatic medulloblastoma are selected for enrollment in these studies. Symptoms are detected using an open field cage evaluation. Symptoms may include head tilt, hunched posture, ataxia, protruding skull, and weight loss. Positive results in this assay would include a

lessening, diminishment, reduction or inhibition of these symptoms by administration of the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants and conjugates thereof.

[0114] Such assays can be used to show the effects of the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants, and conjugates thereof in a variety of cancers, including, for example, cancerous tissue associated with gliomas, astrocytomas medulloblastomas, choroids plexus carcinomas, ependymomas, other brain tumors, neuroblastoma, head and neck cancer, lung cancer, breast cancer, intestinal cancer, pancreatic cancer, liver cancer, kidney cancer, sarcomas, osteosarcoma, rhabdomyosarcoma, Ewing's sarcoma, carcinomas, melanomas (including amelanotic melanoma), ovarian cancer, cervical cancer, lymphoma, thyroid cancer, anal cancer, colo-rectal cancer, endometrial cancer, germ cell tumors, laryngeal cancer, multiple myeloma, prostate cancer, retinoblastoma, gastric cancer, testicular cancer, and Wilm's tumor. In some embodiments, the cancerous tissue can be associated with a glioma, a skin cancer, a lung cancer, a lymphoma, a medulloblastoma, a prostate cancer, a pancreatic cancer, or a combination thereof. In some embodiments, the cancerous tissue can be associated with breast and mammary cancers, colon, skin, lung, lymphoma, glioma, medulloblastoma prostate, pancreatic cancers, oral squamous cell carcinoma, and/or hemangiopericytoma.

[0115] The capacity of the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants, and conjugates thereof to target a detectable label to cancerous or tumor tissue preferentially over normal tissue can be assayed by in vitro binding, ex vivo imaging, animal models, and other assays known in the art and as described above. For example, the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants, can be conjugated to a detectable label, as described herein, to produce a D-amino acid chlorotoxin conjugate or D-amino acid chlorotoxin variant conjugate. An appropriate amount of the D-amino acid chlorotoxin conjugate or D-amino acid chlorotoxin variant conjugate can be injected into the tail vein of mice that show clinical signs consistent with advanced tumors. After three days the mice can be sacrificed and their brains can be imaged using a biophotonic imaging system, such as the Caliper/Xenogen Spectrum or the Odyssey Near-Infrared imaging system. A positive result will show that the peptide conjugate preferentially illuminates or identifies the cancer tissue compared with normal tissue. In an embodiment, the preferential binding of the D-amino acid chlorotoxin conjugate or D-amino acid chlorotoxin variant conjugate to the cancerous tissue over the normal tissue is substantially the same as that of the native chlorotoxin conjugate. In another embodiment, the preferential binding of the D-amino acid chlorotoxin conjugate or D-amino acid chlorotoxin variant conjugate to the cancerous tissue over the normal tissue is less than that of the native chlorotoxin conjugate, but more than that of normal tissue.

[0116] One skilled in the art will be knowledgeable about ex vivo models that are useful for measuring the activity of a D-amino acid peptides, including D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variant, and conjugates thereof described herein. For example, for detection of medulloblastoma, ND2:SmoA1 animals exhibiting symptoms of medulloblastoma are injected with the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants or conjugates through the tail vein. Mice are euthanized using CO₂ inhalation three days after injection and ex vivo biophotonic

images of their brain are obtained using the Xenogen Spectrum Imaging System (Caliper). The brains are then frozen in Tissue-Tek Optimal Cutting Temperature (OCT) Compound (Sakura), sliced in 12 μ m sections and Hemotoxylin and Eosin (H&E) stained according to standard procedures. A positive result will show that the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants, or conjugates thereof preferentially illuminates or identifies the cancer tissue compared with normal tissue. In an embodiment, the preferential binding of the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants, or conjugates thereof to the cancerous tissue over the normal tissue is substantially the same as that of the native chlorotoxin or native chlorotoxin conjugate. In another embodiment, the preferential binding of the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variants, or conjugates thereof to the cancerous tissue over the normal tissue is less than that of the native chlorotoxin or native chlorotoxin conjugate, but more than that of normal tissue.

[0117] Another example relates to prostate tissue imaging. Normal and cancerous human prostate tissue samples are collected and handled in accordance with Human Subjects IRB approved protocols. Sections are incubated with an amount of the D-amino acid chlorotoxin conjugate and/or D-amino acid chlorotoxin variant conjugate, in 5% normal goat serum buffer for 45 minutes. Unbound conjugate is reduced by washing slides 3 times for 5 minutes in PBS buffer. Signal is detected by fluorescence microscopy and correlated to adjacent sections stained with hematoxylin and eosin (H&E). A positive result will show that the D-amino acid chlorotoxin conjugate or the D-amino acid chlorotoxin variant conjugate preferentially illuminates or identifies the cancer tissue compared with normal tissue. In an embodiment, the preferential binding of the D-amino acid chlorotoxin conjugate or D-amino acid chlorotoxin variant conjugate to the cancerous tissue over the normal tissue is substantially the same as that of the native chlorotoxin conjugate. In another embodiment, the preferential binding of the D-amino acid chlorotoxin conjugate or D-amino acid chlorotoxin variant conjugate to the cancerous tissue over the normal tissue is less than that of the native chlorotoxin conjugate, but more than that of normal tissue.

[0118] One skilled in the art will be knowledgeable about in vitro models that are useful for measuring the activity of the D-amino acid chlorotoxin and/or D-amino acid chlorotoxin variant, and conjugates thereof described herein. For example, 9L/lacZ gliosarcoma cells (ATCC, VA) and primary human foreskin fibroblast (HFF) are maintained in DMEM and RPMI medium both supplemented with 1% sodium pyruvate, 1% streptomycin/penicillin and 10% FBS (Hyclone, UT), respectively. 2×10^5 cells are seeded on sterile cover slips 36 hrs prior to labeling and confocal microscopy. Cells are cultured with 1 ml of CTX:CY5.5 conjugate (1 μ M) for 2 hours in a 37° C. humidified incubator maintained at 5% CO₂. Cover slips are washed 2 times in cell culture medium and 2 times in PBS buffer. Following this step, cell membranes are stained with 1 μ M solution of FM 1-43FX (Molecular Probes, OR) for 20 min in dark at room temperature, washed 2 times in PBS and fixed in 4% paraformaldehyde. Cellular nuclei are stained with 4',6-diamidino-2-phenylindole (DAPI, Sigma Aldrich, MO). Confocal images are acquired using a DeltaVision SA3.1 Wide-Field Deconvolution Microscope (Applied Precision, WA) equipped with DAPI, TRITC, and Cy5 filters. Image processing is performed using SoftWoRX (Applied

Precision, WA). A positive result will show that the D-amino acid chlorotoxin or the D-amino acid chlorotoxin variant, or conjugates thereof preferentially illuminates or identifies the cancer tissue compared with normal tissue. In an embodiment, the preferential binding of the D-amino acid chlorotoxin or D-amino acid chlorotoxin variant, or conjugates thereof, to the cancerous tissue over the normal tissue is substantially the same as that of the native chlorotoxin or native chlorotoxin conjugate. In another embodiment, the preferential binding of the D-amino acid chlorotoxin or D-amino acid chlorotoxin variant, or conjugates thereof, to the cancerous tissue over the normal tissue is less than that of the native chlorotoxin or the native chlorotoxin conjugate, but more than that of normal tissue.

[0119] Another model includes the use of subcutaneous xenografts established in nu/nu (nude) mice using 9L, a rat gliosarcoma cell line (ATCC), and RH30, a rhabdomyosarcoma cell line. The xenografts are established using 1 million 9L or RH30 cells suspended in serum free media and matrigel at a 1:1 ratio. Intracranial xenografts are established by stereotaxic injection of 1 million 9L cells suspended in 10 μ l PBS into the brain 3 mm lateral and posterior to the bregma.

Methods of Making Libraries

[0120] In yet another aspect, the present invention includes methods for making potential drug candidates. For example, methods that are generally well known in the art can be used to make small molecule libraries and larger biologics (e.g., antibody) libraries.

[0121] In some embodiments, the present invention includes methods of making peptides for the generated peptide libraries. A wide variety of methods are suitable for making the peptide libraries. As described further herein, the present invention includes scaffolds that can be used as a starting point for generating peptide libraries. These scaffolds as well as a large diversity of scaffold variants can be made using several different approaches. In some aspects, the peptide libraries can be produced using peptide synthesis techniques generally well known in the art. Conventional oligonucleotide synthesis techniques (e.g., chip-based oligonucleotide synthesis) can also be used. In some instances, the synthetic approaches can be combined with a variety of expression systems. In one example embodiment, particular residue positions in a scaffold can be targeted for random mutagenesis using degenerate codons to generate a diverse set of DNAs that can be made using, e.g., chip-based oligonucleotide synthesis and can code for a large library of scaffold variants.

[0122] In some embodiments, the molecules coding for the scaffolds and scaffold variants can be expressed in various expression systems, and can, in some embodiments, be combined as part of a fusion system. The DNA molecules encoding the scaffolds and scaffold variants, e.g., can be combined with fusion systems that can be expressed in several different cell types, e.g., 293 HEK or *E. coli*. Fusions for 293 HEK cells, e.g., can include but are not limited to, IgK leader sequences and/or secreted fusion proteins, such as siderocalin, lipocalin 2, and human serum albumin.

[0123] In some embodiments, the peptides described herein (e.g., knotted peptides) can be expressed as fusions with lipocalin proteins. In one aspect, the present invention includes a method for producing a peptide that can include expressing, in a cell, a fusion protein including a peptide (e.g., a knotted peptide) and a lipocalin protein. The method can

further include separating the peptide from the lipocalin protein, thereby producing the peptide (e.g., the knotted peptide). The present invention further includes compositions of the fusion protein including the lipocalin protein and the peptide (e.g., the knotted peptide). This fusion system offers a variety of advantages for producing peptides (e.g., knotted-peptides) over traditional fusion systems. By way of background, and not to be limiting in any way, the lipocalins are a class of proteins that can have a conserved fold characterized by an eight-stranded beta barrel with a flanking alpha helix. The expression levels of lipocalin proteins, like Lcn2, NGAL and Siderocalin, in mammalian cells equal or surpass many other fusion systems, including Fc fusions.

[0124] The peptides described herein (e.g., knotted peptides) can be expressed using a variety of lipocalin proteins. In some embodiments, siderocalin is used as a secretion partner. Siderocalin is useful as a fusion partner larger proteins because, e.g., of the small size of siderocalin relative to larger proteins (the mature protein is 178 amino acids and has a molecular weight of 20547 Da). Also, a C87S mutation in siderocalin can prevent dimerization and yields pure monomeric fusion protein. A single intramolecular disulfide bond present in siderocalin increases its stability. Also, siderocalin only has a single N-linked glycosylation site, which involves correct processing in the ER before secretion. In some embodiments, the peptides can also be expressed as fusion peptides with Murine Lcn2 (also known as 24p3), which also works very well as a secretion partner. Other homologs can also be used. In addition, the peptides (e.g., knotted peptides) provided herein can also be expressed as fusion systems with the other members of the lipocalin family including Lcn1, Lcn6, Lcn8, Lcn9, Lcn10, Lcn12, Lcn15.

[0125] In some embodiments, the expression of peptides (e.g., knotted peptides) as fusions with Lcn2 can be utilized with a self-cleaving Lcn2, with RARYKR right after the CIDG, and an exogenously cleaved one, with ENLYFQ in that position. The former can be cleaved by the mammalian cells during protein export (e.g., by furin), and the free Lcn2 and knotted peptide can be secreted into surrounding media. ENLYFQ is a tobacco etch protease site, which is not found endogenously in mammalian cells. The constructs in this system can be secreted as fusions, allowing for the knotted peptide to be cleaved off later by adding exogenous TEV protease. This can be useful for recovering the knottins. In some embodiments, purification “handles” such as poly-histidine or poly-arginine can be added to the Lcn2 and subsequently removed by proteolysis. In addition to the knotted peptides, these fusion systems can also be used for difficult-to-express proteins of medical interest such as chemokines, interleukins, and peptide hormones.

[0126] The lipocalin fusions (e.g., siderocalin and/or Lcn2 fused with a knotted peptide) can be used in several ways different ways. It could be used to increase the size of the target protein (for example a potential therapeutic) in order to increase its half-life. It could be used to secrete the target protein where the target protein is naturally expressed in the cytoplasm. Lcn2 also has unique ligand specificity and tightly binds catecholate siderophores (bacterial iron chelators). This opens the possibility of loading the Lcn2 fusion with specific ligands, such as a chemotherapeutic or radioactive reagent or some type of a compound that has beneficial properties. Lcn2, when loaded with siderophores and iron, has a deep red color that can aid in chromatography or other purification steps.

[0127] In addition to several other advantages, the lipocalin fusion systems can be used to make large amounts of protein over relatively short time frames. In some embodiments, the amount of peptide obtained can be less than about 10 mg/L, less than about 20 mg/L, less than about 40 mg/L, less than about 50 mg/L, less than about 100 mg/L, less than about 150 mg/L, less than about 180 mg/L, or less than about 200 mg/L. In some embodiments, the amount of peptide obtained can be between about 10 mg/L and 200 mg/L, between about 50 mg/L and 200 mg/L, between about 100 mg/L and 200 mg/L, and between about 150 mg/L and 200 mg/L.

[0128] In other embodiments, some of the peptides described herein can be expressed in a variety of ways known in the literature. For example, the peptides are expressed in bacterial systems including *E. Coli*, corynebacterium, and pseudomonas fluoresceins. Expression platforms for *E. coli* can include periplasmic expression or cytoplasmic expression. For periplasmic expression, fusions can include pelB, dsbA, and ExFABP fusion. Cytoplasmic fusions can include Small Ubiquitin-like Modifier (SUMO) fusions. The peptides can also be expressed in yeast systems including *Saccharomyces cerevisiae* and *Pichia pastoris*. The peptides can also be expressed in insect cell systems and eukaryotic systems including plant and mammalian systems.

[0129] In some aspects, the peptides disclosed herein can be synthesized by transfection, a technique that involves introduction of foreign DNA into the nucleus of the eukaryotic cells. In some aspects, the peptides can be synthesized by transient transfection (DNA does not integrate with the genome of the eukaryotic cells, but the genes are expressed for 24-96 hours). Various methods can be used to introduce the foreign DNA into the host cells, and transfection can be achieved by chemical-based means including by the calcium phosphate, by dendrimers, by liposomes, and by the use of cationic polymers. Non-chemical methods of transfection include electroporation, sono-poration, optical transfection, protoplast fusion, impalefection, and hydrodynamic delivery. In some embodiments, transfection can be achieved by particle-based methods including gene gun where the DNA is coupled to a nanoparticle of an inert solid which is then “shot” directly into the target cell’s nucleus. Other particle-based transfection methods include magnet assisted transfection and impalefection.

[0130] DNA can also be introduced into cells using virus as a carrier (viral transduction) using reteroviruses or lentiviruses. In some embodiments, the peptides of the present invention can be prepared using a Daedalus expression system. Ashok D. Bandaranayake et al., *Nucleic Acids Res.* 2011 November; 39 (21): e143, which is incorporated herein by reference in its entirety. This technique may also be combined with a serum free mammalian culture system. And, it is also possible to express tagless proteins, which can be purified in a single size exclusion step directly from the media, at high levels.

[0131] In one aspect, the present invention provides a method of making hundreds to thousands or more of peptide variants at high levels. Conventional methods of making knotted peptides can be limited in that activity of knotted peptides can depend on proper folding of the peptides. There has been limited success in making knotted peptides that fold properly during manufacture. The present invention overcomes these problems with other techniques known in the art. FIG. 4 shows an example method for making the peptide libraries of the present invention. As shown, viruses can be

produced by packaging of specific oligonucleotides sequences, transferring the sequences to the viruses, and expressing the peptides. Recovery and scale up of the peptides can be conducted, and then the sample can be purified and assayed. The process can be conducted efficiently (e.g., in three weeks) and large amounts of peptide can be produced (e.g., 200 mg/liter). In some instances, purification by chromatography may not be needed due to the purity of manufacture according to the methods described herein.

[0132] In an example embodiment, the present invention includes fusion proteins of a knotted peptide fused to siderocalin via a cleavable linker. FIG. 5 shows an example fusion system that can be used to make the knotted peptide libraries. As shown, the fusion system includes a sequence including an IgK SP, sFLAG, HIS, siderocalin, TEV, and the knotted peptide sequence of interest. In some embodiments, these fusions can be combined with the Daedalus expression systems. Ashok D. Bandaranayake et al., *Nucleic Acids Res.* 2011 November; 39 (21): e143, which is incorporated herein by reference in its entirety. A lentivirus can be used to gain rapid, stable expression in HEK293 cells, a human kidney cell line. The siderocalin can be highly expressed in this system and, e.g., serves to help the knotted peptide to be expressed as well. The nature of the cleavable linker allows the fusion to be cleaved as the protein is being expressed or later via an exogenously added protease. The siderocalin fusion partner can, e.g., be a generalizable expression enhancement system for any difficult-to-express protein, can be used as a tag to increase the size of a smaller peptide, and/or to improve a peptide's serum half-life (e.g., by increasing the size of the final fusion protein above the glomerular filtration limit).

[0133] Although HEK293 cells are robust and used for general protein expression, the lentivirus can infect a wide variety of cells. Combining this with a system that allows proteins to be cleaved as they are expressed enables a set of powerful assays that rely upon the secreted peptide to act in an autocrine or paracrine manner (i.e. they act on the cell that is secreting them or on nearby cells). An example of this would be to infect cancer target cells with a library of peptide-expressing lentiviruses and then screen those cells by flow cytometry for those that showed signs of apoptosis (e.g. Annexin V expression). The cells showing signs of apoptotic stress could be sorted out and the viruses sequenced, essentially looking for cells that were expressing a peptide that was inducing apoptosis in an autocrine fashion. A related set of screens could be done in a diffusion-limited matrix (e.g., soft agar), where peptide-expressing cells were mixed with target cells and the agar limited diffusion of the peptide. Areas of target cell death would be an indication of an active secreted peptide. Screens done in this manner could employ very large libraries, as the deconvolution would be as simple as sequencing the gene from which the peptide came.

[0134] In some embodiments, the present invention can include methods for producing knottins such that the knottin protein can remain tethered to the surface of the mammalian cell for use in conventional binding screens (e.g., those in which the target molecule is tethered to a column or beads and candidate drugs are identified by affinity to the target). In contrast to other known methods (e.g., phage or yeast display), the methods described herein use fusion systems (e.g., a siderocalin system of the present invention) to express libraries of peptides that have been designed according to the "rules" described above (e.g., ratio of acid/basic amino acids in a peptide) and that can be established through the in vivo

drug discovery process and/or that have already been pre-screened for specific biophysical and pharmacological properties. In these methods, e.g., all DNA sequences and protein products are already known and have already been validated (e.g., the peptides all fold properly and have improved serum half lives). The methods of present invention are in direct contrast to other known display technologies where the displayed proteins are not known and previously validated, and instead have their sequences randomized (using mutagenic oligonucleotides and degenerate NNN codons) yielding libraries of immense size (generally greater than 10^7), where many of the proteins do not fold properly due to deleterious mutations.

Methods of Analyzing Libraries

[0135] The generated libraries provided by the present invention can be analyzed using a variety of methods and used to determine information for a range of applications.

[0136] In one aspect, the present invention includes generating peptide libraries in which some or all of the peptides have a unique mass signature or digest fragment signature (e.g., by trypsin).

[0137] As shown in FIG. 6, a generated peptide library can be analyzed using mass spectrometry to individually identify all the peptides in the library. Each of the peptides can have, e.g., a unique tryptic fragment. Visualization can be obtained for the peptides in several charge states, and MS/MS makes identity unambiguous, thereby inherently adding quality control into the process.

[0138] In some embodiments, pooled, mass-defined libraries of peptides can be used for various applications. Using methods disclosed herein (e.g., chip-based oligonucleotides), libraries of thousands of peptides can be completely defined and used to create libraries in which every member has a unique mass signature or digest fragment mass signature. In some instances, the identity of each peptide can be coded by its mass, allowing for assays, e.g., that cannot be achieved with display methods that encumber the peptide with a bulky identifier. For example, it is not feasible to use display methods to find peptides that cross the blood-brain barrier or interact with intracellular targets due to a lack of cell penetration, as the sheer size of the element (phage, yeast, RNA) displaying the peptide dominates the behavior of the combination. Since the peptides contain little or nothing extra (except, possibly, biotin, dyes, tags or other minor modifications), their behavior is a true measure of the qualities of the peptide itself. This feature is a particular advantage in in vivo systems that can be used to simultaneously test thousands of peptides for oral bioavailability, serum half-life, blood brain barrier penetration, and homing to particular organs or tumors.

[0139] In one aspect, the present invention can include methods of generating mass-defined drug candidate libraries. The methods can include, e.g., producing a plurality of drug candidates, at least some of the drug candidates each having a unique mass signature or digest fragment mass signature. The methods can also include analyzing the plurality of drug candidates using mass spectrometry to measure the unique mass signature or digest fragment mass signature for the at least some of drug candidates. The methods can further include generating a mass-defined drug candidate library comprising the at least some of the plurality of drug candidates, the drug candidate library being generated based on a pharmacological property, wherein the identity of the drug

candidates in the mass-defined drug candidate library can be determined with the unique mass signature or digest fragment mass signature of each of the drug candidates. Pharmacological properties can include, e.g., oral bioavailability, capability to pass the blood-brain barrier, exclusion by the blood-brain barrier, serum half-life, capability to penetrate cells, capability to enter subcellular organelles or other cellular domains, or a combination thereof.

[0140] In some aspects, analysis of large peptide libraries by mass spectroscopy can include analyzing a large number of peptide and/or peptide fragments that can have similar molecular weights, thereby resulting in a relatively crowded mass spectrum that can hinder the unambiguous identification of peptides. In some embodiments, the libraries being analyzed can be divided into sub-libraries having drug candidates (e.g., peptides) that are disparate in mass such that the mass spectrum is less crowded and allowing more efficient identification of the unique mass signatures.

[0141] In some embodiments, the libraries of peptides having unique mass signatures could also be generated by replacing one or more atoms in the scaffold by an isotope having an atomic mass or mass number different from the atomic mass or mass number usually found in nature. In some embodiments, the masses of the peptides in the library can be modified by addition of an isotopically labeled conjugate or tag described further herein. Examples of isotopes that can be incorporated into the present peptide libraries include isotopes of nitrogen, oxygen, hydrogen, carbon, fluorine and chlorine, bromine, sulfur and silicon. For example the peptides of the libraries generates herein differ from each other by least one ^{15}N atom, at least one ^{17}O atom, at least one ^{18}O atom, at least one ^2H atom, at least one ^3H atom, at least one ^{13}C , ^{14}C , ^{18}F , ^{37}Cl , at least one ^{76}Br , at least one ^{33}S atom, ^{34}S atom, ^{36}S atom, at least one ^{29}Si atom, or at least one ^{30}Si atom.

[0142] In some embodiments, the peptides and/or peptide conjugates of the generated libraries can differ from each other by isotopic substitution of at least 1 atom, at least 2 atom, at least 3 atom, at least 4 atoms, at least 5 atoms, at least 6 atoms, at least 7 atoms, at least 8 atoms, at least 9 atoms, or at least 10 atoms. In some embodiments greater than about 10, greater than about 15, greater than about 20, greater than about 25, greater than about 30, greater than about 35, greater than about 40, greater than about 45, greater than about 50 atoms can be substituted by isotopic atoms. In some embodiments, the peptides and/or peptide conjugates can differ by isotopic substitution of all ^1H , ^{12}C , ^{14}N , ^{16}O , or ^{32}S , or any combination thereof.

[0143] The peptide libraries of the present invention can include peptides in which some or all of the peptides have a unique mass signature. The peptide libraries of the invention can be analyzed by MS or tandem mass (MS/MS) analysis. A variety of mass spectrometry methods can be employed in the methods of invention for identifying and/or quantifying the peptides. Different ionization methods specially the soft ionization techniques like electrospray ionization (ESI), matrix assisted laser desorption/ionization (MALDI), desorption/ionization on silicon (DIOS), flat atom bombardment (FAB) and liquid secondary ion mass spectroscopy (LSIMS) can be used for examination of the generated peptide libraries. The ionization techniques can be combined in various ways with different mass analyzers including quadrupole mass analyzers, time of flight (TOF) analyzers, quadrupole ion traps, hybrid quadrupole—TOF and Fourier transform ion cyclo-

tron resonance (FTICR) MS. In some embodiments the peptide libraries can be analyzed using liquid chromatography (e.g., by integrated liquid-chromatography mass spectrometer systems (LC-MS) or LC-MS/MS) and/or other separation techniques

[0144] Using the methods of the present invention, scaffold variant libraries with wide diversity (e.g., thousands) can be produced. Moreover, each of the peptides can be designed to have a unique mass signature or digest fragment mass signature that can be used to screen and individually identify the peptides in the library having at least one of the following pharmacological properties: oral bioavailability, capability to pass the blood-brain barrier, serum half-life, capability to penetrate cells, or capability to enter subcellular organelles or other cellular domains. Conjugates can be employed for mass based analysis of the peptide libraries, as well. In some embodiments, conjugates can be used to label a library with two identical hydrophobic moieties that can be discriminated by mass spectrometry. For example, conjugates having differing isotopes can be used to produce peptides having different mass but, e.g., the same or different amino acid sequence. In some embodiments, conjugates can be added having at least one ^{13}C atom. These heavier conjugates can be otherwise identical to conjugates only having ^{12}C atoms, thereby making a difference that can be detected, e.g., by mass spectrometry. For example, two identical libraries can be administered to a subject; one library can be N-terminally labeled with palmitic acid (a light peptide), and the other can be N-terminally labeled with palmitic acid containing five ^{13}C atoms (a heavy peptide). The library with the ^{13}C atoms will be about five atomic mass units heavier but since it is chemically identical it can have the same ionization properties (peak height) in mass spectrometry. Material recovered from a tissue sample could be doped with some of the heavy library and the 5+ shifted peaks used to quantitate the relative amount of each unshifted peak. This is important because generally peak heights in mass spectrometry are not quantitative but instead reflect the ionizability of a peptide.

[0145] In certain aspects, the present invention can utilize the analysis of light and heavy peptides by a variety of ways. For example, the present invention includes methods of determining a distribution profile of knotted-peptides administered to a subject by different administration pathways. The methods can include, e.g., administering to the subject a light knotted-peptide, the light knotted-peptide being administered by a first route of delivery and having a lower molecular weight than a heavy knotted-peptide having the same sequence as the light knotted-peptide. The methods can also include administering to the subject the heavy knotted-peptide, the heavy knotted-peptide being administered by a second route of delivery that is different than the first route of delivery. The methods can further include comparing a quantity of the light knotted-peptide to a quantity of the heavy knotted-peptide obtained from a tissue or fluid sample of the subject, thereby determining the distribution profile of the light and heavy knotted-peptides in the subject based on the first and second routes of delivery, respectively.

[0146] In some embodiments, the conjugates used for analysis of the peptide libraries can differ from each other by least one ^{15}N atom, at least one ^{17}O atom, at least one ^{18}O atom, at least one ^2H atom, at least one ^3H atom, at least one ^{13}C , ^{14}C , ^{18}F , ^{37}Cl , at least one ^{76}Br , at least one ^{33}S atom, ^{34}S atom, ^{36}S atom, at least one ^{29}Si atom, or at least one ^{30}Si atom. In various embodiments, the conjugates attached to the

peptides of the generated libraries can differ from each other by an isotopic substitution of at least 1 atom, at least 2 atom, at least 3 atom, at least 4 atoms, at least 5 atoms, at least 6 atoms, at least 7 atoms, at least 8 atoms, at least 9 atoms, or at least 10 atoms. In some embodiments, the present invention can include methods that can provide a quantitative assessment of peptides that survive experimental testing (e.g., BBB penetration, oral bioavailability). For these methods, an isotope doping strategy can be employed. Relative masses of the various peptides can be modified by a variety of ways. For example, the present invention includes use of moieties that contain several different isotopic biases in order to compare the relative abundance of peptides in more than two samples. It is also possible to use isotopes other than carbon (e.g., hydrogen/deuterium). In one example, a library of peptides can be split into two parts: some peptides can be conjugated to a moiety of normal isotopic distribution (e.g., mostly or all ^{12}C atoms) while other peptides can be conjugated to an identical but isotopically heavier moiety (e.g., containing a known number of ^{13}C atoms). These two pools of light and heavy peptides can be nearly identical from a chemical perspective (e.g., identical chromatographic elution time), but the ^{13}C containing pool of peptides will be mass shifted to heavier mass. Peak heights in mass spectrometry can be a function not only of peptide abundance but also the ionizability of the peptides. As such, the heavy versions of each peptide can possess the same ionizability as their normal, lighter counterparts. By doping a normal (e.g., mostly or all ^{12}C atoms) peptide sample recovered in an experiment (e.g., from an animal's blood) with a portion of the heavy library just before LC/MS, the mass shifted heavy library will form a set of reference peaks to which to compare the recovered sample. These reference peaks will appear in the same MS plot but will be shifted heavier according to the number of ^{13}C atoms in the conjugate. Because the heavy and normal (light) peptides are chemically similar, they will elute at the same time from the LC (liquid chromatography) column and have the same ionizability. Relative peak heights from the heavy and normal samples will then reflect accurately the relative abundance of the peptides in the two samples. In some embodiments, the peptides can further include moieties to facilitate purification from tissue or fluid from a subject. For example, biotin can be conjugated to the peptides to aid recovery, or fluorescent or "warhead" conjugates could be recovered with an antibody to the conjugate. In certain embodiments, the peptides can also include therapeutic molecules, such as a cytotoxic agent and/or a toxin. In some embodiments, the peptides can also include candidate drugs and/or candidate linkers.

[0147] Conjugation can be done by a variety of ways known in art. In some examples, the conjugates are added to the N-terminal of the peptide libraries. N-conjugation can be achieved by reaction of the peptide libraries with amine reactive esters including succinimidyl esters. In some cases, conjugation can be done by using pH to selectively conjugate to the N-terminus without hitting internal lysine residues in the peptides. In some embodiments, the peptides can be designed so that they do not contain conjugatable lysine residues. In some cases, conjugation of peptides with N-terminal serine or threonine residues can be achieved via oxidative coupling. In some embodiments, the conjugation can be accomplished by conjugating a moiety to a lysine in the peptides in the peptide libraries.

Methods of Using Libraries

[0148] The libraries generated and produced by the methods and systems described herein can be used for a range of applications. For example, the potential drug candidate libraries can be used for therapeutic and/or diagnostic purposes. Some example uses include, but are not limited to, conjugating the peptides to radiolabels and/or fluorescent molecules for bioimaging, linking the peptides to cytotoxic agents, using the peptides for in vitro diagnostics for biochemical assays, as well as, e.g., for veterinary uses, insecticides, antibiotics, herbicides, antifreeze compositions, and antivenoms.

[0149] As will be appreciated by one of ordinary skill in the art, the generated peptide libraries described herein can be tailored for a wide range of targets (e.g., therapeutic targets). In some embodiments, the targets are associated with a variety of diseases or disorders. Some targets, for example, can include but are not limited to glypican-2 (GPC2), protocadherin (1 α (PCDHA1), Ca $_v$ 2.2, K $_v$ 1.3, Na $_v$ 1.2, NaV1.1, NaV1.7, NaV1.8, CIC-3, nAChR, NMDA-R, NPRA, GLP-1R, α_{1B} -AR, NT-R-1, ACE, NET mTor, cMet, VEGF/VEGFR, c-Kit, PDGF/PDGFR, PI3K, HER2, EGFR, Orai1, CD47, Raf, NF κ B, Bromodomains, HATS, HDAC, LDH, IDH2, CD22, MIC, c-Myc, n-Myc, PHF5A, BUB1B, Bcl-2, k-Ras, Notch1, p53, $\alpha 5\beta 3$, NKG2D, CTLA4/CD28, and/or Mcl-1.

[0150] The present invention also provides compositions for administering the drug candidates described herein to a subject to facilitate diagnostic and/or therapeutic applications. In certain embodiments, the compositions can include a pharmaceutically acceptable excipient. Pharmaceutical excipients useful in the present invention include, but are not limited to, binders, fillers, disintegrants, lubricants, coatings, sweeteners, flavors and colors. One of skill in the art will recognize that other pharmaceutical excipients are useful in the present invention. The term "pharmaceutical composition" as used herein includes, e.g., solid and/or liquid dosage forms such as tablet, capsule, pill and the like.

[0151] The drug candidates (e.g., peptides) of the present invention can be administered as frequently as necessary, including hourly, daily, weekly or monthly. The drug candidates (e.g., peptides) utilized in the methods of the invention can be, e.g., administered at dosages that may be varied depending upon the requirements of the method being employed. The drug candidates (e.g., peptides) described herein can be administered to the subject in a variety of ways, including parenterally, subcutaneously, intravenously, intratracheally, intranasally, intradermally, intramuscularly, colonically, rectally, urethrally or intraperitoneally. In some embodiments, the pharmaceutical compositions can be administered parenterally, intravenously, intramuscularly or orally. In some embodiments, the drug candidates can be administered systemically. In some embodiments, the compositions can be administered intratumorally and/or intranodally, such as delivery to a subject's lymph node(s). In certain embodiments, administration can include enteral administration including oral administration, rectal administration, and administration by gastric feeding tube or duodenal feeding tube. Administration can also be including intravenous injection, intra-arterial injection, intra-muscular injection, intracerebral, intracerebroventricular or subcutaneous (under the skin) administration. In some embodiments, administration can be achieved by topical means including epicutaneous (application to skin) and inhalation.

[0152] The oral agents comprising a drug candidates (e.g., peptides) described herein can be in any suitable form for oral

administration, such as liquid, tablets, capsules, or the like. The oral formulations can be further coated or treated to prevent or reduce dissolution in stomach. The compositions of the present invention can be administered to a subject using any suitable methods known in the art. Suitable formulations for use in the present invention and methods of delivery are generally well known in the art. For example, the drug candidates (e.g., peptides) described herein can be formulated as pharmaceutical compositions with a pharmaceutically acceptable diluent, carrier or excipient. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions including pH adjusting and buffering agents, tonicity adjusting agents, wetting agents and the like, such as, for example, sodium acetate, sodium lactate, sodium chloride, potassium chloride, calcium chloride, sorbitan monolaurate, triethanolamine oleate, etc.

[0153] The drug candidates used in the methods of the present invention may be administered by any suitable technique available in the art. In some embodiments, the entire library or sub-library of the drug candidates can be administered into a single subject. Such administration can provide, e.g. a rapid analysis of pharmacological properties of a large number of drug candidates. After administration to a subject, samples can be obtained and analyzed, e.g., by any suitable assay method, such as liquid chromatography coupled to tandem mass spectroscopy.

[0154] The administration of the libraries and the sub-libraries of the present invention can include administration via multiple routes. In some examples, the libraries or sub-libraries of the candidates obtained from the methods of the present invention can be administered by more than one route, e.g., sub-libraries of drug candidates (e.g., knotted peptides) that can be administered both orally and intravenously. Different administration routes of the same drug can, e.g., have different effects including different drug absorption rate, different time of onset and duration of action. In some embodiments, the libraries of the present invention can be administered through more than one route and the results can be compared for desirable properties including maximum drug concentration at the site of action, minimum drug concentration elsewhere, and prolonged drug adsorption and for avoiding fast-pass metabolism. In certain embodiments, a library can be administered to a subject both orally and intravenously and then brain tissue can be analyzed (e.g., by mass spectrometry) to determine the route of administration that leads to improved concentration of the candidates past the blood brain barrier.

[0155] The samples for analysis of the libraries or sub-libraries of the current invention include any sample obtained from a subject. These samples including sample of biological tissue or fluid origin obtained in vivo or in vitro. Non limiting examples of such samples include body fluid (e.g., blood, blood plasma, serum, mucous, spinal fluid, or urine), organs, tissues, fractions, and cells isolated from a subject. Samples also include sections of the biological sample for e.g., sectional portions of an organ or tissue. Tissue samples from subjects can include biopsy and/or necropsy tissue. The tissue samples can also include tissue from a brain, a lung, a kidney, a muscle, a liver, a heart, a stomach, a pancreas, or any other organ. Samples may also include extracts from a biological sample, for example, an antigen from a biological fluid (e.g., blood or urine). In some examples the libraries of the current disclosure can be administered via 3, 4, or 5 administration

routes. In some embodiments, the biological sample is mammalian (e.g., rat, mouse, or human). In certain embodiments, the biological sample is of human origin.

EXAMPLES

Example 1

[0156] This example describes a method for expressing peptide constructs in culture, allowing for the rapid generation of known and novel knottins and greatly facilitating their development, particularly as drugs.

[0157] As shown in FIG. 7, the various knottins can be expressed, e.g., in a lentivirus expression-based method that can include packaging, transfer, and then expression followed by isolation and/or purification of the expressed knottin peptides. Several coding constructs can be used. In this example, the encoding of the knottin peptides included a polynucleotide construct including IgK SP-sFLAG-HIS-Siderocalin-TEV-Knottin. Specific sequences of some example constructs are disclosed in Appendix I below. FIG. 8 shows gel data of a number of example knottins that were made according to the method described in FIG. 7. As shown, chlorotoxin (CTX), chymotrypsin inhibitor (CTI), epiregulin (EPI), hefutoxin (HTX), bubble protein (BUB), potato carboxypeptidase inhibitor (PCI) were properly folded.

[0158] FIG. 9 shows a schematic describing production of a pooled library of knottins. In this example, sequences of thousands of knottins can be encoded in an oligonucleotide pool (1) and selectively amplified using unique primer pairs (2). DNA sublibraries can be cloned into the expression vector, which results in the knottin variants that can, e.g., have unique parental mass signatures and unique tryptic fragment mass signatures that can be resolvable using current techniques, such as mass spectroscopy.

[0159] FIG. 10 includes example knottin variants that describe representative sequencing from a cloned knottin library. The sequences show raw sequencing data from a single round of library cloning. The sequence portions highlighted in grey are full length chlorotoxin variants, and the errors in oligonucleotide synthesis can explain the truncated and extended peptide sequences.

[0160] Using the methods described in this example, variants of several knottin scaffolds were generated and analyzed. FIG. 11 shows an SDS-PAGE analysis of 3000 member knottin libraries for, e.g., hefutoxin, chlorotoxin and chymotrypsin inhibitor. Each column of the SDS-PAGE gel shows a purified sample of a pool of 3000 knottin protein variants run under native and reducing conditions. The migration shift between the paired bands indicates disulfide formation.

[0161] Four scaffolds were selected for the generation of defined libraries: hefutoxin, CTI, chlorotoxin, and epiregulin. A list of target amino acid sequences was generated in silico such that every member of each library would have a tryptic fragment with a unique mass; mutations were selected to be structurally adjacent in order to generate binding epitopes. The cysteines were not mutated, and lysine was specifically avoided in order to make N-terminal conjugation unambiguous. 3000 variants of each scaffold were generated, and each scaffold was flanked by a unique set of PCR primer sites so that each of the four sublibraries could be amplified independently. All constructs had an N-terminal BamHI site and a C-terminal NotI site, and following PCR amplification of each sublibrary from the pool of 12000 oligonucleotides, each sublibrary was restriction digested and cloned into cut

parental vector (both the furin-cleaved and TEV-cleaved versions) as an Lcn2 fusion protein using standard techniques. HEK293 cells were transfected with this plasmid library as well as the accessory plasmids needed for Daedalus expression, and the virus in the media harvested 3-4 days later. Virus was concentrated by centrifugation and used to infect HEK293 cells for protein production using standard procedures. We have found that the TEV-cleavable construct is technically easier to handle when producing libraries because it allows for facile recovery of the fusion by IMAC on nickel resin. Following IMAC, the fusion protein was dialyzed into PBS and allowed to cleave overnight with 6xHis tagged TEV protease, and the Lcn2 and protease were subsequently removed by running the material through nickel resin again. The flow-through containing the cleaved peptide libraries was further purified and buffer exchanged by size exclusion chromatography (SEC) into 10 mM ammonium formate, and the fractions containing the peptides were pooled and lyophilized.

[0162] There were two approaches taken to cloning, Seamless Cloning (Invitrogen) and restriction/ligation based methods. Seamless cloning was employed for making single constructs, typically using synthesized “gBlocks” from IDT. The manufacturer’s instructions were followed. Restriction/ligation methods were standard and were used for cloning libraries as follows: the pooled oligonucleotides from CustomArray were subjected to PCR in order to amplify the relevant sublibrary. The amplified pool was agarose gel purified and cleaned of agarose using a Qiagen column. The purified fragment was digested with FaastDigest (Fermentas) BamHI and NotI and ligated into the parental vector which had been cut with the same two restriction endonucleases. Singleton clones were sequence verified, and 48 members of each library were sequenced in order to verify library quality.

[0163] The cloned knottin or library was cotransfected into HEK293 cells and media was collected as described (Daedalus: a robust, turnkey platform for rapid production of decigram quantities of active recombinant proteins in human cell lines using novel lentiviral vectors. Bandaranayake AD, Correnti C, Ryu B Y, Brault M, Strong R K, Rawlings D J. *Nucleic Acids Res.* 2011 November; 39 (21):e143. doi: 10.1093/nar/gkr706. Epub 2011 Sep. 12.) Fusion protein was isolated using nickel IMAC and cleaved with recombinant TEV protease. Excess siderocalin was removed via size exclusion chromatography, a process which also allowed the buffer to be switched to 10 mM ammonium formate. The knottin containing fractions were then lyophilized. Proper folding and peptide uniformity was demonstrated via SEC chromatography, reverse-phase HPLC, mass spectrometry, and a gel shift in reduced versus non-reduced samples in SDS-PAGE.

[0164] Conjugation to palmitic acid, ICG, or biotimidase-resistant biotin was performed using a 3-10 fold excess of commercially available, activated ester conjugate in PBS. Acetonitrile was added when there were solubility problems. The final material was purified by RP-HPLC for singletons, and excess conjugate was removed from libraries by dialysis.

[0165] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 112

<210> SEQ ID NO 1

<211> LENGTH: 721

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 1

```

gactgagtcg cccgctcgag accatggaga cagacacact cctgctatgg gtactgctgc      60
tctgggttcc aggttcact ggtgactaca aggacgagca tcaccatcat caccatggtg      120
gaagccagga ctccacctca gacctgatcc cagccccacc tctgagcaag gtccctctgc      180
agcagaactt ccaggacaac caattccagg ggaagtggta tgtggtaggc ctggcagggg      240
atgcaattct cagagaagac aaagaccgcg aaaagatgta tgccaccatc tatgagctga      300
aagaagacaa gagctacaat gtcacctccg tcctgttttag gaaaaagaag tgtgactact      360
ggatcaggac ttttgttcca ggttgccagc ccggcgagtt cacgctgggc aacattaaga      420
gttacccctgg attaacgagt tacctcgtcc gagtggtagg caccaactac aaccagcatg      480
ctatggtggt cttcaagaaa gtttctcaaa acaggaggta cttcaagatc accctctacg      540

```

-continued

ggagaaccaa ggagctgact tcggaactaa aggagaactt catccgcttc tccaaatctc 600
tgggcctccc tgaaaaccac atcgtcttcc ctgtcccaat cgaccagtgt atcgacggcg 660
gaggtagcga aaacctgtat tttcagggag gcggccgcta aggatcccgg accgcctctc 720
c 721

<210> SEQ ID NO 2
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 2

aaacctgtatt ttcagggagg c 21

<210> SEQ ID NO 3
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 3

gctaaggatc ccggaccgcc tctcc 25

<210> SEQ ID NO 4
<211> LENGTH: 903
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polynucleotide

<400> SEQUENCE: 4

gactgagtcg cccgctcgag accatggaga cagacacact cctgctatgg gtactgctgc 60
tctgggttcc aggttcact ggtgactaca aggacgagca tcaccatcat caccatggtg 120
gaagccagga ctccacctca gacctgatcc cagccccacc tctgagcaag gtccctctgc 180
agcagaactt ccaggacaac caattccagg ggaagtggta tgtggtaggc ctggcagggg 240
atgcaattct cagagaagac aaagaccgcg aaaagatgta tgccaccatc tatgagctga 300
aagaagacaa gagctacaat gtcacctccg tctgttttag gaaaaagaag tgtgactact 360
ggatcaggac ttttgttcca ggttgccagc ccggcgagtt cacgctgggc aacattaaga 420
gttacctggg attaacgagt tacctcgtcc gagtggtagg caccaactac aaccagcatg 480
ctatggtggt cttcaagaaa gtttctcaaa acagggagta cttcaagatc accctctacg 540
ggagaaccaa ggagctgact tcggaactaa aggagaactt catccgcttc tccaaatctc 600
tgggcctccc tgaaaaccac atcgtcttcc ctgtcccaat cgaccagtgt atcgacggcg 660
gaggtagcga aaacctgtat tttcagggag gcgatacctg cggcagcggc tataatgtgg 720
atcagcgtcg taccaatagc ggctgcaaag cgggcaatgg cgategtcat ttttgcggt 780
gcgacgtac cggcgtggtg gaatgcaaag gcggcaaagtg gaccgaagtg caggattgctg 840
gcagcagcag ctgcaaaggc accagcaatg gcggcgcgac ctgctaatagc taaggatccc 900
gga 903

-continued

```

<210> SEQ ID NO 5
<211> LENGTH: 861
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polynucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(861)

<400> SEQUENCE: 5

atg gag aca gac aca ctc ctg cta tgg gta ctg ctg ctc tgg gtt cca      48
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1          5          10          15

ggt tcc act ggt gac tac aag gac gag cat cac cat cat cac cat ggt      96
Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly
          20          25          30

gga agc cag gac tcc acc tca gac ctg atc cca gcc cca cct ctg agc      144
Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser
          35          40          45

aag gtc cct ctg cag cag aac ttc cag gac aac caa ttc cag ggg aag      192
Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys
          50          55          60

tgg tat gtg gta ggc ctg gca ggg aat gca att ctc aga gaa gac aaa      240
Trp Tyr Val Val Gly Leu Ala Gly Asn Ala Ile Leu Arg Glu Asp Lys
65          70          75          80

gac ccg caa aag atg tat gcc acc atc tat gag ctg aaa gaa gac aag      288
Asp Pro Gln Lys Met Tyr Ala Thr Ile Tyr Glu Leu Lys Glu Asp Lys
          85          90          95

agc tac aat gtc acc tcc gtc ctg ttt agg aaa aag aag tgt gac tac      336
Ser Tyr Asn Val Thr Ser Val Leu Phe Arg Lys Lys Lys Cys Asp Tyr
          100          105          110

tgg atc agg act ttt gtt cca ggt tgc cag ccc ggc gag ttc acg ctg      384
Trp Ile Arg Thr Phe Val Pro Gly Cys Gln Pro Gly Glu Phe Thr Leu
          115          120          125

ggc aac att aag agt tac cct gga tta acg agt tac ctc gtc cga gtg      432
Gly Asn Ile Lys Ser Tyr Pro Gly Leu Thr Ser Tyr Leu Val Arg Val
          130          135          140

gtg agc acc aac tac aac cag cat gct atg gtg ttc ttc aag aaa gtt      480
Val Ser Thr Asn Tyr Asn Gln His Ala Met Val Phe Phe Lys Lys Val
145          150          155          160

tct caa aac agg gag tac ttc aag atc acc ctc tac ggg aga acc aag      528
Ser Gln Asn Arg Glu Tyr Phe Lys Ile Thr Leu Tyr Gly Arg Thr Lys
          165          170          175

gag ctg act tcg gaa cta aag gag aac ttc atc cgc ttc tcc aaa tct      576
Glu Leu Thr Ser Glu Leu Lys Glu Asn Phe Ile Arg Phe Ser Lys Ser
          180          185          190

ctg ggc ctc cct gaa aac cac atc gtc ttc cct gtc cca atc gac cag      624
Leu Gly Leu Pro Glu Asn His Ile Val Phe Pro Val Pro Ile Asp Gln
          195          200          205

tgt atc gac ggc gga ggt agc gaa aac ctg tat ttt cag gga ggc gat      672
Cys Ile Asp Gly Gly Gly Ser Glu Asn Leu Tyr Phe Gln Gly Gly Asp
          210          215          220

acc tgc ggc agc ggc tat aat gtg gat cag cgt cgt acc aat agc ggc      720
Thr Cys Gly Ser Gly Tyr Asn Val Asp Gln Arg Arg Thr Asn Ser Gly
225          230          235          240

tgc aaa gcg ggc aat ggc gat cgt cat ttt tgc ggc tgc gat cgt acc      768
Cys Lys Ala Gly Asn Gly Asp Arg His Phe Cys Gly Cys Asp Arg Thr
          245          250          255

```

-continued

ggc gtg gtg gaa tgc aaa ggc ggc aaa tgg acc gaa gtg cag gat tgc 816
Gly Val Val Glu Cys Lys Gly Gly Lys Trp Thr Glu Val Gln Asp Cys
260 265 270

ggc agc agc agc tgc aaa ggc acc agc aat ggc ggc gcg acc tgc 861
Gly Ser Ser Ser Cys Lys Gly Thr Ser Asn Gly Gly Ala Thr Cys
275 280 285

<210> SEQ ID NO 6
<211> LENGTH: 287
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polypeptide

<400> SEQUENCE: 6

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1 5 10 15

Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly
20 25 30

Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser
35 40 45

Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys
50 55 60

Trp Tyr Val Val Gly Leu Ala Gly Asn Ala Ile Leu Arg Glu Asp Lys
65 70 75 80

Asp Pro Gln Lys Met Tyr Ala Thr Ile Tyr Glu Leu Lys Glu Asp Lys
85 90 95

Ser Tyr Asn Val Thr Ser Val Leu Phe Arg Lys Lys Lys Cys Asp Tyr
100 105 110

Trp Ile Arg Thr Phe Val Pro Gly Cys Gln Pro Gly Glu Phe Thr Leu
115 120 125

Gly Asn Ile Lys Ser Tyr Pro Gly Leu Thr Ser Tyr Leu Val Arg Val
130 135 140

Val Ser Thr Asn Tyr Asn Gln His Ala Met Val Phe Phe Lys Lys Val
145 150 155 160

Ser Gln Asn Arg Glu Tyr Phe Lys Ile Thr Leu Tyr Gly Arg Thr Lys
165 170 175

Glu Leu Thr Ser Glu Leu Lys Glu Asn Phe Ile Arg Phe Ser Lys Ser
180 185 190

Leu Gly Leu Pro Glu Asn His Ile Val Phe Pro Val Pro Ile Asp Gln
195 200 205

Cys Ile Asp Gly Gly Gly Ser Glu Asn Leu Tyr Phe Gln Gly Gly Asp
210 215 220

Thr Cys Gly Ser Gly Tyr Asn Val Asp Gln Arg Arg Thr Asn Ser Gly
225 230 235 240

Cys Lys Ala Gly Asn Gly Asp Arg His Phe Cys Gly Cys Asp Arg Thr
245 250 255

Gly Val Val Glu Cys Lys Gly Gly Lys Trp Thr Glu Val Gln Asp Cys
260 265 270

Gly Ser Ser Ser Cys Lys Gly Thr Ser Asn Gly Gly Ala Thr Cys
275 280 285

<210> SEQ ID NO 7
<211> LENGTH: 885

-continued

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 7

gactgagtcg cccgctcgag accatggaga cagacacact cctgctatgg gtactgctgc 60
tctgggttcc aggttcact ggtgactaca aggacgagca tcaccatcat caccatggtg 120
gaagccagga ctccacctca gacctgatcc cagccccacc tctgagcaag gtccctctgc 180
agcagaactt ccaggacaac caattccagg ggaagtggta tgtggtaggc ctggcagga 240
atgcaattct cagagaagac aaagaccgcg aaaagatgta tgccaccatc tatgagctga 300
aagaagacaa gagctacaat gtcacctccg tcctgttttag gaaaaagaag tgtgactact 360
ggatcaggac ttttgttcca ggttgccagc ccggcgagtt cacgctgggc aacattaaga 420
gttaccttgg attaacgagt tacctcgtcc gagtggtagg caccaactac aaccagcatg 480
ctatgggtgtt cttcaagaaa gtttctcaaa acagggagta cttcaagatc accctctacg 540
ggagaaccaa ggagctgact tcggaactaa aggagaactt catccgcttc tccaaatctc 600
tgggcctccc tgaaaaccac atcgtcttcc ctgtcccaat cgaccagtgt atcgacggcg 660
gaggtagcga aaacctgtat tttcagggag gcgatcagaa ttgcgatatt ggcaatatta 720
ccagccagtg ccagatgcag cataaaaatt gcgaagatgc gaatggctgc gataccatta 780
ttgaagaatg caaaaccagc atggtggaac gttgccagaa tcaggaattt gaaagcgcg 840
cgggcagcac caccctgggc ccgcagtaat gctaaggatc ccgga 885

<210> SEQ ID NO 8
<211> LENGTH: 843
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(843)

<400> SEQUENCE: 8

atg gag aca gac aca ctc ctg cta tgg gta ctg ctg ctc tgg gtt cca 48
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1 5 10 15

ggt tcc act ggt gac tac aag gac gag cat cac cat cat cac cat ggt 96
Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly
20 25 30

gga agc cag gac tcc acc tca gac ctg atc cca gcc cca cct ctg agc 144
Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser
35 40 45

aag gtc cct ctg cag cag aac ttc cag gac aac caa ttc cag ggg aag 192
Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys
50 55 60

tgg tat gtg gta ggc ctg gca ggg aat gca att ctc aga gaa gac aaa 240
Trp Tyr Val Val Gly Leu Ala Gly Asn Ala Ile Leu Arg Glu Asp Lys
65 70 75 80

gac ccg caa aag atg tat gcc acc atc tat gag ctg aaa gaa gac aag 288
Asp Pro Gln Lys Met Tyr Ala Thr Ile Tyr Glu Leu Lys Glu Asp Lys
85 90 95

agc tac aat gtc acc tcc gtc ctg ttt agg aaa aag aag tgt gac tac 336

-continued

Ser	Tyr	Asn	Val	Thr	Ser	Val	Leu	Phe	Arg	Lys	Lys	Lys	Cys	Asp	Tyr	
			100					105					110			
tgg	atc	agg	act	ttt	gtt	cca	ggg	tgc	cag	ccc	ggc	gag	ttc	acg	ctg	384
Trp	Ile	Arg	Thr	Phe	Val	Pro	Gly	Cys	Gln	Pro	Gly	Glu	Phe	Thr	Leu	
		115					120					125				
ggc	aac	att	aag	agt	tac	cct	gga	tta	acg	agt	tac	ctc	gtc	cga	gtg	432
Gly	Asn	Ile	Lys	Ser	Tyr	Pro	Gly	Leu	Thr	Ser	Tyr	Leu	Val	Arg	Val	
	130					135					140					
gtg	agc	acc	aac	tac	aac	cag	cat	gct	atg	gtg	ttc	ttc	aag	aaa	gtt	480
Val	Ser	Thr	Asn	Tyr	Asn	Gln	His	Ala	Met	Val	Phe	Phe	Lys	Lys	Val	
145					150					155					160	
tct	caa	aac	agg	gag	tac	ttc	aag	atc	acc	ctc	tac	ggg	aga	acc	aag	528
Ser	Gln	Asn	Arg	Glu	Tyr	Phe	Lys	Ile	Thr	Leu	Tyr	Gly	Arg	Thr	Lys	
			165					170					175			
gag	ctg	act	tcg	gaa	cta	aag	gag	aac	ttc	atc	cgc	ttc	tcc	aaa	tct	576
Glu	Leu	Thr	Ser	Glu	Leu	Lys	Glu	Asn	Phe	Ile	Arg	Phe	Ser	Lys	Ser	
			180					185					190			
ctg	ggc	ctc	cct	gaa	aac	cac	atc	gtc	ttc	cct	gtc	cca	atc	gac	cag	624
Leu	Gly	Leu	Pro	Glu	Asn	His	Ile	Val	Phe	Pro	Val	Pro	Ile	Asp	Gln	
		195					200					205				
tgt	atc	gac	ggc	gga	ggg	agc	gaa	aac	ctg	tat	ttt	cag	gga	ggc	gat	672
Cys	Ile	Asp	Gly	Gly	Gly	Ser	Glu	Asn	Leu	Tyr	Phe	Gln	Gly	Gly	Asp	
	210					215					220					
cag	aat	tgc	gat	att	ggc	aat	att	acc	agc	cag	tgc	cag	atg	cag	cat	720
Gln	Asn	Cys	Asp	Ile	Gly	Asn	Ile	Thr	Ser	Gln	Cys	Gln	Met	Gln	His	
225					230					235					240	
aaa	aat	tgc	gaa	gat	gcg	aat	ggc	tgc	gat	acc	att	att	gaa	gaa	tgc	768
Lys	Asn	Cys	Glu	Asp	Ala	Asn	Gly	Cys	Asp	Thr	Ile	Ile	Glu	Glu	Cys	
			245					250					255			
aaa	acc	agc	atg	gtg	gaa	cgt	tgc	cag	aat	cag	gaa	ttt	gaa	agc	gcg	816
Lys	Thr	Ser	Met	Val	Glu	Arg	Cys	Gln	Asn	Gln	Glu	Phe	Glu	Ser	Ala	
			260				265					270				
gcg	ggc	agc	acc	acc	ctg	ggc	ccg	cag								843
Ala	Gly	Ser	Thr	Thr	Leu	Gly	Pro	Gln								
		275				280										
<210> SEQ ID NO 9																
<211> LENGTH: 281																
<212> TYPE: PRT																
<213> ORGANISM: Artificial Sequence																
<220> FEATURE:																
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide																
<400> SEQUENCE: 9																
Met	Glu	Thr	Asp	Thr	Leu	Leu	Leu	Trp	Val	Leu	Leu	Leu	Trp	Val	Pro	
1				5					10					15		
Gly	Ser	Thr	Gly	Asp	Tyr	Lys	Asp	Glu	His	His	His	His	His	His	Gly	
			20					25					30			
Gly	Ser	Gln	Asp	Ser	Thr	Ser	Asp	Leu	Ile	Pro	Ala	Pro	Pro	Leu	Ser	
		35					40					45				
Lys	Val	Pro	Leu	Gln	Gln	Asn	Phe	Gln	Asp	Asn	Gln	Phe	Gln	Gly	Lys	
	50					55					60					
Trp	Tyr	Val	Val	Gly	Leu	Ala	Gly	Asn	Ala	Ile	Leu	Arg	Glu	Asp	Lys	
65					70				75						80	
Asp	Pro	Gln	Lys	Met	Tyr	Ala	Thr	Ile	Tyr	Glu	Leu	Lys	Glu	Asp	Lys	
			85					90					95			
Ser	Tyr	Asn	Val	Thr	Ser	Val	Leu	Phe	Arg	Lys	Lys	Lys	Cys	Asp	Tyr	

-continued

100						105						110					
Trp	Ile	Arg	Thr	Phe	Val	Pro	Gly	Cys	Gln	Pro	Gly	Glu	Phe	Thr	Leu		
115						120						125					
Gly	Asn	Ile	Lys	Ser	Tyr	Pro	Gly	Leu	Thr	Ser	Tyr	Leu	Val	Arg	Val		
130						135						140					
Val	Ser	Thr	Asn	Tyr	Asn	Gln	His	Ala	Met	Val	Phe	Phe	Lys	Lys	Val		
145						150						160					
Ser	Gln	Asn	Arg	Glu	Tyr	Phe	Lys	Ile	Thr	Leu	Tyr	Gly	Arg	Thr	Lys		
165						170						175					
Glu	Leu	Thr	Ser	Glu	Leu	Lys	Glu	Asn	Phe	Ile	Arg	Phe	Ser	Lys	Ser		
180						185						190					
Leu	Gly	Leu	Pro	Glu	Asn	His	Ile	Val	Phe	Pro	Val	Pro	Ile	Asp	Gln		
195						200						205					
Cys	Ile	Asp	Gly	Gly	Gly	Ser	Glu	Asn	Leu	Tyr	Phe	Gln	Gly	Gly	Asp		
210						215						220					
Gln	Asn	Cys	Asp	Ile	Gly	Asn	Ile	Thr	Ser	Gln	Cys	Gln	Met	Gln	His		
225						230						240					
Lys	Asn	Cys	Glu	Asp	Ala	Asn	Gly	Cys	Asp	Thr	Ile	Ile	Glu	Glu	Cys		
245						250						255					
Lys	Thr	Ser	Met	Val	Glu	Arg	Cys	Gln	Asn	Gln	Glu	Phe	Glu	Ser	Ala		
260						265						270					
Ala	Gly	Ser	Thr	Thr	Leu	Gly	Pro	Gln									
275						280											

<210> SEQ ID NO 10
<211> LENGTH: 782
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 10

gactgagtcg cccgctcgag accatggaga cagacacact cctgctatgg gtactgctgc 60

tctgggttcc aggttcact ggtgactaca aggacgagca tcaccatcat caccatggtg 120

gaagccagga ctccacctca gacctgatcc cagccccacc tctgagcaag gtccctctgc 180

agcagaactt ccaggacaac caattccagg ggaagtggta tgtggtaggc ctggcaggga 240

atgcaattct cagagaagac aaagaccgcg aaaagatgta tgccaccatc tatgagctga 300

aagaagacaa gagctacaat gtcacctccg tcctgttttag gaaaaagaag tgtgactact 360

ggatcaggac ttttgttcca ggttgccagc ccggcgagtt cacgctgggc aacattaaga 420

gttaccttgg attaacgagt tacctcgtcc gagtggtgag caccaactac aaccagcatg 480

ctatggtggt cttcaagaaa gtttctcaaa acagggagta cttcaagatc accctctacg 540

ggagaaccaa ggagctgact tcggaactaa aggagaactt catccgcttc tccaaatctc 600

tgggcctccc tgaaaaccac atcgtcttcc ctgtcccaat cgaccagtgt atcgacggcg 660

gaggtagcga aaacctgtat tttcagggag gcggccatgc gtgctatcgt aattgctggc 720

gtgaaggcaa tgatgaagaa acctgcaaag aacgttgcta atgctaagga tcccggaccg 780

cc 782

<210> SEQ ID NO 11
<211> LENGTH: 735

-continued

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(735)

<400> SEQUENCE: 11

atg gag aca gac aca ctc ctg cta tgg gta ctg ctg ctc tgg gtt cca	48
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro	
1 5 10 15	
ggt tcc act ggt gac tac aag gac gag cat cac cat cat cac cat ggt	96
Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly	
20 25 30	
gga agc cag gac tcc acc tca gac ctg atc cca gcc cca cct ctg agc	144
Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser	
35 40 45	
aag gtc cct ctg cag cag aac ttc cag gac aac caa ttc cag ggg aag	192
Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys	
50 55 60	
tgg tat gtg gta ggc ctg gca ggg aat gca att ctc aga gaa gac aaa	240
Trp Tyr Val Val Gly Leu Ala Gly Asn Ala Ile Leu Arg Glu Asp Lys	
65 70 75 80	
gac ccg caa aag atg tat gcc acc atc tat gag ctg aaa gaa gac aag	288
Asp Pro Gln Lys Met Tyr Ala Thr Ile Tyr Glu Leu Lys Glu Asp Lys	
85 90 95	
agc tac aat gtc acc tcc gtc ctg ttt agg aaa aag aag tgt gac tac	336
Ser Tyr Asn Val Thr Ser Val Leu Phe Arg Lys Lys Lys Cys Asp Tyr	
100 105 110	
tgg atc agg act ttt gtt cca ggt tgc cag ccc ggc gag ttc acg ctg	384
Trp Ile Arg Thr Phe Val Pro Gly Cys Gln Pro Gly Glu Phe Thr Leu	
115 120 125	
ggc aac att aag agt tac cct gga tta acg agt tac ctc gtc cga gtg	432
Gly Asn Ile Lys Ser Tyr Pro Gly Leu Thr Ser Tyr Leu Val Arg Val	
130 135 140	
gtg agc acc aac tac aac cag cat gct atg gtg ttc ttc aag aaa gtt	480
Val Ser Thr Asn Tyr Asn Gln His Ala Met Val Phe Phe Lys Lys Val	
145 150 155 160	
tct caa aac agg gag tac ttc aag atc acc ctc tac ggg aga acc aag	528
Ser Gln Asn Arg Glu Tyr Phe Lys Ile Thr Leu Tyr Gly Arg Thr Lys	
165 170 175	
gag ctg act tcg gaa cta aag gag aac ttc atc cgc ttc tcc aaa tct	576
Glu Leu Thr Ser Glu Leu Lys Glu Asn Phe Ile Arg Phe Ser Lys Ser	
180 185 190	
ctg ggc ctc cct gaa aac cac atc gtc ttc cct gtc cca atc gac cag	624
Leu Gly Leu Pro Glu Asn His Ile Val Phe Pro Val Pro Ile Asp Gln	
195 200 205	
tgt atc gac ggc gga ggt agc gaa aac ctg tat ttt cag gga ggc ggc	672
Cys Ile Asp Gly Gly Gly Ser Glu Asn Leu Tyr Phe Gln Gly Gly Gly	
210 215 220	
cat gcg tgc tat cgt aat tgc tgg cgt gaa ggc aat gat gaa gaa acc	720
His Ala Cys Tyr Arg Asn Cys Trp Arg Glu Gly Asn Asp Glu Glu Thr	
225 230 235 240	
tgc aaa gaa cgt tgc	735
Cys Lys Glu Arg Cys	
245	

<210> SEQ ID NO 12

-continued

<211> LENGTH: 245
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 12

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1 5 10 15
Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly
20 25 30
Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser
35 40 45
Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys
50 55 60
Trp Tyr Val Val Gly Leu Ala Gly Asn Ala Ile Leu Arg Glu Asp Lys
65 70 75 80
Asp Pro Gln Lys Met Tyr Ala Thr Ile Tyr Glu Leu Lys Glu Asp Lys
85 90 95
Ser Tyr Asn Val Thr Ser Val Leu Phe Arg Lys Lys Lys Cys Asp Tyr
100 105 110
Trp Ile Arg Thr Phe Val Pro Gly Cys Gln Pro Gly Glu Phe Thr Leu
115 120 125
Gly Asn Ile Lys Ser Tyr Pro Gly Leu Thr Ser Tyr Leu Val Arg Val
130 135 140
Val Ser Thr Asn Tyr Asn Gln His Ala Met Val Phe Phe Lys Lys Val
145 150 155 160
Ser Gln Asn Arg Glu Tyr Phe Lys Ile Thr Leu Tyr Gly Arg Thr Lys
165 170 175
Glu Leu Thr Ser Glu Leu Lys Glu Asn Phe Ile Arg Phe Ser Lys Ser
180 185 190
Leu Gly Leu Pro Glu Asn His Ile Val Phe Pro Val Pro Ile Asp Gln
195 200 205
Cys Ile Asp Gly Gly Gly Ser Glu Asn Leu Tyr Phe Gln Gly Gly Gly
210 215 220
His Ala Cys Tyr Arg Asn Cys Trp Arg Glu Gly Asn Asp Glu Glu Thr
225 230 235 240
Cys Lys Glu Arg Cys
245

<210> SEQ ID NO 13
<211> LENGTH: 816
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 13

gactgagtcg cccgctcgag accatggaga cagacacact cctgctatgg gtactgctgc 60
tctgggttcc aggttcact ggtgactaca aggacgagca tcaccatcat caccatggtg 120
gaagccagga ctccacctca gacctgatcc cagccccacc tctgagcaag gtccctctgc 180
agcagaactt ccaggacaac caattccagg ggaagtggta tgtggtaggc ctggcaggga 240
atgcaattct cagagaagac aaagaccgcg aaaagatgta tgccaccatc tatgagctga 300

-continued

aagaagacaa gagctacaat gtcacctccg tcctgttttag gaaaaagaag tgtgactact 360
ggatcaggac ttttgttcca ggttgccagc ccggcgagtt cacgctgggc aacattaaga 420
gttacccctgg attaacgagt tacctcgtcc gagtgggtgag caccaactac aaccagcatg 480
ctatgggtggt cttcaagaaa gttttctcaaa acaggggagta cttcaagatc accctctacg 540
ggagaaccaa ggagctgact tcggaactaa aggagaactt catccgcttc tccaaatctc 600
tgggcctccc tgaaaaccac atcgtcttcc ctgtcccaat cgaccagtgt atcgacggcg 660
gaggtagcga aaacctgtat tttcagggag gcgaatgccg ttatctgttt ggcggtgca 720
aaaccaccag cgattgctgc aaacatctgg gctgcaaatt tcgtgataaa tattgcgcgt 780
gggattttac ctttagctaa tgctaaggat cccgga 816

<210> SEQ ID NO 14
<211> LENGTH: 774
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(774)

<400> SEQUENCE: 14

atg gag aca gac aca ctc ctg cta tgg gta ctg ctg ctc tgg gtt cca 48
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1 5 10 15
ggt tcc act ggt gac tac aag gac gag cat cac cat cat cac cat ggt 96
Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly
20 25 30
gga agc cag gac tcc acc tca gac ctg atc cca gcc cca cct ctg agc 144
Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser
35 40 45
aag gtc cct ctg cag cag aac ttc cag gac aac caa ttc cag ggg aag 192
Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys
50 55 60
tgg tat gtg gta ggc ctg gca ggg aat gca att ctc aga gaa gac aaa 240
Trp Tyr Val Val Gly Leu Ala Gly Asn Ala Ile Leu Arg Glu Asp Lys
65 70 75 80
gac ccg caa aag atg tat gcc acc atc tat gag ctg aaa gaa gac aag 288
Asp Pro Gln Lys Met Tyr Ala Thr Ile Tyr Glu Leu Lys Glu Asp Lys
85 90 95
agc tac aat gtc acc tcc gtc ctg ttt agg aaa aag aag tgt gac tac 336
Ser Tyr Asn Val Thr Ser Val Leu Phe Arg Lys Lys Lys Cys Asp Tyr
100 105 110
tgg atc agg act ttt gtt cca ggt tgc cag ccc ggc gag ttc acg ctg 384
Trp Ile Arg Thr Phe Val Pro Gly Cys Gln Pro Gly Glu Phe Thr Leu
115 120 125
ggc aac att aag agt tac cct gga tta acg agt tac ctc gtc cga gtg 432
Gly Asn Ile Lys Ser Tyr Pro Gly Leu Thr Ser Tyr Leu Val Arg Val
130 135 140
gtg agc acc aac tac aac cag cat gct atg gtg ttc ttc aag aaa gtt 480
Val Ser Thr Asn Tyr Asn Gln His Ala Met Val Phe Phe Lys Lys Val
145 150 155 160
tct caa aac agg gag tac ttc aag atc acc ctc tac ggg aga acc aag 528
Ser Gln Asn Arg Glu Tyr Phe Lys Ile Thr Leu Tyr Gly Arg Thr Lys
165 170 175

-continued																			
gag	ctg	act	tcg	gaa	cta	aag	gag	aac	ttc	atc	cgc	ttc	tcc	aaa	tct				576
Glu	Leu	Thr	Ser	Glu	Leu	Lys	Glu	Asn	Phe	Ile	Arg	Phe	Ser	Lys	Ser				
			180					185					190						
ctg	ggc	ctc	cct	gaa	aac	cac	atc	gtc	ttc	cct	gtc	cca	atc	gac	cag				624
Leu	Gly	Leu	Pro	Glu	Asn	His	Ile	Val	Phe	Pro	Val	Pro	Ile	Asp	Gln				
			195				200					205							
tgt	atc	gac	ggc	gga	ggg	agc	gaa	aac	ctg	tat	ttt	cag	gga	ggc	gaa				672
Cys	Ile	Asp	Gly	Gly	Gly	Ser	Glu	Asn	Leu	Tyr	Phe	Gln	Gly	Gly	Glu				
		210				215					220								
tgc	cgt	tat	ctg	ttt	ggc	ggc	tgc	aaa	acc	acc	agc	gat	tgc	tgc	aaa				720
Cys	Arg	Tyr	Leu	Phe	Gly	Gly	Cys	Lys	Thr	Thr	Ser	Asp	Cys	Cys	Lys				
			225			230					235				240				
cat	ctg	ggc	tgc	aaa	ttt	cgt	gat	aaa	tat	tgc	gcg	tgg	gat	ttt	acc				768
His	Leu	Gly	Cys	Lys	Phe	Arg	Asp	Lys	Tyr	Cys	Ala	Trp	Asp	Phe	Thr				
			245						250					255					
ttt	agc																		774
Phe	Ser																		
<210> SEQ ID NO 15																			
<211> LENGTH: 258																			
<212> TYPE: PRT																			
<213> ORGANISM: Artificial Sequence																			
<220> FEATURE:																			
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide																			
<400> SEQUENCE: 15																			
Met	Glu	Thr	Asp	Thr	Leu	Leu	Leu	Trp	Val	Leu	Leu	Leu	Trp	Val	Pro				
1				5					10					15					
Gly	Ser	Thr	Gly	Asp	Tyr	Lys	Asp	Glu	His	His	His	His	His	His	Gly				
			20					25						30					
Gly	Ser	Gln	Asp	Ser	Thr	Ser	Asp	Leu	Ile	Pro	Ala	Pro	Pro	Leu	Ser				
			35				40					45							
Lys	Val	Pro	Leu	Gln	Gln	Asn	Phe	Gln	Asp	Asn	Gln	Phe	Gln	Gly	Lys				
			50			55					60								
Trp	Tyr	Val	Val	Gly	Leu	Ala	Gly	Asn	Ala	Ile	Leu	Arg	Glu	Asp	Lys				
65				70					75					80					
Asp	Pro	Gln	Lys	Met	Tyr	Ala	Thr	Ile	Tyr	Glu	Leu	Lys	Glu	Asp	Lys				
				85					90					95					
Ser	Tyr	Asn	Val	Thr	Ser	Val	Leu	Phe	Arg	Lys	Lys	Lys	Cys	Asp	Tyr				
			100					105						110					
Trp	Ile	Arg	Thr	Phe	Val	Pro	Gly	Cys	Gln	Pro	Gly	Glu	Phe	Thr	Leu				
			115				120					125							
Gly	Asn	Ile	Lys	Ser	Tyr	Pro	Gly	Leu	Thr	Ser	Tyr	Leu	Val	Arg	Val				
			130			135						140							
Val	Ser	Thr	Asn	Tyr	Asn	Gln	His	Ala	Met	Val	Phe	Phe	Lys	Lys	Val				
145					150					155					160				
Ser	Gln	Asn	Arg	Glu	Tyr	Phe	Lys	Ile	Thr	Leu	Tyr	Gly	Arg	Thr	Lys				
				165					170					175					
Glu	Leu	Thr	Ser	Glu	Leu	Lys	Glu	Asn	Phe	Ile	Arg	Phe	Ser	Lys	Ser				
			180					185					190						
Leu	Gly	Leu	Pro	Glu	Asn	His	Ile	Val	Phe	Pro	Val	Pro	Ile	Asp	Gln				
			195				200						205						
Cys	Ile	Asp	Gly	Gly	Gly	Ser	Glu	Asn	Leu	Tyr	Phe	Gln	Gly	Gly	Glu				
			210			215					220								
Cys	Arg	Tyr	Leu	Phe	Gly	Gly	Cys	Lys	Thr	Thr	Ser	Asp	Cys	Cys	Lys				

-continued

225	230	235	240
His Leu Gly Cys Lys Phe Arg Asp Lys Tyr Cys Ala Trp Asp Phe Thr	245	250	255
Phe Ser			
<210> SEQ ID NO 16			
<211> LENGTH: 819			
<212> TYPE: DNA			
<213> ORGANISM: Artificial Sequence			
<220> FEATURE:			
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide			
<400> SEQUENCE: 16			
gactgagtcg cccgctcgag accatggaga cagacacact cctgctatgg gtactgctgc			60
tctgggttcc aggttccact ggtgactaca aggacgagca tcaccatcat caccatggtg			120
gaagccagga ctccacctca gacctgatcc cagccccacc tctgagcaag gtccctctgc			180
agcagaactt ccaggacaac caattccagg ggaagtggta tgtggtaggc ctggcagggg			240
atgcaattct cagagaagac aaagacccgc aaaagatgta tgccaccatc tatgagctga			300
aagaagacaa gagctacaat gtcacctccg tctgttttag gaaaaagaag tgtgactact			360
ggatcaggac ttttgttcca ggttgccagc ccggcgagtt cacgctgggc aacattaaga			420
gttaccttgg attaacgagt tacctcgtcc gagtggtagg caccaactac aaccagcatg			480
ctatggtggt cttcaagaaa gtttctcaaa acagggagta cttcaagatc accctctacg			540
ggagaaccaa ggagctgact tcggaactaa aggagaactt catccgcttc tccaaatctc			600
tgggcctccc tgaaaaccac atcgtcttcc ctgtcccaat cgaccagtgt atcgacggcg			660
gaggtagcga aaacctgtat tttcagggag gcgaaattag ctgcgaaccg ggcaaacct			720
ttaaagataa atgcaatacc tgccgttgcg gcgcggatgg caaaagcgcg gcgtgcaccc			780
tgaaagcgtg cccgaatcag taatgctaag gatcccgga			819
<210> SEQ ID NO 17			
<211> LENGTH: 777			
<212> TYPE: DNA			
<213> ORGANISM: Artificial Sequence			
<220> FEATURE:			
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide			
<220> FEATURE:			
<221> NAME/KEY: CDS			
<222> LOCATION: (1)..(777)			
<400> SEQUENCE: 17			
atg gag aca gac aca ctc ctg cta tgg gta ctg ctg ctc tgg gtt cca			48
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro			
1 5 10 15			
ggt tcc act ggt gac tac aag gac gag cat cac cat cat cac cat ggt			96
Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly			
20 25 30			
gga agc cag gac tcc acc tca gac ctg atc cca gcc cca cct ctg agc			144
Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser			
35 40 45			
aag gtc cct ctg cag cag aac ttc cag gac aac caa ttc cag ggg aag			192
Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys			
50 55 60			
tgg tat gtg gta ggc ctg gca ggg aat gca att ctc aga gaa gac aaa			240

-continued

Trp	Tyr	Val	Val	Gly	Leu	Ala	Gly	Asn	Ala	Ile	Leu	Arg	Glu	Asp	Lys	
65					70					75					80	
																288
gac	ccg	caa	aag	atg	tat	gcc	acc	atc	tat	gag	ctg	aaa	gaa	gac	aag	
Asp	Pro	Gln	Lys	Met	Tyr	Ala	Thr	Ile	Tyr	Glu	Leu	Lys	Glu	Asp	Lys	
				85					90					95		
																336
agc	tac	aat	gtc	acc	tcc	gtc	ctg	ttt	agg	aaa	aag	aag	tgt	gac	tac	
Ser	Tyr	Asn	Val	Thr	Ser	Val	Leu	Phe	Arg	Lys	Lys	Lys	Cys	Asp	Tyr	
			100					105					110			
																384
tgg	atc	agg	act	ttt	gtt	cca	ggg	tgc	cag	ccc	ggc	gag	ttc	acg	ctg	
Trp	Ile	Arg	Thr	Phe	Val	Pro	Gly	Cys	Gln	Pro	Gly	Glu	Phe	Thr	Leu	
			115				120					125				
																432
ggc	aac	att	aag	agt	tac	cct	gga	tta	acg	agt	tac	ctc	gtc	cga	gtg	
Gly	Asn	Ile	Lys	Ser	Tyr	Pro	Gly	Leu	Thr	Ser	Tyr	Leu	Val	Arg	Val	
			130				135				140					
																480
gtg	agc	acc	aac	tac	aac	cag	cat	gct	atg	gtg	ttc	ttc	aag	aaa	gtt	
Val	Ser	Thr	Asn	Tyr	Asn	Gln	His	Ala	Met	Val	Phe	Phe	Lys	Lys	Val	
					150					155					160	
																528
tct	caa	aac	agg	gag	tac	ttc	aag	atc	acc	ctc	tac	ggg	aga	acc	aag	
Ser	Gln	Asn	Arg	Glu	Tyr	Phe	Lys	Ile	Thr	Leu	Tyr	Gly	Arg	Thr	Lys	
				165					170					175		
																576
gag	ctg	act	tcg	gaa	cta	aag	gag	aac	ttc	atc	cgc	ttc	tcc	aaa	tct	
Glu	Leu	Thr	Ser	Glu	Leu	Lys	Glu	Asn	Phe	Ile	Arg	Phe	Ser	Lys	Ser	
			180					185					190			
																624
ctg	ggc	ctc	cct	gaa	aac	cac	atc	gtc	ttc	cct	gtc	cca	atc	gac	cag	
Leu	Gly	Leu	Pro	Glu	Asn	His	Ile	Val	Phe	Pro	Val	Pro	Ile	Asp	Gln	
			195				200					205				
																672
tgt	atc	gac	ggc	gga	ggg	agc	gaa	aac	ctg	tat	ttt	cag	gga	ggc	gaa	
Cys	Ile	Asp	Gly	Gly	Gly	Ser	Glu	Asn	Leu	Tyr	Phe	Gln	Gly	Gly	Glu	
			210				215				220					
																720
att	agc	tgc	gaa	ccg	ggc	aaa	acc	ttt	aaa	gat	aaa	tgc	aat	acc	tgc	
Ile	Ser	Cys	Glu	Pro	Gly	Lys	Thr	Phe	Lys	Asp	Lys	Cys	Asn	Thr	Cys	
					230					235					240	
																768
cgt	tgc	ggc	gcg	gat	ggc	aaa	agc	gcg	gcg	tgc	acc	ctg	aaa	gcg	tgc	
Arg	Cys	Gly	Ala	Asp	Gly	Lys	Ser	Ala	Ala	Cys	Thr	Leu	Lys	Ala	Cys	
				245					250					255		
																777
ccg	aat	cag														
Pro	Asn	Gln														
<210> SEQ ID NO 18																
<211> LENGTH: 259																
<212> TYPE: PRT																
<213> ORGANISM: Artificial Sequence																
<220> FEATURE:																
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide																
<400> SEQUENCE: 18																
Met	Glu	Thr	Asp	Thr	Leu	Leu	Leu	Trp	Val	Leu	Leu	Leu	Trp	Val	Pro	
1				5					10					15		
Gly	Ser	Thr	Gly	Asp	Tyr	Lys	Asp	Glu	His	His	His	His	His	His	Gly	
			20					25					30			
Gly	Ser	Gln	Asp	Ser	Thr	Ser	Asp	Leu	Ile	Pro	Ala	Pro	Pro	Leu	Ser	
			35				40					45				
Lys	Val	Pro	Leu	Gln	Gln	Asn	Phe	Gln	Asp	Asn	Gln	Phe	Gln	Gly	Lys	
			50			55					60					
Trp	Tyr	Val	Val	Gly	Leu	Ala	Gly	Asn	Ala	Ile	Leu	Arg	Glu	Asp	Lys	
65					70					75					80	
Asp	Pro	Gln	Lys	Met	Tyr	Ala	Thr	Ile	Tyr	Glu	Leu	Lys	Glu	Asp	Lys	

-continued

85					90					95					
Ser	Tyr	Asn	Val	Thr	Ser	Val	Leu	Phe	Arg	Lys	Lys	Lys	Cys	Asp	Tyr
100					105					110					
Trp	Ile	Arg	Thr	Phe	Val	Pro	Gly	Cys	Gln	Pro	Gly	Glu	Phe	Thr	Leu
115			120					125							
Gly	Asn	Ile	Lys	Ser	Tyr	Pro	Gly	Leu	Thr	Ser	Tyr	Leu	Val	Arg	Val
130			135					140							
Val	Ser	Thr	Asn	Tyr	Asn	Gln	His	Ala	Met	Val	Phe	Phe	Lys	Lys	Val
145		150					155			160					
Ser	Gln	Asn	Arg	Glu	Tyr	Phe	Lys	Ile	Thr	Leu	Tyr	Gly	Arg	Thr	Lys
165				170					175						
Glu	Leu	Thr	Ser	Glu	Leu	Lys	Glu	Asn	Phe	Ile	Arg	Phe	Ser	Lys	Ser
180			185					190							
Leu	Gly	Leu	Pro	Glu	Asn	His	Ile	Val	Phe	Pro	Val	Pro	Ile	Asp	Gln
195		200					205								
Cys	Ile	Asp	Gly	Gly	Gly	Ser	Glu	Asn	Leu	Tyr	Phe	Gln	Gly	Gly	Glu
210		215					220								
Ile	Ser	Cys	Glu	Pro	Gly	Lys	Thr	Phe	Lys	Asp	Lys	Cys	Asn	Thr	Cys
225		230					235			240					
Arg	Cys	Gly	Ala	Asp	Gly	Lys	Ser	Ala	Ala	Cys	Thr	Leu	Lys	Ala	Cys
245			250					255							
Pro Asn Gln															

<210> SEQ ID NO 19
<211> LENGTH: 822
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 19

gactgagtcg cccgctcgag accatggaga cagacacact cctgctatgg gtactgctgc 60
tctgggttcc aggttcact ggtgactaca aggacgagca tcaccatcat caccatggtg 120
gaagccagga ctccacctca gacctgatcc cagccccacc tctgagcaag gtccctctgc 180
agcagaactt ccaggacaac caattccagg ggaagtggta tgtggtaggc ctggcaggga 240
atgcaattct cagagaagac aaagaccgcg aaaagatgta tgccaccatc tatgagctga 300
aagaagacaa gagctacaat gtcacctccg tctgttttag gaaaaagaag tgtgactact 360
ggatcaggac ttttgttcca ggttgccagc ccggcgagtt cacgctgggc aacattaaga 420
gttacccctgg attaacgagt tacctcgtcc gagtggtagg caccaactac aaccagcatg 480
ctatggtggt cttcaagaaa gtttctcaaa acagggagta cttcaagatc accctctacg 540
ggagaaccaa ggagctgact tcggaactaa aggagaactt catccgcttc tccaaatctc 600
tgggcctccc tgaaaaccac atcgtcttcc ctgtcccaat cgaccagtgt atcgacggcg 660
gaggtagcga aaacctgtat tttcagggag gcgtgtgccg tgattggttt aaagaaaccg 720
cgtgccgtca tgcgaaaagc ctgggcaatt gccgtaccag ccagaaatat cgtgcgaatt 780
gcgcgaaaac ctgcgaactg tgctaattgct aaggatcccg ga 822

<210> SEQ ID NO 20
<211> LENGTH: 780
<212> TYPE: DNA

-continued

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(780)

<400> SEQUENCE: 20

atg gag aca gac aca ctc ctg cta tgg gta ctg ctg ctc tgg gtt cca	48
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro	
1 5 10 15	
ggt tcc act ggt gac tac aag gac gag cat cac cat cat cac cat ggt	96
Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly	
20 25 30	
gga agc cag gac tcc acc tca gac ctg atc cca gcc cca cct ctg agc	144
Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser	
35 40 45	
aag gtc cct ctg cag cag aac ttc cag gac aac caa ttc cag ggg aag	192
Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys	
50 55 60	
tgg tat gtg gta ggc ctg gca ggg aat gca att ctc aga gaa gac aaa	240
Trp Tyr Val Val Gly Leu Ala Gly Asn Ala Ile Leu Arg Glu Asp Lys	
65 70 75 80	
gac ccg caa aag atg tat gcc acc atc tat gag ctg aaa gaa gac aag	288
Asp Pro Gln Lys Met Tyr Ala Thr Ile Tyr Glu Leu Lys Glu Asp Lys	
85 90 95	
agc tac aat gtc acc tcc gtc ctg ttt agg aaa aag aag tgt gac tac	336
Ser Tyr Asn Val Thr Ser Val Leu Phe Arg Lys Lys Lys Cys Asp Tyr	
100 105 110	
tgg atc agg act ttt gtt cca ggt tgc cag ccc ggc gag ttc acg ctg	384
Trp Ile Arg Thr Phe Val Pro Gly Cys Gln Pro Gly Glu Phe Thr Leu	
115 120 125	
ggc aac att aag agt tac cct gga tta acg agt tac ctc gtc cga gtg	432
Gly Asn Ile Lys Ser Tyr Pro Gly Leu Thr Ser Tyr Leu Val Arg Val	
130 135 140	
gtg agc acc aac tac aac cag cat gct atg gtg ttc ttc aag aaa gtt	480
Val Ser Thr Asn Tyr Asn Gln His Ala Met Val Phe Phe Lys Lys Val	
145 150 155 160	
tct caa aac agg gag tac ttc aag atc acc ctc tac ggg aga acc aag	528
Ser Gln Asn Arg Glu Tyr Phe Lys Ile Thr Leu Tyr Gly Arg Thr Lys	
165 170 175	
gag ctg act tcg gaa cta aag gag aac ttc atc cgc ttc tcc aaa tct	576
Glu Leu Thr Ser Glu Leu Lys Glu Asn Phe Ile Arg Phe Ser Lys Ser	
180 185 190	
ctg ggc ctc cct gaa aac cac atc gtc ttc cct gtc cca atc gac cag	624
Leu Gly Leu Pro Glu Asn His Ile Val Phe Pro Val Pro Ile Asp Gln	
195 200 205	
tgt atc gac ggc gga ggt agc gaa aac ctg tat ttt cag gga ggc gtg	672
Cys Ile Asp Gly Gly Gly Ser Glu Asn Leu Tyr Phe Gln Gly Gly Val	
210 215 220	
tgc cgt gat tgg ttt aaa gaa acc gcg tgc cgt cat gcg aaa agc ctg	720
Cys Arg Asp Trp Phe Lys Glu Thr Ala Cys Arg His Ala Lys Ser Leu	
225 230 235 240	
ggc aat tgc cgt acc agc cag aaa tat cgt gcg aat tgc gcg aaa acc	768
Gly Asn Cys Arg Thr Ser Gln Lys Tyr Arg Ala Asn Cys Ala Lys Thr	
245 250 255	
tgc gaa ctg tgc	780
Cys Glu Leu Cys	
260	

-continued

<210> SEQ ID NO 21
<211> LENGTH: 260
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 21

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1 5 10 15
Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly
20 25 30
Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser
35 40 45
Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys
50 55 60
Trp Tyr Val Val Gly Leu Ala Gly Asn Ala Ile Leu Arg Glu Asp Lys
65 70 75 80
Asp Pro Gln Lys Met Tyr Ala Thr Ile Tyr Glu Leu Lys Glu Asp Lys
85 90 95
Ser Tyr Asn Val Thr Ser Val Leu Phe Arg Lys Lys Lys Cys Asp Tyr
100 105 110
Trp Ile Arg Thr Phe Val Pro Gly Cys Gln Pro Gly Glu Phe Thr Leu
115 120 125
Gly Asn Ile Lys Ser Tyr Pro Gly Leu Thr Ser Tyr Leu Val Arg Val
130 135 140
Val Ser Thr Asn Tyr Asn Gln His Ala Met Val Phe Phe Lys Lys Val
145 150 155 160
Ser Gln Asn Arg Glu Tyr Phe Lys Ile Thr Leu Tyr Gly Arg Thr Lys
165 170 175
Glu Leu Thr Ser Glu Leu Lys Glu Asn Phe Ile Arg Phe Ser Lys Ser
180 185 190
Leu Gly Leu Pro Glu Asn His Ile Val Phe Pro Val Pro Ile Asp Gln
195 200 205
Cys Ile Asp Gly Gly Gly Ser Glu Asn Leu Tyr Phe Gln Gly Gly Val
210 215 220
Cys Arg Asp Trp Phe Lys Glu Thr Ala Cys Arg His Ala Lys Ser Leu
225 230 235 240
Gly Asn Cys Arg Thr Ser Gln Lys Tyr Arg Ala Asn Cys Ala Lys Thr
245 250 255
Cys Glu Leu Cys
260

<210> SEQ ID NO 22
<211> LENGTH: 849
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 22

gactgagtcg cccgctcgag accatggaga cagacacact cctgctatgg gtactgctgc 60
tctgggttcc aggttcact ggtgactaca aggacgagca tcaccatcat caccatggtg 120

-continued

gaagccagga ctccacctca gacctgatcc cagccccacc tctgagcaag gtccctctgc 180
agcagaactt ccaggacaac caattccagg ggaagtggta tgtggtaggc ctggcagggga 240
atgcaattct cagagaagac aaagacccgc aaaagatgta tgccaccatc tatgagctga 300
aagaagacaa gagctacaat gtcacctccg tcctgttttag gaaaaagaag tgtgactact 360
ggatcaggac ttttgttcca ggttgccagc ccggcgagtt cacgctgggc aacattaaga 420
gttaccctgg attaacgagt tacctcgtec gagtggtagg caccaactac aaccagcatg 480
ctatggtggt cttcaagaaa gtttctcaaa acagggagta cttcaagatc accctctacg 540
ggagaaccaa ggagctgact tcggaactaa aggagaactt catccgcttc tccaaatctc 600
tgggcctccc tgaaaaccac atcgtcttcc ctgtcccaat cgaccagtgt atcgacggcg 660
gaggtagcga aaacctgtat tttcagggag gcgtgagcat taccaaatgc agcagcgata 720
tgaatggcta ttgctgcat ggccagtgc tttatctggt ggatatgagc cagaattatt 780
gccgttgcca agtgggctat accggcgtgc gttgcgaaca ttttttctg taatgctaag 840
gatcccgga 849

<210> SEQ ID NO 23
<211> LENGTH: 807
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(807)

<400> SEQUENCE: 23

atg gag aca gac aca ctc ctg cta tgg gta ctg ctg ctc tgg gtt cca 48
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1 5 10 15

ggt tcc act ggt gac tac aag gac gag cat cac cat cat cac cat ggt 96
Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly
20 25 30

gga agc cag gac tcc acc tca gac ctg atc cca gcc cca cct ctg agc 144
Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser
35 40 45

aag gtc cct ctg cag cag aac ttc cag gac aac caa ttc cag ggg aag 192
Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys
50 55 60

tgg tat gtg gta ggc ctg gca ggg aat gca att ctc aga gaa gac aaa 240
Trp Tyr Val Val Gly Leu Ala Gly Asn Ala Ile Leu Arg Glu Asp Lys
65 70 75 80

gac ccg caa aag atg tat gcc acc atc tat gag ctg aaa gaa gac aag 288
Asp Pro Gln Lys Met Tyr Ala Thr Ile Tyr Glu Leu Lys Glu Asp Lys
85 90 95

agc tac aat gtc acc tcc gtc ctg ttt agg aaa aag aag tgt gac tac 336
Ser Tyr Asn Val Thr Ser Val Leu Phe Arg Lys Lys Lys Cys Asp Tyr
100 105 110

tgg atc agg act ttt gtt cca ggt tgc cag ccc ggc gag ttc acg ctg 384
Trp Ile Arg Thr Phe Val Pro Gly Cys Gln Pro Gly Glu Phe Thr Leu
115 120 125

ggc aac att aag agt tac cct gga tta acg agt tac ctc gtc cga gtg 432
Gly Asn Ile Lys Ser Tyr Pro Gly Leu Thr Ser Tyr Leu Val Arg Val
130 135 140

-continued

gtg agc acc aac tac aac cag cat gct atg gtg ttc ttc aag aaa gtt	480
Val Ser Thr Asn Tyr Asn Gln His Ala Met Val Phe Phe Lys Lys Val	
145 150 155 160	
tct caa aac agg gag tac ttc aag atc acc ctc tac ggg aga acc aag	528
Ser Gln Asn Arg Glu Tyr Phe Lys Ile Thr Leu Tyr Gly Arg Thr Lys	
165 170 175	
gag ctg act tcg gaa cta aag gag aac ttc atc cgc ttc tcc aaa tct	576
Glu Leu Thr Ser Glu Leu Lys Glu Asn Phe Ile Arg Phe Ser Lys Ser	
180 185 190	
ctg ggc ctc cct gaa aac cac atc gtc ttc cct gtc cca atc gac cag	624
Leu Gly Leu Pro Glu Asn His Ile Val Phe Pro Val Pro Ile Asp Gln	
195 200 205	
tgt atc gac ggc gga ggt agc gaa aac ctg tat ttt cag gga ggc gtg	672
Cys Ile Asp Gly Gly Gly Ser Glu Asn Leu Tyr Phe Gln Gly Gly Val	
210 215 220	
agc att acc aaa tgc agc agc gat atg aat ggc tat tgc ctg cat ggc	720
Ser Ile Thr Lys Cys Ser Ser Asp Met Asn Gly Tyr Cys Leu His Gly	
225 230 235 240	
cag tgc att tat ctg gtg gat atg agc cag aat tat tgc cgt tgc gaa	768
Gln Cys Ile Tyr Leu Val Asp Met Ser Gln Asn Tyr Cys Arg Cys Glu	
245 250 255	
gtg ggc tat acc ggc gtg cgt tgc gaa cat ttt ttt ctg	807
Val Gly Tyr Thr Gly Val Arg Cys Glu His Phe Phe Leu	
260 265	
<210> SEQ ID NO 24	
<211> LENGTH: 269	
<212> TYPE: PRT	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide	
<400> SEQUENCE: 24	
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro	
1 5 10 15	
Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly	
20 25 30	
Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser	
35 40 45	
Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys	
50 55 60	
Trp Tyr Val Val Gly Leu Ala Gly Asn Ala Ile Leu Arg Glu Asp Lys	
65 70 75 80	
Asp Pro Gln Lys Met Tyr Ala Thr Ile Tyr Glu Leu Lys Glu Asp Lys	
85 90 95	
Ser Tyr Asn Val Thr Ser Val Leu Phe Arg Lys Lys Lys Cys Asp Tyr	
100 105 110	
Trp Ile Arg Thr Phe Val Pro Gly Cys Gln Pro Gly Glu Phe Thr Leu	
115 120 125	
Gly Asn Ile Lys Ser Tyr Pro Gly Leu Thr Ser Tyr Leu Val Arg Val	
130 135 140	
Val Ser Thr Asn Tyr Asn Gln His Ala Met Val Phe Phe Lys Lys Val	
145 150 155 160	
Ser Gln Asn Arg Glu Tyr Phe Lys Ile Thr Leu Tyr Gly Arg Thr Lys	
165 170 175	
Glu Leu Thr Ser Glu Leu Lys Glu Asn Phe Ile Arg Phe Ser Lys Ser	

	180		185		190
Leu Gly Leu Pro Glu Asn His Ile Val Phe Pro Val Pro Ile Asp Gln 195 200 205					
Cys Ile Asp Gly Gly Gly Ser Glu Asn Leu Tyr Phe Gln Gly Gly Val 210 215 220					
Ser Ile Thr Lys Cys Ser Ser Asp Met Asn Gly Tyr Cys Leu His Gly 225 230 235 240					
Gln Cys Ile Tyr Leu Val Asp Met Ser Gln Asn Tyr Cys Arg Cys Glu 245 250 255					
Val Gly Tyr Thr Gly Val Arg Cys Glu His Phe Phe Leu 260 265					

```
<210> SEQ ID NO 25<211> LENGTH: 801<212> TYPE: DNA<213> ORGANISM: Artificial Sequence<220> FEATURE:<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide<400> SEQUENCE: 25gactgagtcg cccgcctcgag accatggaga cagacacact cctgctatgg gtactgctgc      60tctgggtttc aggttccact ggtgactaca aggacgagca tcaccatcat caccatggtg     120gaagccagga ctccacctca gacctgatcc cagccccacc tctgagcaag gtcctctctgc    180agcagaactt ccaggacaac caattccagg ggaagtggta tgtggtaggc ctggcaggga     240atgcaattct cagagaagac aaagaccogc aaaagatgta tgcccaccatc tatgagctga    300aagaagaaaa gagctacaat gtcacctecg tcctgtttag gaaaagaag tgtgactact     360ggatcaggac ttttgttcca gggtgccage cgggcgagtt cacgctgggc aacattaaga    420gttacctctg attaacgagt tacctcgtec gagtgggtgag caccaactac aaccagcatg    480ctatggtgtt cttcaagaaa gtttctcaaa acagggagta cttcaagatc acctctacg     540ggagaaccaaa ggagctgact tcggaactaa aggagaactt catccgcttc tccaaatctc    600tgggcctccc tgaaaaccac atcgtcttcc ctgtcccaat cgaccagtgt atcgacggcg    660gaggtagcga aaacctgtat ttccaggagg gcggcattcc gtgcggcgaa agctgcgtgt    720ggattccgtg cattagcgcg gcgctgggct gcagctgcaa aaataaagtg tgctatcgta    780attaatgcta aggatcccg a                                801<210> SEQ ID NO 26<211> LENGTH: 759<212> TYPE: DNA<213> ORGANISM: Artificial Sequence<220> FEATURE:<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide<220> FEATURE:<221> NAME/KEY: CDS<222> LOCATION: (1)..(759)<400> SEQUENCE: 26atg gag aca gac aca ctc ctg cta tgg gta ctg ctg ctc tgg gtt cca          48Met Glu Thr Asp Thr Leu Leu Trp Val Leu Leu Trp Val Pro                    1         5             10              15ggt tcc act ggt gac tac aag gac gag cat cac cat cat cac cat ggt          96Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly            20        25           30
```

-continued

gga agc cag gac tcc acc tca gac ctg atc cca gcc cca cct ctg agc	144
Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser	
35 40 45	
aag gtc cct ctg cag cag aac ttc cag gac aac caa ttc cag ggg aag	192
Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys	
50 55 60	
tgg tat gtg gta ggc ctg gca ggg aat gca att ctc aga gaa gac aaa	240
Trp Tyr Val Val Gly Leu Ala Gly Asn Ala Ile Leu Arg Glu Asp Lys	
65 70 75 80	
gac ccg caa aag atg tat gcc acc atc tat gag ctg aaa gaa gac aag	288
Asp Pro Gln Lys Met Tyr Ala Thr Ile Tyr Glu Leu Lys Glu Asp Lys	
85 90 95	
agc tac aat gtc acc tcc gtc ctg ttt agg aaa aag aag tgt gac tac	336
Ser Tyr Asn Val Thr Ser Val Leu Phe Arg Lys Lys Lys Cys Asp Tyr	
100 105 110	
tgg atc agg act ttt gtt cca ggt tgc cag ccc ggc gag ttc acg ctg	384
Trp Ile Arg Thr Phe Val Pro Gly Cys Gln Pro Gly Glu Phe Thr Leu	
115 120 125	
ggc aac att aag agt tac cct gga tta acg agt tac ctc gtc cga gtg	432
Gly Asn Ile Lys Ser Tyr Pro Gly Leu Thr Ser Tyr Leu Val Arg Val	
130 135 140	
gtg agc acc aac tac aac cag cat gct atg gtg ttc ttc aag aaa gtt	480
Val Ser Thr Asn Tyr Asn Gln His Ala Met Val Phe Phe Lys Lys Val	
145 150 155 160	
tct caa aac agg gag tac ttc aag atc acc ctc tac ggg aga acc aag	528
Ser Gln Asn Arg Glu Tyr Phe Lys Ile Thr Leu Tyr Gly Arg Thr Lys	
165 170 175	
gag ctg act tcg gaa cta aag gag aac ttc atc cgc ttc tcc aaa tct	576
Glu Leu Thr Ser Glu Leu Lys Glu Asn Phe Ile Arg Phe Ser Lys Ser	
180 185 190	
ctg ggc ctc cct gaa aac cac atc gtc ttc cct gtc cca atc gac cag	624
Leu Gly Leu Pro Glu Asn His Ile Val Phe Pro Val Pro Ile Asp Gln	
195 200 205	
tgt atc gac ggc gga ggt agc gaa aac ctg tat ttt cag gga ggc ggc	672
Cys Ile Asp Gly Gly Gly Ser Glu Asn Leu Tyr Phe Gln Gly Gly Gly	
210 215 220	
att ccg tgc ggc gaa agc tgc gtg tgg att ccg tgc att agc gcg gcg	720
Ile Pro Cys Gly Glu Ser Cys Val Trp Ile Pro Cys Ile Ser Ala Ala	
225 230 235 240	
ctg ggc tgc agc tgc aaa aat aaa gtg tgc tat cgt aat	759
Leu Gly Cys Ser Cys Lys Asn Lys Val Cys Tyr Arg Asn	
245 250	

<210> SEQ ID NO 27
<211> LENGTH: 253
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 27

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro	
1 5 10 15	
Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly	
20 25 30	
Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser	
35 40 45	

-continued

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

<210> SEQ ID NO 28
<211> LENGTH: 873
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 28

gactgagtcg cccgctcgag accatggaga cagacacact cctgctatgg gtactgctgc	60
tctgggttcc aggttccact ggtgactaca aggacgagca tcaccatcat caccatggtg	120
gaagccagga ctccacctca gacctgatcc cagccccacc tctgagcaag gtccctctgc	180
agcagaactt ccaggacaac caattccagg ggaagtggta tgtggtaggc ctggcagggg	240
atgcaattct cagagaagac aaagaccgcg aaaagatgta tgccaccatc tatgagctga	300
aagaagaaa gagctacaat gtcacctccg tcctgttttag gaaaaagaag tgtgactact	360
ggatcaggac ttttgttcca ggttgccagc ccggcgagtt cacgctgggc aacattaaga	420
gttaccctgg attaacgagt tacctcgtec gagtggtagg caccaactac aaccagcatg	480
ctatggtggt cttcaagaaa gtttctcaaa acaggggagta cttcaagatc accctctacg	540
ggagaaccaa ggagctgact toggaactaa aggagaactt catccgcttc tccaaatctc	600
tgggcctccc tgaaaaccac atcgtcttcc ctgtcccaat cgaccagtgt atcgacggcg	660
gaggtagcga aaacctgtat tttcaggagg gccaggataa atgcaaaaaa gtgtatgaaa	720
attatccggt gagcaaatgc cagctggcga atcagtgcaa ttatgattgc aaactggata	780
aacatgcgcg tagcggcgaa tgcttttatg atgaaaaacg taatctgcag tgcatttgcg	840

-continued

attattgcga atattaatgc taaggatccc gga

873

<210> SEQ ID NO 29

<211> LENGTH: 831

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<220> FEATURE:

<221> NAME/KEY: CDS

<222> LOCATION: (1)..(831)

<400> SEQUENCE: 29

atg	gag	aca	gac	aca	ctc	ctg	cta	tgg	gta	ctg	ctg	ctc	tgg	gtt	cca	48
Met	Glu	Thr	Asp	Thr	Leu	Leu	Leu	Trp	Val	Leu	Leu	Leu	Trp	Val	Pro	
1				5				10						15		

ggt	tcc	act	ggt	gac	tac	aag	gac	gag	cat	cac	cat	cat	cac	cat	ggt	96
Gly	Ser	Thr	Gly	Asp	Tyr	Lys	Asp	Glu	His	His	His	His	His	His	Gly	
			20					25					30			

gga	agc	cag	gac	tcc	acc	tca	gac	ctg	atc	cca	gcc	cca	cct	ctg	agc	144
Gly	Ser	Gln	Asp	Ser	Thr	Ser	Asp	Leu	Ile	Pro	Ala	Pro	Pro	Leu	Ser	
		35					40					45				

aag	gtc	cct	ctg	cag	cag	aac	ttc	cag	gac	aac	caa	ttc	cag	ggg	aag	192
Lys	Val	Pro	Leu	Gln	Gln	Asn	Phe	Gln	Asp	Asn	Gln	Phe	Gln	Gly	Lys	
	50					55					60					

tgg	tat	gtg	gta	ggc	ctg	gca	ggg	aat	gca	att	ctc	aga	gaa	gac	aaa	240
Trp	Tyr	Val	Val	Gly	Leu	Ala	Gly	Asn	Ala	Ile	Leu	Arg	Glu	Asp	Lys	
65				70				75						80		

gac	ccg	caa	aag	atg	tat	gcc	acc	atc	tat	gag	ctg	aaa	gaa	gac	aag	288
Asp	Pro	Gln	Lys	Met	Tyr	Ala	Thr	Ile	Tyr	Glu	Leu	Lys	Glu	Asp	Lys	
			85					90					95			

agc	tac	aat	gtc	acc	tcc	gtc	ctg	ttt	agg	aaa	aag	aag	tgt	gac	tac	336
Ser	Tyr	Asn	Val	Thr	Ser	Val	Leu	Phe	Arg	Lys	Lys	Lys	Cys	Asp	Tyr	
			100					105						110		

tgg	atc	agg	act	ttt	gtt	cca	ggt	tgc	cag	ccc	ggc	gag	ttc	acg	ctg	384
Trp	Ile	Arg	Thr	Phe	Val	Pro	Gly	Cys	Gln	Pro	Gly	Glu	Phe	Thr	Leu	
			115				120					125				

ggc	aac	att	aag	agt	tac	cct	gga	tta	acg	agt	tac	ctc	gtc	cga	gtg	432
Gly	Asn	Ile	Lys	Ser	Tyr	Pro	Gly	Leu	Thr	Ser	Tyr	Leu	Val	Arg	Val	
	130					135					140					

gtg	agc	acc	aac	tac	aac	cag	cat	gct	atg	gtg	ttc	ttc	aag	aaa	gtt	480
Val	Ser	Thr	Asn	Tyr	Asn	Gln	His	Ala	Met	Val	Phe	Phe	Lys	Lys	Val	
145					150					155				160		

tct	caa	aac	agg	gag	tac	ttc	aag	atc	acc	ctc	tac	ggg	aga	acc	aag	528
Ser	Gln	Asn	Arg	Glu	Tyr	Phe	Lys	Ile	Thr	Leu	Tyr	Gly	Arg	Thr	Lys	
			165					170						175		

gag	ctg	act	tcg	gaa	cta	aag	gag	aac	ttc	atc	cgc	ttc	tcc	aaa	tct	576
Glu	Leu	Thr	Ser	Glu	Leu	Lys	Glu	Asn	Phe	Ile	Arg	Phe	Ser	Lys	Ser	
			180					185					190			

ctg	ggc	ctc	cct	gaa	aac	cac	atc	gtc	ttc	cct	gtc	cca	atc	gac	cag	624
Leu	Gly	Leu	Pro	Glu	Asn	His	Ile	Val	Phe	Pro	Val	Pro	Ile	Asp	Gln	
		195				200						205				

tgt	atc	gac	ggc	gga	ggt	agc	gaa	aac	ctg	tat	ttt	cag	gga	ggc	cag	672
Cys	Ile	Asp	Gly	Gly	Gly	Ser	Glu	Asn	Leu	Tyr	Phe	Gln	Gly	Gly	Gln	
	210					215					220					

gat	aaa	tgc	aaa	aaa	gtg	tat	gaa	aat	tat	ccg	gtg	agc	aaa	tgc	cag	720
Asp	Lys	Cys	Lys	Lys	Val	Tyr	Glu	Asn	Tyr	Pro	Val	Ser	Lys	Cys	Gln	
225					230					235				240		

```
<210> SEQ ID NO 30
<211> LENGTH: 277
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polypeptide
```

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1 5 10 15

Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly
20 25 30

Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser
35 40 45

Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys
50 55 60

Trp Tyr Val Val Gly Leu Ala Gly Asn Ala Ile Leu Arg Glu Asp Lys
65 70 75 80

Asp Pro Gln Lys Met Tyr Ala Thr Ile Tyr Glu Leu Lys Glu Asp Lys
85 90 95

Ser Tyr Asn Val Thr Ser Val Leu Phe Arg Lys Lys Lys Cys Asp Tyr
100 105 110

Trp Ile Arg Thr Phe Val Pro Gly Cys Gln Pro Gly Glu Phe Thr Leu
115 120 125

Gly Asn Ile Lys Ser Tyr Pro Gly Leu Thr Ser Tyr Leu Val Arg Val
130 135 140

Val	Ser	Thr	Asn	Tyr	Asn	Gln	His	Ala	Met	Val	Phe	Phe	Lys	Lys	Val
145					150					155					160

Ser Gln Asn Arg Glu Tyr Phe Lys Ile Thr Leu Tyr Gly Arg Thr Lys
165 170 175

Glu Leu Thr Ser Glu Leu Lys Glu Asn Phe Ile Arg Phe Ser Lys Ser
180 185 190

Leu Gly Leu Pro Glu Asn His Ile Val Phe Pro Val Pro Ile Asp Gln
105 200 305

Cys Ile Asp Gly Gly Gly Ser Glu Asn Leu Tyr Phe Gln Gly Gly Gln

Asp Lys Cys Lys Lys Val Tyr Glu Asn Tyr Pro Val Ser Lys Cys Gln
225 230 235 240

Leu Ala Asn Gln Cys Asn Tyr Asp Cys Lys Leu Asp Lys His Ala Arg

Ser Gly Glu Cys Phe Tyr Asp Glu Lys Arg Asn Leu Gln Cys Ile Cys
260 265 270

Asp Tyr Cys Glu Tyr

-continued

<210> SEQ ID NO 31
<211> LENGTH: 819
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 31

gactgagtcg cccgctcgag accatggaga cagacacact cctgctatgg gtactgctgc	60
tctgggttcc aggttcact ggtgactaca aggacgagca tcaccatcat caccatggtg	120
gaagccagga ctccacctca gacctgatcc cagccccacc tctgagcaag gtccctctgc	180
agcagaactt ccaggacaac caattccagg ggaagtggta tgtggtaggc ctggcaggga	240
atgcaattct cagagaagac aaagaccgcg aaaagatgta tgccaccatc tatgagctga	300
aagaagacaa gagctacaat gtcacctccg tcctgttttag gaaaaagaag tgtgactact	360
ggatcaggac ttttgttcca ggttgccagc ccggcgagtt cacgctgggc aacattaaga	420
gttacccctgg attaacgagt tacctcgtcc gagtggtagg caccaactac aaccagcatg	480
ctatggtggt cttcaagaaa gtttctcaaa acagggagta cttcaagatc accctctacg	540
ggagaaccaa ggagctgact tcggaactaa aggagaactt catccgcttc tccaaatctc	600
tgggcctccc tgaaaaccac atcgtcttcc ctgtcccaat cgaccagtgt atcgacggcg	660
gaggtagcga aaacctgtat tttcagggag gcatgtgcat gccgtgcttt accaccgatc	720
atcagatggc gcgtaaatgc gatgattgct gcggcgccaa aggccgtggc aaatgctatg	780
gcccgcagtg cctgtgccgt taatgctaag gatccccga	819

<210> SEQ ID NO 32
<211> LENGTH: 777
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(777)

<400> SEQUENCE: 32

atg gag aca gac aca ctc ctg cta tgg gta ctg ctg ctc tgg gtt cca	48
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro	
1 5 10 15	
ggt tcc act ggt gac tac aag gac gag cat cac cat cat cac cat ggt	96
Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly	
20 25 30	
gga agc cag gac tcc acc tca gac ctg atc cca gcc cca cct ctg agc	144
Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser	
35 40 45	
aag gtc cct ctg cag cag aac ttc cag gac aac caa ttc cag ggg aag	192
Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys	
50 55 60	
tgg tat gtg gta ggc ctg gca ggg aat gca att ctc aga gaa gac aaa	240
Trp Tyr Val Val Gly Leu Ala Gly Asn Ala Ile Leu Arg Glu Asp Lys	
65 70 75 80	
gac ccg caa aag atg tat gcc acc atc tat gag ctg aaa gaa gac aag	288
Asp Pro Gln Lys Met Tyr Ala Thr Ile Tyr Glu Leu Lys Glu Asp Lys	
85 90 95	

-continued

agc tac aat gtc acc tcc gtc ctg ttt agg aaa aag aag tgt gac tac	336
Ser Tyr Asn Val Thr Ser Val Leu Phe Arg Lys Lys Lys Cys Asp Tyr	
100 105 110	
tgg atc agg act ttt gtt cca ggt tgc cag ccc ggc gag ttc acg ctg	384
Trp Ile Arg Thr Phe Val Pro Gly Cys Gln Pro Gly Glu Phe Thr Leu	
115 120 125	
ggc aac att aag agt tac cct gga tta acg agt tac ctc gtc cga gtg	432
Gly Asn Ile Lys Ser Tyr Pro Gly Leu Thr Ser Tyr Leu Val Arg Val	
130 135 140	
gtg agc acc aac tac aac cag cat gct atg gtg ttc ttc aag aaa gtt	480
Val Ser Thr Asn Tyr Asn Gln His Ala Met Val Phe Phe Lys Lys Val	
145 150 155 160	
tct caa aac agg gag tac ttc aag atc acc ctc tac ggg aga acc aag	528
Ser Gln Asn Arg Glu Tyr Phe Lys Ile Thr Leu Tyr Gly Arg Thr Lys	
165 170 175	
gag ctg act tcg gaa cta aag gag aac ttc atc cgc ttc tcc aaa tct	576
Glu Leu Thr Ser Glu Leu Lys Glu Asn Phe Ile Arg Phe Ser Lys Ser	
180 185 190	
ctg ggc ctc cct gaa aac cac atc gtc ttc cct gtc cca atc gac cag	624
Leu Gly Leu Pro Glu Asn His Ile Val Phe Pro Val Pro Ile Asp Gln	
195 200 205	
tgt atc gac ggc gga ggt agc gaa aac ctg tat ttt cag gga ggc atg	672
Cys Ile Asp Gly Gly Gly Ser Glu Asn Leu Tyr Phe Gln Gly Gly Met	
210 215 220	
tgc atg ccg tgc ttt acc acc gat cat cag atg gcg cgt aaa tgc gat	720
Cys Met Pro Cys Phe Thr Thr Asp His Gln Met Ala Arg Lys Cys Asp	
225 230 235 240	
gat tgc tgc ggc ggc aaa ggc cgt ggc aaa tgc tat ggc ccg cag tgc	768
Asp Cys Cys Gly Gly Lys Gly Arg Gly Lys Cys Tyr Gly Pro Gln Cys	
245 250 255	
ctg tgc cgt	777
Leu Cys Arg	
<210> SEQ ID NO 33	
<211> LENGTH: 259	
<212> TYPE: PRT	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide	
<400> SEQUENCE: 33	
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro	
1 5 10 15	
Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly	
20 25 30	
Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser	
35 40 45	
Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys	
50 55 60	
Trp Tyr Val Val Gly Leu Ala Gly Asn Ala Ile Leu Arg Glu Asp Lys	
65 70 75 80	
Asp Pro Gln Lys Met Tyr Ala Thr Ile Tyr Glu Leu Lys Glu Asp Lys	
85 90 95	
Ser Tyr Asn Val Thr Ser Val Leu Phe Arg Lys Lys Lys Cys Asp Tyr	
100 105 110	
Trp Ile Arg Thr Phe Val Pro Gly Cys Gln Pro Gly Glu Phe Thr Leu	
115 120 125	

-continued

Gly Asn Ile Lys Ser Tyr Pro Gly Leu Thr Ser Tyr Leu Val Arg Val
130 135 140

Val Ser Thr Asn Tyr Asn Gln His Ala Met Val Phe Phe Lys Lys Val
145 150 155 160

Ser Gln Asn Arg Glu Tyr Phe Lys Ile Thr Leu Tyr Gly Arg Thr Lys
165 170 175

Glu Leu Thr Ser Glu Leu Lys Glu Asn Phe Ile Arg Phe Ser Lys Ser
180 185 190

Leu Gly Leu Pro Glu Asn His Ile Val Phe Pro Val Pro Ile Asp Gln
195 200 205

Cys Ile Asp Gly Gly Gly Ser Glu Asn Leu Tyr Phe Gln Gly Gly Met
210 215 220

Cys Met Pro Cys Phe Thr Thr Asp His Gln Met Ala Arg Lys Cys Asp
225 230 235 240

Asp Cys Cys Gly Gly Lys Gly Arg Gly Lys Cys Tyr Gly Pro Gln Cys
245 250 255

Leu Cys Arg

<210> SEQ ID NO 34
<211> LENGTH: 734
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 34

gactgagtcg cccgctcgag accatggaga cagacacact cctgctatgg gtactgctgc 60

tctgggttcc aggttcact ggtgactaca aggacgagca tcaccatcat caccatggtg 120

gaagccagga ctccacctca gacctgatcc cagccccacc tctgagcaag gtccctctgc 180

agcagaactt ccaggacaac caattccagg ggaagtggta tgtggtaggc ctggcaggga 240

atgcaattct cagagaagac aaagaccgcg aaaagatgta tgccaccatc tatgagctga 300

aagaagacaa gagctacaat gtcacctccg tcctgttttag gaaaaagaag tgtgactact 360

ggatcaggac ttttgttcca gggtgccagc ccggcgagtt cacgctgggc aacattaaga 420

gttaccttgg attaacgagt tacctcgtcc gagtgggtgag caccaactac aaccagcatg 480

ctatggtggt cttcaagaaa gtttctcaaa acagggagta cttcaagatc acctctacg 540

ggagaaccaa ggagctgact tcggaactaa aggagaactt catccgcttc tccaaatctc 600

tgggcctccc tgaaaaccac atcgtcttcc ctgtcccaat cgaccagtgt atcgacggcg 660

gaggtagcga aaacctgtat tttcagggat cctaattgtg gccatgatgt taggcggccg 720

ctaaggatcc cgga 734

<210> SEQ ID NO 35
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 35

Arg Ala Arg Tyr Lys Arg Ser
1 5

-continued

<210> SEQ ID NO 36
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 36

Arg Ala Arg Tyr Lys Arg Gly Ser
1 5

<210> SEQ ID NO 37
<211> LENGTH: 734
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 37

gactgagtcg cccgctcgag accatggaga cagacacact cctgctatgg gtactgctgc 60
tctgggttcc aggttccact ggtgactaca aggacgagca tcaccatcat caccatggtg 120
gaagccagga ctccacctca gacctgatcc cagccccacc tctgagcaag gtccctctgc 180
agcagaactt ccaggacaac caattccagg ggaagtggta tgtggtaggc ctggcagggg 240
atgcaattct cagagaagac aaagaccgcg aaaagatgta tgccaccatc tatgagctga 300
aagaagacaa gagctacaat gtcacctccg tctgttttag gaaaaagaag tgtgactact 360
ggatcaggac ttttgttcca ggttgccagc ccggcgagtt cacgctgggc aacattaaga 420
gttaccctgg attaacgagt tacctcgtcc gagtggtagg caccaactac aaccagcatg 480
ctatgggtgtt cttcaagaaa gtttctcaaa acagggagta cttcaagatc accctctacg 540
ggagaaccaa ggagctgact tcggaactaa aggagaactt catccgcttc tccaaatctc 600
tgggcctccc tgaaaaccac atcgctcttc ctgtcccaat cgaccagtgt atcgacggcg 660
gaggtagccg cgcgcgctat aaacgcggat cctaattgtg gccatgatgt taggcggccg 720
ctaaggatcc cgga 734

<210> SEQ ID NO 38
<211> LENGTH: 848
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 38

gactgagtcg cccgctcgag accatggaga cagacacact cctgctatgg gtactgctgc 60
tctgggttcc aggttccact ggtgactaca aggacgagca tcaccatcat caccatggtg 120
gaagccagga ctccacctca gacctgatcc cagccccacc tctgagcaag gtccctctgc 180
agcagaactt ccaggacaac caattccagg ggaagtggta tgtggtaggc ctggcagggg 240
atgcaattct cagagaagac aaagaccgcg aaaagatgta tgccaccatc tatgagctga 300
aagaagacaa gagctacaat gtcacctccg tctgttttag gaaaaagaag tgtgactact 360
ggatcaggac ttttgttcca ggttgccagc ccggcgagtt cacgctgggc aacattaaga 420

-continued

```

gttacccctgg attaacgagt tacctcgtcc gagtgggtgag caccaactac aaccagcatg 480
ctatgggtggt cttcaagaaa gttttctcaaa acaggggagta cttcaagatc accctctacg 540
ggagaaccaa ggagctgact tcggaactaa aggagaactt catccgcttc tccaaatctc 600
tgggcctccc tgaaaaccac atcgtcttcc ctgtcccaat cgaccagtgt atcgacggcg 660
gaggtagcga aaacctgtat tttcagagcg attgcaaata taaatttgaa aactggggcg 720
cgtgcgatgg cggcaccggc accaaagtgc gccagggcac cctgaaaaaa gcgcgctata 780
acgcgcagtg ccaggaacc attcgcgtga ccaaaccgtg ctaatgctgg atcccggacc 840
gcctctcc 848

```

```

<210> SEQ ID NO 39
<211> LENGTH: 798
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        polynucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(798)

```

```

<400> SEQUENCE: 39

```

```

atg gag aca gac aca ctc ctg cta tgg gta ctg ctg ctc tgg gtt cca 48
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1 5 10 15

ggt tcc act ggt gac tac aag gac gag cat cac cat cat cac cat ggt 96
Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly
20 25 30

gga agc cag gac tcc acc tca gac ctg atc cca gcc cca cct ctg agc 144
Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser
35 40 45

aag gtc cct ctg cag cag aac ttc cag gac aac caa ttc cag ggg aag 192
Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys
50 55 60

tgg tat gtg gta ggc ctg gca ggg aat gca att ctc aga gaa gac aaa 240
Trp Tyr Val Val Gly Leu Ala Gly Asn Ala Ile Leu Arg Glu Asp Lys
65 70 75 80

gac ccg caa aag atg tat gcc acc atc tat gag ctg aaa gaa gac aag 288
Asp Pro Gln Lys Met Tyr Ala Thr Ile Tyr Glu Leu Lys Glu Asp Lys
85 90 95

agc tac aat gtc acc tcc gtc ctg ttt agg aaa aag aag tgt gac tac 336
Ser Tyr Asn Val Thr Ser Val Leu Phe Arg Lys Lys Lys Cys Asp Tyr
100 105 110

tgg atc agg act ttt gtt cca ggt tgc cag ccc ggc gag ttc acg ctg 384
Trp Ile Arg Thr Phe Val Pro Gly Cys Gln Pro Gly Glu Phe Thr Leu
115 120 125

ggc aac att aag agt tac cct gga tta acg agt tac ctc gtc cga gtg 432
Gly Asn Ile Lys Ser Tyr Pro Gly Leu Thr Ser Tyr Leu Val Arg Val
130 135 140

gtg agc acc aac tac aac cag cat gct atg gtg ttc ttc aag aaa gtt 480
Val Ser Thr Asn Tyr Asn Gln His Ala Met Val Phe Phe Lys Lys Val
145 150 155 160

tct caa aac agg gag tac ttc aag atc acc ctc tac ggg aga acc aag 528
Ser Gln Asn Arg Glu Tyr Phe Lys Ile Thr Leu Tyr Gly Arg Thr Lys
165 170 175

gag ctg act tcg gaa cta aag gag aac ttc atc cgc ttc tcc aaa tct 576
Glu Leu Thr Ser Glu Leu Lys Glu Asn Phe Ile Arg Phe Ser Lys Ser
180 185 190

```

-continued

```
ctg ggc ctc cct gaa aac cac atc gtc ttc cct gtc cca atc gac cag      624
Leu Gly Leu Pro Glu Asn His Ile Val Phe Pro Val Pro Ile Asp Gln
      195                      200                      205

tgt atc gac ggc gga ggt agc gaa aac ctg tat ttt cag agc gat tgc      672
Cys Ile Asp Gly Gly Gly Ser Glu Asn Leu Tyr Phe Gln Ser Asp Cys
      210                      215                      220

aaa tat aaa ttt gaa aac tgg ggc gcg tgc gat ggc ggc acc ggc acc      720
Lys Tyr Lys Phe Glu Asn Trp Gly Ala Cys Asp Gly Gly Thr Gly Thr
      225                      230                      235                      240

aaa gtg cgc cag ggc acc ctg aaa aaa gcg cgc tat aac gcg cag tgc      768
Lys Val Arg Gln Gly Thr Leu Lys Lys Ala Arg Tyr Asn Ala Gln Cys
      245                      250                      255

cag gaa acc att cgc gtg acc aaa ccg tgc      798
Gln Glu Thr Ile Arg Val Thr Lys Pro Cys
      260                      265
```

<210> SEQ ID NO 40
<211> LENGTH: 266
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 40

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1      5      10      15

Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly
      20      25      30

Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser
      35      40      45

Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys
      50      55      60

Trp Tyr Val Val Gly Leu Ala Gly Asn Ala Ile Leu Arg Glu Asp Lys
65      70      75      80

Asp Pro Gln Lys Met Tyr Ala Thr Ile Tyr Glu Leu Lys Glu Asp Lys
      85      90      95

Ser Tyr Asn Val Thr Ser Val Leu Phe Arg Lys Lys Lys Cys Asp Tyr
      100     105     110

Trp Ile Arg Thr Phe Val Pro Gly Cys Gln Pro Gly Glu Phe Thr Leu
      115     120     125

Gly Asn Ile Lys Ser Tyr Pro Gly Leu Thr Ser Tyr Leu Val Arg Val
      130     135     140

Val Ser Thr Asn Tyr Asn Gln His Ala Met Val Phe Phe Lys Lys Val
145     150     155     160

Ser Gln Asn Arg Glu Tyr Phe Lys Ile Thr Leu Tyr Gly Arg Thr Lys
      165     170     175

Glu Leu Thr Ser Glu Leu Lys Glu Asn Phe Ile Arg Phe Ser Lys Ser
      180     185     190

Leu Gly Leu Pro Glu Asn His Ile Val Phe Pro Val Pro Ile Asp Gln
      195     200     205

Cys Ile Asp Gly Gly Gly Ser Glu Asn Leu Tyr Phe Gln Ser Asp Cys
      210     215     220

Lys Tyr Lys Phe Glu Asn Trp Gly Ala Cys Asp Gly Gly Thr Gly Thr
225     230     235     240
```

-continued

Lys	Val	Arg	Gln	Gly	Thr	Leu	Lys	Lys	Ala	Arg	Tyr	Asn	Ala	Gln	Cys
			245						250					255	
Gln	Glu	Thr	Ile	Arg	Val	Thr	Lys	Pro	Cys						
		260					265								
<210> SEQ ID NO 41															
<211> LENGTH: 796															
<212> TYPE: DNA															
<213> ORGANISM: Artificial Sequence															
<220> FEATURE:															
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide															
<400> SEQUENCE: 41															
gactgagtcg cccgctcgag accatggaga cagacacact cctgctatgg gtactgctgc														60	
tctgggttcc aggttccact ggtgactaca aggacgagca tcaccatcat caccatgggtg														120	
gaagccagga ctccacctca gacctgatcc cagccccacc tctgagcaag gtccctctgc														180	
agcagaactt ccaggacaac caattccagg ggaagtggta tgtggtaggc ctggcagggg														240	
atgcaattct cagagaagac aaagaccgcg aaaagatgta tgccaccatc tatgagctga														300	
aagaagacaa gagctacaat gtcacctccg tcctgttttag gaaaaagaag tgtgactact														360	
ggatcaggac ttttgttcca ggttgccagc ccggcgagtt cacgctgggc aacattaaga														420	
gttaccttgg attaacgagt tacctcgtcc gagtggtagg caccaactac aaccagcatg														480	
ctatgggtgtt cttcaagaaa gtttctcaaa acagggagta cttcaagatc accctctacg														540	
ggagaaccaa ggagctgact tcggaactaa aggagaactt catccgcttc tccaaatctc														600	
tgggcctccc tgaaaaccac atcgtcttcc ctgtcccaat cgaccagtgt atcgacggcg														660	
gaggtagcga aaacctgtat tttcagggag gcagcgccat cagctgcggc gagacctgct														720	
tcaagttcaa gtgctacacc ccagatgca gctgcagcta ccccggtgtgc aagtaagcta														780	
aggatcccg accgcc														796	
<210> SEQ ID NO 42															
<211> LENGTH: 750															
<212> TYPE: DNA															
<213> ORGANISM: Artificial Sequence															
<220> FEATURE:															
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide															
<220> FEATURE:															
<221> NAME/KEY: CDS															
<222> LOCATION: (1)..(750)															
<400> SEQUENCE: 42															
atg	gag	aca	gac	aca	ctc	ctg	cta	tgg	gta	ctg	ctg	ctc	tgg	gtt	cca
Met	Glu	Thr	Asp	Thr	Leu	Leu	Leu	Trp	Val	Leu	Leu	Leu	Trp	Val	Pro
1				5				10					15		
ggt	tcc	act	ggt	gac	tac	aag	gac	gag	cat	cac	cat	cat	cac	cat	ggt
Gly	Ser	Thr	Gly	Asp	Tyr	Lys	Asp	Glu	His	His	His	His	His	His	Gly
			20				25					30			
gga	agc	cag	gac	tcc	acc	tca	gac	ctg	atc	cca	gcc	cca	cct	ctg	agc
Gly	Ser	Gln	Asp	Ser	Thr	Ser	Asp	Leu	Ile	Pro	Ala	Pro	Pro	Leu	Ser
		35					40				45				
aag	gtc	cct	ctg	cag	cag	aac	ttc	cag	gac	aac	caa	ttc	cag	ggg	aag
Lys	Val	Pro	Leu	Gln	Gln	Asn	Phe	Gln	Asp	Asn	Gln	Phe	Gln	Gly	Lys
	50				55				60						
tgg	tat	gtg	gta	ggc	ctg	gca	ggg	aat	gca	att	ctc	aga	gaa	gac	aaa
Trp	Tyr	Val	Val	Gly	Leu	Ala	Gly	Asn	Ala	Ile	Leu	Arg	Glu	Asp	Lys

-continued																				
65					70					75					80					
gac	ccg	caa	aag	atg	tat	gcc	acc	atc	tat	gag	ctg	aaa	gaa	gac	aag	288				
Asp	Pro	Gln	Lys	Met	Tyr	Ala	Thr	Ile	Tyr	Glu	Leu	Lys	Glu	Asp	Lys					
				85					90				95							
agc	tac	aat	gtc	acc	tcc	gtc	ctg	ttt	agg	aaa	aag	aag	tgt	gac	tac	336				
Ser	Tyr	Asn	Val	Thr	Ser	Val	Leu	Phe	Arg	Lys	Lys	Lys	Cys	Asp	Tyr					
				100					105				110							
tgg	atc	agg	act	ttt	gtt	cca	ggg	tgc	cag	ccc	ggc	gag	ttc	acg	ctg	384				
Trp	Ile	Arg	Thr	Phe	Val	Pro	Gly	Cys	Gln	Pro	Gly	Glu	Phe	Thr	Leu					
				115					120				125							
ggc	aac	att	aag	agt	tac	cct	gga	tta	acg	agt	tac	ctc	gtc	cga	gtg	432				
Gly	Asn	Ile	Lys	Ser	Tyr	Pro	Gly	Leu	Thr	Ser	Tyr	Leu	Val	Arg	Val					
				130					135				140							
gtg	agc	acc	aac	tac	aac	cag	cat	gct	atg	gtg	ttc	ttc	aag	aaa	gtt	480				
Val	Ser	Thr	Asn	Tyr	Asn	Gln	His	Ala	Met	Val	Phe	Phe	Lys	Lys	Val					
				145					150				155				160			
tct	caa	aac	agg	gag	tac	ttc	aag	atc	acc	ctc	tac	ggg	aga	acc	aag	528				
Ser	Gln	Asn	Arg	Glu	Tyr	Phe	Lys	Ile	Thr	Leu	Tyr	Gly	Arg	Thr	Lys					
				165					170				175							
gag	ctg	act	tcg	gaa	cta	aag	gag	aac	ttc	atc	cgc	ttc	tcc	aaa	tct	576				
Glu	Leu	Thr	Ser	Glu	Leu	Lys	Glu	Asn	Phe	Ile	Arg	Phe	Ser	Lys	Ser					
				180					185				190							
ctg	ggc	ctc	cct	gaa	aac	cac	atc	gtc	ttc	cct	gtc	cca	atc	gac	cag	624				
Leu	Gly	Leu	Pro	Glu	Asn	His	Ile	Val	Phe	Pro	Val	Pro	Ile	Asp	Gln					
				195					200				205							
tgt	atc	gac	ggc	gga	ggg	agc	gaa	aac	ctg	tat	ttt	cag	gga	ggc	agc	672				
Cys	Ile	Asp	Gly	Gly	Gly	Ser	Glu	Asn	Leu	Tyr	Phe	Gln	Gly	Gly	Ser					
				210					215				220							
gcc	atc	agc	tgc	ggc	gag	acc	tgc	ttc	aag	ttc	aag	tgc	tac	acc	ccc	720				
Ala	Ile	Ser	Cys	Gly	Glu	Thr	Cys	Phe	Lys	Phe	Lys	Cys	Tyr	Thr	Pro					
				225					230				235				240			
aga	tgc	agc	tgc	agc	tac	ccc	gtg	tgc	aag							750				
Arg	Cys	Ser	Cys	Ser	Tyr	Pro	Val	Cys	Lys											
				245					250											
<210> SEQ ID NO 43																				
<211> LENGTH: 250																				
<212> TYPE: PRT																				
<213> ORGANISM: Artificial Sequence																				
<220> FEATURE:																				
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide																				
<400> SEQUENCE: 43																				
Met	Glu	Thr	Asp	Thr	Leu	Leu	Leu	Trp	Val	Leu	Leu	Leu	Trp	Val	Pro					
1					5					10				15						
Gly	Ser	Thr	Gly	Asp	Tyr	Lys	Asp	Glu	His	His	His	His	His	His	Gly					
				20					25				30							
Gly	Ser	Gln	Asp	Ser	Thr	Ser	Asp	Leu	Ile	Pro	Ala	Pro	Pro	Leu	Ser					
				35					40				45							
Lys	Val	Pro	Leu	Gln	Gln	Asn	Phe	Gln	Asp	Asn	Gln	Phe	Gln	Gly	Lys					
				50					55				60							
Trp	Tyr	Val	Val	Gly	Leu	Ala	Gly	Asn	Ala	Ile	Leu	Arg	Glu	Asp	Lys					
65					70				75				80							
Asp	Pro	Gln	Lys	Met	Tyr	Ala	Thr	Ile	Tyr	Glu	Leu	Lys	Glu	Asp	Lys					
				85					90				95							
Ser	Tyr	Asn	Val	Thr	Ser	Val	Leu	Phe	Arg	Lys	Lys	Lys	Cys	Asp	Tyr					
				100					105				110							

-continued

Trp Ile Arg Thr Phe Val Pro Gly Cys Gln Pro Gly Glu Phe Thr Leu
115 120 125

Gly Asn Ile Lys Ser Tyr Pro Gly Leu Thr Ser Tyr Leu Val Arg Val
130 135 140

Val Ser Thr Asn Tyr Asn Gln His Ala Met Val Phe Phe Lys Lys Val
145 150 155 160

Ser Gln Asn Arg Glu Tyr Phe Lys Ile Thr Leu Tyr Gly Arg Thr Lys
165 170 175

Glu Leu Thr Ser Glu Leu Lys Glu Asn Phe Ile Arg Phe Ser Lys Ser
180 185 190

Leu Gly Leu Pro Glu Asn His Ile Val Phe Pro Val Pro Ile Asp Gln
195 200 205

Cys Ile Asp Gly Gly Gly Ser Glu Asn Leu Tyr Phe Gln Gly Gly Ser
210 215 220

Ala Ile Ser Cys Gly Glu Thr Cys Phe Lys Phe Lys Cys Tyr Thr Pro
225 230 235 240

Arg Cys Ser Cys Ser Tyr Pro Val Cys Lys
245 250

<210> SEQ ID NO 44
<211> LENGTH: 745
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 44

gactgagtcg cccgctcgag accatggaga cagacacact cctgctatgg gtactgctgc 60

tctgggttcc aggttccact ggtgactaca aggacgagca tcaccatcat caccatggtg 120

gaagccagga ctccacctca gacctgatcc cagccccacc tctgagcaag gtccctctgc 180

agcagaactt ccaggacaac caattccagg ggaagtggta tgtggtaggc ctggcaggga 240

atgcaattct cagagaagac aaagaccgcg aaaagatgta tgccaccatc tatgagctga 300

aagaagacaa gagctacaat gtcacctccg tcctgttttag gaaaaagaag tgtgactact 360

ggatcaggac ttttgttcca gggtgccagc ccggcgagtt cacgctgggc aacattaaga 420

gttaccctgg attaacgagt tacctcgtcc gagtggtagg caccaactac aaccagcatg 480

ctatgggtgtt cttcaagaaa gtttctcaaa acagggagta cttcaagatc accctctacg 540

ggagaaccaa ggagctgact tcggaactaa aggagaactt catccgcttc tccaaatctc 600

tgggcctccc tgaaaaccac atcgtcttcc ctgtcccaat cgaccagtgt atcgacggcg 660

gaggtagcga aaacctgtat tttcagggag gcgtgtgctg cggctacaag ctgtgccacc 720

cctgctaagc taaggatccc ggacc 745

<210> SEQ ID NO 45
<211> LENGTH: 702
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(702)

-continued

<400> SEQUENCE: 45	
atg gag aca gac aca ctc ctg cta tgg gta ctg ctg ctc tgg gtt cca Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro 1 5 10 15	48
ggt tcc act ggt gac tac aag gac gag cat cac cat cat cac cat ggt Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly 20 25 30	96
gga agc cag gac tcc acc tca gac ctg atc cca gcc cca cct ctg agc Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser 35 40 45	144
aag gtc cct ctg cag cag aac ttc cag gac aac caa ttc cag ggg aag Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys 50 55 60	192
tgg tat gtg gta ggc ctg gca ggg aat gca att ctc aga gaa gac aaa Trp Tyr Val Val Gly Leu Ala Gly Asn Ala Ile Leu Arg Glu Asp Lys 65 70 75 80	240
gac ccg caa aag atg tat gcc acc atc tat gag ctg aaa gaa gac aag Asp Pro Gln Lys Met Tyr Ala Thr Ile Tyr Glu Leu Lys Glu Asp Lys 85 90 95	288
agc tac aat gtc acc tcc gtc ctg ttt agg aaa aag aag tgt gac tac Ser Tyr Asn Val Thr Ser Val Leu Phe Arg Lys Lys Lys Cys Asp Tyr 100 105 110	336
tgg atc agg act ttt gtt cca ggt tgc cag ccc ggc gag ttc acg ctg Trp Ile Arg Thr Phe Val Pro Gly Cys Gln Pro Gly Glu Phe Thr Leu 115 120 125	384
ggc aac att aag agt tac cct gga tta acg agt tac ctc gtc cga gtg Gly Asn Ile Lys Ser Tyr Pro Gly Leu Thr Ser Tyr Leu Val Arg Val 130 135 140	432
gtg agc acc aac tac aac cag cat gct atg gtg ttc ttc aag aaa gtt Val Ser Thr Asn Tyr Asn Gln His Ala Met Val Phe Phe Lys Lys Val 145 150 155 160	480
tct caa aac agg gag tac ttc aag atc acc ctc tac ggg aga acc aag Ser Gln Asn Arg Glu Tyr Phe Lys Ile Thr Leu Tyr Gly Arg Thr Lys 165 170 175	528
gag ctg act tcg gaa cta aag gag aac ttc atc cgc ttc tcc aaa tct Glu Leu Thr Ser Glu Leu Lys Glu Asn Phe Ile Arg Phe Ser Lys Ser 180 185 190	576
ctg ggc ctc cct gaa aac cac atc gtc ttc cct gtc cca atc gac cag Leu Gly Leu Pro Glu Asn His Ile Val Phe Pro Val Pro Ile Asp Gln 195 200 205	624
tgt atc gac ggc gga ggt agc gaa aac ctg tat ttt cag gga ggc gtg Cys Ile Asp Gly Gly Gly Ser Glu Asn Leu Tyr Phe Gln Gly Gly Val 210 215 220	672
tgc tgc ggc tac aag ctg tgc cac ccc tgc Cys Cys Gly Tyr Lys Leu Cys His Pro Cys 225 230	702
<210> SEQ ID NO 46	
<211> LENGTH: 234	
<212> TYPE: PRT	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide	
<400> SEQUENCE: 46	
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro 1 5 10 15	
Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly	

-continued																
20					25					30						
Gly	Ser	Gln	Asp	Ser	Thr	Ser	Asp	Leu	Ile	Pro	Ala	Pro	Pro	Leu	Ser	
35					40					45						
Lys	Val	Pro	Leu	Gln	Gln	Asn	Phe	Gln	Asp	Asn	Gln	Phe	Gln	Gly	Lys	
50					55					60						
Trp	Tyr	Val	Val	Gly	Leu	Ala	Gly	Asn	Ala	Ile	Leu	Arg	Glu	Asp	Lys	
65					70					75						
Asp	Pro	Gln	Lys	Met	Tyr	Ala	Thr	Ile	Tyr	Glu	Leu	Lys	Glu	Asp	Lys	
85					90					95						
Ser	Tyr	Asn	Val	Thr	Ser	Val	Leu	Phe	Arg	Lys	Lys	Lys	Cys	Asp	Tyr	
100					105					110						
Trp	Ile	Arg	Thr	Phe	Val	Pro	Gly	Cys	Gln	Pro	Gly	Glu	Phe	Thr	Leu	
115					120					125						
Gly	Asn	Ile	Lys	Ser	Tyr	Pro	Gly	Leu	Thr	Ser	Tyr	Leu	Val	Arg	Val	
130					135					140						
Val	Ser	Thr	Asn	Tyr	Asn	Gln	His	Ala	Met	Val	Phe	Phe	Lys	Lys	Val	
145					150					155						
Ser	Gln	Asn	Arg	Glu	Tyr	Phe	Lys	Ile	Thr	Leu	Tyr	Gly	Arg	Thr	Lys	
165					170					175						
Glu	Leu	Thr	Ser	Glu	Leu	Lys	Glu	Asn	Phe	Ile	Arg	Phe	Ser	Lys	Ser	
180					185					190						
Leu	Gly	Leu	Pro	Glu	Asn	His	Ile	Val	Phe	Pro	Val	Pro	Ile	Asp	Gln	
195					200					205						
Cys	Ile	Asp	Gly	Gly	Gly	Ser	Glu	Asn	Leu	Tyr	Phe	Gln	Gly	Gly	Val	
210					215					220						
Cys	Cys	Gly	Tyr	Lys	Leu	Cys	His	Pro	Cys							
225					230											
<210> SEQ ID NO 47																
<211> LENGTH: 751																
<212> TYPE: DNA																
<213> ORGANISM: Artificial Sequence																
<220> FEATURE:																
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide																
<400> SEQUENCE: 47																
gactgagtcg cccgctcgag accatggaga cagacacact cctgctatgg gtactgctgc															60	
tctgggttcc aggttcact ggtgactaca aggacgagca tcaccatcat caccatggtg															120	
gaagccagga ctccacctca gacctgatcc cagccccacc tctgagcaag gtccctctgc															180	
agcagaactt ccaggacaac caattccagg ggaagtggta tgtggtaggc ctggcaggga															240	
atgcaattct cagagaagac aaagaccgcg aaaagatgta tgccaccatc tatgagctga															300	
aagaagacaa gagctacaat gtcacctccg tcctgttttag gaaaaagaag tgtgactact															360	
ggatcaggac ttttgttcca gggtgccagc ccggcgagtt cacgctgggc aacattaaga															420	
gttaccctgg attaacgagt tacctcgtcc gagtggtagg caccaactac aaccagcatg															480	
ctatgggtgtt cttcaagaaa gtttctcaaa acagggagta cttcaagatc acctctacg															540	
ggagaaccaa ggagctgact tcggaactaa aggagaactt catccgcttc tccaaatctc															600	
tgggcctccc tgaaaaccac atcgtcttcc ctgtcccaat cgaccagtgt atcgacggcg															660	
gaggtagcga aaacctgtat ttccagggag gcaacggcgt gtgctgcggc tacaagctgt															720	
gccaccctg ctaagctaag gatcccgac c															751	

-continued

<210> SEQ ID NO 48
<211> LENGTH: 708
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..(708)

<400> SEQUENCE: 48

atg gag aca gac aca ctc ctg cta tgg gta ctg ctg ctc tgg gtt cca 48
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1 5 10 15

ggt tcc act ggt gac tac aag gac gag cat cac cat cat cac cat ggt 96
Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly
20 25 30

gga agc cag gac tcc acc tca gac ctg atc cca gcc cca cct ctg agc 144
Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser
35 40 45

aag gtc cct ctg cag cag aac ttc cag gac aac caa ttc cag ggg aag 192
Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys
50 55 60

tgg tat gtg gta ggc ctg gca ggg aat gca att ctc aga gaa gac aaa 240
Trp Tyr Val Val Gly Leu Ala Gly Asn Ala Ile Leu Arg Glu Asp Lys
65 70 75 80

gac ccg caa aag atg tat gcc acc atc tat gag ctg aaa gaa gac aag 288
Asp Pro Gln Lys Met Tyr Ala Thr Ile Tyr Glu Leu Lys Glu Asp Lys
85 90 95

agc tac aat gtc acc tcc gtc ctg ttt agg aaa aag aag tgt gac tac 336
Ser Tyr Asn Val Thr Ser Val Leu Phe Arg Lys Lys Lys Cys Asp Tyr
100 105 110

tgg atc agg act ttt gtt cca ggt tgc cag ccc ggc gag ttc acg ctg 384
Trp Ile Arg Thr Phe Val Pro Gly Cys Gln Pro Gly Glu Phe Thr Leu
115 120 125

ggc aac att aag agt tac cct gga tta acg agt tac ctc gtc cga gtg 432
Gly Asn Ile Lys Ser Tyr Pro Gly Leu Thr Ser Tyr Leu Val Arg Val
130 135 140

gtg agc acc aac tac aac cag cat gct atg gtg ttc ttc aag aaa gtt 480
Val Ser Thr Asn Tyr Asn Gln His Ala Met Val Phe Phe Lys Lys Val
145 150 155 160

tct caa aac agg gag tac ttc aag atc acc ctc tac ggg aga acc aag 528
Ser Gln Asn Arg Glu Tyr Phe Lys Ile Thr Leu Tyr Gly Arg Thr Lys
165 170 175

gag ctg act tcg gaa cta aag gag aac ttc atc cgc ttc tcc aaa tct 576
Glu Leu Thr Ser Glu Leu Lys Glu Asn Phe Ile Arg Phe Ser Lys Ser
180 185 190

ctg ggc ctc cct gaa aac cac atc gtc ttc cct gtc cca atc gac cag 624
Leu Gly Leu Pro Glu Asn His Ile Val Phe Pro Val Pro Ile Asp Gln
195 200 205

tgt atc gac ggc gga ggt agc gaa aac ctg tat ttt cag gga ggc aac 672
Cys Ile Asp Gly Gly Gly Ser Glu Asn Leu Tyr Phe Gln Gly Gly Asn
210 215 220

ggc gtg tgc tgc ggc tac aag ctg tgc cac ccc tgc 708
Gly Val Cys Cys Gly Tyr Lys Leu Cys His Pro Cys
225 230 235

<210> SEQ ID NO 49

-continued

<211> LENGTH: 236
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 49

Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1 5 10 15
Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly
20 25 30
Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser
35 40 45
Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys
50 55 60
Trp Tyr Val Val Gly Leu Ala Gly Asn Ala Ile Leu Arg Glu Asp Lys
65 70 75 80
Asp Pro Gln Lys Met Tyr Ala Thr Ile Tyr Glu Leu Lys Glu Asp Lys
85 90 95
Ser Tyr Asn Val Thr Ser Val Leu Phe Arg Lys Lys Lys Cys Asp Tyr
100 105 110
Trp Ile Arg Thr Phe Val Pro Gly Cys Gln Pro Gly Glu Phe Thr Leu
115 120 125
Gly Asn Ile Lys Ser Tyr Pro Gly Leu Thr Ser Tyr Leu Val Arg Val
130 135 140
Val Ser Thr Asn Tyr Asn Gln His Ala Met Val Phe Phe Lys Lys Val
145 150 155 160
Ser Gln Asn Arg Glu Tyr Phe Lys Ile Thr Leu Tyr Gly Arg Thr Lys
165 170 175
Glu Leu Thr Ser Glu Leu Lys Glu Asn Phe Ile Arg Phe Ser Lys Ser
180 185 190
Leu Gly Leu Pro Glu Asn His Ile Val Phe Pro Val Pro Ile Asp Gln
195 200 205
Cys Ile Asp Gly Gly Gly Ser Glu Asn Leu Tyr Phe Gln Gly Gly Asn
210 215 220
Gly Val Cys Cys Gly Tyr Lys Leu Cys His Pro Cys
225 230 235

<210> SEQ ID NO 50
<211> LENGTH: 831
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 50

gactgagtcg cccgctcgag accatggaga cagacacact cctgctatgg gtactgctgc 60
tctgggttcc aggttcact ggtgactaca aggacgagca tcaccatcat caccatggtg 120
gaagccagga ctccacctca gacctgatcc cagccccacc tctgagcaag gtccctctgc 180
agcagaactt ccaggacaac caattccagg ggaagtggta tgtggtaggc ctggcagggg 240
atgcaattct cagagaagac aaagaccgcg aaaagatgta tgccaccatc tatgagctga 300
aagaagacaa gagctacaat gtcacctccg tcctgttttag gaaaaagaag tgtgactact 360

-continued

```

ggatcaggac ttttgttcca ggttgccagc ccggcgagtt cacgctgggc aacattaaga 420
gttaccctgg attaacgagt tacctcgtec gagtgggtgag caccaactac aaccagcatg 480
ctatggtggt cttcaagaaa gttttctcaaa acagggagta cttcaagatc accctctacg 540
ggagaaccaa ggagctgact tcggaactaa aggagaactt catccgcttc tccaaatctc 600
tgggcctccc tgaaaaccac atcgtcttcc ctgtcccaat cgaccagtgt atcgacggcg 660
gaggtagcga aaacctgtat tttcagggag gccagcagca tgcggatccg atttgcaaca 720
aaccgtgcaa aacccatgat gattgcagcg gcgcgtgggt ttgccaggcg tgctggaaca 780
gcgcgcgcac ctgcggcccg tatgtgggct aatgctaagg atcccgacc g 831

```

<210> SEQ ID NO 51

<211> LENGTH: 789

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<220> FEATURE:

<221> NAME/KEY: CDS

<222> LOCATION: (1)..(786)

<400> SEQUENCE: 51

```

atg gag aca gac aca ctc ctg cta tgg gta ctg ctg ctc tgg gtt cca 48
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1 5 10 15

ggt tcc act ggt gac tac aag gac gag cat cac cat cat cac cat ggt 96
Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly
20 25 30

gga agc cag gac tcc acc tca gac ctg atc cca gcc cca cct ctg agc 144
Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser
35 40 45

aag gtc cct ctg cag cag aac ttc cag gac aac caa ttc cag ggg aag 192
Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys
50 55 60

tgg tat gtg gta ggc ctg gca ggg aat gca att ctc aga gaa gac aaa 240
Trp Tyr Val Val Gly Leu Ala Gly Asn Ala Ile Leu Arg Glu Asp Lys
65 70 75 80

gac ccg caa aag atg tat gcc acc atc tat gag ctg aaa gaa gac aag 288
Asp Pro Gln Lys Met Tyr Ala Thr Ile Tyr Glu Leu Lys Glu Asp Lys
85 90 95

agc tac aat gtc acc tcc gtc ctg ttt agg aaa aag aag tgt gac tac 336
Ser Tyr Asn Val Thr Ser Val Leu Phe Arg Lys Lys Lys Cys Asp Tyr
100 105 110

tgg atc agg act ttt gtt cca ggt tgc cag ccc ggc gag ttc acg ctg 384
Trp Ile Arg Thr Phe Val Pro Gly Cys Gln Pro Gly Glu Phe Thr Leu
115 120 125

ggc aac att aag agt tac cct gga tta acg agt tac ctc gtc cga gtg 432
Gly Asn Ile Lys Ser Tyr Pro Gly Leu Thr Ser Tyr Leu Val Arg Val
130 135 140

gtg agc acc aac tac aac cag cat gct atg gtg ttc ttc aag aaa gtt 480
Val Ser Thr Asn Tyr Asn Gln His Ala Met Val Phe Phe Lys Lys Val
145 150 155 160

tct caa aac agg gag tac ttc aag atc acc ctc tac ggg aga acc aag 528
Ser Gln Asn Arg Glu Tyr Phe Lys Ile Thr Leu Tyr Gly Arg Thr Lys
165 170 175

gag ctg act tcg gaa cta aag gag aac ttc atc cgc ttc tcc aaa tct 576
Glu Leu Thr Ser Glu Leu Lys Glu Asn Phe Ile Arg Phe Ser Lys Ser
180 185 190

```

-continued

```
ctg ggc ctc cct gaa aac cac atc gtc ttc cct gtc cca atc gac cag      624
Leu Gly Leu Pro Glu Asn His Ile Val Phe Pro Val Pro Ile Asp Gln
      195                      200                      205

tgt atc gac ggc gga ggt agc gaa aac ctg tat ttt cag gga ggc cag      672
Cys Ile Asp Gly Gly Gly Ser Glu Asn Leu Tyr Phe Gln Gly Gly Gln
      210                      215                      220

cag cat gcg gat ccg att tgc aac aaa ccg tgc aaa acc cat gat gat      720
Gln His Ala Asp Pro Ile Cys Asn Lys Pro Cys Lys Thr His Asp Asp
      225                      230                      235                      240

tgc agc ggc gcg tgg ttt tgc cag gcg tgc tgg aac agc gcg cgc acc      768
Cys Ser Gly Ala Trp Phe Cys Gln Ala Cys Trp Asn Ser Ala Arg Thr
      245                      250                      255

tgc ggc ccg tat gtg ggc taa      789
Cys Gly Pro Tyr Val Gly
      260
```

<210> SEQ ID NO 52
<211> LENGTH: 262
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 52

```
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro
1      5      10      15

Gly Ser Thr Gly Asp Tyr Lys Asp Glu His His His His His His Gly
      20      25      30

Gly Ser Gln Asp Ser Thr Ser Asp Leu Ile Pro Ala Pro Pro Leu Ser
      35      40      45

Lys Val Pro Leu Gln Gln Asn Phe Gln Asp Asn Gln Phe Gln Gly Lys
      50      55      60

Trp Tyr Val Val Gly Leu Ala Gly Asn Ala Ile Leu Arg Glu Asp Lys
65      70      75      80

Asp Pro Gln Lys Met Tyr Ala Thr Ile Tyr Glu Leu Lys Glu Asp Lys
      85      90      95

Ser Tyr Asn Val Thr Ser Val Leu Phe Arg Lys Lys Lys Cys Asp Tyr
      100     105     110

Trp Ile Arg Thr Phe Val Pro Gly Cys Gln Pro Gly Glu Phe Thr Leu
      115     120     125

Gly Asn Ile Lys Ser Tyr Pro Gly Leu Thr Ser Tyr Leu Val Arg Val
      130     135     140

Val Ser Thr Asn Tyr Asn Gln His Ala Met Val Phe Phe Lys Lys Val
145     150     155     160

Ser Gln Asn Arg Glu Tyr Phe Lys Ile Thr Leu Tyr Gly Arg Thr Lys
      165     170     175

Glu Leu Thr Ser Glu Leu Lys Glu Asn Phe Ile Arg Phe Ser Lys Ser
      180     185     190

Leu Gly Leu Pro Glu Asn His Ile Val Phe Pro Val Pro Ile Asp Gln
      195     200     205

Cys Ile Asp Gly Gly Gly Ser Glu Asn Leu Tyr Phe Gln Gly Gly Gln
      210     215     220

Gln His Ala Asp Pro Ile Cys Asn Lys Pro Cys Lys Thr His Asp Asp
225     230     235     240
```

-continued

Cys Ser Gly Ala Trp Phe Cys Gln Ala Cys Trp Asn Ser Ala Arg Thr
245 250 255

Cys Gly Pro Tyr Val Gly
260

<210> SEQ ID NO 53
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
peptide

<400> SEQUENCE: 53

Ser Ala Trp Ser His Pro Gln Phe Glu Lys
1 5 10

<210> SEQ ID NO 54
<211> LENGTH: 36
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (1)..(36)
<223> OTHER INFORMATION: D-amino acids

<400> SEQUENCE: 54

Met Cys Met Pro Cys Phe Thr Thr Asp His Gln Met Ala Arg Lys Cys
1 5 10 15

Asp Asp Cys Cys Gly Gly Lys Gly Arg Gly Lys Cys Tyr Gly Pro Gln
20 25 30

Cys Leu Cys Arg
35

<210> SEQ ID NO 55
<211> LENGTH: 36
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polypeptide
<220> FEATURE:
<223> OTHER INFORMATION: N-term H
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (1)..(36)
<223> OTHER INFORMATION: D-amino acids
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (15)..(15)
<223> OTHER INFORMATION: dArg, dAla or dLys
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (23)..(23)
<223> OTHER INFORMATION: dArg, dAla or dLys
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: dArg, dAla or dLys
<220> FEATURE:
<223> OTHER INFORMATION: C-term OH

<400> SEQUENCE: 55

Met Cys Met Pro Cys Phe Thr Thr Asp His Gln Met Ala Arg Xaa Cys
1 5 10 15

-continued

Asp Asp Cys Cys Gly Gly Xaa Gly Arg Gly Xaa Cys Tyr Gly Pro Gln
20 25 30
Cys Leu Cys Arg
35

<210> SEQ ID NO 56
<211> LENGTH: 36
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<223> OTHER INFORMATION: N-term H
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (1)..(36)
<223> OTHER INFORMATION: D-amino acids
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (15)..(15)
<223> OTHER INFORMATION: dArg, dAla or dLys
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (23)..(23)
<223> OTHER INFORMATION: dArg, dAla or dLys
<220> FEATURE:
<223> OTHER INFORMATION: C-term OH

<400> SEQUENCE: 56

Met Cys Met Pro Cys Phe Thr Thr Asp His Gln Met Ala Arg Xaa Cys
1 5 10 15
Asp Asp Cys Cys Gly Gly Xaa Gly Arg Gly Lys Cys Tyr Gly Pro Gln
20 25 30
Cys Leu Cys Arg
35

<210> SEQ ID NO 57
<211> LENGTH: 36
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<223> OTHER INFORMATION: N-term H
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (1)..(36)
<223> OTHER INFORMATION: D-amino acids
<220> FEATURE:
<223> OTHER INFORMATION: C-term OH

<400> SEQUENCE: 57

Met Cys Met Pro Cys Phe Thr Thr Asp His Gln Met Ala Arg Arg Cys
1 5 10 15
Asp Asp Cys Cys Gly Gly Arg Gly Arg Gly Lys Cys Tyr Gly Pro Gln
20 25 30
Cys Leu Cys Arg
35

<210> SEQ ID NO 58
<211> LENGTH: 36
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<223> OTHER INFORMATION: N-term H
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (1)..(36)
<223> OTHER INFORMATION: D-amino acids
<220> FEATURE:
<223> OTHER INFORMATION: C-term OH

<400> SEQUENCE: 58

Met Cys Met Pro Cys Phe Thr Thr Asp His Gln Met Ala Arg Arg Cys
1 5 10 15

Asp Asp Cys Cys Gly Gly Ala Gly Arg Gly Lys Cys Tyr Gly Pro Gln
20 25 30

Cys Leu Cys Arg
35

<210> SEQ ID NO 59
<211> LENGTH: 36
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<223> OTHER INFORMATION: N-term H
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (1)..(36)
<223> OTHER INFORMATION: D-amino acids
<220> FEATURE:
<223> OTHER INFORMATION: C-term OH

<400> SEQUENCE: 59

Met Cys Met Pro Cys Phe Thr Thr Asp His Gln Met Ala Arg Ala Cys
1 5 10 15

Asp Asp Cys Cys Gly Gly Ala Gly Arg Gly Lys Cys Tyr Gly Pro Gln
20 25 30

Cys Leu Cys Arg
35

<210> SEQ ID NO 60
<211> LENGTH: 36
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<223> OTHER INFORMATION: N-term H
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (1)..(36)
<223> OTHER INFORMATION: D-amino acids
<220> FEATURE:
<223> OTHER INFORMATION: C-term OH

<400> SEQUENCE: 60

Met Cys Met Pro Cys Phe Thr Thr Asp His Gln Met Ala Arg Ala Cys
1 5 10 15

Asp Asp Cys Cys Gly Gly Arg Gly Arg Gly Lys Cys Tyr Gly Pro Gln
20 25 30

Cys Leu Cys Arg
35

-continued

<210> SEQ ID NO 61
<211> LENGTH: 36
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<223> OTHER INFORMATION: N-term H
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (1)..(36)
<223> OTHER INFORMATION: D-amino acids
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (15)..(15)
<223> OTHER INFORMATION: dArg, dAla or dLys
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (23)..(23)
<223> OTHER INFORMATION: dArg, dAla or dLys
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: dLys(6-ICGhexanoyl)
<220> FEATURE:
<223> OTHER INFORMATION: C-term OH

<400> SEQUENCE: 61

Met Cys Met Pro Cys Phe Thr Thr Asp His Gln Met Ala Arg Xaa Cys
1 5 10 15

Asp Asp Cys Cys Gly Gly Xaa Gly Arg Gly Lys Cys Tyr Gly Pro Gln
 20 25 30

Cys Leu Cys Arg
 35

<210> SEQ ID NO 62
<211> LENGTH: 36
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<223> OTHER INFORMATION: N-term H
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (1)..(36)
<223> OTHER INFORMATION: D-amino acids
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: dLys(6-ICGhexanoyl)
<220> FEATURE:
<223> OTHER INFORMATION: C-term OH

<400> SEQUENCE: 62

Met Cys Met Pro Cys Phe Thr Thr Asp His Gln Met Ala Arg Arg Cys
1 5 10 15

Asp Asp Cys Cys Gly Gly Arg Gly Arg Gly Lys Cys Tyr Gly Pro Gln
 20 25 30

Cys Leu Cys Arg
 35

<210> SEQ ID NO 63
<211> LENGTH: 36
<212> TYPE: PRT

-continued

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<223> OTHER INFORMATION: N-term H
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (1)..(36)
<223> OTHER INFORMATION: D-amino acids
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: dLys(6-ICGhexanoyl)
<220> FEATURE:
<223> OTHER INFORMATION: C-term OH

<400> SEQUENCE: 63

Met Cys Met Pro Cys Phe Thr Thr Asp His Gln Met Ala Arg Arg Cys
1 5 10 15

Asp Asp Cys Cys Gly Gly Ala Gly Arg Gly Lys Cys Tyr Gly Pro Gln
20 25 30

Cys Leu Cys Arg
35

<210> SEQ ID NO 64
<211> LENGTH: 36
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<223> OTHER INFORMATION: N-term H
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (1)..(36)
<223> OTHER INFORMATION: D-amino acids
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: dLys(6-ICGhexanoyl)
<220> FEATURE:
<223> OTHER INFORMATION: C-term OH

<400> SEQUENCE: 64

Met Cys Met Pro Cys Phe Thr Thr Asp His Gln Met Ala Arg Ala Cys
1 5 10 15

Asp Asp Cys Cys Gly Gly Ala Gly Arg Gly Lys Cys Tyr Gly Pro Gln
20 25 30

Cys Leu Cys Arg
35

<210> SEQ ID NO 65
<211> LENGTH: 36
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<223> OTHER INFORMATION: N-term H
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (1)..(36)
<223> OTHER INFORMATION: D-amino acids
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (27)..(27)

-continued

<223> OTHER INFORMATION: dLys(6-ICGhexanoyl)

<220> FEATURE:

<223> OTHER INFORMATION: C-term OH

<400> SEQUENCE: 65

Met Cys Met Pro Cys Phe Thr Thr Asp His Gln Met Ala Arg Ala Cys
1 5 10 15

Asp Asp Cys Cys Gly Gly Arg Gly Arg Gly Lys Cys Tyr Gly Pro Gln
20 25 30

Cys Leu Cys Arg
35

<210> SEQ ID NO 66

<211> LENGTH: 36

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<220> FEATURE:

<221> NAME/KEY: misc_feature

<222> LOCATION: (1)..(36)

<223> OTHER INFORMATION: D-amino acids

<220> FEATURE:

<221> NAME/KEY: MOD_RES

<222> LOCATION: (15)..(15)

<223> OTHER INFORMATION: Lys, Ala or Arg

<220> FEATURE:

<221> NAME/KEY: MOD_RES

<222> LOCATION: (23)..(23)

<223> OTHER INFORMATION: Lys, Ala or Arg

<220> FEATURE:

<221> NAME/KEY: MOD_RES

<222> LOCATION: (27)..(27)

<223> OTHER INFORMATION: Lys, Ala or Arg

<220> FEATURE:

<221> NAME/KEY: misc_feature

<222> LOCATION: (1)..(36)

<223> OTHER INFORMATION: Some or all positions may disclose D-amino acids

<220> FEATURE:

<223> OTHER INFORMATION: See specification as filed for detailed description of substitutions and preferred embodiments

<400> SEQUENCE: 66

Met Cys Met Pro Cys Phe Thr Thr Asp His Gln Met Ala Arg Xaa Cys
1 5 10 15

Asp Asp Cys Cys Gly Gly Xaa Gly Arg Gly Xaa Cys Tyr Gly Pro Gln
20 25 30

Cys Leu Cys Arg
35

<210> SEQ ID NO 67

<211> LENGTH: 6

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 67

Arg Ala Arg Tyr Lys Arg
1 5

<210> SEQ ID NO 68

<211> LENGTH: 4

-continued

<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 68

Cys Ile Asp Gly
1

<210> SEQ ID NO 69
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 69

Glu Asn Leu Tyr Phe Gln
1 5

<210> SEQ ID NO 70
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic 6xHis tag

<400> SEQUENCE: 70

His His His His His His
1 5

<210> SEQ ID NO 71
<211> LENGTH: 36
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 71

Met Cys Met Pro Cys Phe Thr Thr Asp His Gln Met Ala Arg Ala Cys
1 5 10 15
Asp Asp Cys Cys Gly Gly Ala Gly Arg Gly Lys Cys Tyr Gly Pro Gln
20 25 30
Cys Leu Cys Arg
35

<210> SEQ ID NO 72
<211> LENGTH: 191
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 72

ctgtaacctg atcctgctag cgggtctcaa ggtggatcca tgtgcatgcc gtgcttcacc 60
accgataacg aaatggcggc gaactgcgat cgctgctgcg gtggccgtgg ccgtggccgc 120
tgctatggcc cgcagtgccg gtgccgctaa tgcggccgct agtgagaccg cttgcatggc 180
tagtacctac g 191

-continued

<210> SEQ ID NO 73
<211> LENGTH: 191
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 73

ctgtaacctg atcctgctag cgggtctcaa ggtggatcca tgtgcatgcc gtgcttcacc 60
accgataacg atatggcgga tgaatgcgat cgctgctgcg gtggccgtgg ccgtggccgc 120
tgctatggcc cgcagtgcct gtgccgctaa tgcggccgct agtgagaccg cttgcatggc 180
tagtacctac g 191

<210> SEQ ID NO 74
<211> LENGTH: 36
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (21)..(27)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (29)..(32)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (34)..(34)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (36)..(36)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 74

Met Cys Met Pro Cys Phe Thr Thr Asp Asn Glu Met Ala Ala Asn Cys
1 5 10 15
Asp Arg Cys Cys Xaa Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa
20 25 30
Cys Xaa Cys Xaa
35

<210> SEQ ID NO 75
<211> LENGTH: 36
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (21)..(27)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (29)..(32)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (34)..(34)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:

-continued

<221> NAME/KEY: MOD_RES
<222> LOCATION: (36)..(36)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 75

Met Cys Met Pro Cys Phe Thr Thr Asp Asn Asp Met Ala Asp Glu Cys
1 5 10 15

Asp Arg Cys Cys Xaa Xaa Xaa Xaa Xaa Xaa Xaa Cys Xaa Xaa Xaa Xaa
20 25 30

Cys Xaa Cys Xaa
35

<210> SEQ ID NO 76
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (13)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 76

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Asp Gly Xaa Met Ala Asn
1 5 10 15

Ile Cys Asp Arg Cys Cys Gly Gly Arg Gly Gly Gly Arg Cys Tyr Gly
20 25 30

Pro Gln Cys Leu Cys Arg
35

<210> SEQ ID NO 77
<211> LENGTH: 45
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (12)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 77

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Phe Xaa Xaa Met Ala Arg
1 5 10 15

Arg Cys Asp Asp Cys Cys Gly Gly Arg Gly Gly Ala Ala Ala Met Ala
20 25 30

Arg Ser Ala Cys Ala Ala Asn Ala Ala Ala His His His
35 40 45

<210> SEQ ID NO 78
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (11)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 78

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Xaa Xaa Xaa Met Glu Tyr
1 5 10 15
Arg Cys Asp Asp Cys Cys Gly Gly Arg Gly Ile Gly Arg Cys Tyr Gly
20 25 30
Pro Gln Cys Leu Cys Arg
35

<210> SEQ ID NO 79
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (12)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 79

Gly Ser Glu Cys Met Pro Cys Phe Thr Thr Asp Xaa Xaa Met Ala Arg
1 5 10 15
Arg Cys Asp Glu Cys Cys Gly Gly Arg Gly Arg Gly Arg Cys Tyr Gly
20 25 30
Pro Gln Cys Leu Cys Arg
35

<210> SEQ ID NO 80
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (11)..(11)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (13)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 80

Gly Ser Met Cys Met Pro Cys Phe Thr Thr Xaa Pro Xaa Met Ala Gln
1 5 10 15
Arg Cys Asp Asp Cys Cys Gly Gly Arg Gly Arg Gly Arg Cys Tyr Gly
20 25 30
Pro Gln Cys Leu Cys Arg
35

<210> SEQ ID NO 81
<211> LENGTH: 38

-continued

<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (11)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 81

Gly Ser Met Cys Met Pro Cys Phe Thr Thr Xaa Xaa Xaa Met Ile Trp
1 5 10 15
Arg Cys Asp Asp Cys Cys Gly Gly Arg Leu Ile Gly Tyr Cys Tyr Gly
20 25 30
Pro Gln Cys Leu Cys Arg
35

<210> SEQ ID NO 82
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (12)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 82

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Ala Xaa Xaa Met Ala Ala
1 5 10 15
Ile Cys Asp Asp Cys Cys Gly Gly Arg Gly Arg Gly Arg Cys Tyr Gly
20 25 30
Pro Gln Cys Leu Cys Arg
35

<210> SEQ ID NO 83
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (12)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 83

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Leu Xaa Xaa Met Ala Ala
1 5 10 15
Arg Cys Asp Asp Cys Cys Gly Glu Arg His Arg Gly Arg Cys Tyr Gly
20 25 30
Pro Gln Cys Leu Cys Arg
35

-continued

<210> SEQ ID NO 84
<211> LENGTH: 45
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (13)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 84

Gly Ser Arg Cys Met Pro Cys Phe Thr Thr Ile Thr Xaa Met Ala Arg
1 5 10 15

Arg Cys Asp Asp Cys Cys Gly Gly Arg Gly Arg Ala Ala Ala Met Ala
20 25 30

Arg Ser Ala Cys Ala Ala Asn Ala Ala Ala His His His
35 40 45

<210> SEQ ID NO 85
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (12)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 85

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Val Xaa Xaa Met Ala Trp
1 5 10 15

Arg Cys Asp Asp Cys Cys Gly Gly Arg Gly Arg Gly Arg Cys Tyr Gly
20 25 30

Pro Gln Cys Leu Cys Arg
35

<210> SEQ ID NO 86
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (2)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (11)..(13)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (27)..(27)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 86

Gly Xaa Xaa Ser Met Pro Cys Phe Thr Thr Xaa Xaa Xaa Met Ala Tyr
1 5 10 15

-continued

Phe Cys Asp Asp Cys Cys Gly Gly Arg Ser Xaa Gly Ile Cys Val Gly
20 25 30

Pro Gln Cys Leu Cys Arg
35

<210> SEQ ID NO 87
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (12)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 87

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Asp Xaa Xaa Met Ala Arg
1 5 10 15

Arg Cys Asp Asp Cys Cys Gly Asp Tyr Gly Asp Gly Asp Cys Gln Gly
20 25 30

Pro Gln Cys Leu Cys Arg
35

<210> SEQ ID NO 88
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (13)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 88

Gly Ser Cys Ala Cys Arg Ala Ser Pro Pro Ile Ile Xaa Trp Arg Ala
1 5 10 15

Ala Ala Met Ile Ala Ala Ala Ala Ala Ala Met Val Ser Ala Met Ala
20 25 30

Arg Ile Ala Cys Ala Ala
35

<210> SEQ ID NO 89
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (11)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 89

-continued

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Xaa Xaa Xaa Met Ala Arg
1 5 10 15
Arg Cys Asp Asp Cys Cys Gly Gly Tyr Leu His Gly Glu Cys Leu Gly
20 25 30
Pro Gln Cys Leu Cys Arg
35

<210> SEQ ID NO 90
<211> LENGTH: 45
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (11)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 90

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Xaa Xaa Xaa Met Gln Gln
1 5 10 15
Ser Cys Asp Arg Cys Cys Gly Gly Arg Ala Val Ala Ala Ala Met Ala
20 25 30
Arg Ser Val Cys Ala Ala Asn Ala Ala Ala His His His
35 40 45

<210> SEQ ID NO 91
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (11)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 91

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Xaa Xaa Xaa Met Ala Arg
1 5 10 15
Arg Cys Asp Asp Cys Cys Gly Gly Asp Gly Asp Gly Asn Cys Arg Gly
20 25 30
Pro Gln Cys Leu Cys Arg
35

<210> SEQ ID NO 92
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (2)..(3)

-continued

<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (12)..(14)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 92

Gly Xaa Xaa Cys Met Pro Cys Phe Thr Thr Asn Xaa Xaa Xaa Ala Leu
1 5 10 15
Ala Cys Asp Arg Cys Cys Gly Gly Arg Gly Arg Gly Arg Cys Tyr Gly
20 25 30
Pro Gln Cys Leu Cys Arg
35

<210> SEQ ID NO 93
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (12)..(13)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (35)..(36)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 93

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Asp Xaa Xaa Met Ala Ile
1 5 10 15
Ile Cys Asp Arg Cys Cys Gly Gly Arg Gly Arg Gly Arg Cys Tyr Gly
20 25 30
Pro Gln Xaa Xaa Cys Arg
35

<210> SEQ ID NO 94
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (11)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 94

Gly Ser Xaa Cys Met Pro Ser Ser Pro Pro Xaa Xaa Xaa
1 5 10

<210> SEQ ID NO 95
<211> LENGTH: 45
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (11)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 95

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Xaa Xaa Xaa Met Ala Ser
1 5 10 15
His Cys Asp Asp Cys Cys Gly Gly Arg Gly Leu Gly Gly Cys Trp Gly
20 25 30
Arg Ser Ala Cys Ala Ala Asn Ala Ala Ala His His His
35 40 45

<210> SEQ ID NO 96
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 96

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Asp His Gln Met Ala Ala
1 5 10 15
Glu Cys Asp Arg Cys Cys Gly Ser Arg Gly Arg Gly Arg Cys Tyr Gly
20 25 30
Pro Gln Cys Leu Cys Arg
35

<210> SEQ ID NO 97
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (11)..(12)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 97

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Xaa Xaa Trp Met Ala Trp
1 5 10 15
Arg Cys Asp Arg Cys Cys Gly Gly Gly Gly Arg Gly Arg Cys Tyr Gly
20 25 30
Pro Gln Tyr Leu Cys Arg
35

<210> SEQ ID NO 98
<211> LENGTH: 38

-continued

<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(4)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (12)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 98

Gly Ser Xaa Xaa Met Pro Cys Phe Thr Thr Asp Xaa Xaa Met Glu His
1 5 10 15

Tyr Cys Asp Arg Cys Cys Gly Gly Asp Gly Arg Gly Arg Cys Tyr Gly
20 25 30

Pro Gln Cys Leu Cys Arg
35

<210> SEQ ID NO 99
<211> LENGTH: 49
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (11)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 99

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Xaa Xaa Xaa Met Ala Arg
1 5 10 15

Arg Ser Met Ile Ala Ala Ala Ala Leu Ala Met Ala Ala Ala Met Ala
20 25 30

Arg Ser Ala Cys Ala Val Met Arg Pro Leu Ile Thr Ile Asn His His
35 40 45

His

<210> SEQ ID NO 100
<211> LENGTH: 49
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (2)..(4)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (11)..(14)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (26)..(26)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES

-continued

<222> LOCATION: (34)..(34)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 100

Gly Xaa Xaa Xaa Met Pro Cys Phe Thr Thr Xaa Xaa Xaa Xaa Ala Arg
1 5 10 15

Ala Cys Asp Phe Cys Ser Val Ala Ala Xaa Val Ala Val Leu Trp Pro
20 25 30

Ala Xaa Pro Val Pro Leu Met Arg Pro Leu Ile Thr Ile Asn His His
35 40 45

His

<210> SEQ ID NO 101
<211> LENGTH: 49
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 101

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Asp His Ile Met Met Thr
1 5 10 15

Trp Cys Asp Glu Cys Cys Gly Pro Arg Gln Trp Pro Leu Leu Trp Pro
20 25 30

Ala Val Pro Val Pro Leu Met Arg Pro Leu Ile Thr Ile Asn His His
35 40 45

His

<210> SEQ ID NO 102
<211> LENGTH: 45
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (11)..(12)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 102

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Xaa Xaa Val Met Ile Ala
1 5 10 15

Ile Cys Asp Ile Cys Cys Gly Gly Arg Gly Glu Gly Ala Cys Tyr Gly
20 25 30

Pro Leu Cys Leu Ala Ala Asn Ala Ala Ala His His His
35 40 45

<210> SEQ ID NO 103
<211> LENGTH: 45
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic

-continued

```

    polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (12)..(12)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 103

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Asp Xaa Gly Met Ala Gly
1          5          10          15

Ser Cys Asp Arg Cys Cys Gly Ile Arg Ala Val Ala Ala Ala Met Ala
          20          25          30

Arg Ala Ala Cys Ala Ala Asn Ala Ala Ala His His His
          35          40          45

<210> SEQ ID NO 104
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
    polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (12)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 104

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Phe Xaa Xaa Met Tyr Ile
1          5          10          15

Arg Cys Asp Asp Cys Cys Gly Gly Arg Gly Arg Gly Arg Cys Asn Gly
          20          25          30

Pro Gln Cys Leu Cys Arg
          35

<210> SEQ ID NO 105
<211> LENGTH: 50
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
    polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 105

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Arg Ser Tyr Asp Gly Val
1          5          10          15

Leu Leu Arg Leu Val Leu Arg Trp Pro Trp Pro Arg Arg Val Leu Arg
          20          25          30

Pro Val Val Pro Val Pro Leu Met Arg Pro Leu Ile Thr Ile Asn His
          35          40          45

His His
          50
```

-continued

<210> SEQ ID NO 106
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (12)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 106

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Tyr Xaa Xaa Met Trp Ile
1 5 10 15

Arg Cys Asp Asp Cys Cys Gly Gly Arg Gly Arg Gly Arg Cys Tyr Gly
20 25 30

Pro Gln Cys Leu Cys Arg
35

<210> SEQ ID NO 107
<211> LENGTH: 48
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (12)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 107

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Asp Xaa Xaa Met Ala Arg
1 5 10 15

Arg Cys Asp His Cys Cys Gly Gly Arg Gly Ile Cys Tyr Gly Pro Gln
20 25 30

Leu Pro Val Pro Leu Met Arg Pro Leu Ile Thr Ile Asn His His His
35 40 45

<210> SEQ ID NO 108
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (11)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 108

Gly Ser Met Cys Met Ser Cys Phe Thr Thr Xaa Xaa Xaa Met Ala Arg
1 5 10 15

Arg Cys Asp Asp Cys Cys Gly Gly Ile Gly Phe Gly Ile Cys Gln Gly
20 25 30

Pro Gln Cys Leu Cys Arg

-continued

35

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Asp Xaa Xaa Met Ala Arg
1 5 10 15
Arg Cys Asp Asp Cys Cys Gly Gly Asp Gly Tyr Gly Ile Cys His Gly
20 25 30
Pro Gln Cys Leu Cys Arg
35

Gly Ser Xaa Cys Met Pro Cys Phe Thr Thr Xaa Xaa Xaa Met Ala Glu
1 5 10 15
Thr Cys Asp Asp Cys Cys Gly Gly Arg Gly Arg Gly Arg Cys Tyr Gly
20 25 30
Pro Gln Cys Leu Cys Arg
35

```

<210> SEQ ID NO 111
<211> LENGTH: 38
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polypeptide
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)..(3)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (11)..(13)
<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 111

```

-continued

Gly	Ser	Xaa	Cys	Met	Pro	Cys	Phe	Thr	Thr	Xaa	Xaa	Xaa	Met	Ala	Ile
1				5					10					15	

Gln	Cys	Asp	Arg	Cys	Cys	Ser	Gly	Asp	Gly	Arg	Gly	Arg	Cys	Tyr	Gly
		20					25						30		

Pro	Gln	Cys	Leu	Cys	Arg
		35			

<210> SEQ ID NO 112

<211> LENGTH: 38

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<220> FEATURE:

<221> NAME/KEY: MOD_RES

<222> LOCATION: (11)..(13)

<223> OTHER INFORMATION: Any amino acid

<400> SEQUENCE: 112

Gly	Ser	Met	Cys	Met	Pro	Cys	Phe	Thr	Thr	Xaa	Xaa	Xaa	Met	Ala	Arg
1				5					10					15	

Arg	Cys	Asp	Asp	Cys	Cys	Gly	Ile	Tyr	Gly	His	Gly	His	Cys	Trp	Gly
		20					25						30		

Pro	Gln	Cys	Leu	Cys	Arg
		35			

1. A method for identifying drug candidates having a pharmacological property, the method comprising:

administering, to a subject, a composition comprising a plurality of drug candidates;

obtaining, from the subject, a sample comprising at least some of the drug candidates in the plurality; and

analyzing the sample to determine the identity of the at least some of the drug candidates having the pharmacological property.

2. The method of claim 1, wherein the drug candidates are selected from the group consisting of small chemical molecules, biologics, and peptides.

3. The method of claim 1, wherein the plurality of drug candidates comprises a plurality of peptides.

4. The method of claim 1, wherein the pharmacological property comprises oral bioavailability, capability to pass the blood-brain barrier, exclusion by the blood-brain barrier, serum half-life, capability to penetrate cells, capability to enter subcellular organelles or other cellular domains, capability to target organs or tissue, capability to target cancerous tissue, or a combination thereof.

5. The method of claim 3, further comprising generating a peptide library comprising the at least some of the peptides having the pharmacological property.

6. The method of claim 3, further comprising screening the at least some of the peptides to identify which peptides exhibit an activity for inhibiting a protein:protein interaction, inhibiting antagonism of a receptor, inhibiting binding of an agonist to a receptor, modulating an ion channel, inhibiting a signaling pathway, activating a signaling pathway, and/or a inhibiting a protein: small molecule interaction.

7. The method of claim 6, wherein the protein:protein interaction is associated with a disease or disorder selected

from the group consisting of a cancer, an infectious disease, an inflammatory disease, an immune disease, a metabolic disease, a cardiac disease, an aging-related disease, and a neurologic disease.

8. (canceled)

9. The method of claim 3, further comprising obtaining from the subject a plurality of samples, each sample being obtained at a different time point after administering the plurality of peptides.

10. The method of claim 9, analyzing at least some samples of the plurality of samples to determine the identity of the at least some of the peptides having the pharmacological property in each sample of the plurality.

11. The method of claim 3, wherein the identity of the at least some of the peptides is determined using mass spectrometry, wherein each of the peptides comprise a unique mass signature or digest fragment mass signature detected by a mass spectrometer.

12. The method of claim 1, wherein the plurality of drug candidates comprises greater than five drug candidates.

13. The method of claim 1, wherein the plurality of drug candidates comprises greater than 10 drug candidates, greater than 100 drug candidates, greater than 1000 drug candidates, or greater than 10000 drug candidates.

14-19. (canceled)

20. The method of claim 3, wherein at least one peptide of the plurality of peptides comprises a detectable label for tracing the at least one peptide in the subject.

21. The method of claim 20, wherein the detectable label comprises a near infrared dye.

22. The method of claim 3, wherein at least one peptide of the plurality of peptides comprises a hydrophobic moiety conjugated to the N-terminus of the at least one peptide,

wherein the at least one peptide exhibits an increased half-life as compared to the at least one peptide lacking the hydrophobic moiety.

23. The method of claim **22**, wherein the hydrophobic moiety comprises a hydrophobic fluorescent dye or a saturated or unsaturated alkyl group.

24. The method of claim **1**, wherein the sample comprises a tissue sample or a fluid sample.

25. The method of claim **24**, wherein the fluid sample comprises blood, urine, mucous or spinal fluid.

26. (canceled)

27. The method of claim **24**, wherein the tissue sample comprises biopsy or necropsy tissue.

28. The method of claim **24**, wherein the tissue comprises tissue from a brain, a lung, a kidney, a muscle, a liver, a heart, a stomach, a pancreas, or an organ.

29. The method of claim **1**, further comprising using the pharmacological property common to the drug candidates to generate additional drug candidates having the pharmacological property.

30. (canceled)

31. The method of claim **1**, wherein the subject is a human.

32-126. (canceled)

* * * * *