METHODS FOR SCREENING

CROSS REFERENCES TO RELATED APPLICATIONS

[0001] The present application claims the benefit of U.S. Provisional Application No. 61/735,516 filed December 10, 2012 and U.S. Provisional Application No. 61/794,685 filed March 15, 2013, both of which are 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 numbers R01CA135491, NIH ROI AI059543, AI094419, and AI097786 by the National Institutes of Health.

SEQUENCE LISTING

[0002.1] The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on March 7, 2014, is named 44189-701.601_SL.txt and is 152,593 bytes in size.

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, the 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 new 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). In various aspects, the present disclosure relates to methods for identifying a drug candidate having a pharmacological property, the method comprising, analyzing an isolated sample from a subject following administration of a plurality of drug candidates to the subject; and identifying in the isolated sample at least one drug candidate having the pharmacological property.

[0007] In some aspects, the present disclosure relates to methods 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.

[0008] In some aspects, the present disclosure relates to methods of generating a mass-defined drug candidate library where the method comprises producing a plurality of drug candidates, at least some of the drug candidates each having a unique mass signature or digest fragment mass signature and 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. In certain embodiments, the methods 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.

[0009] In some aspects, the present disclosure relates to methods for identifying a library of drug candidates having a pharmacological property where the method comprises analyzing an isolated sample from a subject following administration of a plurality of drug candidates to the subject. In certain embodiments, the library of drug candidates are from the mass-defined drug candidate

library provided herein. In certain embodiments, the methods include identifying in the isolated sample at least one drug candidate having the pharmacological property.

[0010] In some aspects, the present disclosure relates to methods for identifying library drug candidates having a pharmacological property where the method comprises administering to a subject a plurality of library drug candidates, wherein the library drug candidates are from the mass-defined drug candidate library provided herein, obtaining, from the subject, a sample comprising at least some of the plurality of library drug candidates; and analyzing the sample to determine the identity of the at least some of the plurality of library drug candidates having the pharmacological property.

[0011] In some aspects, the present disclosure provides methods of determining a distribution profile of knotted-peptides administered to a subject by different administration pathways where the method comprises 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; 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; and 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.

[0012] In some aspects, the present disclosure provides peptides for imaging a tumor in a subject, the peptide comprising a chlorotoxin comprising 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 comprises a detectable label.

[0013] In some aspects of the present disclosure, methods are provided for treating a disease associated with cells expressing a chlorotoxin target, the method comprising: administering, to a subject in need thereof, a therapeutically effective amount of a pharmaceutical composition comprising a peptide according to the present disclosure or a composition comprising a peptide of the present disclosure, thereby treating the disease.

INCORPORATION BY REFERENCE

[0014] 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

[0015] 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:

[0016] FIG. 1 provides a gel image showing a variety of scaffolds, in accordance with an embodiment of the present invention.

[0017] 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. FIG. 2 discloses SEQ ID NO: 74.

[0018] FIG. 3 illustrates an example method of generating peptide libraries, in accordance with an embodiment of the present invention.

[0019] FIG. 4 shows an example method of producing peptide libraries, in accordance with an embodiment of the present invention.

[0020] 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.

[0021] FIG. 6 shows analysis of a large library using mass spectrometry, in accordance with an embodiment of the present invention.

[0022] FIG. 7 shows an example method of using siderocalin fusions to express knottin variants, in accordance with an embodiment of the present invention.

[0023] FIG. 8 depicts SDS-PAGE analysis of expressing knottin scaffolds, in accordance with embodiments of the present invention.

[0024] FIG. 9 provides a schematic of pooled library production, in accordance with an embodiment of the present invention. FIG. 9B is an expanded view of FIG. 9A. FIG. 9A discloses SEQ ID NOS 75-78, respectively, in order of appearance, and FIG. 9B discloses SEQ ID NOS 79-80, respectively, in order of appearance.

[0025] FIG. 10 describes representative sequencing data from a cloned knottin library, in accordance with an embodiment of the present invention. FIG. 10 discloses SEQ ID NOS 81-109, respectively, in order of appearance.

[0026] FIG. 11 shows SDS-PAGE analysis of 3000 member knottin libraries, in accordance with embodiments of the present invention.

[0027] FIG. 12 shows an assessment of mutational tolerance by reverse proteomics, in accordance with embodiments of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0028] The present disclosure relates to systems and methods for drug discovery. In certain aspects, the present disclosure relates to methods for identifying drug candidates having a desired pharmacological property. In further aspects, the present disclosure relates to methods of generating mass-defined drug candidate libraries. In other aspects, the present disclosure relates to methods of determining a distribution profile of knotted-peptides administered to a subject by different administration pathways.

Generating Selected Libraries

[0029] In one aspect, the present invention relates to methods and systems for generating libraries of potential drug candidates (e.g., libraries of small chemical molecules, biologics, peptides, and any derivatives or fragments thereof) that can be screened to identify sub-libraries of potential drug candidates (e.g., sub-libraries of small chemical molecules, biologics, peptides, and any derivatives or fragments thereof) 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 pharmacological 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. In some embodiments, 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. In certain embodiments, the hydrophobic moiety comprises a hydrophobic fluorescent dye or a saturated or unsaturated alkyl group.

[0030] In some aspects of the present disclosure, the plurality of drug candidates comprises a plurality of peptides.

[0031] In some aspects of the present disclosure, the plurality of drug candidates comprises a plurality of small chemical molecules.

[0032] In some aspects of the present disclosure, the plurality of drug candidates comprises a plurality of biologics.

[0033] In some embodiments, 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, or a combination thereof. In certain embodiments, the pharmacological property comprises oral bioavailability.

[0034] As used herein, the term "drug candidate" means any agent that could potentially elicit an effect on a target or subject. In some aspects, the drug candidate may have a therapeutic effect. Exemplary drug candidates may include, but are not limited to, small chemical molecules, biologics, peptides, and any derivatives or fragments thereof.

[0035] As used herein, a "pharmacological property" is a property of a drug candidate that can be used to characterize the potential of a drug candidate to elicit an effect on a target or subject. Exemplary pharmacologic properties may include, but are not limited to, 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, and the like.

[0036] 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.

[0037] Large numbers of drug candidates can be administered to a subject and screened to determine, 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.

[0038] 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.

[0039] As used herein, the term "protein:protein interaction" means a physical contact between one protein and another which can be transient or permanent. The interaction of a first protein with a second protein may cause an effect on either the first or the second protein such that the effect changes either the first or the second or both proteins in a detectable manner.

[0040] 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., small chemical molecules, biologics, peptides, and any derivatives or fragments thereof) having one or more drug-like properties, such as those described above. For example, in a first screen, thousands of candidates (e.g., small chemical molecules, biologics, peptides, and any derivatives or fragments thereof) 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.

[0041] 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.

[0042] In some aspects, the drug candidate is a small chemical molecule. With small chemical molecule libraries, e.g., the small chemical can have a variety of common characteristics that cause the small chemical molecules to have certain drug-like properties. Rules can be identified to help define which of the small chemical molecules may be more likely to have the drug-like properties. In various aspects, rules can include parameters for absorption, distribution, metabolism, and excretion, which are indicative of utility as a potential drug. For example, rules can include hydrogen bonding, lipophilicity, molecular weight, pKa, polar surface area, shape, reactivity, solubility, permeability, chemical stability, metabolism (phases I and II), protein and tissue binding, transport (e.g., uptake, efflux), clearance, half-life, bioavailability, drug-drug interactions, LD_{50}, and the like.

[0043] 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.

[0044] In some embodiments, the rules could also include identifying the number of mutations per peptide construct. In some aspects, a rule might include a number of mutations in the range of 4-8 amino acid mutations per peptide construct. In some aspects, a rule could include about 4 amino acid mutations, 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 aspects, 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 or aid identification by MS.

[0045] As used herein, the term "fragment" refers to a portion of a peptide or a protein which is generated by cleavage of a peptide or protein from which the fragment was derived.

[0046] As used herein, the term "mass signature" refers to the pattern of peaks output by a mass spectrometer by performing an analysis of fragments of at least one protein or at least one peptide. The mass signature may be unique for each fragment or the mass signature may be the same for some fragments.

[0047] 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)

[0048] 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.

[0049] 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 molecules 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.

[0050] 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.

[0051] 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.

[0052] The small molecules to be screened as potential drug candidates can have molecular weights in the range if 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.

[0053] In some aspects, the present invention includes methods and systems for generating large libraries of knotted peptides (also referred to as "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).

[0054] 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 β- 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.

[0055] 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 predefined 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.

[0056] 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 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.

[0057] 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.

[0058] 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 10-50 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.

[0059] 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.

[0060] 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 10-35 amino acids, about 10-20 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.

[0061] 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.

[0062] 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 Ka, 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.

[0063] 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.

In some aspects, the present disclosure provides peptides for imaging a tumor in a subject, the peptide comprising a chlorotoxin comprising 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 comprises a detectable label.

In further aspects of the present disclosure, 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 peptide are D-amino acids.

In some aspects, the peptides of the present invention can have an amino acid sequence at least 80%, 83%, 86%, 89%, 90% or 92 % identical to the following sequence of MCMPCFTTDHQMARXCDDCCGGXGRGXCYGPQCLCR (SEQ ID NO: 47), wherein X is selected from K, A and R.

In some aspects, the peptides of the present invention, the peptides comprise an amino acid sequence having at least 80% identical to the following sequence of MCMPCFTTDHQMARXCDDCCGGXGRGXCYGPQCLCR (SEQ ID NO: 47), wherein at least three of the amino acids in the amino acid sequence are D-amino acids, wherein X is selected from K, A and R.

In further aspects, the peptides of the present disclosure are at least 50%, at least 60%, or at least 70% of the amino acids in the peptide are D-amino acids.

In some aspects, the peptides of the present disclosure further comprise a detectable label.

In some aspects of the present disclosure, all the amino acids in a given peptide are D-amino acids.

In some aspects of the present disclosure, the detectable label is conjugated to the N-terminus of the peptide or conjugated to a lysine residue in the peptide.

In other aspects of the present disclosure, the detectable label comprises a near-infrared dye.

In other aspects of the present disclosure, the detectable label comprises a cyanine dye.

In various aspects of the present disclosure, a composition can be produced comprising any of the peptides described herein, or a combination thereof.

In some aspects of the present disclosure, methods are provided for detecting a peptide in a subject, the method comprising: administering to the subject an effective amount of a peptide of the present disclosure, or a composition thereof, and detecting a detectable label contained therein.

In further aspects, the present disclosure further comprises obtaining an image of a region in the subject by detecting the detectable label. In further aspects, the detecting comprises intra-operative visualization of cancerous tissue. In still further aspects, visualization of the detectable label guides surgical removal of a tumor in the subject.

In some aspects of the present disclosure, methods are provided for treating a disease associated with cells expressing a chlorotoxin target, the method comprising: administering, to a subject in need thereof, a therapeutically effective amount of a pharmaceutical composition comprising a peptide according to the present disclosure or a composition comprising a peptide of the present disclosure, thereby treating the disease.

In further aspects, the peptide composition further comprises a cytotoxic agent, a toxin, an antisense nucleotide, a cancer drug, a nucleotide drug, a metabolic modulator, a radiosensitizer, a peptide therapeutic, a peptide-drug conjugate, or a combination thereof. In still further aspects, the disease comprises cancerous tissue associated with glioma, skin cancer, lung cancer, lymphoma, medulloblastoma, prostate cancer, pancreatic cancer, breast cancer, mammary cancer, colon cancer, sarcoma, oral squamous cell carcinoma, hemangiopericytoma, or a combination thereof.

[0064] 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.

[0065] 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.

[0066] 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.

[0067] 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.

[0068] 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.

[0069] 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).

[0070] 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).

[0071] 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).

[0072] 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 (SAWSHPQFEK) (SEQ ID NO: 34)). 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.

[0073] 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

[0074] 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.

[0075] In certain aspects, 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 palmitylation) 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.

[0076] 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.

[0077] 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 780nm, or between a range of 720 to 750nm. 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.

[0078] As used herein, the term "detectable label" means a tag or modification that can be attached to a small chemical molecule, peptide, protein, or a fragment or a portion thereof such

that the small chemical molecule, peptide, protein, or a fragment thereof is recognizable using a device, apparatus or method that permits the detection of the tag or modification.

[0079] In some aspects, the detectable label is a fluorescent dye. 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, aminorhodol, carboxyrhodol, chlororhodol, methylrhodol, sulforhodol; aminorhodamine, carboxyrhodamine, chlororhodamine, methylrhodamine, sulforhodamine, and thiorhodamine, cyanine, indocarbocyanine, oxacarbocyanine, thiacarbocyanine, 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, phtalocyanine, 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.

[0080] 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-dPEG$_4$-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 flour 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.

[0081] 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 B_2, vitamin B_3, vitamin B_6, vitamin B_{12}, 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.

[0082] 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

[0083] 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.

[0084] 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: MCMPCFTTDHQMARKCDDCCGGKGRGKCYGPQCLCR (SEQ ID NO: 35). 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.

[0085] 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.

[0086] 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 MCMPCFTTDHQMARKCDDCCGGKGRGKCYGPQCLCR (SEQ ID NO: 35), 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.

[0087] 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 (SEQ ID NO: 36).

[0088] 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 (SEQ ID NO: 37).

[0089] 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 (SEQ ID NO: 38).

[0090] 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) (SEQ ID NO: 39).

[0091] 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) (SEQ ID NO: 40).

[0092] 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) (SEQ ID NO: 41).

[0093] 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 (SEQ ID NO: 42).

[0094] 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 (SEQ ID NO: 43).

[0095] 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) (SEQ ID NO: 44).

[0096] 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) (SEQ ID NO: 45).

[0097] 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) (SEQ ID NO: 46).

[0098] 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).

[0099] 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 MCMPCFTTDHQMARXCDDCCGGXGRGXCYGPQCLCR (SEQ ID NO: 47), 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.

[00100] 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.

[00101] 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).

[00102] 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 780nm, or between a range of 720 to 750nm. 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.

[00103] 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, aminorhodol, carboxyrhodol, chlororhodol, methylrhodol, sulforhodol; aminorhodamine, carboxyrhodamine, chlororhodamine, methylrhodamine, sulforhodamine, and thiorhodamine, cyanine, indocarbocyanine, oxacarbocyanine, thiacarbocyanine, 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, phtalocyanine, 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.

[00104] 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.

[00105] 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 flour 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.

[00106] 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.

[00107] 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.

[00108] 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.

[00109] 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.

[00110] 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.

[00111] 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.

[00112] 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.

[00113] 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.

[00114] 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.

[00115] 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 biological 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.

[00116] "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).

[00117] 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).

[00118] 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.

[00119] 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.

[00120] 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.

[00121] 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.

[00122] 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, carminomycin, daunorubicin, aminopterin, methotrexate, methopterin, dichloromethotrexate, mitomycin C, porfiromycin, 5-fluorouracil, 6-mercaptopurine, gemcitabine, cytosine arabinoside, podophyllotoxin or podo- phyllotoxin derivatives, such as etoposide, etoposide phosphate or teniposide, melphalan, vinblastine, vincristine, leurosidine, vindesine, leurosine, paclitaxel and therapeutically effective analogs and derivatives of the same. Other useful antineoplastic agents include estramustine, cisplatin, carboplatin, cyclophosphamide, bleomycin, gemcitibine, ifosamide, melphalan, hexamethyl melamine, thiotepa, cytarabin, idatrexate, trimetrexate, dacarbazine, L- asparaginase, camptothecin, CPT-11, topotecan, ara-C, bicalutamide, flutamide, leuprolide, pyridobenzoindole derivatives, interferons and interleukins.

[00123] Suitable metabolic modulators can include, but are not limited to, lonidamine, dichloroacetate, alpha-tocopheryl succinate, methyl jasmonate, betulinic acid, and resveratrol

[00124] 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, EO9, RB 6145, nicotinamide, 5-bromodeoxyuridine (BUdR), 5-iododeoxyuridine (IUdR), bromodeoxycytidine, fluorodeoxyuridine (FudR), hydroxyurea, cisplatin, and therapeutically effective analogs and derivatives of the same.

[00125] 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.

[00126] In certain embodiments, the present invention can include treating diseases, disorders, and/or conditions, such as 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, 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 including 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.

[00127] 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.

[00128] 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.

[00129] 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.

[00130] 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.

[00131] 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.

[00132] 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.

[00133] 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.

[00134] One example of a mouse model of cancer is the autochthonous mouse model of medulloblastoma, ND2:SmoAl. See A.R. Hallahan, et al., "The SmoAl 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:SmoAl) 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.

[00135] 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.

[00136] 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.

[00137] 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:SmoAl 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 $C0_2$ 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 μm 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.

[00138] 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.

[00139] 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 x 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 uM) for 2 hours in a 37°C. humidified incubator maintained at 5% CO_2. 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 uM 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-phenylndole (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.

[00140] 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 µl PBS into the brain 3 mm lateral and posterior to the bregma.

Methods of Making Libraries

[00141] 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.

[00142] 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.

In some embodiments, the method further comprises generating a peptide library comprising the at least some of the peptides having the pharmacological property. In certain embodiments, the methods further comprise 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. In some embodiments, 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. In certain embodiments, the protein:protein interaction is associated with cancer. In some embodiments, the method further comprises obtaining from the subject a plurality of samples, each sample being obtained at a different time point after administering the plurality of peptides. In certain embodiments, at least one peptide of the plurality of peptides comprises a detectable label for tracing the at least one peptide in the subject. In certain embodiments, the detectable label comprises a near infrared dye.

[00143] 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.

[00144] 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.

[00145] 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.

[00146] 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 (SEQ ID NO: 48) right after the CIDG (SEQ ID NO: 49), and an exogenously cleaved one, with ENLYFQ (SEQ ID NO: 50) 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 (SEQ ID NO: 50) 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 used for difficult-to-express proteins of medical interest such as chemokines, interleukins, and peptide hormones.

[00147] 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 or 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.

[00148] 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.

[00149] 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.

[00150] 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.

[00151] 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.

[00152] 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 the methods described herein.

[00153] 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.

[00154] 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.

[00155] 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 prescreened 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.

[00156] In some embodiments, the present invention relates to methods of generating a mass-defined drug candidate library where the method comprises producing a plurality of drug candidates, at least some of the drug candidates each having a unique mass signature or digest fragment mass signature and 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. In certain embodiments, the methods 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.

[00157] The invention further provides methods of generating mass-defined drug candidate libraries, where the method comprises, producing a plurality of drug candidates, at least some of the drug candidates each having a unique mass signature or digest fragment mass signature; 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; and 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.

[00158] In some embodiments, at least one of the drug candidates in the plurality comprises a pre-defined number of a heavy isotope atom to modify the unique mass signature or digest fragment mass signature of the at least one drug candidate. In certain embodiments, the heavy isotope atom comprises ^{13}C or deuterium. In some embodiments, the unique mass signature or digest fragment mass signature of at least one of the knotted-peptides in the plurality is defined by a moiety conjugated to the at least one knotted-peptide. In certain embodiments, the moiety is conjugated to the N-terminus of the at least one knotted-peptide. In some embodiments, the moiety comprises a pre-defined number of a heavy isotope atom to modify the unique mass signature or digest fragment mass signature of the at least one knotted-peptide. In certain embodiments, the heavy isotope atom comprises ^{13}C or deuterium. In some embodiments, the plurality of knotted peptides comprises greater than 100 peptides, greater than 1000 peptides, or greater than 10000 peptides.

[00159] In some embodiments, the plurality of drug candidates comprise knotted-peptides. In certain embodiments, the unique mass signature or digest fragment mass signature of the knotted-

peptides are defined by the natural amino acid sequence of the knotted peptides. In certain embodiments, at least one of the drug candidates in the plurality comprises a pre-defined number of a heavy isotope atom to modify the unique mass signature or digest fragment mass signature of the at least one drug candidate. In certain embodiments, the heavy isotope atom comprises ^{13}C or deuterium.

[00160] In some embodiments, the unique mass signature or digest fragment mass signature of at least one of the knotted-peptides in the plurality is defined by a moiety conjugated to the at least one knotted-peptide. In certain embodiments, the moiety is conjugated to the N-terminus of the at least one knotted-peptide. In certain embodiments, the moiety comprises a pre-defined number of a heavy isotope atom to modify the unique mass signature or digest fragment mass signature of the at least one knotted-peptide. In certain embodiments, the heavy isotope atom comprises ^{13}C or deuterium. In certain embodiments, the plurality of knotted peptides comprises greater than 100 peptides. In certain embodiments, the plurality of knotted peptides comprises greater than 1000 peptides. In certain embodiments, the plurality of knotted peptides comprises greater than 10000 peptides.

[00161] The invention further provides methods of determining a distribution profile of knotted-peptides administered to a subject by different administration pathways where the method comprises 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; 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; and 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.

[00162] In some embodiments, the light knotted peptide comprises fewer heavier isotopes than the light knotted peptide. In some embodiments, the heavy knotted-peptide comprises at least one more ^{13}C atom or deuterium atom than the light knotted-peptide. In certain embodiments the first knotted-peptide is conjugated to a first moiety and the heavy knotted-peptide is conjugated to a second moiety. In other embodiments, the first moiety, second moiety, or both is conjugated to the N-terminus of the light and heavy knotted peptide, respectively. In certain embodiments, the first moiety, second moiety, or both comprises a hydrophobic moiety.

[00163] In some embodiments, the light knotted peptide comprises fewer heavier isotopes than the light knotted peptide. In other embodiments, the heavy knotted-peptide comprises at least one

more ^{13}C atom or deuterium atom than the light knotted-peptide. In certain embodiments, the first knotted-peptide is conjugated to a first moiety and the heavy knotted-peptide is conjugated to a second moiety. In certain embodiments, the first moiety, second moiety, or both is conjugated to the N-terminus of the light and heavy knotted peptide, respectively. In some embodiments, the first moiety, second moiety, or both comprises a hydrophobic moiety. In some embodiments, at least one ^{13}C atom is present in the first moiety, the second moiety, or both. In some embodiments, the light knotted-peptide, the heavy knotted-peptide, or both have unique mass signature or digest mass signature when analyzed by mass spectrometry.

[00164] In some embodiments, at least one ^{13}C atom is present in the first moiety, the second moiety, or both. In other embodiments, wherein the light knotted-peptide, the heavy knotted-peptide, or both have unique mass signature or digest mass signature when analyzed by mass spectrometry.

[00165] In some embodiments, 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. In certain embodiments, the hydrophobic moiety comprises a hydrophobic fluorescent dye or a saturated or unsaturated alkyl group.

[00166] In some embodiments, at least one peptide of the plurality of peptides comprises a detectable label for tracing the at least one peptide in the subject. In certain embodiments, the detectable label comprises a near infrared dye.

Methods of Analyzing Libraries

[00167] 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. The methods provided herein further comprise 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. In some embodiments, 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. In other embodiments, the plurality of drug candidates comprises greater than 5 drug candidates, greater than 10 drug candidates, greater than 100 drug candidates, greater than 1000 drug candidates, or greater than 10000 drug candidates. In certain embodiments, at least one peptide of the plurality of peptides comprises a detectable label for tracing the at least one peptide in the subject. In certain embodiments, the detectable label comprises a near infrared dye.

[00168] In one aspect, the methods include generating peptide libraries in which some or all of the peptides in a library have a unique mass signature or digest fragment signature (e.g., by trypsin). In certain aspects, peptides of a generated peptide library are digested into smaller, more manageable fragments using enzymes such as trypsin. By doing so, peptides samples can more readily be analyzed by mass spectrometry.

[00169] 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.

[00170] 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.

[00171] 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.

[00172] 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.

[00173] 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 ^{2}H atom, at least one ^{3}H 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 ^{29Si} atom, or at least one ^{30}Si atom.

[00174] 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 ^{1}H, ^{12}C, ^{14}N, ^{16}O, or ^{32}S, or any combination thereof.

[00175] 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 spectral (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 cyclotron 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

[00176] 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.

[00177] In some aspects of the present disclosure, peptides in a peptide library are conjugated. Conjugation can occur at any suitable location within a peptide, for example, the peptide can be conjugated at any amino acid residue making up the peptide or at either its N-terminus or C-terminus. In various aspects, peptides of the present disclosure are conjugated at their N-termini. In some aspects of the present disclosure, peptide libraries can include conjugated peptides.

[00178] 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 aspects, 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 aspects, 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.

[00179] In some aspects of the present disclosure, conjugated peptides can be isolated from a pool of tryptically digested peptide fragments in the peptide library by binding the conjugate. For example, conjugated peptides can be immobilized onto a surface by selectively binding the conjugate. In certain aspects, the N-terminus of a peptide is conjugated with biotin or palmitic acid for recovery following proteolysis with trypsin. Any suitable conjugation-binding agent combination can be used to isolate peptides of the present disclosure.

[00180] In various aspects of the present disclosure, the conjugated peptide fragments have unique mass spectral signatures relative to other conjugated peptide fragments in the library. In further aspects, the mass spectral signature of a conjugated peptide is different from the mass spectral signature of the corresponding unconjugated peptide. In various aspects, the mass spectral signature of a peptide lacking N-terminal conjugation is different than the mass spectral signature of the corresponding peptide having N-terminus conjugation.

[00181] In various aspects, conjugation of a peptide can be performed at a single amino acid. In some aspects, the amino acid may be cysteine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, valine, alanine, arginine, asparagines, aspartic acid, glutamic acid, glutamine, glycine, ornithine, proline, selenocysteine, serine or tyrosine. In an exemplary case, the amino acid may be cysteine. In other aspects, the conjugation can be performed at more than one amino acid. In some aspects, the more than one amino acid may be cysteine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, valine, alanine, arginine, asparagines, aspartic acid, glutamic acid, glutamine, glycine, ornithine, proline, selenocysteine, serine or tyrosine.

[00182] In some aspects of the present disclosure, peptides can be conjugated to prevent oxidation. For example, cysteine residues can be conjugated following reduction in order to prevent re-oxidation of the cysteine residues. Any suitable thiol-alkylating agent can be used to prevent re-oxidation of cysteine residues following reduction. In certain aspects, cysteine residues are conjugated with iodoacetamide, N-ethylmaleimide, or iodoacetic acid to prevent re-oxidation of cysteine residues following reduction. In various aspects, cysteine residues are conjugated with iodoacetamide. In certain aspects, peptides can be conjugated with "heavy" iodoacetamide, thereby enabling its use as a reference standard for material conjugated to standard (i.e., "light") iodoacetamide. This strategy is particularly useful for knottin peptide libraries due to the large numbers of cysteine residues in those peptides.

[00183] In various aspects of the present disclosure, the conjugate atoms can be either "heavy" or "light," depending on the relative atomic mass of the atom. For example, the ^{13}C atom is a "heavy" atom and the ^{12}C atom is a "light" atom. In some aspects, for example, a conjugate that

contains at least one ^{13}C atom is heavy relative to a conjugate that contains only ^{12}C atoms. For example, in some aspects, a conjugate that contains more than two ^{13}C atoms is heavy relative to a conjugate that contains only one ^{13}C atom. In some aspects, atoms besides carbon may be used for heavy and light conjugates.

[00184] 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).

[00185] 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.

[00186] In various aspects of the present disclosure, a heavy library can be spiked into samples containing an unknown quantity of light library members. The heavy library peaks observed during mass spectral analysis can then be used to help quantitate the light library peaks appearing in the same sample.

[00187] A heavy library is also useful to monitor the relative amounts of candidates in a library following a particular procedure. For example, a small amount of heavy library can be mixed into a standard library to determine a baseline of relative peak heights. Because the heavy library is chemically identical but different in mass, in this mixture every member of the standard library will be seen by LC/MS/MS to have a heavy peak adjacent to it. The heights of these heavy peaks will form a gauge by which the amounts of members of the standard library can be judged. So the standard library could be, for example, incubated in serum to test for susceptibility to serum proteases. Following incubation in serum, the material could be heated to inactivate the serum proteases, and the heavy library doped in as a standard. The LC/MS/MS peak heights of the heavy material can then be compared to the peak heights of the standard library to determine which library members were subject to proteolysis. The central point here is that heavy libraries can be doped into standard ones to provide reference peaks. The different peptides will "fly" differently just based upon their innate qualities, affecting their peak heights independent of concentration, but they will fly the same as their heavy counterparts, except that the heavy ones

will be shifted by their mass. So the peak heights of the heavy ones can be used to calibrate the peak heights of the standard ones to provide concentration from peak heights.

[00188] 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.

[00189] 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 ^{2}H atom, at least one ^{3}H 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 ^{29Si} 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 or candidate linkers.

Methods of Using Libraries

[00190] 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.

[00191] 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 (1α(PCDHA1), $Ca_v2.2$, $K_v1.3$, $Na_v1.2$, 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, NFκB, Bromodomains, HATS, HDAC, LDH, IDH2, CD22, MIC, c-Myc, n-Myc, PHF5A, BUB1B, Bcl-2, k-Ras, Notch1, p53, α5β3, NKG2D, CTLA4/CD28, and/or Mcl-1.

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.

[00192] In some embodiments, the drug candidates are selected from the group consisting of small chemical molecules, biologics, and peptides. In certain embodiments, the plurality of drug candidates comprises a plurality of peptides.

[00193] 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.

[00194] 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.

[00195] 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.

[00196] 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.

[00197] The present invention further relates to methods for identifying a library of drug candidates having a pharmacological property where the method comprises analyzing an isolated sample from a subject following administration of a plurality of drug candidates to the subject. In certain embodiments, the library of drug candidates are from the mass-defined drug candidate library provided herein. In certain embodiments, the methods include identifying in the isolated sample at least one drug candidate having the pharmacological property.

[00198] In some embodiments, the methods further comprise using the pharmacological property common to the drug candidates to generate additional drug candidates having the pharmacological property.

[00199] In some embodiments, 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, or a combination thereof. In certain embodiments, the pharmacological property comprises oral bioavailability.

[00200] In some embodiments, the present invention relates to methods for identifying library drug candidates having a pharmacological property where the method comprises administering to a subject a plurality of library drug candidates, wherein the library drug candidates are from the mass-defined drug candidate library provided herein, obtaining, from the subject, a sample comprising at least some of the plurality of library drug candidates; and analyzing the sample to determine the identity of the at least some of the plurality of library drug candidates having the pharmacological property.

[00201] In some embodiments, the invention provides methods 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.

[00202] In some embodiments, 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. In certain embodiments, the method further comprises generating a peptide library comprising the at least some of the peptides having the pharmacological property.

[00203] In some embodiments, the first route and second route of delivery are independently selected from an oral route, a topical route, a transmucosal route, an intravenous route, an intramuscular route, and an inhalation route. In certain embodiments, the either the first route or the second route of delivery comprises an oral route.

[00204] 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.

[00205] In some embodiments, the method further comprises using the pharmacological property common to the drug candidates to generate additional drug candidates having the pharmacological property. In some embodiments, the subject is an animal. In certain embodiments, the subject is a human.

[00206] In some embodiments, the sample comprises a tissue sample or a fluid sample. In certain embodiments, the fluid sample comprises blood, urine, mucous or spinal fluid. In certain embodiments, the sample comprises blood. In some embodiments, the tissue sample comprises biopsy or necropsy tissue. In certain embodiments, the tissue comprises tissue from a brain, a lung, a kidney, a muscle, a liver, a heart, a stomach, a pancreas, or an organ. In some embodiments, the subject is an animal. In some embodiments, the subject is a human.

[00207] In some embodiments, the sample comprises a tissue sample or a fluid sample. In some embodiments, the fluid sample comprises blood, urine, mucous or spinal fluid. In certain embodiments, the sample comprises blood. In some embodiments, the tissue sample comprises biopsy or necropsy tissue. In other embodiments, the tissue comprises tissue from a brain, a lung, a kidney, a muscle, a liver, a heart, a stomach, a pancreas, or an organ.

[00208] In some aspects of the present disclosure, the method further comprises establishing a relationship between the structure and the pharmacological property of the drug candidates, and in further aspects, a subset of drug candidates based on that relationship.

[00209] In some aspects of the present disclosure, the methods further comprise obtaining a plurality of samples at a selected time interval for a selected duration of time following administration. In certain aspects, the time intervals are 5 minutes, 10 minutes, 20 minutes, 30 minutes, 40 minutes, 50 minutes, or 60 minutes. In still further aspects, the selected duration of time is 1 hour, 2 hours, 3 hours, 4 hours, 5 hours, 6 hours, 7 hours, 8 hours, 9 hours, 10 hours, 11 hours, 12 hours, 15 hours, 30 hours, 45 hours, 60 hours, 75 hours, 90 hours, 105 hours, or 120 hours.

[00210] In some embodiments, the methods further comprise obtaining from the subject a plurality of samples, each sample being obtained at a different time point after administering the plurality of peptides. In some embodiments, the methods further comprise 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. In certain

embodiments, 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. In certain embodiments, the plurality of drug candidates comprises greater than five drug candidates. In other embodiments, the plurality of drug candidates comprises greater than 10 drug candidates. In other embodiments, the plurality of drug candidates comprises greater than 100 drug candidates. In other embodiments, the plurality of drug candidates comprises greater than 1000 drug candidates. In other embodiments, the plurality of drug candidates comprises greater than 10000 drug candidates. In other embodiments, the plurality of peptides comprises greater than 100 peptides. In other embodiments, the plurality of peptides comprises greater than 1000 peptides. In other embodiments, the plurality of peptides comprises greater than 10000 peptides.

[00211] In some embodiments, the plurality of drug candidates comprises greater than five drug candidates, greater than 10 drug candidates, greater than 100 drug candidates, greater than 1000 drug candidates, or greater than 10000 drug candidates. In some embodiments, the plurality of drug candidates comprise knotted-peptides. In certain embodiments the unique mass signature or digest fragment mass signature of the knotted-peptides are defined by the natural amino acid sequence of the knotted peptides.

[00212] A method for identifying a drug candidate having a pharmacological property, the method comprising, analyzing an isolated sample from a subject following administration of a plurality of drug candidates to the subject; and identifying in the isolated sample at least one drug candidate having the pharmacological property. In some embodiments, the drug candidates are selected from the group consisting of small chemical molecules, biologics, and peptides. In certain embodiments, the plurality of drug candidates comprises a plurality of peptides. In other embodiments, 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.

[00213] 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. In some embodiments, the drug candidates are selected from the group consisting of small chemical molecules, biologics, and peptides. In certain embodiments, the plurality of drug candidates comprises a plurality of peptides. In other

embodiments, 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. In some embodiments, the method further comprises obtaining from the subject a plurality of samples, each sample being obtained at a different time point after administering the plurality of peptides. In certain embodiments, at least one peptide of the plurality of peptides comprises a detectable label for tracing the at least one peptide in the subject. In certain embodiments, the detectable label comprises a near infrared dye.

[00214] In some embodiments, the first route and second route of delivery are independently selected from an oral route, a topical route, a transmucosal route, an intravenous route, an intramuscular route, and an inhalation route. In certain embodiments, the either the first route or the second route of delivery comprises an oral route.

[00215] In some embodiments, the method further comprises 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. In other embodiments, 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. In certain embodiments, the protein:protein interaction is associated with cancer.

EXEMPLARY ASPECTS

EXAMPLE 1

Expressing Peptide Constructs for Knottin Generation

[00216] This example describes a method for expressing peptide constructs in culture and greatly facilitating their development, particularly as drugs.

[00217] 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 the SEQUENCE section 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.

[00218] 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.

[00219] 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.

[00220] 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.

[00221] 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 (SEQ ID NO: 51) 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 10mM ammonium formate, and the fractions containing the peptides were pooled and lyophilized.

[00222] 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.

[00223] 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 BY, Brault M, Strong RK, Rawlings DJ. Nucleic Acids Res. 2011 Nov;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 10mM 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.

[00224] Conjugation to palmitic acid, ICG, or biotinidase-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.

EXAMPLE 2

Generating Mass-Indexed Peptide Libraries and Peptides for Drug Development

[00225] This example describes a method for generating peptide libraries that contain mass-indexed peptides for the generation of knottins. This example also describes the analysis of mass-indexed peptides in a sample.

[00226] A mass-indexed library of 3,000 variants of the peptide Chayote Trypsin Inhibitor was created *in silico*, All lysine residues in the peptide were changed to arginine residues. The resulting sequence was: CPRILMRCRLDTDCFPTCTCRPSGFCG (SEQ ID NO: 52). G1 and S2 of this sequence were added to the molecule's N-terminus. The molecular structure of the scaffold was analyzed manually to identify portions of the sequence that may be altered without disrupting scaffold structure. The identified amino acid resides included M8, R9, L12, D13, T14, F17, P18, T19, T21, R23, P24, S25, G26. The identified amino acids were R23, P24, S25, and G26 and were located at positions 1 through 4 of a beta turn, respectively.

[00227] Python software was used to create a pool of five million sequences, such that each sequence was a unique variant. The sequences were created by altering the amino acids of each member in the pool such that each member contained a fully-tryptic peptide in the range of 7-30 amino acids in length. The fully-tryptic peptide was within the mass range 800-3500Da when reduced and alkylated with iodoacetamide.

[00228] The number of amino acid alterations from the starting amino acid sequence for each variant was chosen from a distribution of between 4 and 8. The distribution was weighted as follows: 4: 15%, 5: 50%, 6: 20%, 7: 10%, 8: 5%. The positions that were altered were chosen randomly from the list of possibly alterable positions. For each amino acid alteration, the amino acid to be substituted was selected from a set of amino acids. The set of amino acids was appropriate for the structure of the scaffold in a given position and weighted appropriately to favor amino acids that were more likely to favor proper folding. The weights for each amino acid in the different structural classifications were as follows:

Beta turn position 1: D: 15, G: 11, H: 13, N: 10, Q: 15, S: 13, T: 10, Y: 10

Beta turn position 2: D: 12, E: 14, G: 10, N: 10, Q: 10, S: 12, P: 35

Beta turn position 3: D: 18, G: 21, H: 12, N: 10, Q: 21, S: 10

Beta turn position 4: G: 16, N: 10, Q: 10, S: 10, T: 12, Y: 10

All other positions: A: 4, D: 8, E: 4, F: 7, G: 8, H: 8, I: 8, L: 2, M: 4, N: 4, Q: 4, R: 8, S: 7, T: 4, V: 4, W: 6, Y: 10

[00229] There were no variants with more than 3 tryptophan residues retained in the pool. Reverse translations of each member of the pool were chosen for DNA synthesis. During DNA synthesis, the coding of the n-terminal GS portion was changed to GGATCC. Further, the 5' sequence AACTGCCATGTGCAACTCGTAAG (SEQ ID NO: 53) and the 3' sequence

TAATGCGGCCGCGTTCTTAGTCACCTTGCATGGAC (SEQ ID NO: 54) were added to each member of the pool. The reverse translations were chosen so as not to contain BamHI (GGATCC) or NotI (GCGGCCGC) restriction sites unless selected for specific members of the pool.

[00230] From the pool of possible variants, the members of the mass-indexed library were randomly chosen. The members were added to the library one at a time until the desired number of members was reached. Members were only added if addition of a member preserved the following criteria for the library:

- each member had at most 2 neighbors within 0.15 Daltons and 2.75 minutes of predicted liquid chromatography (LC) retention time in a 90-minute gradient;
- every member had at least one tryptic peptide of 7-25 amino acids that also had at most 2 neighbors within 2.1 Thompsons and 2.75 minutes of predicted LC retention time that is unique in sequence within the library.

Subsequently, 3,000 sense-strand oligonucleotides as well as 3,000 full reverse-complement oligonucleotides were ordered for library members.

[00231] Members of the libraries were screened for protein expression. Samples of the members of the library were prepared as protein mixtures. The samples of the protein mixture was reduced with 20 mM DTT and blocked with 25 mM iodoacetamide. The proteins were digested with trypsin at a trypsin to protein ratio: 1/50 overnight (or approximately16–18 hours) and purified using a C18 desalting column. One μg of purified peptide sample was injected into an Orbitrap Elite hybrid mass spectrometer (Thermo Fisher Scientific, Waltham, MA) that was coupled with nano-flow HPLC. The liquid chromatography/mass spectrometry apparatus and method consisted of a trap column (100μm × 1.5cm) made from an IntegraFrit (New Objective, Woburn, MA) packed with Magic C18AQ resin (5μm, 200Å particles; Michrom Bioresources, Auburn CA), followed by an analytical column (75μm × 27cm) made from a PicoFrit (New Objective) packed with Magic C18AQ resin (5μm, 100Å particles; Michrom Bioresources). The columns were connected in-line to an Eksigent 2D nano-HPLC (Eksigent Technologies, Dublin, CA) in a vented column configuration to allow fast sample loading at 3uL/min.

[00232] The peptide samples were analyzed by LC-MS/MS using a 90-minute non-linear gradient as follows: start at 5% acetonitrile with 0.1% formic acid (against water with 0.1% formic acid), change to 7% over 2 minutes, then to 35% over 90 minutes, then to 50% over 1 minute, hold at 50% for 9 minutes, change to 95% over 1 minute, hold at 95% for 5 minutes, drop to 5% over 1 minute and recondition at 5%. The flow rate for the peptide separation was 300 nL/min.

[00233] A spray voltage of 2.25 kV was applied to the nanospray tip. The mass spectrometry experiment consisted of a full MS scan in the Orbitrap (AGC target value 1e6, resolution 60K, and one microscan, FT preview scan on) followed by up to 5 MS/MS spectra acquisitions in the linear ion trap. The five most intense ions from the Orbitrap scan were selected for MS/MS using collision-induced dissociation (isolation width 2 m/z, target value 1e4, collision energy 35%, max injection time 100 ms). Lower abundance peptide ions were interrogated using dynamic exclusion (repeat count 1, repeat duration 30 sec., exclusion list size 100, exclusion time 45 sec., exclusion mass width −0.55 m/z low to 1.55 m/z high). Charge state screen was used, allowing for MS/MS of any ions with identifiable charge states +2, +3, and +4 and higher.

[00234] For samples that contained the peptides of interest, the following data analysis methods were used. A FASTA database was constructed for database search, containing the entire UniProt human proteome (2012/12/19), the expected 3,000 proteins from the chayote trypsin inhibitor library, and a set of commonly observed contaminant proteins. Raw machine output files from all MS runs were converted to mzXML files and searched with X!Tandem (version 2011.12.01.1) configured with the k-score scoring algorithm, against this FASTA database. The search parameters were as follows: enzyme, trypsin; maximum missed cleavages, 2; fixed modification, carboxamidomethylation on cysteine; potential modification, oxidization on methionine; parent monositopic mass error, -1.5Da to 2.5Da.

[00235] Peptide identifications were assigned probability by PeptideProphet (Trans-Proteomic Pipeline version 4.6), and all identifications assigned probability <0.95 were discarded. Variants present in the sample were inferred based on the peptide identifications of the tryptic peptides unique to those variants. For mouse plasma and tissue samples, the methods above were followed, using a FASTA database containing the sequences described above in addition to the entire UniProt mouse proteome (2012/12/19).

EXAMPLE 3

Generating Mass-Indexed Peptide Libraries Using Conjugated Peptides

[00236] This example describes a method for generating peptide libraries that contain conjugated mass-indexed peptides that can be used for the generation of knottins. This example further describes the analysis of mass-indexed conjugated peptides in a sample.

[00237] Peptides were conjugated to biotin and retrieved from a peptide library following proteolysis with trypsin. A library of 1000 Chayote Trypsin Inhibitor variants at 10mg/ml was allowed to react with a 3x molar excess of N-hydroxy succinimide (NHS) ester of biotin in phosphate buffered saline. The conjugation proceeded strictly at the N-terminus of the peptides because the library was engineered to contain no lysine residues.

[00238] Excess biotin and NHS were removed by ultrafiltration, and the resultant library administered to mice with xenografted human tumors at 100mg/kg via tail vein injection. Mice were sacrificed at 5, 10, 30, and 60 minutes and the tumor, brain, blood, and kidneys were collected and mechanically homogenized. Cells and clotted material were removed by centrifugation and the resultant supernatant was incubated with avidin immobilized on beads for 30 minutes at room temperature with shaking. The beads were then washed extensively with PBS and boiled to release the biotinylated peptides. These peptides were then subjected to reduction, alkylation (of cysteine residues), trypsinization, and LC/MS/MS analysis to determine the identities of the peptides in the various tissues.

[00239] In a variation of this method, trypsin was added prior to the mechanical homogenization to break up protein-protein interactions with knottin peptides, which may have formed during sample trypsinization. Recovery of the peptides was the same, with avidin beads, except only the N-terminus of the peptide was recovered.

[00240] In addition to conjugation with biotin, additional peptide libraries were generated by isotopic labeling. Both types of libraries were quantified via mass spectrometry. A library of 1000 Chayote Trypsin Inhibitor variants at 10mg/ml was split into two halves. One half reacted with a 3x molar excess of N-hydroxy succinimide (NHS) ester of isotopically normal palmitic acid in phosphate buffered saline. The other half was treated in the same manner with palmitic acid containing at four C13 (isotopically heavy) carbons. DMSO or DMF was added sufficient to keep the palmitic acid in solution. Excess palmitic acid was removed by extraction with ethyl acetate after an overnight (12-18 hour) incubation.

[00241] The isotopically heavy library is chemically identical to the isotopically normal library but was differentiated by mass. The heavy library was administered to animals by IV dosing at 100mg/kg and the isotopically normal library administered in the same animals by oral lavage. The animals were sacrificed 20 minutes after library administration. The blood was collected and allowed to clot. Cells and clotted material were removed by centrifugation and the resultant plasma was reduced with 5mM TCEP, alkylated with iodoacetamide, and applied to a C18 solid phase extraction column. The column was then rinsed with increasing percentages of acetonitrile in 0.5% TFA. The N-terminally palmitylated peptides elute at 80% or higher acetonitrile. This eluted material was then subjected to trypsinization and LC/MS/MS analysis, and the isotopically normal and heavy peaks were compared directly. Normal peaks were observed and corresponded to peptides with oral bioavailability.

[00242] In a variation of this protocol, the plasma was treated with trypsin after iodoacetamide treatment and prior to applying it to the C18 column. With this method, only the N-terminal

tryptic peptide of each library member was recovered and distinguishable based on unique properties.

[00243] A library composed of 28 knottin peptides was screened for those that bind with high affinity to the human derived rhabdomyosarcoma (e.g., cancer) cell line, A-204. Twenty-four hours prior to exposure, the cells were plated on Poly-L-Lysine coated 6-well plates (Corning) and grown in RPMI-1640 media (Life Technology) supplemented with 10% FBS (Hyclone) to reach ~85% of confluency. The adherent cells were washed 3x with 3ml D-PBS (Life Technology) and incubated with the knottin library (30μg/ml) in D-PBS for 1hr at 37C. The unbound peptides were aspirated, and cells were washed 3x with D-PBS in which the last wash was recovered for analysis. To elute the bound knottins, the cells were incubated in 1ml 280mM NaCl in D-PBS for 5min at 37C, followed by 560mM and 1.12M NaCl. Following the final elution, the cells were collected in HyPure Molecular Biology Grade Water (Hyclone).
The cells were lysed with three freeze thaw cycles, and 5min in a bath sonicator. The protein concentration of the lysate was quantified based on the absorbance at 280nm using a Nanodrop 1000 and the sample was diluted to 100μg/ml total protein in water. The lysate and elution samples were desalted using a Strata-X desalting column (Phenomenex). The recovered peptides were then reduced, alkylated and trypsinized before being submitted for LC/MS-MS analysis. Individual knottins were identified from each sample by their unique tryptic peptide signature.

EXAMPLE 4

Identifying Knottins with Serum Stability, Half Life and Tissue Homing Capabilities

[00244] This example describes a method for identifying knottin peptides that contain mass-indexed peptides for the generation of knottins. This example further describes the analysis of mass-indexed peptides in a sample.

[00245] Six mice were given a library of mass-indexed knottins via tail-vein injection. Three mice were sacrificed at 5 minutes, three at 30 minutes. The blood was collected by cardiac puncture and the organs (brain, liver, kidneys) dissected for analysis. Tissue was flash-frozen in liquid nitrogen. Tissue was subsequently thawed, homogenized, and subjected to analysis by mass spectrometry to determine tissue distribution and half-life estimation.

EXAMPLE 5

Parallel Assessment of Chymotrypsin Inhibitor Mutational Tolerance

[00246] This example describes the assessment of mutational tolerance of chymotrypsin inhibitor (CTI) according to an aspect of the present disclosure.

[00247] Chip-based oligonucleotide synthesis was performed in order to generate a set of 3000 mutants of CTI. The mutants were amplified by PCR and cloned by restriction digestion and ligation into the lentiviral system.

[00248] FIG. 12 depicts the assessment of mutational tolerance of CTI by reverse proteomics 3000 mutants of CTI with 4-8 mutations, each with a unique MS signature, were generated. These viruses were used to inflect HEK293 cells as a pool, and the resulting peptides were isolated, digested with trypsin, treated with iodoacetamide, and applied to an LC/MS/MS Orbitrap Elite system, and 1595 proteins were observed. The over- or underrepresentation of amino acids at each position, in the expressed and observed proteins versus the full set of proteins ordered, is shown. Amino acids below the line at the bottom were not observed at all in the set of produced and observed proteins. Globally, arginine and tryptophan amino acids were less likely to be present in the produced and observed proteins than in the set of ordered proteins, and glutamic acid was more likely to be present.

[00249] 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.

SEQUENCES

[00250] The following are DNA and/or amino acid sequences of genes of interest and constructs identified herein.

Construction of parental construct for seamless cloning:
SEQ ID NO:1 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-PARENTAL *for Xho/Bam cut from pUC57 and ligation into pCVL*

GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGCGGCCGC**TAAGGATCCCGGACCGCCTCTCC

NotI cut is AACCTGTATTTTCAGGGA**GGC** (SEQ ID NO: 55) -
GCTAAGGATCCCGGACCGCCTCTCC (SEQ ID NO: 56)
Fusion protein Sequences – original set of 10 – cloned into NotI cut parent above by seamless cloning:

SEQ ID NO:2 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-BubbleProtein
GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGCGATACCTGCGGCAGCGGCTATAATGTGGATCAGCGTCGTACCA**
ATAGCGGCTGCAAAGCGGGCAATGGCGATCGTCATTTTTGCGGCTGCGATCGT
ACCGGCGTGGTGGAATGCAAAGGCGGCAAATGGACCGAAGTGCAGGATTGCG
GCAGCAGCAGCTGCAAAGGCACCAGCAATGGCGGCGCGACCTGCTAATGCTAA
GGATCCCGGA

SEQ ID NOS3 & 57 -

```
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
 M  E  T  D  T  L  L  L  W  V  L  L  L  W  V  P  G  S  T  G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
 D  Y  K  D  E  H  H  H  H  H  H  G  G  S  Q  D  S  T  S  D
ctgatcccagccccacctctgagcaaggtccctctgcagcagaacttccaggacaaccaa
 L  I  P  A  P  P  L  S  K  V  P  L  Q  Q  N  F  Q  D  N  Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
 F  Q  G  K  W  Y  V  V  G  L  A  G  N  A  I  L  R  E  D  K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
 D  P  Q  K  M  Y  A  T  I  Y  E  L  K  E  D  K  S  Y  N  V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
 T  S  V  L  F  R  K  K  K  C  D  Y  W  I  R  T  F  V  P  G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
 C  Q  P  G  E  F  T  L  G  N  I  K  S  Y  P  G  L  T  S  Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
 L  V  R  V  V  S  T  N  Y  N  Q  H  A  M  V  F  F  K  K  V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
 S  Q  N  R  E  Y  F  K  I  T  L  Y  G  R  T  K  E  L  T  S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
 E  L  K  E  N  F  I  R  F  S  K  S  L  G  L  P  E  N  H  I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
 V  F  P  V  P  I  D  Q  C  I  D  G  G  G  S  E  N  L  Y  F
cagggaggcgatacctgcggcagcggctataatgtggatcagcgtcgtaccaatagcggc
 Q  G  G  D  T  C  G  S  G  Y  N  V  D  Q  R  R  T  N  S  G
tgcaaagcgggcaatggcgatcgtcatttttgcggctgcgatcgtaccggcgtggtggaa
```

C K A G N G D R H F C G C D R T G V V E
tgcaaaggcggcaaatggaccgaagtgcaggattgcggcagcagcagctgcaaaggcacc
C K G G K W T E V Q D C G S S S C K G T
agcaatggcggcgcgacctgc
S N G G A T C

SEQ ID NO:4 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-Attractin
GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGCGATCAGAATTGCGATATTGGCAATATTACCAGCCAGTGCCAGA**
TGCAGCATAAAAATTGCGAAGATGCGAATGGCTGCGATACCATTATTGAAGAAT
GCAAAACCAGCATGGTGGAACGTTGCCAGAATCAGGAATTTGAAAGCGCGGCG
GGCAGCACCACCCTGGGCCCGCAGTAATGCTAAGGATCCCGGA

SEQ ID NOS 5 & 58 -
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
M E T D T L L L W V L L L W V P G S T G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
D Y K D E H H H H H H G G S Q D S T S D
ctgatcccagccccacctctgagcaaggtccctctgcagcagaacttccaggacaaccaa
L I P A P P L S K V P L Q Q N F Q D N Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
F Q G K W Y V V G L A G N A I L R E D K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
D P Q K M Y A T I Y E L K E D K S Y N V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
T S V L F R K K K C D Y W I R T F V P G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
C Q P G E F T L G N I K S Y P G L T S Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
L V R V V S T N Y N Q H A M V F F K K V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
S Q N R E Y F K I T L Y G R T K E L T S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
E L K E N F I R F S K S L G L P E N H I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
V F P V P I D Q C I D G G G S E N L Y F
cagggaggcgatcagaattgcgatattggcaatattaccagccagtgccagatgcagcat
Q G G D Q N C D I G N I T S Q C Q M Q H
aaaaattgcgaagatgcgaatggctgcgataccattattgaagaatgcaaaaccagcatg
K N C E D A N G C D T I I E E C K T S M

```
gtggaacgttgccagaatcaggaatttgaaagcgcggcgggcagcaccaccctgggcccg
 V  E  R  C  Q  N  Q  E  F  E  S  A  A  G  S  T  T  L  G  P
cag
 Q
```

SEQ ID NO:6 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-Hefutoxin

GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGCGGCCATGCGTGCTATCGTAATTGCTGGCGTGAAGGCAATGATG**
AAGAAACCTGCAAAGAACGTTGCTAATGCTAAGGATCCCGGA*CCGCC*

SEQ ID NOS7 & 59 -

```
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
 M  E  T  D  T  L  L  L  W  V  L  L  L  W  V  P  G  S  T  G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
 D  Y  K  D  E  H  H  H  H  H  H  G  G  S  Q  D  S  T  S  D
ctgatcccagccccacctctgagcaaggtccctctgcagcagaacttccaggacaaccaa
 L  I  P  A  P  P  L  S  K  V  P  L  Q  Q  N  F  Q  D  N  Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
 F  Q  G  K  W  Y  V  V  G  L  A  G  N  A  I  L  R  E  D  K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
 D  P  Q  K  M  Y  A  T  I  Y  E  L  K  E  D  K  S  Y  N  V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
 T  S  V  L  F  R  K  K  K  C  D  Y  W  I  R  T  F  V  P  G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
 C  Q  P  G  E  F  T  L  G  N  I  K  S  Y  P  G  L  T  S  Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
 L  V  R  V  V  S  T  N  Y  N  Q  H  A  M  V  F  F  K  K  V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
 S  Q  N  R  E  Y  F  K  I  T  L  Y  G  R  T  K  E  L  T  S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
 E  L  K  E  N  F  I  R  F  S  K  S  L  G  L  P  E  N  H  I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
 V  F  P  V  P  I  D  Q  C  I  D  G  G  G  S  E  N  L  Y  F
cagggaggcggccatgcgtgctatcgtaattgctggcgtgaaggcaatgatgaagaaacc
 Q  G  G  G  H  A  C  Y  R  N  C  W  R  E  G  N  D  E  E  T
tgcaaagaacgttgc
 C  K  E  R  C
```

SEQ ID NO:8 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-Hanatoxin

GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGCGAATGCCGTTATCTGTTTGGCGGCTGCAAAACCACCAGCGATT**
GCTGCAAACATCTGGGCTGCAAATTTCGTGATAAATATTGCGCGTGGGATTTTA
CCTTTAGCTAATGCTAAGGATCCCGGA

SEQ ID NOS9 & 60 -

```
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
 M  E  T  D  T  L  L  L  W  V  L  L  L  W  V  P  G  S  T  G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
 D  Y  K  D  E  H  H  H  H  H  H  G  G  S  Q  D  S  T  S  D
ctgatcccagccccacctctgagcaaggtccctctgcagcagaacttccaggacaaccaa
 L  I  P  A  P  P  L  S  K  V  P  L  Q  Q  N  F  Q  D  N  Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
 F  Q  G  K  W  Y  V  V  G  L  A  G  N  A  I  L  R  E  D  K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
 D  P  Q  K  M  Y  A  T  I  Y  E  L  K  E  D  K  S  Y  N  V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
 T  S  V  L  F  R  K  K  K  C  D  Y  W  I  R  T  F  V  P  G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
 C  Q  P  G  E  F  T  L  G  N  I  K  S  Y  P  G  L  T  S  Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
 L  V  R  V  V  S  T  N  Y  N  Q  H  A  M  V  F  F  K  K  V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
 S  Q  N  R  E  Y  F  K  I  T  L  Y  G  R  T  K  E  L  T  S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
 E  L  K  E  N  F  I  R  F  S  K  S  L  G  L  P  E  N  H  I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
 V  F  P  V  P  I  D  Q  C  I  D  G  G  G  S  E  N  L  Y  F
cagggaggcgaatgccgttatctgtttggcggctgcaaaaccaccagcgattgctgcaaa
 Q  G  G  E  C  R  Y  L  F  G  G  C  K  T  T  S  D  C  C  K
catctgggctgcaaatttcgtgataaatattgcgcgtgggattttacctttagc
 H  L  G  C  K  F  R  D  K  Y  C  A  W  D  F  T  F  S
```

SEQ ID NO:10 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-ChymotrypsinInhibitor

GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG

GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGC**<u>**GAAATTAGCTGCGAACCGGGCAAAACCTTTAAAGATAAATGCA**
ATACCTGCCGTTGCGGCGCGGATGGCAAAAGCGCGGCGTGCACCCTGAAAGCG
TGCCCGAATCAG</u>**TAATGCTAAGGAT**CCCGGA

SEQ ID NOS11& 61 -

```
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
 M  E  T  D  T  L  L  L  W  V  L  L  L  W  V  P  G  S  T  G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
 D  Y  K  D  E  H  H  H  H  H  H  G  G  S  Q  D  S  T  S  D
ctgatcccagccccacctctgagcaaggtcctctgcagcagaacttccaggacaaccaa
 L  I  P  A  P  P  L  S  K  V  P  L  Q  Q  N  F  Q  D  N  Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
 F  Q  G  K  W  Y  V  V  G  L  A  G  N  A  I  L  R  E  D  K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
 D  P  Q  K  M  Y  A  T  I  Y  E  L  K  E  D  K  S  Y  N  V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
 T  S  V  L  F  R  K  K  K  C  D  Y  W  I  R  T  F  V  P  G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
 C  Q  P  G  E  F  T  L  G  N  I  K  S  Y  P  G  L  T  S  Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
 L  V  R  V  V  S  T  N  Y  N  Q  H  A  M  V  F  F  K  K  V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
 S  Q  N  R  E  Y  F  K  I  T  L  Y  G  R  T  K  E  L  T  S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
 E  L  K  E  N  F  I  R  F  S  K  S  L  G  L  P  E  N  H  I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
 V  F  P  V  P  I  D  Q  C  I  D  G  G  G  S  E  N  L  Y  F
cagggaggcgaaattagctgcgaaccgggcaaaacctttaaagataaatgcaatacctgc
 Q  G  G  E  I  S  C  E  P  G  K  T  F  K  D  K  C  N  T  C
cgttgcggcgcggatggcaaaagcgcggcgtgcaccctgaaagcgtgcccgaatcag
 R  C  G  A  D  G  K  S  A  A  C  T  L  K  A  C  P  N  Q
```

SEQ ID NO:12 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-<u>ToxinK</u>

GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG

TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGCGTGTGCCGTGATTGGTTTAAAGAAACCGCGTGCCGTCATGCGA**
AAAGCCTGGGCAATTGCCGTACCAGCCAGAAATATCGTGCGAATTGCGCGAAA
ACCTGCGAACTGTGCTAATGCTAAGGATCCCGGA

SEQ ID NOS13& 62 -

```
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
 M  E  T  D  T  L  L  L  W  V  L  L  L  W  V  P  G  S  T  G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
 D  Y  K  D  E  H  H  H  H  H  H  G  G  S  Q  D  S  T  S  D
ctgatcccagccccacctctgagcaaggtcctctgcagcagaacttccaggacaaccaa
 L  I  P  A  P  P  L  S  K  V  P  L  Q  Q  N  F  Q  D  N  Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
 F  Q  G  K  W  Y  V  V  G  L  A  G  N  A  I  L  R  E  D  K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
 D  P  Q  K  M  Y  A  T  I  Y  E  L  K  E  D  K  S  Y  N  V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
 T  S  V  L  F  R  K  K  K  C  D  Y  W  I  R  T  F  V  P  G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
 C  Q  P  G  E  F  T  L  G  N  I  K  S  Y  P  G  L  T  S  Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
 L  V  R  V  V  S  T  N  Y  N  Q  H  A  M  V  F  F  K  K  V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
 S  Q  N  R  E  Y  F  K  I  T  L  Y  G  R  T  K  E  L  T  S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
 E  L  K  E  N  F  I  R  F  S  K  S  L  G  L  P  E  N  H  I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
 V  F  P  V  P  I  D  Q  C  I  D  G  G  G  S  E  N  L  Y  F
cagggaggcgtgtgccgtgattggtttaaagaaaccgcgtgccgtcatgcgaaaagcctg
 Q  G  G  V  C  R  D  W  F  K  E  T  A  C  R  H  A  K  S  L
ggcaattgccgtaccagccagaaatatcgtgcgaattgcgcgaaaacctgcgaactgtgc
 G  N  C  R  T  S  Q  K  Y  R  A  N  C  A  K  T  C  E  L  C
```

SEQ ID NO:14 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-EGFepiregulinCore

GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGCGTGAGCATTACCAAATGCAGCAGCGATATGAATGGCTATTGCC**
TGCATGGCCAGTGCATTTATCTGGTGGATATGAGCCAGAATTATTGCCGTTGCG
AAGTGGGCTATACCGGCGTGCGTTGCGAACATTTTTTTCTGTAATGCTAAGGAT
CCCGGA

SEQ ID NOS15 & 63-

atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
M E T D T L L L W V L L L W V P G S T G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
D Y K D E H H H H H H G G S Q D S T S D
ctgatcccagccccacctctgagcaaggtcctctgcagcagaacttccaggacaaccaa
L I P A P P L S K V P L Q Q N F Q D N Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
F Q G K W Y V V G L A G N A I L R E D K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
D P Q K M Y A T I Y E L K E D K S Y N V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
T S V L F R K K K C D Y W I R T F V P G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
C Q P G E F T L G N I K S Y P G L T S Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
L V R V V S T N Y N Q H A M V F F K K V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
S Q N R E Y F K I T L Y G R T K E L T S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
E L K E N F I R F S K S L G L P E N H I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
V F P V P I D Q C I D G G G S E N L Y F
cagggaggcgtgagcattaccaaatgcagcagcgatatgaatggctattgcctgcatggc
Q G G V S I T K C S S D M N G Y C L H G
cagtgcatttatctggtggatatgagccagaattattgccgttgcgaagtgggctatacc
Q C I Y L V D M S Q N Y C R C E V G Y T
ggcgtgcgttgcgaacatttttttctg
G V R C E H F F L

SEQ ID NO:16 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-Circulin

GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGCGGCATTCCGTGCGGCGAAAGCTGCGTGTGGATTCCGTGCATTA**
GCGCGGCGCTGGGCTGCAGCTGCAAAAATAAAGTGTGCTATCGTAATTAATGC
TAAGGATCCCGGA

SEQ ID NOS17 & 64 -

atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
M E T D T L L L W V L L L W V P G S T G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac

D Y K D E H H H H H H G G S Q D S T S D
ctgatcccagccccacctctgagcaaggtcctctgcagcagaacttccaggacaaccaa
L I P A P P L S K V P L Q Q N F Q D N Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
F Q G K W Y V V G L A G N A I L R E D K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
D P Q K M Y A T I Y E L K E D K S Y N V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
T S V L F R K K K C D Y W I R T F V P G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
C Q P G E F T L G N I K S Y P G L T S Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
L V R V V S T N Y N Q H A M V F F K K V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
S Q N R E Y F K I T L Y G R T K E L T S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
E L K E N F I R F S K S L G L P E N H I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
V F P V P I D Q C I D G G G S E N L Y F
cagggaggcggcattccgtgcggcgaaagctgcgtgtggattccgtgcattagcgcggcg
Q G G G I P C G E S C V W I P C I S A A
ctgggctgcagctgcaaaaataaagtgtgctatcgtaat
L G C S C K N K V C Y R N

SEQ ID NO:18 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-Brazzein
GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGC**<u>**CAGGATAAATGCAAAAAAGTGTATGAAAATTATCCGGTGAGCA**
AATGCCAGCTGGCGAATCAGTGCAATTATGATTGCAAACTGGATAAACATGCGC
GTAGCGGCGAATGCTTTTATGATGAAAAACGTAATCTGCAGTGCATTTGCGATT
ATTGCGAATAT</u>**TAATGCTAAGGAT**CCCGGA

SEQ ID NOS19 & 65-
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
M E T D T L L L W V L L L W V P G S T G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
D Y K D E H H H H H H G G S Q D S T S D
ctgatcccagccccacctctgagcaaggtcctctgcagcagaacttccaggacaaccaa
L I P A P P L S K V P L Q Q N F Q D N Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
F Q G K W Y V V G L A G N A I L R E D K

gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
D P Q K M Y A T I Y E L K E D K S Y N V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
T S V L F R K K K C D Y W I R T F V P G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
C Q P G E F T L G N I K S Y P G L T S Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
L V R V V S T N Y N Q H A M V F F K K V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
S Q N R E Y F K I T L Y G R T K E L T S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
E L K E N F I R F S K S L G L P E N H I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
V F P V P I D Q C I D G G G S E N L Y F
cagggaggccaggataaatgcaaaaaagtgtatgaaaattatccggtgagcaaatgccag
Q G G Q D K C K K V Y E N Y P V S K C Q
ctggcgaatcagtgcaattatgattgcaaactggataaacatgcgcgtagcggcgaatgc
L A N Q C N Y D C K L D K H A R S G E C
ttttatgatgaaaaacgtaatctgcagtgcatttgcgattattgcgaatat
F Y D E K R N L Q C I C D Y C E Y

SEQ ID NO:20 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-Chlorotoxin

GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGCATGTGCATGCCGTGCTTTACCACCGATCATCAGATGGCGCGTA**
AATGCGATGATTGCTGCGGCGGCAAAGGCCGTGGCAAATGCTATGGCCCGCAG
TGCCTGTGCCGTTAATGCTAAGGATCCCGGA

SEQ ID NOS21 & 66-

atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
M E T D T L L L W V L L L W V P G S T G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
D Y K D E H H H H H H G G S Q D S T S D
ctgatcccagccccacctctgagcaaggtccctctgcagcagaacttccaggacaaccaa
L I P A P P L S K V P L Q Q N F Q D N Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
F Q G K W Y V V G L A G N A I L R E D K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
D P Q K M Y A T I Y E L K E D K S Y N V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
T S V L F R K K K C D Y W I R T F V P G

```
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
 C  Q  P  G  E  F  T  L  G  N  I  K  S  Y  P  G  L  T  S  Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
 L  V  R  V  V  S  T  N  Y  N  Q  H  A  M  V  F  F  K  K  V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
 S  Q  N  R  E  Y  F  K  I  T  L  Y  G  R  T  K  E  L  T  S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
 E  L  K  E  N  F  I  R  F  S  K  S  L  G  L  P  E  N  H  I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
 V  F  P  V  P  I  D  Q  C  I  D  G  G  G  S  E  N  L  Y  F
cagggaggcatgtgcatgccgtgctttaccaccgatcatcagatggcgcgtaaatgcgat
 Q  G  G  M  C  M  P  C  F  T  T  D  H  Q  M  A  R  K  C  D
gattgctgcggcggcaaaggccgtggcaaatgctatggcccgcagtgcctgtgccgt
 D  C  C  G  G  K  G  R  G  K  C  Y  G  P  Q  C  L  C  R
```

Construction of parental construct for BamH1/NotI cloning:

SEQ ID NO:22 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GS-PARENTAL
GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAG**GGATCC**TAATGTTGGCCATGATGTTAG**GCGGCCGC**TAAGGATCCCGGA

BamHI site: GGATCC
NotI site: GCGGCCGC

A BamHI site adds "GS" before a knottin. This construct can be used for cloning libraries.

Construction of parental construct for furin cleavage, BamHI/NotI cloning can include an idealized furin cut site is RARYKRS (SEQ ID NO: 67) –RARYKRGS (SEQ ID NO: 68) can be used for a Bam HI site.

SEQ ID NO:23 - IgK-SF-H6-GGS-lcn2C-GGS-furin-GS-PARENTAL
GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA

CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCcgcgcgcgctataaacgc**G**
GATCCTAATGTTGGCCATGATGTTAG**GCGGCCGC**TAAGGATCCCGGA

SEQ ID NO:24 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GS-MIDKINE
GACTGAGTCGCCCG*CTCGAG*ACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAG**AGC**GATTGCAAATATAAATTTGAAAACTGGGGCGCGTGCGATGGCGGCACC
GGCACCAAAGTGCGCCAGGGCACCCTGAAAAAAGCGCGCTATAACGCGCAGTGCCA
GGAAACCATTCGCGTGACCAAACCGTGCTAAT**GCT***GGATCC*CGGACCGCCTCTCC

SEQ ID NOS25 & 69-

```
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
 M  E  T  D  T  L  L  L  W  V  L  L  L  W  V  P  G  S  T  G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
 D  Y  K  D  E  H  H  H  H  H  H  G  G  S  Q  D  S  T  S  D
ctgatcccagccccacctctgagcaaggtccctctgcagcagaacttccaggacaaccaa
 L  I  P  A  P  P  L  S  K  V  P  L  Q  Q  N  F  Q  D  N  Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
 F  Q  G  K  W  Y  V  V  G  L  A  G  N  A  I  L  R  E  D  K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
 D  P  Q  K  M  Y  A  T  I  Y  E  L  K  E  D  K  S  Y  N  V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
 T  S  V  L  F  R  K  K  K  C  D  Y  W  I  R  T  F  V  P  G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
 C  Q  P  G  E  F  T  L  G  N  I  K  S  Y  P  G  L  T  S  Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
 L  V  R  V  V  S  T  N  Y  N  Q  H  A  M  V  F  F  K  K  V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
 S  Q  N  R  E  Y  F  K  I  T  L  Y  G  R  T  K  E  L  T  S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
 E  L  K  E  N  F  I  R  F  S  K  S  L  G  L  P  E  N  H  I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
 V  F  P  V  P  I  D  Q  C  I  D  G  G  G  S  E  N  L  Y  F
cagagcgattgcaaatataaatttgaaaactggggcgcgtgcgatggcggcaccggcacc
 Q  S  D  C  K  Y  K  F  E  N  W  G  A  C  D  G  G  T  G  T
aaagtgcgccagggcaccctgaaaaaagcgcgctataacgcgcagtgccaggaaaccatt
 K  V  R  Q  G  T  L  K  K  A  R  Y  N  A  Q  C  Q  E  T  I
cgcgtgaccaaaccgtgc
 R  V  T  K  P  C
```

SEQ ID NO:26 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-Violacin A

GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGC**AGCGCCATCAGCTGCGGCGAGACCTGCTTCAAGTTCAAGTGCTAC
ACCCCCAGATGCAGCTGCAGCTACCCCGTGTGCAAGTAAGCTAAGGATCCCGGACC
GCC

SEQ ID NOS27 & 70-

```
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
 M  E  T  D  T  L  L  L  W  V  L  L  L  W  V  P  G  S  T  G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
 D  Y  K  D  E  H  H  H  H  H  H  G  G  S  Q  D  S  T  S  D
ctgatcccagccccacctctgagcaaggtccctctgcagcagaacttccaggacaaccaa
 L  I  P  A  P  P  L  S  K  V  P  L  Q  Q  N  F  Q  D  N  Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
 F  Q  G  K  W  Y  V  V  G  L  A  G  N  A  I  L  R  E  D  K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
 D  P  Q  K  M  Y  A  T  I  Y  E  L  K  E  D  K  S  Y  N  V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
 T  S  V  L  F  R  K  K  K  C  D  Y  W  I  R  T  F  V  P  G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
 C  Q  P  G  E  F  T  L  G  N  I  K  S  Y  P  G  L  T  S  Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
 L  V  R  V  V  S  T  N  Y  N  Q  H  A  M  V  F  F  K  K  V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
 S  Q  N  R  E  Y  F  K  I  T  L  Y  G  R  T  K  E  L  T  S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
 E  L  K  E  N  F  I  R  F  S  K  S  L  G  L  P  E  N  H  I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
 V  F  P  V  P  I  D  Q  C  I  D  G  G  G  S  E  N  L  Y  F
cagggaggcagcgccatcagctgcggcgagacctgcttcaagttcaagtgctacaccccc
 Q  G  G  S  A  I  S  C  G  E  T  C  F  K  F  K  C  Y  T  P
agatgcagctgcagctaccccgtgtgcaag
 R  C  S  C  S  Y  P  V  C  K
```

SEQ ID NO:28 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-Lambda Toxin

GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA

GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGC**GTGTGCTGCGGCTACAAGCTGTGCCACCCCTGCTAAGCTAAGGATC
CCGGACC

SEQ ID NOS29 & 71-

```
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
 M  E  T  D  T  L  L  L  W  V  L  L  L  W  V  P  G  S  T  G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
 D  Y  K  D  E  H  H  H  H  H  H  G  G  S  Q  D  S  T  S  D
ctgatcccagccccacctctgagcaaggtcctctgcagcagaacttccaggacaaccaa
 L  I  P  A  P  P  L  S  K  V  P  L  Q  Q  N  F  Q  D  N  Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
 F  Q  G  K  W  Y  V  V  G  L  A  G  N  A  I  L  R  E  D  K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
 D  P  Q  K  M  Y  A  T  I  Y  E  L  K  E  D  K  S  Y  N  V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
 T  S  V  L  F  R  K  K  K  C  D  Y  W  I  R  T  F  V  P  G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
 C  Q  P  G  E  F  T  L  G  N  I  K  S  Y  P  G  L  T  S  Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
 L  V  R  V  V  S  T  N  Y  N  Q  H  A  M  V  F  F  K  K  V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
 S  Q  N  R  E  Y  F  K  I  T  L  Y  G  R  T  K  E  L  T  S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
 E  L  K  E  N  F  I  R  F  S  K  S  L  G  L  P  E  N  H  I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
 V  F  P  V  P  I  D  Q  C  I  D  G  G  G  S  E  N  L  Y  F
cagggaggcgtgtgctgcggctacaagctgtgccacccctgc
 Q  G  G  V  C  C  G  Y  K  L  C  H  P  C
```

SEQ ID NO:30 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-Lambda Toxin NG

GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT

TTCAGGGA**GGC**AACGGCGTGTGCTGCGGCTACAAGCTGTGCCACCCCTGCTAAGCT
AAGGATCCCGGACC

SEQ ID NOS 31 & 72-

```
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
 M  E  T  D  T  L  L  L  W  V  L  L  L  W  V  P  G  S  T  G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
 D  Y  K  D  E  H  H  H  H  H  H  G  G  S  Q  D  S  T  S  D
ctgatcccagccccacctctgagcaaggtccctctgcagcagaacttccaggacaaccaa
 L  I  P  A  P  P  L  S  K  V  P  L  Q  Q  N  F  Q  D  N  Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
 F  Q  G  K  W  Y  V  V  G  L  A  G  N  A  I  L  R  E  D  K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
 D  P  Q  K  M  Y  A  T  I  Y  E  L  K  E  D  K  S  Y  N  V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
 T  S  V  L  F  R  K  K  K  C  D  Y  W  I  R  T  F  V  P  G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
 C  Q  P  G  E  F  T  L  G  N  I  K  S  Y  P  G  L  T  S  Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
 L  V  R  V  V  S  T  N  Y  N  Q  H  A  M  V  F  F  K  K  V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
 S  Q  N  R  E  Y  F  K  I  T  L  Y  G  R  T  K  E  L  T  S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
 E  L  K  E  N  F  I  R  F  S  K  S  L  G  L  P  E  N  H  I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
 V  F  P  V  P  I  D  Q  C  I  D  G  G  G  S  E  N  L  Y  F
cagggaggcaacggcgtgtgctgcggctacaagctgtgccacccctgc
 Q  G  G  N  G  V  C  C  G  Y  K  L  C  H  P  C
```

SEQ ID NO:32 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-Potato Carboxypeptidase Inhibitor

GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGC**cagcagcatgcggatccgatttgcaacaaaccgtgcaaaacccatgatgattgcagcggcgcgtggttttgcc
aggcgtgctggaacagcgcgcgcacctgcggcccgtatgtgggcTAATGCTAAGGATCCCGGACCG

SEQ ID NOS33 & 73-

```
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
 M  E  T  D  T  L  L  L  W  V  L  L  L  W  V  P  G  S  T  G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
 D  Y  K  D  E  H  H  H  H  H  H  G  G  S  Q  D  S  T  S  D
ctgatcccagccccacctctgagcaaggtccctctgcagcagaacttccaggacaaccaa
```

```
 L  I  P  A  P  P  L  S  K  V  P  L  Q  Q  N  F  Q  D  N  Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
 F  Q  G  K  W  Y  V  V  G  L  A  G  N  A  I  L  R  E  D  K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
 D  P  Q  K  M  Y  A  T  I  Y  E  L  K  E  D  K  S  Y  N  V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
 T  S  V  L  F  R  K  K  K  C  D  Y  W  I  R  T  F  V  P  G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
 C  Q  P  G  E  F  T  L  G  N  I  K  S  Y  P  G  L  T  S  Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
 L  V  R  V  V  S  T  N  Y  N  Q  H  A  M  V  F  F  K  K  V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
 S  Q  N  R  E  Y  F  K  I  T  L  Y  G  R  T  K  E  L  T  S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
 E  L  K  E  N  F  I  R  F  S  K  S  L  G  L  P  E  N  H  I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
 V  F  P  V  P  I  D  Q  C  I  D  G  G  G  S  E  N  L  Y  F
cagggaggccagcagcatgcggatccgatttgcaacaaaccgtgcaaaacccatgatgat
 Q  G  G  Q  Q  H  A  D  P  I  C  N  K  P  C  K  T  H  D  D
tgcagcggcgcgtggttttgccaggcgtgctggaacagcgcgcgcacctgcggcccgtat
 C  S  G  A  W  F  C  Q  A  C  W  N  S  A  R  T  C  G  P  Y
gtgggctaa
 V  G
```

CLAIMS

WHAT IS CLAIMED IS:

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

analyzing an isolated sample from a subject following administration of a plurality of drug candidates to the subject; and

identifying in the isolated sample at least one drug candidate having the pharmacological property.

2. 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.

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

4. The method of claim 1 or 2, wherein the plurality of drug candidates comprises a plurality of peptides.

5. The method of claim 1 or 2, wherein the plurality of drug candidates comprises a plurality of small chemical molecules.

6. The method of claim 1 or 2, wherein the plurality of drug candidates comprises a plurality of biologics.

7. The method of any one of the preceding claims, 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.

8. The method of claim 6, further comprising generating a peptide library comprising the at least some of the peptides having the pharmacological property.

9. The method of any one of the preceding claims, 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.

10. The method of claim 9, 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.

11. The method of claim 9 or 10, wherein the protein:protein interaction is associated with cancer.

12. The method of any one of the preceding claims, 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.

13. The method of claim 12, further comprising 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.

14. The method of any one of the preceding claims, 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.

15. The method of any one of the preceding claims, wherein the plurality of drug candidates comprises greater than 5 drug candidates, greater than 10 drug candidates, greater than 100 drug candidates, greater than 1000 drug candidates, or greater than 10000 drug candidates.

16. The method of any one of the preceding claims, wherein at least one peptide of the plurality of peptides comprises a detectable label for tracing the at least one peptide in the subject.

17. The method of claim 16, wherein the detectable label comprises a near infrared dye.

18. The method of any one of the preceding claims, 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.

19. The method of claim 18, wherein the hydrophobic moiety comprises a hydrophobic fluorescent dye or a saturated or unsaturated alkyl group.

20. The method of any one of the preceding claims, wherein the sample comprises a tissue sample or a fluid sample.

21. The method of claim 20, wherein the fluid sample comprises blood, urine, mucous or spinal fluid.

22. The method of claim 21, wherein the sample comprises blood.

23. The method of claim 20, wherein the tissue sample comprises biopsy or necropsy tissue.

24. The method of claim 20 or 23, 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.

25. The method of any one of the preceding claims, further comprising using the pharmacological property common to the drug candidates to generate additional drug candidates having the pharmacological property.

26. The method of any one of the preceding claims, wherein the subject is an animal.

27. The method of claim 26, wherein the subject is a human.

28. The method of any one of the preceding claims, further comprising establishing a relationship between the structure and the pharmacological property of the drug candidates.

29. The method of claim 28, further comprising selecting a subset of drug candidates based on the relationship.

30. The method of any one of the preceding claims, further comprising obtaining a plurality of samples at a selected time interval for a selected duration of time following administration.

31. The method of claim 30, wherein selected time intervals are 5 minutes, 10 minutes, 20 minutes, 30 minutes, 40 minutes, 50 minutes, or 60 minutes.

32. The method of claim 30, wherein selected duration of time is 1 hour, 2 hours, 3 hours, 4 hours, 5 hours, 6 hours, 7 hours, 8 hours, 9 hours, 10 hours, 11 hours, 12 hours, 15 hours, 30 hours, 45 hours, 60 hours, 75 hours, 90 hours, 105 hours, or 120 hours.

33. A method of generating a mass-defined drug candidate library, the method comprising:

producing a plurality of drug candidates, at least some of the drug candidates each having a unique mass signature or digest fragment mass signature;

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; and

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.

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

analyzing an isolated sample from a subject following administration of a plurality of drug candidates to the subject, wherein the library drug candidates are from the mass-defined drug candidate library of claim 33;

identifying in the isolated sample at least one drug candidate having the pharmacological property.

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

administering to a subject a plurality of library drug candidates, wherein the library drug candidates are from the mass-defined drug candidate library of claim 26;

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

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

36. The method of any one of claims 33 to 35, 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, or a combination thereof.

37. The method of claim 36, wherein the pharmacological property comprises oral bioavailability.

38. The method of any one of claims 33 to 37, wherein the plurality of drug candidates comprises greater than five drug candidates, greater than 10 drug candidates, greater than 100 drug candidates, greater than 1000 drug candidates, or greater than 10000 drug candidates.

39. The method of any one of claims 33 to 35, wherein the plurality of drug candidates comprises knotted-peptides.

40. The method of claim 39, wherein the unique mass signature or digest fragment mass signature of the knotted-peptides are defined by the natural amino acid sequence of the knotted peptides.

41. The method of any one of claims 33 to 35, wherein at least one of the drug candidates in the plurality comprises a pre-defined number of a heavy isotope atom to modify the unique mass signature or digest fragment mass signature of the at least one drug candidate.

42. The method of claim 41, wherein the heavy isotope atom comprises ^{13}C or deuterium.

43. The method of claim 39 or 40, wherein the unique mass signature or digest fragment mass signature of at least one of the knotted-peptides in the plurality is defined by a moiety conjugated to the at least one knotted-peptide.

44. The method of claim 40, wherein the moiety is conjugated to the N-terminus of the at least one knotted-peptide.

45. The method of claim 43 or 44, wherein the moiety comprises a pre-defined number of a heavy isotope atom to modify the unique mass signature or digest fragment mass signature of the at least one knotted-peptide.

46. The method of claim 45, wherein the heavy isotope atom comprises ^{13}C or deuterium.

47. The method of claim 39, wherein the plurality of knotted-peptides comprises greater than 100 peptides, greater than 1000 peptides, or greater than 10000 peptides.

48. The method of claims 33 to 47, further comprising establishing a relationship between the structure and the pharmacological property of the drug candidates.

49. The method of claim 48, further comprising selecting a subset of drug candidates based on the relationship.

50. The method of any one of claims 33 to 49, further comprising obtaining a plurality of samples at a selected time interval for a selected duration of time following administration.

51. The method of claim 50, wherein selected time intervals are 5 minutes, 10 minutes, 20 minutes, 30 minutes, 40 minutes, 50 minutes, or 60 minutes.

52. The method of claim 50, wherein selected duration of time is 1 hour, 2 hours, 3 hours, 4 hours, 5 hours, 6 hours, 7 hours, 8 hours, 9 hours, 10 hours, 11 hours, 12 hours, 15 hours, 30 hours, 45 hours, 60 hours, 75 hours, 90 hours, 105 hours, or 120 hours.

53. A method of determining a distribution profile of knotted-peptides administered to a subject by different administration pathways, the method comprising:

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;

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; and

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.

54. A method of determining a distribution profile of knotted-peptides administered to a subject by different administration pathways, the method comprising:

analyzing an isolated sample from a subject following administration to the subject of a composition comprising 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;

analyzing an isolated sample from a subject following administration to the subject of a composition comprising 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; and

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.

55. The method of claim 53 or 54, wherein the light knotted peptide comprises fewer heavier isotopes than the light knotted peptide.

56. The method of any one of claims 53 to 55, wherein the heavy knotted-peptide comprises at least one more ^{13}C atom or deuterium atom than the light knotted-peptide.

57. The method of any one of claims 53 to 56, wherein the first knotted-peptide is conjugated to a first moiety and the heavy knotted-peptide is conjugated to a second moiety.

58. The method of claim 57, wherein the first moiety, second moiety, or both is conjugated to the N-terminus of the light and heavy knotted peptide, respectively.

59. The method of claim 57, wherein the first moiety, second moiety, or both comprises a hydrophobic moiety.

60. The method of any one of claims 53 to 59, wherein the first route and second route of delivery are independently selected from an oral route, a topical route, a transmucosal route, an intravenous route, an intramuscular route, and an inhalation route.

61. The method of any one of claims 53 to 60, wherein the either the first route or the second route of delivery comprises an oral route.

62. The method of claim 57, wherein at least one ^{13}C atom is present in the first moiety, the second moiety, or both.

63. The method of any one of claims 53 to 62, wherein the light knotted-peptide, the heavy knotted-peptide, or both have unique mass signature or digest mass signature when analyzed by mass spectrometry.

64. The method of claims 53 to 63, further comprising establishing a relationship between the structure and the pharmacological property of the drug candidates.

65. The method of claim 64, further comprising selecting a subset of drug candidates based on the relationship.

66. The method of any one of claims 53 to 65, further comprising obtaining a plurality of samples at a selected time interval for a selected duration of time following administration.

67. The method of claim 66, wherein selected time intervals are 5 minutes, 10 minutes, 20 minutes, 30 minutes, 40 minutes, 50 minutes, or 60 minutes.

68. The method of claim 66, wherein selected duration of time is 1 hour, 2 hours, 3 hours, 4 hours, 5 hours, 6 hours, 7 hours, 8 hours, 9 hours, 10 hours, 11 hours, 12 hours, 15 hours, 30 hours, 45 hours, 60 hours, 75 hours, 90 hours, 105 hours, or 120 hours.

69. 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.

70. The method of claim 69, wherein the drug candidates are selected from the group consisting of small chemical molecules, biologics, and peptides.

71. The method of claim 69, wherein the plurality of drug candidates comprises a plurality of peptides.

72. The method of claim 69, 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.

73. The method of claim 71, further comprising generating a peptide library comprising the at least some of the peptides having the pharmacological property.

74. The method of claim 71, 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.

75. The method of claim 74, 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.

76. The method of claim 75, wherein the protein:protein interaction is associated with cancer.

77. The method of claim 71, 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.

78. The method of claim 77, 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.

79. The method of claim 71, 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.

80. The method of claim 69, wherein the plurality of drug candidates comprises greater than five drug candidates.

81. The method of claim 69, wherein the plurality of drug candidates comprises greater than 10 drug candidates.

82. The method of claim 69, wherein the plurality of drug candidates comprises greater than 100 drug candidates.

83. The method of claim 69, wherein the plurality of drug candidates comprises greater than 1000 drug candidates.

84. The method of claim 69, wherein the plurality of drug candidates comprises greater than 10000 drug candidates.

85. The method of claim 71, wherein the plurality of peptides comprises greater than 100 peptides.

86. The method of claim 71, wherein the plurality of peptides comprises greater than 1000 peptides.

87. The method of claim 71, wherein the plurality of peptides comprises greater than 10000 peptides.

88. The method of claim 71, wherein at least one peptide of the plurality of peptides comprises a detectable label for tracing the at least one peptide in the subject.

89. The method of claim 88, wherein the detectable label comprises a near infrared dye.

90. The method of claim 71, 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.

91. The method of claim 90, wherein the hydrophobic moiety comprises a hydrophobic fluorescent dye or a saturated or unsaturated alkyl group.

92. The method of claim 69, wherein the sample comprises a tissue sample or a fluid sample.

93. The method of claim 92, wherein the fluid sample comprises blood, urine, mucous or spinal fluid.

94. The method of claim 69, wherein the sample comprises blood.

95. The method of claim 92, wherein the tissue sample comprises biopsy or necropsy tissue.

96. The method of claim 92, 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.

97. The method of claim 69, further comprising using the pharmacological property common to the drug candidates to generate additional drug candidates having the pharmacological property.

98. The method of claim 69, wherein the subject is an animal.

99. The method of claim 69, wherein the subject is a human.

100. A method of generating mass-defined drug candidate libraries, the method comprising:

producing a plurality of drug candidates, at least some of the drug candidates each having a unique mass signature or digest fragment mass signature;

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; and

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.

101. The method of claim 100, 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, or a combination thereof.

102. The method of claim 100, wherein the pharmacological property comprises oral bioavailability.

103. The method of claim 100, wherein the drug candidates are selected from the group consisting of small chemical molecules, biologics, and peptides.

104. The method of claim 100, wherein the plurality of drug candidates comprises greater than five drug candidates.

105. The method of claim 100, wherein the plurality of drug candidates comprises greater than 10 drug candidates.

106. The method of claim 100, wherein the plurality of drug candidates comprises greater than 100 drug candidates.

107. The method of claim 100, wherein the plurality of drug candidates comprises greater than 1000 drug candidates.

108. The method of claim 100, wherein the plurality of drug candidates comprises greater than 10000 drug candidates.

109. The method of claim 100, wherein the plurality of drug candidates comprise knotted-peptides.

110. The method of claim 109, wherein the unique mass signature or digest fragment mass signature of the knotted-peptides are defined by the natural amino acid sequence of the knotted peptides.

111. The method of claim 100, wherein at least one of the drug candidates in the plurality comprises a pre-defined number of a heavy isotope atom to modify the unique mass signature or digest fragment mass signature of the at least one drug candidate.

112. The method of claim 101, wherein the heavy isotope atom comprises ^{13}C or deuterium.

113. The method of claim 109, wherein the unique mass signature or digest fragment mass signature of at least one of the knotted-peptides in the plurality is defined by a moiety conjugated to the at least one knotted-peptide.

114. The method of claim 113, wherein the moiety is conjugated to the N-terminus of the at least one knotted-peptide.

115. The method of claim 113, wherein the moiety comprises a pre-defined number of a heavy isotope atom to modify the unique mass signature or digest fragment mass signature of the at least one knotted-peptide.

116. The method of claim 115, wherein the heavy isotope atom comprises ^{13}C or deuterium.

117. The method of claim 109, wherein the plurality of knotted peptides comprises greater than 100 peptides.

118. The method of claim 109, wherein the plurality of knotted peptides comprises greater than 1000 peptides.

119. The method of claim 109, wherein the plurality of knotted peptides comprises greater than 10000 peptides.

120. A method of determining a distribution profile of knotted-peptides administered to a subject by different administration pathways, the method comprising:

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;

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; and

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.

121. The method of claim 120, wherein the light knotted peptide comprises fewer heavier isotopes than the light knotted peptide.

122. The method of claim 120, wherein the heavy knotted-peptide comprises at least one more ^{13}C atom or deuterium atom than the light knotted-peptide.

123. The method of claim 120, wherein the first knotted-peptide is conjugated to a first moiety and the heavy knotted-peptide is conjugated to a second moiety.

124. The method of claim 123, wherein the first moiety, second moiety, or both is conjugated to the N-terminus of the light and heavy knotted peptide, respectively.

125. The method of claim 123, wherein the first moiety, second moiety, or both comprises a hydrophobic moiety.

126. The method of claim 120, wherein the first route and second route of delivery are independently selected from an oral route, a topical route, a transmucosal route, an intravenous route, an intramuscular route, and an inhalation route.

127. The method of claim 120, wherein the either the first route or the second route of delivery comprises an oral route.

128. The method of claim 120, wherein at least one ^{13}C atom is present in the first moiety, the second moiety, or both.

129. The method of claim 120, wherein the light knotted-peptide, the heavy knotted-peptide, or both have unique mass signature or digest mass signature when analyzed by mass spectrometry.

130. A peptide for imaging a tumor in a subject, the peptide comprising a chlorotoxin comprising 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 comprises a detectable label.

131. The peptide of claim 130, wherein 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 peptide are D-amino acids.

132. The peptide of claim 130 or 131, having an amino acid sequence at least 80%, 83%, 86%, 89%, 90% or 92 % identical to the following sequence of MCMPCFTTDHQMARXCDDCCGGXGRGXCYGPQCLCR (SEQ ID NO: 47), wherein X is selected from K, A and R.

133. A peptide comprising an amino acid sequence having at least 80% identical to the following sequence of MCMPCFTTDHQMARXCDDCCGGXGRGXCYGPQCLCR (SEQ ID NO: 47), wherein at least three of the amino acids in the amino acid sequence are D-amino acids, wherein X is selected from K, A and R.

134. The peptide of claim 133, wherein at least 50%, at least 60%, or at least 70% of the amino acids in the peptide are D-amino acids.

135. The peptide of any of claims 133-134, further comprising a detectable label.

136. The peptide of any of claim 130-135, wherein all the amino acids in the peptide are D-amino acids.

137. The peptide any of claims 130-132 or 135-137, wherein the detectable label is conjugated to the N-terminus of the peptide or conjugated to a lysine residue in the peptide.

138. The peptide of claim 137, wherein the detectable label comprises a near-infrared dye.

139. The peptide of claim 137, wherein the detectable label comprises a cyanine dye.

140. A composition comprising the peptide of any one of claims 130-139, or a combination thereof.

141. A method of detecting a peptide in a subject, the method comprising:

administering to the subject an effective amount of the peptide according to any of claims 130-132 or the peptide according to any of claims 135-139 or the composition according to claim 126, and detecting detectable label.

142. The method of claim 141, wherein the detecting comprises obtaining an image of a region in the subject by detecting the detectable label.

143. The method of claim 142, wherein the detecting comprises intra-operative visualization of cancerous tissue.

144. The method of claim 142, wherein visualization of the detectable label guides surgical removal of a tumor in the subject.

145. A method of treating a disease associated with cells expressing a chlorotoxin target, the method comprising:

administering, to a subject in need thereof, a therapeutically effective amount of a pharmaceutical composition comprising the peptide according to claims 130-139 or the composition of claim 140, thereby treating the disease.

146. The method of claim 145, wherein the peptide of the composition further comprises a cytotoxic agent, a toxin, an antisense nucleotide, a cancer drug, a nucleotide drug, a metabolic modulator, a radiosensitizer, a peptide therapeutic, a peptide-drug conjugate, or a combination thereof.

147. The method of claim 145 or of claim 146, wherein the disease comprises cancerous tissue associated with glioma, skin cancer, lung cancer, lymphoma, medulloblastoma, prostate cancer, pancreatic cancer, breast cancer, mammary cancer, colon cancer, sarcoma, oral squamous cell carcinoma, hemangiopericytoma, or a combination thereof.

ABSTRACT OF THE DISCLOSURE

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.

METHODS FOR SCREENING

CROSS REFERENCES TO RELATED APPLICATIONS

[0001] The present application claims the benefit of U.S. Provisional Application No. 61/735,516 filed December 10, 2012 and U.S. Provisional Application No. 61/794,685 filed March 15, 2013, both of which are 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 numbers R01CA135491, NIH ROI AI059543, AI094419, and AI097786 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, the 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 new 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). In various aspects, the present disclosure relates to methods for identifying a drug candidate having a pharmacological property, the method comprising, analyzing an isolated sample from a subject following administration of a plurality of drug candidates to the subject; and identifying in the isolated sample at least one drug candidate having the pharmacological property.

[0007] In some aspects, the present disclosure relates to methods 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.

[0008] In some aspects, the present disclosure relates to methods of generating a mass-defined drug candidate library where the method comprises producing a plurality of drug candidates, at least some of the drug candidates each having a unique mass signature or digest fragment mass signature and 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. In certain embodiments, the methods 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.

[0009] In some aspects, the present disclosure relates to methods for identifying a library of drug candidates having a pharmacological property where the method comprises analyzing an isolated sample from a subject following administration of a plurality of drug candidates to the subject. In certain embodiments, the library of drug candidates are from the mass-defined drug candidate library provided herein. In certain embodiments, the methods include identifying in the isolated sample at least one drug candidate having the pharmacological property.

[0010] In some aspects, the present disclosure relates to methods for identifying library drug candidates having a pharmacological property where the method comprises administering to a subject a plurality of library drug candidates, wherein the library drug candidates are from the

mass-defined drug candidate library provided herein, obtaining, from the subject, a sample comprising at least some of the plurality of library drug candidates; and analyzing the sample to determine the identity of the at least some of the plurality of library drug candidates having the pharmacological property.

[0011] In some aspects, the present disclosure provides methods of determining a distribution profile of knotted-peptides administered to a subject by different administration pathways where the method comprises 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; 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; and 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.

[0012] In some aspects, the present disclosure provides peptides for imaging a tumor in a subject, the peptide comprising a chlorotoxin comprising 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 comprises a detectable label.

[0013] In some aspects of the present disclosure, methods are provided for treating a disease associated with cells expressing a chlorotoxin target, the method comprising: administering, to a subject in need thereof, a therapeutically effective amount of a pharmaceutical composition comprising a peptide according to the present disclosure or a composition comprising a peptide of the present disclosure, thereby treating the disease.

INCORPORATION BY REFERENCE

[0014] 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

[0015] 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:

[0016] FIG. 1 provides a gel image showing a variety of scaffolds, in accordance with an embodiment of the present invention.

[0017] 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.

[0018] FIG. 3 illustrates an example method of generating peptide libraries, in accordance with an embodiment of the present invention.

[0019] FIG. 4 shows an example method of producing peptide libraries, in accordance with an embodiment of the present invention.

[0020] 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.

[0021] FIG. 6 shows analysis of a large library using mass spectrometry, in accordance with an embodiment of the present invention.

[0022] FIG. 7 shows an example method of using siderocalin fusions to express knottin variants, in accordance with an embodiment of the present invention.

[0023] FIG. 8 depicts SDS-PAGE analysis of expressing knottin scaffolds, in accordance with embodiments of the present invention.

[0024] FIG. 9 provides a schematic of pooled library production, in accordance with an embodiment of the present invention. FIG. 9B is an expanded view of FIG. 9A.

[0025] FIG. 10 describes representative sequencing data from a cloned knottin library, in accordance with an embodiment of the present invention.

[0026] FIG. 11 shows SDS-PAGE analysis of 3000 member knottin libraries, in accordance with embodiments of the present invention.

[0027] FIG. 12 shows an assessment of mutational tolerance by reverse proteomics, in accordance with embodiments of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

[0028] The present disclosure relates to systems and methods for drug discovery. In certain aspects, the present disclosure relates to methods for identifying drug candidates having a desired pharmacological property. In further aspects, the present disclosure relates to methods of generating mass-defined drug candidate libraries. In other aspects, the present disclosure relates to methods of determining a distribution profile of knotted-peptides administered to a subject by different administration pathways.

Generating Selected Libraries

[0029] In one aspect, the present invention relates to methods and systems for generating libraries of potential drug candidates (e.g., libraries of small chemical molecules, biologics, peptides, and any derivatives or fragments thereof) that can be screened to identify sub-libraries of potential drug candidates (e.g., sub-libraries of small chemical molecules, biologics, peptides, and any derivatives or fragments thereof) 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 pharmacological 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. In some embodiments, 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. In certain embodiments, the hydrophobic moiety comprises a hydrophobic fluorescent dye or a saturated or unsaturated alkyl group.

[0030] In some aspects of the present disclosure, the plurality of drug candidates comprises a plurality of peptides.

[0031] In some aspects of the present disclosure, the plurality of drug candidates comprises a plurality of small chemical molecules.

[0032] In some aspects of the present disclosure, the plurality of drug candidates comprises a plurality of biologics.

[0033] In some embodiments, 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, or a combination thereof. In certain embodiments, the pharmacological property comprises oral bioavailability.

[0034] As used herein, the term "drug candidate" means any agent that could potentially elicit an effect on a target or subject. In some aspects, the drug candidate may have a therapeutic effect. Exemplary drug candidates may include, but are not limited to, small chemical molecules, biologics, peptides, and any derivatives or fragments thereof.

[0035] As used herein, a "pharmacological property" is a property of a drug candidate that can be used to characterize the potential of a drug candidate to elicit an effect on a target or subject. Exemplary pharmacologic properties may include, but are not limited to, 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, and the like.

[0036] 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.

[0037] Large numbers of drug candidates can be administered to a subject and screened to determine, 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.

[0038] 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.

[0039] As used herein, the term "protein:protein interaction" means a physical contact between one protein and another which can be transient or permanent. The interaction of a first protein with a second protein may cause an effect on either the first or the second protein such that the effect changes either the first or the second or both proteins in a detectable manner.

[0040] 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., small chemical molecules, biologics, peptides, and any derivatives or fragments thereof) having one or more drug-like properties, such as those described above. For example, in a first screen, thousands of candidates (e.g., small chemical molecules, biologics, peptides, and any derivatives or fragments thereof) 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.

[0041] 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.

[0042] In some aspects, the drug candidate is a small chemical molecule. With small chemical molecule libraries, e.g., the small chemical can have a variety of common characteristics that cause the small chemical molecules to have certain drug-like properties. Rules can be identified to help define which of the small chemical molecules may be more likely to have the drug-like properties. In various aspects, rules can include parameters for absorption, distribution, metabolism, and excretion, which are indicative of utility as a potential drug. For example, rules can include hydrogen bonding, lipophilicity, molecular weight, pKa, polar surface area, shape, reactivity, solubility, permeability, chemical stability, metabolism (phases I and II), protein and tissue binding, transport (e.g., uptake, efflux), clearance, half-life, bioavailability, drug-drug interactions, LD_{50}, and the like.

[0043] 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.

[0044] In some embodiments, the rules could also include identifying the number of mutations per peptide construct. In some aspects, a rule might include a number of mutations in the range of 4-8 amino acid mutations per peptide construct. In some aspects, a rule could include about 4 amino acid mutations, 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 aspects, 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 or aid identification by MS.

[0045] As used herein, the term "fragment" refers to a portion of a peptide or a protein which is generated by cleavage of a peptide or protein from which the fragment was derived.

[0046] As used herein, the term "mass signature" refers to the pattern of peaks output by a mass spectrometer by performing an analysis of fragments of at least one protein or at least one peptide. The mass signature may be unique for each fragment or the mass signature may be the same for some fragments.

[0047] 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)

[0048] 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.

[0049] 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 molecules 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.

[0050] 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.

[0051] 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.

[0052] The small molecules to be screened as potential drug candidates can have molecular weights in the range if 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.

[0053] In some aspects, the present invention includes methods and systems for generating large libraries of knotted peptides (also referred to as "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).

[0054] 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 β- 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.

[0055] 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 predefined 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.

[0056] 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 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.

[0057] 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.

[0058] 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 10-50 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.

[0059] 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.

[0060] 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 10-35 amino acids, about 10-20 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.

[0061] 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.

[0062] 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 Ka, 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.

[0063] 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.

In some aspects, the present disclosure provides peptides for imaging a tumor in a subject, the peptide comprising a chlorotoxin comprising 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 comprises a detectable label.

In further aspects of the present disclosure, 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 peptide are D-amino acids.

In some aspects, the peptides of the present invention can have an amino acid sequence at least 80%, 83%, 86%, 89%, 90% or 92 % identical to the following sequence of MCMPCFTTDHQMARXCDDCCGGXGRGXCYGPQCLCR, wherein X is selected from K, A and R.

In some aspects, the peptides of the present invention, the peptides comprise an amino acid sequence having at least 80% identical to the following sequence of MCMPCFTTDHQMARXCDDCCGGXGRGXCYGPQCLCR, wherein at least three of the amino acids in the amino acid sequence are D-amino acids, wherein X is selected from K, A and R.

In further aspects, the peptides of the present disclosure are at least 50%, at least 60%, or at least 70% of the amino acids in the peptide are D-amino acids.

In some aspects, the peptides of the present disclosure further comprise a detectable label.

In some aspects of the present disclosure, all the amino acids in a given peptide are D-amino acids.

In some aspects of the present disclosure, the detectable label is conjugated to the N-terminus of the peptide or conjugated to a lysine residue in the peptide.

In other aspects of the present disclosure, the detectable label comprises a near-infrared dye.

In other aspects of the present disclosure, the detectable label comprises a cyanine dye.

In various aspects of the present disclosure, a composition can be produced comprising any of the peptides described herein, or a combination thereof.

In some aspects of the present disclosure, methods are provided for detecting a peptide in a subject, the method comprising: administering to the subject an effective amount of a peptide of the present disclosure, or a composition thereof, and detecting a detectable label contained therein.

In further aspects, the present disclosure further comprises obtaining an image of a region in the subject by detecting the detectable label. In further aspects, the detecting comprises intra-operative visualization of cancerous tissue. In still further aspects, visualization of the detectable label guides surgical removal of a tumor in the subject.

In some aspects of the present disclosure, methods are provided for treating a disease associated with cells expressing a chlorotoxin target, the method comprising: administering, to a subject in need thereof, a therapeutically effective amount of a pharmaceutical composition comprising a peptide according to the present disclosure or a composition comprising a peptide of the present disclosure, thereby treating the disease.

In further aspects, the peptide composition further comprises a cytotoxic agent, a toxin, an antisense nucleotide, a cancer drug, a nucleotide drug, a metabolic modulator, a radiosensitizer, a peptide therapeutic, a peptide-drug conjugate, or a combination thereof. In still further aspects, the disease comprises cancerous tissue associated with glioma, skin cancer, lung cancer, lymphoma, medulloblastoma, prostate cancer, pancreatic cancer, breast cancer, mammary cancer, colon cancer, sarcoma, oral squamous cell carcinoma, hemangiopericytoma, or a combination thereof.

[0064] 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.

[0065] 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.

[0066] 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.

[0067] 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.

[0068] 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.

[0069] 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).

[0070] 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).

[0071] 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).

[0072] 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 (SAWSHPQFEK)). 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.

[0073] 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

[0074] 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.

[0075] In certain aspects, 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 palmitylation) 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.

[0076] 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.

[0077] 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 780nm, or between a range of 720 to 750nm. 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.

[0078] As used herein, the term "detectable label" means a tag or modification that can be attached to a small chemical molecule, peptide, protein, or a fragment or a portion thereof such that the small chemical molecule, peptide, protein, or a fragment thereof is recognizable using a device, apparatus or method that permits the detection of the tag or modification.

[0079] In some aspects, the detectable label is a fluorescent dye. 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, aminorhodol, carboxyrhodol, chlororhodol, methylrhodol, sulforhodol; aminorhodamine, carboxyrhodamine, chlororhodamine, methylrhodamine, sulforhodamine, and thiorhodamine, cyanine, indocarbocyanine, oxacarbocyanine, thiacarbocyanine, 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, phtalocyanine, 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.

[0080] 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-dPEG$_4$-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 flour 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.

[0081] 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 B_2, vitamin B_3, vitamin B_6, vitamin B_{12}, 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.

[0082] 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

[0083] 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.

[0084] 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: MCMPCFTTDHQMARKCDDCCGGKGRGKCYGPQCLCR. 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.

[0085] 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.

[0086] 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 MCMPCFTTDHQMARKCDDCCGGKGRGKCYGPQCLCR, 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.

[0087] 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.

[0088] 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.

[0089] 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.

[0090] 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).

[0091] 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).

[0092] 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).

[0093] 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.

[0094] 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.

[0095] 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).

[0096] 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).

[0097] 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).

[0098] 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).

[0099] 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 MCMPCFTTDHQMARXCDDCCGGXGRGXCYGPQCLCR, 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.

[00100] 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.

[00101] 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).

[00102] 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 780nm, or between a range of 720 to 750nm. 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.

[00103] 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, aminorhodol,

carboxyrhodol, chlororhodol, methylrhodol, sulforhodol; aminorhodamine, carboxyrhodamine, chlororhodamine, methylrhodamine, sulforhodamine, and thiorhodamine, cyanine, indocarbocyanine, oxacarbocyanine, thiacarbocyanine, 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, phtalocyanine, 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.

[00104] 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.

[00105] 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 flour 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.

[00106] 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.

[00107] 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.

[00108] 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.

[00109] 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.

[00110] 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.

[00111] 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.

[00112] 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.

[00113] 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.

[00114] 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.

[00115] 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 biological 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.

[00116] "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).

[00117] 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).

[00118] 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.

[00119] 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.

[00120] 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.

[00121] 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.

[00122] 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, carminomycin, daunorubicin, aminopterin, methotrexate, methopterin, dichloromethotrexate, mitomycin C, porfiromycin, 5-fluorouracil, 6-mercaptopurine, gemcitabine, cytosine arabinoside, podophyllotoxin or podo- phyllotoxin derivatives, such as etoposide, etoposide phosphate or teniposide, melphalan, vinblastine, vincristine, leurosidine, vindesine, leurosine, paclitaxel and therapeutically effective analogs and derivatives of the same. Other useful antineoplastic agents include estramustine, cisplatin, carboplatin, cyclophosphamide, bleomycin, gemcitibine, ifosamide, melphalan, hexamethyl melamine, thiotepa, cytarabin, idatrexate, trimetrexate, dacarbazine, L- asparaginase,

camptothecin, CPT-11, topotecan, ara-C, bicalutamide, flutamide, leuprolide, pyridobenzoindole derivatives, interferons and interleukins.

[00123] Suitable metabolic modulators can include, but are not limited to, lonidamine, dichloroacetate, alpha-tocopheryl succinate, methyl jasmonate, betulinic acid, and resveratrol

[00124] 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, EO9, RB 6145, nicotinamide, 5-bromodeoxyuridine (BUdR), 5-iododeoxyuridine (IUdR), bromodeoxycytidine, fluorodeoxyuridine (FudR), hydroxyurea, cisplatin, and therapeutically effective analogs and derivatives of the same.

[00125] 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.

[00126] In certain embodiments, the present invention can include treating diseases, disorders, and/or conditions, such as 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, 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 including 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.

[00127] 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.

[00128] 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.

[00129] 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.

[00130] 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.

[00131] 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.

[00132] 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.

[00133] 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.

[00134] One example of a mouse model of cancer is the autochthonous mouse model of medulloblastoma, ND2:SmoAl. See A.R. Hallahan, et al., "The SmoAl 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:SmoAl) 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.

[00135] 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.

[00136] 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.

[00137] 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 $C0_2$ 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 μm 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.

[00138] 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.

[00139] 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 x 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 uM) for 2 hours in a 37°C. humidified incubator maintained at 5% CO_2. 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 uM 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-phenylndole (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.

[00140] 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 µl PBS into the brain 3 mm lateral and posterior to the bregma.

Methods of Making Libraries

[00141] 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.

[00142] 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.

In some embodiments, the method further comprises generating a peptide library comprising the at least some of the peptides having the pharmacological property. In certain embodiments, the

methods further comprise 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. In some embodiments, 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. In certain embodiments, the protein:protein interaction is associated with cancer. In some embodiments, the method further comprises obtaining from the subject a plurality of samples, each sample being obtained at a different time point after administering the plurality of peptides. In certain embodiments, at least one peptide of the plurality of peptides comprises a detectable label for tracing the at least one peptide in the subject. In certain embodiments, the detectable label comprises a near infrared dye.

[00143] 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.

[00144] 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.

[00145] 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.

[00146] 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 used for difficult-to-express proteins of medical interest such as chemokines, interleukins, and peptide hormones.

[00147] 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 or 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.

[00148] 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.

[00149] 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 fluorescens. 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.

[00150] 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.

[00151] 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.

[00152] 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 the methods described herein.

[00153] 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.

[00154] 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.

[00155] 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 prescreened 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.

[00156] In some embodiments, the present invention relates to methods of generating a mass-defined drug candidate library where the method comprises producing a plurality of drug candidates, at least some of the drug candidates each having a unique mass signature or digest fragment mass signature and 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. In certain embodiments, the methods 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.

[00157] The invention further provides methods of generating mass-defined drug candidate libraries, where the method comprises, producing a plurality of drug candidates, at least some of the drug candidates each having a unique mass signature or digest fragment mass signature; 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; and 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.

[00158] In some embodiments, at least one of the drug candidates in the plurality comprises a pre-defined number of a heavy isotope atom to modify the unique mass signature or digest fragment mass signature of the at least one drug candidate. In certain embodiments, the heavy isotope atom comprises ^{13}C or deuterium. In some embodiments, the unique mass signature or digest fragment mass signature of at least one of the knotted-peptides in the plurality is defined by a moiety conjugated to the at least one knotted-peptide. In certain embodiments, the moiety is conjugated to the N-terminus of the at least one knotted-peptide. In some embodiments, the moiety comprises a pre-defined number of a heavy isotope atom to modify the unique mass signature or digest fragment mass signature of the at least one knotted-peptide. In certain embodiments, the heavy isotope atom comprises ^{13}C or deuterium. In some embodiments, the plurality of knotted peptides comprises greater than 100 peptides, greater than 1000 peptides, or greater than 10000 peptides.

[00159] In some embodiments, the plurality of drug candidates comprise knotted-peptides. In certain embodiments, the unique mass signature or digest fragment mass signature of the knotted-peptides are defined by the natural amino acid sequence of the knotted peptides. In certain embodiments, at least one of the drug candidates in the plurality comprises a pre-defined number of a heavy isotope atom to modify the unique mass signature or digest fragment mass signature of the at least one drug candidate. In certain embodiments, the heavy isotope atom comprises ^{13}C or deuterium.

[00160] In some embodiments, the unique mass signature or digest fragment mass signature of at least one of the knotted-peptides in the plurality is defined by a moiety conjugated to the at least

one knotted-peptide. In certain embodiments, the moiety is conjugated to the N-terminus of the at least one knotted-peptide. In certain embodiments, the moiety comprises a pre-defined number of a heavy isotope atom to modify the unique mass signature or digest fragment mass signature of the at least one knotted-peptide. In certain embodiments, the heavy isotope atom comprises ^{13}C or deuterium. In certain embodiments, the plurality of knotted peptides comprises greater than 100 peptides. In certain embodiments, the plurality of knotted peptides comprises greater than 1000 peptides. In certain embodiments, the plurality of knotted peptides comprises greater than 10000 peptides.

[00161] The invention further provides methods of determining a distribution profile of knotted-peptides administered to a subject by different administration pathways where the method comprises 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; 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; and 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.

[00162] In some embodiments, the light knotted peptide comprises fewer heavier isotopes than the light knotted peptide. In some embodiments, the heavy knotted-peptide comprises at least one more ^{13}C atom or deuterium atom than the light knotted-peptide. In certain embodiments the first knotted-peptide is conjugated to a first moiety and the heavy knotted-peptide is conjugated to a second moiety. In other embodiments, the first moiety, second moiety, or both is conjugated to the N-terminus of the light and heavy knotted peptide, respectively. In certain embodiments, the first moiety, second moiety, or both comprises a hydrophobic moiety.

[00163] In some embodiments, the light knotted peptide comprises fewer heavier isotopes than the light knotted peptide. In other embodiments, the heavy knotted-peptide comprises at least one more ^{13}C atom or deuterium atom than the light knotted-peptide. In certain embodiments, the first knotted-peptide is conjugated to a first moiety and the heavy knotted-peptide is conjugated to a second moiety. In certain embodiments, the first moiety, second moiety, or both is conjugated to the N-terminus of the light and heavy knotted peptide, respectively. In some embodiments, the first moiety, second moiety, or both comprises a hydrophobic moiety. In some embodiments, at least one ^{13}C atom is present in the first moiety, the second moiety, or both. In some

embodiments, the light knotted-peptide, the heavy knotted-peptide, or both have unique mass signature or digest mass signature when analyzed by mass spectrometry.

[00164] In some embodiments, at least one ^{13}C atom is present in the first moiety, the second moiety, or both. In other embodiments, wherein the light knotted-peptide, the heavy knotted-peptide, or both have unique mass signature or digest mass signature when analyzed by mass spectrometry.

[00165] In some embodiments, 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. In certain embodiments, the hydrophobic moiety comprises a hydrophobic fluorescent dye or a saturated or unsaturated alkyl group.

[00166] In some embodiments, at least one peptide of the plurality of peptides comprises a detectable label for tracing the at least one peptide in the subject. In certain embodiments, the detectable label comprises a near infrared dye.

Methods of Analyzing Libraries

[00167] 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. The methods provided herein further comprise 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. In some embodiments, 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. In other embodiments, the plurality of drug candidates comprises greater than 5 drug candidates, greater than 10 drug candidates, greater than 100 drug candidates, greater than 1000 drug candidates, or greater than 10000 drug candidates. In certain embodiments, at least one peptide of the plurality of peptides comprises a detectable label for tracing the at least one peptide in the subject. In certain embodiments, the detectable label comprises a near infrared dye.

[00168] In one aspect, the methods include generating peptide libraries in which some or all of the peptides in a library have a unique mass signature or digest fragment signature (e.g., by trypsin). In certain aspects, peptides of a generated peptide library are digested into smaller, more manageable fragments using enzymes such as trypsin. By doing so, peptides samples can more readily be analyzed by mass spectrometry.

[00169] 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.

[00170] 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.

[00171] 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.

[00172] 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.

[00173] 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 ^{2}H atom, at least one ^{3}H 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 ^{29Si} atom, or at least one ^{30}Si atom.

[00174] 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 ^{1}H, ^{12}C, ^{14}N, ^{16}O, or ^{32}S, or any combination thereof.

[00175] 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 spectral (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 cyclotron 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.

[00176] 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.

[00177] In some aspects of the present disclosure, peptides in a peptide library are conjugated. Conjugation can occur at any suitable location within a peptide, for example, the peptide can be conjugated at any amino acid residue making up the peptide or at either its N-terminus or C-terminus. In various aspects, peptides of the present disclosure are conjugated at their N-termini. In some aspects of the present disclosure, peptide libraries can include conjugated peptides.

[00178] 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 aspects, 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 aspects, 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.

[00179] In some aspects of the present disclosure, conjugated peptides can be isolated from a pool of tryptically digested peptide fragments in the peptide library by binding the conjugate.

For example, conjugated peptides can be immobilized onto a surface by selectively binding the conjugate. In certain aspects, the N-terminus of a peptide is conjugated with biotin or palmitic acid for recovery following proteolysis with trypsin. Any suitable conjugation-binding agent combination can be used to isolate peptides of the present disclosure.

[00180] In various aspects of the present disclosure, the conjugated peptide fragments have unique mass spectral signatures relative to other conjugated peptide fragments in the library. In further aspects, the mass spectral signature of a conjugated peptide is different from the mass spectral signature of the corresponding unconjugated peptide. In various aspects, the mass spectral signature of a peptide lacking N-terminal conjugation is different than the mass spectral signature of the corresponding peptide having N-terminus conjugation.

[00181] In various aspects, conjugation of a peptide can be performed at a single amino acid. In some aspects, the amino acid may be cysteine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, valine, alanine, arginine, asparagines, aspartic acid, glutamic acid, glutamine, glycine, ornithine, proline, selenocysteine, serine or tyrosine. In an exemplary case, the amino acid may be cysteine. In other aspects, the conjugation can be performed at more than one amino acid. In some aspects, the more than one amino acid may be cysteine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, valine, alanine, arginine, asparagines, aspartic acid, glutamic acid, glutamine, glycine, ornithine, proline, selenocysteine, serine or tyrosine.

[00182] In some aspects of the present disclosure, peptides can be conjugated to prevent oxidation. For example, cysteine residues can be conjugated following reduction in order to prevent re-oxidation of the cysteine residues. Any suitable thiol-alkylating agent can be used to prevent re-oxidation of cysteine residues following reduction. In certain aspects, cysteine residues are conjugated with iodoacetamide, N-ethylmaleimide, or iodoacetic acid to prevent re-oxidation of cysteine residues following reduction. In various aspects, cysteine residues are conjugated with iodoacetamide. In certain aspects, peptides can be conjugated with "heavy" iodoacetamide, thereby enabling its use as a reference standard for material conjugated to standard (i.e., "light") iodoacetamide. This strategy is particularly useful for knottin peptide libraries due to the large numbers of cysteine residues in those peptides.

[00183] In various aspects of the present disclosure, the conjugate atoms can be either "heavy" or "light," depending on the relative atomic mass of the atom. For example, the ^{13}C atom is a "heavy" atom and the ^{12}C atom is a "light" atom. In some aspects, for example, a conjugate that contains at least one ^{13}C atom is heavy relative to a conjugate that contains only ^{12}C atoms. For

example, in some aspects, a conjugate that contains more than two ^{13}C atoms is heavy relative to a conjugate that contains only one ^{13}C atom. In some aspects, atoms besides carbon may be used for heavy and light conjugates.

[00184] 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).

[00185] 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.

[00186] In various aspects of the present disclosure, a heavy library can be spiked into samples containing an unknown quantity of light library members. The heavy library peaks observed during mass spectral analysis can then be used to help quantitate the light library peaks appearing in the same sample.

[00187] A heavy library is also useful to monitor the relative amounts of candidates in a library following a particular procedure. For example, a small amount of heavy library can be mixed into a standard library to determine a baseline of relative peak heights. Because the heavy library is chemically identical but different in mass, in this mixture every member of the standard library will be seen by LC/MS/MS to have a heavy peak adjacent to it. The heights of these heavy peaks will form a gauge by which the amounts of members of the standard library can be judged. So the standard library could be, for example, incubated in serum to test for susceptibility to serum proteases. Following incubation in serum, the material could be heated to inactivate the serum proteases, and the heavy library doped in as a standard. The LC/MS/MS peak heights of the heavy material can then be compared to the peak heights of the standard library to determine which library members were subject to proteolysis. The central point here is that heavy libraries can be doped into standard ones to provide reference peaks. The different peptides will "fly" differently just based upon their innate qualities, affecting their peak heights independent of concentration, but they will fly the same as their heavy counterparts, except that the heavy ones

will be shifted by their mass. So the peak heights of the heavy ones can be used to calibrate the peak heights of the standard ones to provide concentration from peak heights.

[00188] 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.

[00189] 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 ^{2}H atom, at least one ^{3}H 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 ^{29Si} 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 or candidate linkers.

Methods of Using Libraries

[00190] 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.

[00191] 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 (1α(PCDHA1), $Ca_v2.2$, $K_v1.3$, $Na_v1.2$, 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, Orail, CD47, Raf, NFκB, Bromodomains, HATS, HDAC, LDH, IDH2, CD22, MIC, c-

Myc, n-Myc, PHF5A, BUB1B, Bcl-2, k-Ras, Notch1, p53, α5β3, NKG2D, CTLA4/CD28, and/or Mcl-1.

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.

[00192] In some embodiments, the drug candidates are selected from the group consisting of small chemical molecules, biologics, and peptides. In certain embodiments, the plurality of drug candidates comprises a plurality of peptides.

[00193] 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.

[00194] 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.

[00195] 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.

[00196] 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.

[00197] The present invention further relates to methods for identifying a library of drug candidates having a pharmacological property where the method comprises analyzing an isolated sample from a subject following administration of a plurality of drug candidates to the subject. In certain embodiments, the library of drug candidates are from the mass-defined drug candidate library provided herein. In certain embodiments, the methods include identifying in the isolated sample at least one drug candidate having the pharmacological property.

[00198] In some embodiments, the methods further comprise using the pharmacological property common to the drug candidates to generate additional drug candidates having the pharmacological property.

[00199] In some embodiments, 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, or a combination thereof. In certain embodiments, the pharmacological property comprises oral bioavailability.

[00200] In some embodiments, the present invention relates to methods for identifying library drug candidates having a pharmacological property where the method comprises administering to a subject a plurality of library drug candidates, wherein the library drug candidates are from the mass-defined drug candidate library provided herein, obtaining, from the subject, a sample comprising at least some of the plurality of library drug candidates; and analyzing the sample to determine the identity of the at least some of the plurality of library drug candidates having the pharmacological property.

[00201] In some embodiments, the invention provides methods 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.

[00202] In some embodiments, 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. In certain embodiments, the method further comprises generating a peptide library comprising the at least some of the peptides having the pharmacological property.

[00203] In some embodiments, the first route and second route of delivery are independently selected from an oral route, a topical route, a transmucosal route, an intravenous route, an intramuscular route, and an inhalation route. In certain embodiments, the either the first route or the second route of delivery comprises an oral route.

[00204] 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.

[00205] In some embodiments, the method further comprises using the pharmacological property common to the drug candidates to generate additional drug candidates having the pharmacological property. In some embodiments, the subject is an animal. In certain embodiments, the subject is a human.

[00206] In some embodiments, the sample comprises a tissue sample or a fluid sample. In certain embodiments, the fluid sample comprises blood, urine, mucous or spinal fluid. In certain embodiments, the sample comprises blood. In some embodiments, the tissue sample comprises biopsy or necropsy tissue. In certain embodiments, the tissue comprises tissue from a brain, a lung, a kidney, a muscle, a liver, a heart, a stomach, a pancreas, or an organ. In some embodiments, the subject is an animal. In some embodiments, the subject is a human.

[00207] In some embodiments, the sample comprises a tissue sample or a fluid sample. In some embodiments, the fluid sample comprises blood, urine, mucous or spinal fluid. In certain embodiments, the sample comprises blood. In some embodiments, the tissue sample comprises biopsy or necropsy tissue. In other embodiments, the tissue comprises tissue from a brain, a lung, a kidney, a muscle, a liver, a heart, a stomach, a pancreas, or an organ.

[00208] In some aspects of the present disclosure, the method further comprises establishing a relationship between the structure and the pharmacological property of the drug candidates, and in further aspects, a subset of drug candidates based on that relationship.

[00209] In some aspects of the present disclosure, the methods further comprise obtaining a plurality of samples at a selected time interval for a selected duration of time following administration. In certain aspects, the time intervals are 5 minutes, 10 minutes, 20 minutes, 30 minutes, 40 minutes, 50 minutes, or 60 minutes. In still further aspects, the selected duration of time is 1 hour, 2 hours, 3 hours, 4 hours, 5 hours, 6 hours, 7 hours, 8 hours, 9 hours, 10 hours, 11

hours, 12 hours, 15 hours, 30 hours, 45 hours, 60 hours, 75 hours, 90 hours, 105 hours, or 120 hours.

[00210] In some embodiments, the methods further comprise obtaining from the subject a plurality of samples, each sample being obtained at a different time point after administering the plurality of peptides. In some embodiments, the methods further comprise 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. In certain embodiments, 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. In certain embodiments, the plurality of drug candidates comprises greater than five drug candidates. In other embodiments, the plurality of drug candidates comprises greater than 10 drug candidates. In other embodiments, the plurality of drug candidates comprises greater than 100 drug candidates. In other embodiments, the plurality of drug candidates comprises greater than 1000 drug candidates. In other embodiments, the plurality of drug candidates comprises greater than 10000 drug candidates. In other embodiments, the plurality of peptides comprises greater than 100 peptides. In other embodiments, the plurality of peptides comprises greater than 1000 peptides. In other embodiments, the plurality of peptides comprises greater than 10000 peptides.

[00211] In some embodiments, the plurality of drug candidates comprises greater than five drug candidates, greater than 10 drug candidates, greater than 100 drug candidates, greater than 1000 drug candidates, or greater than 10000 drug candidates. In some embodiments, the plurality of drug candidates comprise knotted-peptides. In certain embodiments the unique mass signature or digest fragment mass signature of the knotted-peptides are defined by the natural amino acid sequence of the knotted peptides.

[00212] A method for identifying a drug candidate having a pharmacological property, the method comprising, analyzing an isolated sample from a subject following administration of a plurality of drug candidates to the subject; and identifying in the isolated sample at least one drug candidate having the pharmacological property. In some embodiments, the drug candidates are selected from the group consisting of small chemical molecules, biologics, and peptides. In certain embodiments, the plurality of drug candidates comprises a plurality of peptides. In other embodiments, 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.

[00213] 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. In some embodiments, the drug candidates are selected from the group consisting of small chemical molecules, biologics, and peptides. In certain embodiments, the plurality of drug candidates comprises a plurality of peptides. In other embodiments, 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. In some embodiments, the method further comprises obtaining from the subject a plurality of samples, each sample being obtained at a different time point after administering the plurality of peptides. In certain embodiments, at least one peptide of the plurality of peptides comprises a detectable label for tracing the at least one peptide in the subject. In certain embodiments, the detectable label comprises a near infrared dye.

[00214] In some embodiments, the first route and second route of delivery are independently selected from an oral route, a topical route, a transmucosal route, an intravenous route, an intramuscular route, and an inhalation route. In certain embodiments, the either the first route or the second route of delivery comprises an oral route.

[00215] In some embodiments, the method further comprises 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. In other embodiments, 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. In certain embodiments, the protein:protein interaction is associated with cancer.

EXEMPLARY ASPECTS

EXAMPLE 1

Expressing Peptide Constructs for Knottin Generation

[00216] This example describes a method for expressing peptide constructs in culture and greatly facilitating their development, particularly as drugs.

[00217] 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 the SEQUENCE section 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.

[00218] 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.

[00219] 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.

[00220] 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.

[00221] 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 10mM ammonium formate, and the fractions containing the peptides were pooled and lyophilized.

[00222] 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.

[00223] 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 BY, Brault M, Strong RK, Rawlings DJ. Nucleic Acids Res. 2011 Nov;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 10mM 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.

[00224] Conjugation to palmitic acid, ICG, or biotinidase-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.

EXAMPLE 2

Generating Mass-Indexed Peptide Libraries and Peptides for Drug Development

[00225] This example describes a method for generating peptide libraries that contain mass-indexed peptides for the generation of knottins. This example also describes the analysis of mass-indexed peptides in a sample.

[00226] A mass-indexed library of 3,000 variants of the peptide Chayote Trypsin Inhibitor was created *in silico*, All lysine residues in the peptide were changed to arginine residues. The resulting sequence was: CPRILMRCRLDTDCFPTCTCRPSGFCG. G1 and S2 of this sequence were added to the molecule's N-terminus. The molecular structure of the scaffold was analyzed manually to identify portions of the sequence that may be altered without disrupting scaffold structure. The identified amino acid resides included M8, R9, L12, D13, T14, F17, P18, T19, T21, R23, P24, S25, G26. The identified amino acids were R23, P24, S25, and G26 and were located at positions 1 through 4 of a beta turn, respectively.

[00227] Python software was used to create a pool of five million sequences, such that each sequence was a unique variant. The sequences were created by altering the amino acids of each member in the pool such that each member contained a fully-tryptic peptide in the range of 7-30 amino acids in length. The fully-tryptic peptide was within the mass range 800-3500Da when reduced and alkylated with iodoacetamide.

[00228] The number of amino acid alterations from the starting amino acid sequence for each variant was chosen from a distribution of between 4 and 8. The distribution was weighted as follows: 4: 15%, 5: 50%, 6: 20%, 7: 10%, 8: 5%. The positions that were altered were chosen randomly from the list of possibly alterable positions. For each amino acid alteration, the amino acid to be substituted was selected from a set of amino acids. The set of amino acids was appropriate for the structure of the scaffold in a given position and weighted appropriately to favor amino acids that were more likely to favor proper folding. The weights for each amino acid in the different structural classifications were as follows:

Beta turn position 1: D: 15, G: 11, H: 13, N: 10, Q: 15, S: 13, T: 10, Y: 10

Beta turn position 2: D: 12, E: 14, G: 10, N: 10, Q: 10, S: 12, P: 35

Beta turn position 3: D: 18, G: 21, H: 12, N: 10, Q: 21, S: 10

Beta turn position 4: G: 16, N: 10, Q: 10, S: 10, T: 12, Y: 10

All other positions: A: 4, D: 8, E: 4, F: 7, G: 8, H: 8, I: 8, L: 2, M: 4, N: 4, Q: 4, R: 8, S: 7, T: 4, V: 4, W: 6, Y: 10

[00229] There were no variants with more than 3 tryptophan residues retained in the pool. Reverse translations of each member of the pool were chosen for DNA synthesis. During DNA synthesis, the coding of the n-terminal GS portion was changed to GGATCC. Further, the 5' sequence AACTGCCATGTGCAACTCGTAAG and the 3' sequence TAATGCGGCCGCGTTCTTAGTCACCTTGCATGGAC were added to each member of the pool. The reverse translations were chosen so as not to contain BamHI (GGATCC) or NotI (GCGGCCGC) restriction sites unless selected for specific members of the pool.

[00230] From the pool of possible variants, the members of the mass-indexed library were randomly chosen. The members were added to the library one at a time until the desired number of members was reached. Members were only added if addition of a member preserved the following criteria for the library:

- each member had at most 2 neighbors within 0.15 Daltons and 2.75 minutes of predicted liquid chromatography (LC) retention time in a 90-minute gradient;
- every member had at least one tryptic peptide of 7-25 amino acids that also had at most 2 neighbors within 2.1 Thompsons and 2.75 minutes of predicted LC retention time that is unique in sequence within the library.

Subsequently, 3,000 sense-strand oligonucleotides as well as 3,000 full reverse-complement oligonucleotides were ordered for library members.

[00231] Members of the libraries were screened for protein expression. Samples of the members of the library were prepared as protein mixtures. The samples of the protein mixture was reduced with 20 mM DTT and blocked with 25 mM iodoacetamide. The proteins were digested with trypsin at a trypsin to protein ratio: 1/50 overnight (or approximately16–18 hours) and purified using a C18 desalting column. One μg of purified peptide sample was injected into an Orbitrap Elite hybrid mass spectrometer (Thermo Fisher Scientific, Waltham, MA) that was coupled with nano-flow HPLC. The liquid chromatography/mass spectrometry apparatus and method consisted of a trap column (100μm × 1.5cm) made from an IntegraFrit (New Objective, Woburn, MA) packed with Magic C18AQ resin (5μm, 200Å particles; Michrom Bioresources, Auburn

CA), followed by an analytical column (75μm × 27cm) made from a PicoFrit (New Objective) packed with Magic C18AQ resin (5μm, 100Å particles; Michrom Bioresources). The columns were connected in-line to an Eksigent 2D nano-HPLC (Eksigent Technologies, Dublin, CA) in a vented column configuration to allow fast sample loading at 3uL/min.

[00232] The peptide samples were analyzed by LC-MS/MS using a 90-minute non-linear gradient as follows: start at 5% acetonitrile with 0.1% formic acid (against water with 0.1% formic acid), change to 7% over 2 minutes, then to 35% over 90 minutes, then to 50% over 1 minute, hold at 50% for 9 minutes, change to 95% over 1 minute, hold at 95% for 5 minutes, drop to 5% over 1 minute and recondition at 5%. The flow rate for the peptide separation was 300 nL/min.

[00233] A spray voltage of 2.25 kV was applied to the nanospray tip. The mass spectrometry experiment consisted of a full MS scan in the Orbitrap (AGC target value 1e6, resolution 60K, and one microscan, FT preview scan on) followed by up to 5 MS/MS spectra acquisitions in the linear ion trap. The five most intense ions from the Orbitrap scan were selected for MS/MS using collision-induced dissociation (isolation width 2 m/z, target value 1e4, collision energy 35%, max injection time 100 ms). Lower abundance peptide ions were interrogated using dynamic exclusion (repeat count 1, repeat duration 30 sec., exclusion list size 100, exclusion time 45 sec., exclusion mass width −0.55 m/z low to 1.55 m/z high). Charge state screen was used, allowing for MS/MS of any ions with identifiable charge states +2, +3, and +4 and higher.

[00234] For samples that contained the peptides of interest, the following data analysis methods were used. A FASTA database was constructed for database search, containing the entire UniProt human proteome (2012/12/19), the expected 3,000 proteins from the chayote trypsin inhibitor library, and a set of commonly observed contaminant proteins. Raw machine output files from all MS runs were converted to mzXML files and searched with X!Tandem (version 2011.12.01.1) configured with the k-score scoring algorithm, against this FASTA database. The search parameters were as follows: enzyme, trypsin; maximum missed cleavages, 2; fixed modification, carboxamidomethylation on cysteine; potential modification, oxidization on methionine; parent monositopic mass error, -1.5Da to 2.5Da.

[00235] Peptide identifications were assigned probability by PeptideProphet (Trans-Proteomic Pipeline version 4.6), and all identifications assigned probability <0.95 were discarded. Variants present in the sample were inferred based on the peptide identifications of the tryptic peptides unique to those variants. For mouse plasma and tissue samples, the methods above were

followed, using a FASTA database containing the sequences described above in addition to the entire UniProt mouse proteome (2012/12/19).

EXAMPLE 3

Generating Mass-Indexed Peptide Libraries Using Conjugated Peptides

[00236] This example describes a method for generating peptide libraries that contain conjugated mass-indexed peptides that can be used for the generation of knottins. This example further describes the analysis of mass-indexed conjugated peptides in a sample.

[00237] Peptides were conjugated to biotin and retrieved from a peptide library following proteolysis with trypsin. A library of 1000 Chayote Trypsin Inhibitor variants at 10mg/ml was allowed to react with a 3x molar excess of N-hydroxy succinimide (NHS) ester of biotin in phosphate buffered saline. The conjugation proceeded strictly at the N-terminus of the peptides because the library was engineered to contain no lysine residues.

[00238] Excess biotin and NHS were removed by ultrafiltration, and the resultant library administered to mice with xenografted human tumors at 100mg/kg via tail vein injection. Mice were sacrificed at 5, 10, 30, and 60 minutes and the tumor, brain, blood, and kidneys were collected and mechanically homogenized. Cells and clotted material were removed by centrifugation and the resultant supernatant was incubated with avidin immobilized on beads for 30 minutes at room temperature with shaking. The beads were then washed extensively with PBS and boiled to release the biotinylated peptides. These peptides were then subjected to reduction, alkylation (of cysteine residues), trypsinization, and LC/MS/MS analysis to determine the identities of the peptides in the various tissues.

[00239] In a variation of this method, trypsin was added prior to the mechanical homogenization to break up protein-protein interactions with knottin peptides, which may have formed during sample trypsinization. Recovery of the peptides was the same, with avidin beads, except only the N-terminus of the peptide was recovered.

[00240] In addition to conjugation with biotin, additional peptide libraries were generated by isotopic labeling. Both types of libraries were quantified via mass spectrometry. A library of 1000 Chayote Trypsin Inhibitor variants at 10mg/ml was split into two halves. One half reacted with a 3x molar excess of N-hydroxy succinimide (NHS) ester of isotopically normal palmitic acid in phosphate buffered saline. The other half was treated in the same manner with palmitic acid containing at four C13 (isotopically heavy) carbons. DMSO or DMF was added sufficient to keep the palmitic acid in solution. Excess palmitic acid was removed by extraction with ethyl acetate after an overnight (12-18 hour) incubation.

[00241] The isotopically heavy library is chemically identical to the isotopically normal library but was differentiated by mass. The heavy library was administered to animals by IV dosing at 100mg/kg and the isotopically normal library administered in the same animals by oral lavage. The animals were sacrificed 20 minutes after library administration. The blood was collected and allowed to clot. Cells and clotted material were removed by centrifugation and the resultant plasma was reduced with 5mM TCEP, alkylated with iodoacetamide, and applied to a C18 solid phase extraction column. The column was then rinsed with increasing percentages of acetonitrile in 0.5% TFA. The N-terminally palmitylated peptides elute at 80% or higher acetonitrile. This eluted material was then subjected to trypsinization and LC/MS/MS analysis, and the isotopically normal and heavy peaks were compared directly. Normal peaks were observed and corresponded to peptides with oral bioavailability.

[00242] In a variation of this protocol, the plasma was treated with trypsin after iodoacetamide treatment and prior to applying it to the C18 column. With this method, only the N-terminal tryptic peptide of each library member was recovered and distinguishable based on unique properties.

[00243] A library composed of 28 knottin peptides was screened for those that bind with high affinity to the human derived rhabdomyosarcoma (e.g., cancer) cell line, A-204. Twenty-four hours prior to exposure, the cells were plated on Poly-L-Lysine coated 6-well plates (Corning) and grown in RPMI-1640 media (Life Technology) supplemented with 10% FBS (Hyclone) to reach ~85% of confluency. The adherent cells were washed 3x with 3ml D-PBS (Life Technology) and incubated with the knottin library (30μg/ml) in D-PBS for 1hr at 37C. The unbound peptides were aspirated, and cells were washed 3x with D-PBS in which the last wash was recovered for analysis. To elute the bound knottins, the cells were incubated in 1ml 280mM NaCl in D-PBS for 5min at 37C, followed by 560mM and 1.12M NaCl. Following the final elution, the cells were collected in HyPure Molecular Biology Grade Water (Hyclone).
The cells were lysed with three freeze thaw cycles, and 5min in a bath sonicator. The protein concentration of the lysate was quantified based on the absorbance at 280nm using a Nanodrop 1000 and the sample was diluted to 100μg/ml total protein in water. The lysate and elution samples were desalted using a Strata-X desalting column (Phenomenex). The recovered peptides were then reduced, alkylated and trypsinized before being submitted for LC/MS-MS analysis. Individual knottins were identified from each sample by their unique tryptic peptide signature.

EXAMPLE 4

Identifying Knottins with Serum Stability, Half Life and Tissue Homing Capabilities

[00244] This example describes a method for identifying knottin peptides that contain mass-indexed peptides for the generation of knottins. This example further describes the analysis of mass-indexed peptides in a sample.

[00245] Six mice were given a library of mass-indexed knottins via tail-vein injection. Three mice were sacrificed at 5 minutes, three at 30 minutes. The blood was collected by cardiac puncture and the organs (brain, liver, kidneys) dissected for analysis. Tissue was flash-frozen in liquid nitrogen. Tissue was subsequently thawed, homogenized, and subjected to analysis by mass spectrometry to determine tissue distribution and half-life estimation.

EXAMPLE 5

Parallel Assessment of Chymotrypsin Inhibitor Mutational Tolerance

[00246] This example describes the assessment of mutational tolerance of chymotrypsin inhibitor (CTI) according to an aspect of the present disclosure.

[00247] Chip-based oligonucleotide synthesis was performed in order to generate a set of 3000 mutants of CTI. The mutants were amplified by PCR and cloned by restriction digestion and ligation into the lentiviral system.

[00248] FIG. 12 depicts the assessment of mutational tolerance of CTI by reverse proteomics 3000 mutants of CTI with 4-8 mutations, each with a unique MS signature, were generated. These viruses were used to inflect HEK293 cells as a pool, and the resulting peptides were isolated, digested with trypsin, treated with iodoacetamide, and applied to an LC/MS/MS Orbitrap Elite system, and 1595 proteins were observed. The over- or underrepresentation of amino acids at each position, in the expressed and observed proteins versus the full set of proteins ordered, is shown. Amino acids below the line at the bottom were not observed at all in the set of produced and observed proteins. Globally, arginine and tryptophan amino acids were less likely to be present in the produced and observed proteins than in the set of ordered proteins, and glutamic acid was more likely to be present.

[00249] 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.

SEQUENCES

[00250] The following are DNA and/or amino acid sequences of genes of interest and constructs identified herein.

Construction of parental construct for seamless cloning:
SEQ ID NO:1 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-PARENTAL *for Xho/Bam cut from pUC57 and ligation into pCVL*
GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGCGGCCGC**TAAGGATCCCGGACCGCCTCTCC

NotI cut is AACCTGTATTTTCAGGGA**GGC - GC**TAAGGATCCCGGACCGCCTCTCC
Fusion protein Sequences – original set of 10 – cloned into NotI cut parent above by seamless cloning:

SEQ ID NO:2 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-BubbleProtein
GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGC**<u>**GATACCTGCGGCAGCGGCTATAATGTGGATCAGCGTCGTACCA**
ATAGCGGCTGCAAAGCGGGCAATGGCGATCGTCATTTTTGCGGCTGCGATCGT
ACCGGCGTGGTGGAATGCAAAGGCGGCAAATGGACCGAAGTGCAGGATTGCG
GCAGCAGCAGCTGCAAAGGCACCAGCAATGGCGGCGCGACCTGC</u>**TAATGCTAA**
GGATCCCGGA

SEQ ID NO:3 -
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt

M E T D T L L L W V L L L W V P G S T G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
D Y K D E H H H H H H G G S Q D S T S D
ctgatcccagccccacctctgagcaaggtccctctgcagcagaacttccaggacaaccaa
L I P A P P L S K V P L Q Q N F Q D N Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
F Q G K W Y V V G L A G N A I L R E D K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
D P Q K M Y A T I Y E L K E D K S Y N V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
T S V L F R K K K C D Y W I R T F V P G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
C Q P G E F T L G N I K S Y P G L T S Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
L V R V V S T N Y N Q H A M V F F K K V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
S Q N R E Y F K I T L Y G R T K E L T S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
E L K E N F I R F S K S L G L P E N H I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
V F P V P I D Q C I D G G G S E N L Y F
cagggaggcgatacctgcggcagcggctataatgtggatcagcgtcgtaccaatagcggc
Q G G D T C G S G Y N V D Q R R T N S G
tgcaaagcgggcaatggcgatcgtcattttgcggctgcgatcgtaccggcgtggtggaa
C K A G N G D R H F C G C D R T G V V E
tgcaaaggcggcaaatggaccgaagtgcaggattgcggcagcagcagctgcaaaggcacc
C K G G K W T E V Q D C G S S S C K G T
agcaatggcggcgcgacctgc
S N G G A T C

SEQ ID NO:4 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-Attractin
GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGCGATCAGAATTGCGATATTGGCAATATTACCAGCCAGTGCCAGA**
TGCAGCATAAAAATTGCGAAGATGCGAATGGCTGCGATACCATTATTGAAGAAT
GCAAAACCAGCATGGTGGAACGTTGCCAGAATCAGGAATTTGAAAGCGCGGCG
GGCAGCACCACCCTGGGCCCGCAGTAATGCTAAGGATCCCGGA

SEQ ID NO 5 -
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt

M E T D T L L L W V L L L W V P G S T G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
D Y K D E H H H H H H G G S Q D S T S D
ctgatcccagccccacctctgagcaaggtcctctgcagcagaacttccaggacaaccaa
L I P A P P L S K V P L Q Q N F Q D N Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
F Q G K W Y V V G L A G N A I L R E D K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
D P Q K M Y A T I Y E L K E D K S Y N V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
T S V L F R K K K C D Y W I R T F V P G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
C Q P G E F T L G N I K S Y P G L T S Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
L V R V V S T N Y N Q H A M V F F K K V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
S Q N R E Y F K I T L Y G R T K E L T S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
E L K E N F I R F S K S L G L P E N H I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
V F P V P I D Q C I D G G G S E N L Y F
cagggaggcgatcagaattgcgatattggcaatattaccagccagtgccagatgcagcat
Q G G D Q N C D I G N I T S Q C Q M Q H
aaaaattgcgaagatgcgaatggctgcgataccattattgaagaatgcaaaaccagcatg
K N C E D A N G C D T I I E E C K T S M
gtggaacgttgccagaatcaggaatttgaaagcgcggcgggcagcaccaccctgggcccg
V E R C Q N Q E F E S A A G S T T L G P
cag
Q

SEQ ID NO:6 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-Hefutoxin
GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGCGGCCATGCGTGCTATCGTAATTGCTGGCGTGAAGGCAATGATG**
AAGAAACCTGCAAAGAACGTTGCTAATGCTAAGGATCCCGGA*CCGCC*

SEQ ID NO:7 -
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
M E T D T L L L W V L L L W V P G S T G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac

D Y K D E H H H H H H G G S Q D S T S D
ctgatcccagccccacctctgagcaaggtccctctgcagcagaacttccaggacaaccaa
L I P A P P L S K V P L Q Q N F Q D N Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
F Q G K W Y V V G L A G N A I L R E D K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
D P Q K M Y A T I Y E L K E D K S Y N V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
T S V L F R K K K C D Y W I R T F V P G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
C Q P G E F T L G N I K S Y P G L T S Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
L V R V V S T N Y N Q H A M V F F K K V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
S Q N R E Y F K I T L Y G R T K E L T S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
E L K E N F I R F S K S L G L P E N H I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
V F P V P I D Q C I D G G G S E N L Y F
cagggaggcggccatgcgtgctatcgtaattgctggcgtgaaggcaatgatgaagaaacc
Q G G G H A C Y R N C W R E G N D E E T
tgcaaagaacgttgc
C K E R C

SEQ ID NO:8 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-Hanatoxin
GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGCGAATGCCGTTATCTGTTTGGCGGCTGCAAAACCACCAGCGATT**
GCTGCAAACATCTGGGCTGCAAATTTCGTGATAAATATTGCGCGTGGGATTTTA
CCTTTAGCTAATGCTAAGGATCCCGGA

SEQ ID NO:9 -
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
M E T D T L L L W V L L L W V P G S T G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
D Y K D E H H H H H H G G S Q D S T S D
ctgatcccagccccacctctgagcaaggtccctctgcagcagaacttccaggacaaccaa
L I P A P P L S K V P L Q Q N F Q D N Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
F Q G K W Y V V G L A G N A I L R E D K

gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
D P Q K M Y A T I Y E L K E D K S Y N V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
T S V L F R K K K C D Y W I R T F V P G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
C Q P G E F T L G N I K S Y P G L T S Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
L V R V V S T N Y N Q H A M V F F K K V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
S Q N R E Y F K I T L Y G R T K E L T S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
E L K E N F I R F S K S L G L P E N H I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
V F P V P I D Q C I D G G G S E N L Y F
cagggaggcgaatgccgttatctgtttggcggctgcaaaaccaccagcgattgctgcaaa
Q G G E C R Y L F G G C K T T S D C C K
catctgggctgcaaatttcgtgataaatattgcgcgtgggatttacctttagc
H L G C K F R D K Y C A W D F T F S

SEQ ID NO:10 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-ChymotrypsinInhibitor
GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGC**<u>**GAAATTAGCTGCGAACCGGGCAAAACCTTTAAAGATAAATGCA**
ATACCTGCCGTTGCGGCGCGGATGGCAAAAGCGCGGCGTGCACCCTGAAAGCG
TGCCCGAATCAG</u>**TAATGCTAAGGAT**CCCGGA

SEQ ID NO:11 -
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
M E T D T L L L W V L L L W V P G S T G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
D Y K D E H H H H H H G G S Q D S T S D
ctgatcccagccccacctctgagcaaggtcctctgcagcagaacttccaggacaaccaa
L I P A P P L S K V P L Q Q N F Q D N Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
F Q G K W Y V V G L A G N A I L R E D K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
D P Q K M Y A T I Y E L K E D K S Y N V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
T S V L F R K K K C D Y W I R T F V P G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac

C Q P G E F T L G N I K S Y P G L T S Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
L V R V V S T N Y N Q H A M V F F K K V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
S Q N R E Y F K I T L Y G R T K E L T S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
E L K E N F I R F S K S L G L P E N H I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
V F P V P I D Q C I D G G G S E N L Y F
cagggaggcgaaattagctgcgaaccgggcaaaacctttaaagataaatgcaatacctgc
Q G G E I S C E P G K T F K D K C N T C
cgttgcggcgcggatggcaaaagcgcggcgtgcaccctgaaagcgtgcccgaatcag
R C G A D G K S A A C T L K A C P N Q

SEQ ID NO:12 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-ToxinK

GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGC**<u>**GTGTGCCGTGATTGGTTTAAAGAAACCGCGTGCCGTCATGCGA**
AAAGCCTGGGCAATTGCCGTACCAGCCAGAAATATCGTGCGAATTGCGCGAAA
ACCTGCGAACTGTGC</u>**TAATGCTAAGGATCCCGGA**

SEQ ID NO:13 -
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
M E T D T L L L W V L L L W V P G S T G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
D Y K D E H H H H H H G G S Q D S T S D
ctgatcccagccccacctctgagcaaggtccctctgcagcagaacttccaggacaaccaa
L I P A P P L S K V P L Q Q N F Q D N Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
F Q G K W Y V V G L A G N A I L R E D K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
D P Q K M Y A T I Y E L K E D K S Y N V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
T S V L F R K K K C D Y W I R T F V P G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
C Q P G E F T L G N I K S Y P G L T S Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
L V R V V S T N Y N Q H A M V F F K K V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg

S Q N R E Y F K I T L Y G R T K E L T S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
E L K E N F I R F S K S L G L P E N H I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
V F P V P I D Q C I D G G G S E N L Y F
cagggaggcgtgtgccgtgattggtttaaagaaaccgcgtgccgtcatgcgaaaagcctg
Q G G V C R D W F K E T A C R H A K S L
ggcaattgccgtaccagccagaaatatcgtgcgaattgcgcgaaaacctgcgaactgtgc
G N C R T S Q K Y R A N C A K T C E L C

SEQ ID NO:14 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-EGFepiregulinCore
GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGCGTGAGCATTACCAAATGCAGCAGCGATATGAATGGCTATTGCC**
TGCATGGCCAGTGCATTTATCTGGTGGATATGAGCCAGAATTATTGCCGTTGCG
AAGTGGGCTATACCGGCGTGCGTTGCGAACATTTTTTTCTGTAATGCTAAGGAT
CCCGGA

SEQ ID NO:15 -
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
M E T D T L L L W V L L L W V P G S T G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
D Y K D E H H H H H H G G S Q D S T S D
ctgatcccagccccacctctgagcaaggtccctctgcagcagaacttccaggacaaccaa
L I P A P P L S K V P L Q Q N F Q D N Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
F Q G K W Y V V G L A G N A I L R E D K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
D P Q K M Y A T I Y E L K E D K S Y N V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
T S V L F R K K K C D Y W I R T F V P G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
C Q P G E F T L G N I K S Y P G L T S Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
L V R V V S T N Y N Q H A M V F F K K V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
S Q N R E Y F K I T L Y G R T K E L T S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
E L K E N F I R F S K S L G L P E N H I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt

V F P V P I D Q C I D G G G S E N L Y F
cagggaggcgtgagcattaccaaatgcagcagcgatatgaatggctattgcctgcatggc
Q G G V S I T K C S S D M N G Y C L H G
cagtgcatttatctggtggatatgagccagaattattgccgttgcgaagtgggctatacc
Q C I Y L V D M S Q N Y C R C E V G Y T
ggcgtgcgttgcgaacatttttttctg
G V R C E H F F L

SEQ ID NO:16 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-Circulin
GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGC**<u>**GGCATTCCGTGCGGCGAAAGCTGCGTGTGGATTCCGTGCATTA**
GCGCGGCGCTGGGCTGCAGCTGCAAAAATAAGTGTGCTATCGTAATTAATGC</u>
TAAGGATCCCGGA

SEQ ID NO:17 -
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
M E T D T L L L W V L L L W V P G S T G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
D Y K D E H H H H H H G G S Q D S T S D
ctgatcccagccccacctctgagcaaggtccctctgcagcagaacttccaggacaaccaa
L I P A P P L S K V P L Q Q N F Q D N Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
F Q G K W Y V V G L A G N A I L R E D K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
D P Q K M Y A T I Y E L K E D K S Y N V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
T S V L F R K K K C D Y W I R T F V P G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
C Q P G E F T L G N I K S Y P G L T S Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
L V R V V S T N Y N Q H A M V F F K K V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
S Q N R E Y F K I T L Y G R T K E L T S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
E L K E N F I R F S K S L G L P E N H I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
V F P V P I D Q C I D G G G S E N L Y F
cagggaggcggcattccgtgcggcgaaagctgcgtgtggattccgtgcattagcgcggcg
Q G G G I P C G E S C V W I P C I S A A

ctgggctgcagctgcaaaaataaagtgtgctatcgtaat
L G C S C K N K V C Y R N

SEQ ID NO:18 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-Brazzein
GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGC**<u>**CAGGATAAATGCAAAAAAGTGTATGAAAATTATCCGGTGAGCA**
AATGCCAGCTGGCGAATCAGTGCAATTATGATTGCAAACTGGATAAACATGCGC
GTAGCGGCGAATGCTTTTATGATGAAAAACGTAATCTGCAGTGCATTTGCGATT
ATTGCGAATATT</u>**AATGCTAAGGAT**CCCGGA

SEQ ID NO:19 -
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
M E T D T L L L W V L L L W V P G S T G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
D Y K D E H H H H H H G G S Q D S T S D
ctgatcccagccccacctctgagcaaggtccctctgcagcagaacttccaggacaaccaa
L I P A P P L S K V P L Q Q N F Q D N Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
F Q G K W Y V V G L A G N A I L R E D K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
D P Q K M Y A T I Y E L K E D K S Y N V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
T S V L F R K K K C D Y W I R T F V P G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
C Q P G E F T L G N I K S Y P G L T S Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
L V R V V S T N Y N Q H A M V F F K K V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
S Q N R E Y F K I T L Y G R T K E L T S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
E L K E N F I R F S K S L G L P E N H I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
V F P V P I D Q C I D G G G S E N L Y F
cagggaggccaggataaatgcaaaaaagtgtatgaaaattatccggtgagcaaatgccag
Q G G Q D K C K K V Y E N Y P V S K C Q
ctggcgaatcagtgcaattatgattgcaaactggataaacatgcgcgtagcggcgaatgc
L A N Q C N Y D C K L D K H A R S G E C
ttttatgatgaaaaacgtaatctgcagtgcatttgcgattattgcgaatat
F Y D E K R N L Q C I C D Y C E Y

SEQ ID NO:20 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-Chlorotoxin

GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGCATGTGCATGCCGTGCTTTACCACCGATCATCAGATGGCGCGTA**
AATGCGATGATTGCTGCGGCGGCAAAGGCCGTGGCAAATGCTATGGCCCGCAG
TGCCTGTGCCGTTAATGCTAAGGATCCCGGA

SEQ ID NO:21 -

```
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
 M  E  T  D  T  L  L  L  W  V  L  L  L  W  V  P  G  S  T  G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
 D  Y  K  D  E  H  H  H  H  H  H  G  G  S  Q  D  S  T  S  D
ctgatcccagccccacctctgagcaaggtcctctgcagcagaacttccaggacaaccaa
 L  I  P  A  P  P  L  S  K  V  P  L  Q  Q  N  F  Q  D  N  Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
 F  Q  G  K  W  Y  V  V  G  L  A  G  N  A  I  L  R  E  D  K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
 D  P  Q  K  M  Y  A  T  I  Y  E  L  K  E  D  K  S  Y  N  V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
 T  S  V  L  F  R  K  K  K  C  D  Y  W  I  R  T  F  V  P  G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
 C  Q  P  G  E  F  T  L  G  N  I  K  S  Y  P  G  L  T  S  Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
 L  V  R  V  V  S  T  N  Y  N  Q  H  A  M  V  F  F  K  K  V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
 S  Q  N  R  E  Y  F  K  I  T  L  Y  G  R  T  K  E  L  T  S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
 E  L  K  E  N  F  I  R  F  S  K  S  L  G  L  P  E  N  H  I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
 V  F  P  V  P  I  D  Q  C  I  D  G  G  G  S  E  N  L  Y  F
cagggaggcatgtgcatgccgtgctttaccaccgatcatcagatggcgcgtaaatgcgat
 Q  G  G  M  C  M  P  C  F  T  T  D  H  Q  M  A  R  K  C  D
gattgctgcggcggcaaaggccgtggcaaatgctatggcccgcagtgcctgtgccgt
 D  C  C  G  G  K  G  R  G  K  C  Y  G  P  Q  C  L  C  R
```

Construction of parental construct for BamH1/NotI cloning:

SEQ ID NO:22 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GS-PARENTAL

GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAG**GGATCC**TAATGTTGGCCATGATGTTAG**GCGGCCGC**TAAGGATCCCGGA

BamHI site: GGATCC
NotI site: GCGGCCGC

A BamHI site adds "GS" before a knottin. This construct can be used for cloning libraries.

Construction of parental construct for furin cleavage, BamHI/NotI cloning can include an idealized furin cut site is RARYKRS –RARYKRGS can be used for a Bam HI site.

SEQ ID NO:23 - IgK-SF-H6-GGS-lcn2C-GGS-furin-GS-PARENTAL
GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCcgcgcgcgctataaacgc**G**
GATCCTAATGTTGGCCATGATGTTAG**GCGGCCGC**TAAGGATCCCGGA

SEQ ID NO:24 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GS-MIDKINE
GACTGAGTCGCCCG*CTCGAG*ACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT

TTCAGAGCGATTGCAAATATAAATTTGAAAACTGGGGCGCGTGCGATGGCGGCACC
GGCACCAAAGTGCGCCAGGGCACCCTGAAAAAAGCGCGCTATAACGCGCAGTGCCA
GGAAACCATTCGCGTGACCAAACCGTGCTAATGCT*GGATCC*CGGACCGCCTCTCC

SEQ ID NO:25 -

```
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
 M  E  T  D  T  L  L  L  W  V  L  L  L  W  V  P  G  S  T  G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcaggac
 D  Y  K  D  E  H  H  H  H  H  H  G  G  S  Q  D  S  T  S  D
ctgatcccagccccacctctgagcaaggtcctctgcagcagaacttccaggacaaccaa
 L  I  P  A  P  P  L  S  K  V  P  L  Q  Q  N  F  Q  D  N  Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
 F  Q  G  K  W  Y  V  V  G  L  A  G  N  A  I  L  R  E  D  K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
 D  P  Q  K  M  Y  A  T  I  Y  E  L  K  E  D  K  S  Y  N  V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggactttgttccaggt
 T  S  V  L  F  R  K  K  K  C  D  Y  W  I  R  T  F  V  P  G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
 C  Q  P  G  E  F  T  L  G  N  I  K  S  Y  P  G  L  T  S  Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
 L  V  R  V  V  S  T  N  Y  N  Q  H  A  M  V  F  F  K  K  V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
 S  Q  N  R  E  Y  F  K  I  T  L  Y  G  R  T  K  E  L  T  S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
 E  L  K  E  N  F  I  R  F  S  K  S  L  G  L  P  E  N  H  I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
 V  F  P  V  P  I  D  Q  C  I  D  G  G  G  S  E  N  L  Y  F
cagagcgattgcaaatataaatttgaaaactggggcgcgtgcgatggcggcaccggcacc
 Q  S  D  C  K  Y  K  F  E  N  W  G  A  C  D  G  G  T  G  T
aaagtgcgccagggcaccctgaaaaaagcgcgctataacgcgcagtgccaggaaaccatt
 K  V  R  Q  G  T  L  K  K  A  R  Y  N  A  Q  C  Q  E  T  I
cgcgtgaccaaaccgtgc
 R  V  T  K  P  C
```

SEQ ID NO:26 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-Violacin A

GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGAGGCAGCGCCATCAGCTGCGGCGAGACCTGCTTCAAGTTCAAGTGCTAC
ACCCCCAGATGCAGCTGCAGCTACCCCGTGTGCAAGTAAGCTAAGGATCCCGGACC
GCC

SEQ ID NO:27 -

atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
M E T D T L L L W V L L L W V P G S T G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
D Y K D E H H H H H H G G S Q D S T S D
ctgatcccagccccacctctgagcaaggtccctctgcagcagaacttccaggacaaccaa
L I P A P P L S K V P L Q Q N F Q D N Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
F Q G K W Y V V G L A G N A I L R E D K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
D P Q K M Y A T I Y E L K E D K S Y N V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
T S V L F R K K K C D Y W I R T F V P G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
C Q P G E F T L G N I K S Y P G L T S Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
L V R V V S T N Y N Q H A M V F F K K V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
S Q N R E Y F K I T L Y G R T K E L T S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
E L K E N F I R F S K S L G L P E N H I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
V F P V P I D Q C I D G G G S E N L Y F
cagggaggcagcgccatcagctgcggcgagacctgcttcaagttcaagtgctacaccccc
Q G G G S A I S C G E T C F K F K C Y T P
agatgcagctgcagctaccccgtgtgcaag
R C S C S Y P V C K

SEQ ID NO:28 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-Lambda Toxin

GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGAGGCGTGTGCTGCGGCTACAAGCTGTGCCACCCCTGCTAAGCTAAGGATC
CCGGACC

SEQ ID NO:29 -

atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
M E T D T L L L W V L L L W V P G S T G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
D Y K D E H H H H H H G G S Q D S T S D

ctgatcccagccccacctctgagcaaggtcctctgcagcagaacttccaggacaaccaa
L I P A P P L S K V P L Q Q N F Q D N Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
F Q G K W Y V V G L A G N A I L R E D K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
D P Q K M Y A T I Y E L K E D K S Y N V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
T S V L F R K K K C D Y W I R T F V P G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
C Q P G E F T L G N I K S Y P G L T S Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
L V R V V S T N Y N Q H A M V F F K K V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
S Q N R E Y F K I T L Y G R T K E L T S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
E L K E N F I R F S K S L G L P E N H I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
V F P V P I D Q C I D G G G S E N L Y F
cagggaggcgtgtgctgcggctacaagctgtgccacccctgc
Q G G V C C G Y K L C H P C

SEQ ID NO:30 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-Lambda Toxin NG
GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGC**AACGGCGTGTGCTGCGGCTACAAGCTGTGCCACCCCTGCTAAGCT
AAGGATCCCGGACC

SEQ ID NO:31 -
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
M E T D T L L L W V L L L W V P G S T G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
D Y K D E H H H H H H H G G S Q D S T S D
ctgatcccagccccacctctgagcaaggtcctctgcagcagaacttccaggacaaccaa
L I P A P P L S K V P L Q Q N F Q D N Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
F Q G K W Y V V G L A G N A I L R E D K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
D P Q K M Y A T I Y E L K E D K S Y N V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
T S V L F R K K K C D Y W I R T F V P G

tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
C Q P G E F T L G N I K S Y P G L T S Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
L V R V V S T N Y N Q H A M V F F K K V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
S Q N R E Y F K I T L Y G R T K E L T S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
E L K E N F I R F S K S L G L P E N H I
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
V F P V P I D Q C I D G G G S E N L Y F
cagggaggcaacggcgtgtgctgcggctacaagctgtgccacccctgc
Q G G N G V C C G Y K L C H P C

SEQ ID NO:32 - IgK-SF-H6-GGS-lcn2C-GGS-ENLYFQ-GG-Potato Carboxypeptidase Inhibitor
GACTGAGTCGCCCGCTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGC
TGCTCTGGGTTCCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACC
ATGGTGGAAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAG
GTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGC
CACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAG
GAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGGCGA
GTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTCCGAGT
GGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTTTCTCAAAA
CAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCTGACTTCGGAAC
TAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCCTGAAAACCACATCG
TCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGGTAGCGAAAACCTGTATT
TTCAGGGA**GGC**cagcagcatgcggatccgatttgcaacaaaccgtgcaaaacccatgatgattgcagcggcgcgtggttttgcc
aggcgtgctggaacagcgcgcgcacctgcggcccgtatgtgggcTAATGCTAAGGATCCCGGACCG

SEQ ID NO:33 -
atggagacagacacactcctgctatgggtactgctgctctgggttccaggttccactggt
M E T D T L L L W V L L L W V P G S T G
gactacaaggacgagcatcaccatcatcaccatggtggaagccaggactccacctcagac
D Y K D E H H H H H H G G S Q D S T S D
ctgatcccagccccacctctgagcaaggtccctctgcagcagaacttccaggacaaccaa
L I P A P P L S K V P L Q Q N F Q D N Q
ttccaggggaagtggtatgtggtaggcctggcagggaatgcaattctcagagaagacaaa
F Q G K W Y V V G L A G N A I L R E D K
gacccgcaaaagatgtatgccaccatctatgagctgaaagaagacaagagctacaatgtc
D P Q K M Y A T I Y E L K E D K S Y N V
acctccgtcctgtttaggaaaaagaagtgtgactactggatcaggacttttgttccaggt
T S V L F R K K K C D Y W I R T F V P G
tgccagcccggcgagttcacgctgggcaacattaagagttaccctggattaacgagttac
C Q P G E F T L G N I K S Y P G L T S Y
ctcgtccgagtggtgagcaccaactacaaccagcatgctatggtgttcttcaagaaagtt
L V R V V S T N Y N Q H A M V F F K K V
tctcaaaacagggagtacttcaagatcaccctctacgggagaaccaaggagctgacttcg
S Q N R E Y F K I T L Y G R T K E L T S
gaactaaaggagaacttcatccgcttctccaaatctctgggcctccctgaaaaccacatc
E L K E N F I R F S K S L G L P E N H I

```
gtcttccctgtcccaatcgaccagtgtatcgacggcggaggtagcgaaaacctgtatttt
 V  F  P  V  P  I  D  Q  C  I  D  G  G  G  S  E  N  L  Y  F
cagggaggccagcagcatgcggatccgatttgcaacaaaccgtgcaaaacccatgatgat
 Q  G  G  Q  Q  H  A  D  P  I  C  N  K  P  C  K  T  H  D  D
tgcagcggcgcgtggttttgccaggcgtgctggaacagcgcgcgcacctgcggcccgtat
 C  S  G  A  W  F  C  Q  A  C  W  N  S  A  R  T  C  G  P  Y
gtgggctaa
 V  G
```

CLAIMS

WHAT IS CLAIMED IS:

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

analyzing an isolated sample from a subject following administration of a plurality of drug candidates to the subject; and

identifying in the isolated sample at least one drug candidate having the pharmacological property.

2. 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.

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

4. The method of claim 1 or 2, wherein the plurality of drug candidates comprises a plurality of peptides.

5. The method of claim 1 or 2, wherein the plurality of drug candidates comprises a plurality of small chemical molecules.

6. The method of claim 1 or 2, wherein the plurality of drug candidates comprises a plurality of biologics.

7. The method of any one of the preceding claims, 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.

8. The method of claim 6, further comprising generating a peptide library comprising the at least some of the peptides having the pharmacological property.

9. The method of any one of the preceding claims, 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.

10. The method of claim 9, 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.

11. The method of claim 9 or 10, wherein the protein:protein interaction is associated with cancer.

12. The method of any one of the preceding claims, 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.

13. The method of claim 12, further comprising 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.

14. The method of any one of the preceding claims, 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.

15. The method of any one of the preceding claims, wherein the plurality of drug candidates comprises greater than 5 drug candidates, greater than 10 drug candidates, greater than 100 drug candidates, greater than 1000 drug candidates, or greater than 10000 drug candidates.

16. The method of any one of the preceding claims, wherein at least one peptide of the plurality of peptides comprises a detectable label for tracing the at least one peptide in the subject.

17. The method of claim 16, wherein the detectable label comprises a near infrared dye.

18. The method of any one of the preceding claims, 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.

19. The method of claim 18, wherein the hydrophobic moiety comprises a hydrophobic fluorescent dye or a saturated or unsaturated alkyl group.

20. The method of any one of the preceding claims, wherein the sample comprises a tissue sample or a fluid sample.

21. The method of claim 20, wherein the fluid sample comprises blood, urine, mucous or spinal fluid.

22. The method of claim 21, wherein the sample comprises blood.

23. The method of claim 20, wherein the tissue sample comprises biopsy or necropsy tissue.

24. The method of claim 20 or 23, 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.

25. The method of any one of the preceding claims, further comprising using the pharmacological property common to the drug candidates to generate additional drug candidates having the pharmacological property.

26. The method of any one of the preceding claims, wherein the subject is an animal.

27. The method of claim 26, wherein the subject is a human.

28. The method of any one of the preceding claims, further comprising establishing a relationship between the structure and the pharmacological property of the drug candidates.

29. The method of claim 28, further comprising selecting a subset of drug candidates based on the relationship.

30. The method of any one of the preceding claims, further comprising obtaining a plurality of samples at a selected time interval for a selected duration of time following administration.

31. The method of claim 30, wherein selected time intervals are 5 minutes, 10 minutes, 20 minutes, 30 minutes, 40 minutes, 50 minutes, or 60 minutes.

32. The method of claim 30, wherein selected duration of time is 1 hour, 2 hours, 3 hours, 4 hours, 5 hours, 6 hours, 7 hours, 8 hours, 9 hours, 10 hours, 11 hours, 12 hours, 15 hours, 30 hours, 45 hours, 60 hours, 75 hours, 90 hours, 105 hours, or 120 hours.

33. A method of generating a mass-defined drug candidate library, the method comprising:

producing a plurality of drug candidates, at least some of the drug candidates each having a unique mass signature or digest fragment mass signature;

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; and

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.

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

analyzing an isolated sample from a subject following administration of a plurality of drug candidates to the subject, wherein the library drug candidates are from the mass-defined drug candidate library of claim 33;

identifying in the isolated sample at least one drug candidate having the pharmacological property.

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

administering to a subject a plurality of library drug candidates, wherein the library drug candidates are from the mass-defined drug candidate library of claim 26;

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

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

36. The method of any one of claims 33 to 35, 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, or a combination thereof.

37. The method of claim 36, wherein the pharmacological property comprises oral bioavailability.

38. The method of any one of claims 33 to 37, wherein the plurality of drug candidates comprises greater than five drug candidates, greater than 10 drug candidates, greater than 100 drug candidates, greater than 1000 drug candidates, or greater than 10000 drug candidates.

39. The method of any one of claims 33 to 35, wherein the plurality of drug candidates comprises knotted-peptides.

40. The method of claim 39, wherein the unique mass signature or digest fragment mass signature of the knotted-peptides are defined by the natural amino acid sequence of the knotted peptides.

41. The method of any one of claims 33 to 35, wherein at least one of the drug candidates in the plurality comprises a pre-defined number of a heavy isotope atom to modify the unique mass signature or digest fragment mass signature of the at least one drug candidate.

42. The method of claim 41, wherein the heavy isotope atom comprises ^{13}C or deuterium.

43. The method of claim 39 or 40, wherein the unique mass signature or digest fragment mass signature of at least one of the knotted-peptides in the plurality is defined by a moiety conjugated to the at least one knotted-peptide.

44. The method of claim 40, wherein the moiety is conjugated to the N-terminus of the at least one knotted-peptide.

45. The method of claim 43 or 44, wherein the moiety comprises a pre-defined number of a heavy isotope atom to modify the unique mass signature or digest fragment mass signature of the at least one knotted-peptide.

46. The method of claim 45, wherein the heavy isotope atom comprises ^{13}C or deuterium.

47. The method of claim 39, wherein the plurality of knotted-peptides comprises greater than 100 peptides, greater than 1000 peptides, or greater than 10000 peptides.

48. The method of claims 33 to 47, further comprising establishing a relationship between the structure and the pharmacological property of the drug candidates.

49. The method of claim 48, further comprising selecting a subset of drug candidates based on the relationship.

50. The method of any one of claims 33 to 49, further comprising obtaining a plurality of samples at a selected time interval for a selected duration of time following administration.

51. The method of claim 50, wherein selected time intervals are 5 minutes, 10 minutes, 20 minutes, 30 minutes, 40 minutes, 50 minutes, or 60 minutes.

52. The method of claim 50, wherein selected duration of time is 1 hour, 2 hours, 3 hours, 4 hours, 5 hours, 6 hours, 7 hours, 8 hours, 9 hours, 10 hours, 11 hours, 12 hours, 15 hours, 30 hours, 45 hours, 60 hours, 75 hours, 90 hours, 105 hours, or 120 hours.

53. A method of determining a distribution profile of knotted-peptides administered to a subject by different administration pathways, the method comprising:

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;

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; and

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.

54. A method of determining a distribution profile of knotted-peptides administered to a subject by different administration pathways, the method comprising:

analyzing an isolated sample from a subject following administration to the subject of a composition comprising 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;

analyzing an isolated sample from a subject following administration to the subject of a composition comprising 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; and

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.

55. The method of claim 53 or 54, wherein the light knotted peptide comprises fewer heavier isotopes than the light knotted peptide.

56. The method of any one of claims 53 to 55, wherein the heavy knotted-peptide comprises at least one more ^{13}C atom or deuterium atom than the light knotted-peptide.

57. The method of any one of claims 53 to 56, wherein the first knotted-peptide is conjugated to a first moiety and the heavy knotted-peptide is conjugated to a second moiety.

58. The method of claim 57, wherein the first moiety, second moiety, or both is conjugated to the N-terminus of the light and heavy knotted peptide, respectively.

59. The method of claim 57, wherein the first moiety, second moiety, or both comprises a hydrophobic moiety.

60. The method of any one of claims 53 to 59, wherein the first route and second route of delivery are independently selected from an oral route, a topical route, a transmucosal route, an intravenous route, an intramuscular route, and an inhalation route.

61. The method of any one of claims 53 to 60, wherein the either the first route or the second route of delivery comprises an oral route.

62. The method of claim 57, wherein at least one ^{13}C atom is present in the first moiety, the second moiety, or both.

63. The method of any one of claims 53 to 62, wherein the light knotted-peptide, the heavy knotted-peptide, or both have unique mass signature or digest mass signature when analyzed by mass spectrometry.

64. The method of claims 53 to 63, further comprising establishing a relationship between the structure and the pharmacological property of the drug candidates.

65. The method of claim 64, further comprising selecting a subset of drug candidates based on the relationship.

66. The method of any one of claims 53 to 65, further comprising obtaining a plurality of samples at a selected time interval for a selected duration of time following administration.

67. The method of claim 66, wherein selected time intervals are 5 minutes, 10 minutes, 20 minutes, 30 minutes, 40 minutes, 50 minutes, or 60 minutes.

68. The method of claim 66, wherein selected duration of time is 1 hour, 2 hours, 3 hours, 4 hours, 5 hours, 6 hours, 7 hours, 8 hours, 9 hours, 10 hours, 11 hours, 12 hours, 15 hours, 30 hours, 45 hours, 60 hours, 75 hours, 90 hours, 105 hours, or 120 hours.

69. 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.

70. The method of claim 69, wherein the drug candidates are selected from the group consisting of small chemical molecules, biologics, and peptides.

71. The method of claim 69, wherein the plurality of drug candidates comprises a plurality of peptides.

72. The method of claim 69, 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.

73. The method of claim 71, further comprising generating a peptide library comprising the at least some of the peptides having the pharmacological property.

74. The method of claim 71, 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.

75. The method of claim 74, 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.

76. The method of claim 75, wherein the protein:protein interaction is associated with cancer.

77. The method of claim 71, 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.

78. The method of claim 77, 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.

79. The method of claim 71, 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.

80. The method of claim 69, wherein the plurality of drug candidates comprises greater than five drug candidates.

81. The method of claim 69, wherein the plurality of drug candidates comprises greater than 10 drug candidates.

82. The method of claim 69, wherein the plurality of drug candidates comprises greater than 100 drug candidates.

83. The method of claim 69, wherein the plurality of drug candidates comprises greater than 1000 drug candidates.

84. The method of claim 69, wherein the plurality of drug candidates comprises greater than 10000 drug candidates.

85. The method of claim 71, wherein the plurality of peptides comprises greater than 100 peptides.

86. The method of claim 71, wherein the plurality of peptides comprises greater than 1000 peptides.

87. The method of claim 71, wherein the plurality of peptides comprises greater than 10000 peptides.

88. The method of claim 71, wherein at least one peptide of the plurality of peptides comprises a detectable label for tracing the at least one peptide in the subject.

89. The method of claim 88, wherein the detectable label comprises a near infrared dye.

90. The method of claim 71, 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.

91. The method of claim 90, wherein the hydrophobic moiety comprises a hydrophobic fluorescent dye or a saturated or unsaturated alkyl group.

92. The method of claim 69, wherein the sample comprises a tissue sample or a fluid sample.

93. The method of claim 92, wherein the fluid sample comprises blood, urine, mucous or spinal fluid.

94. The method of claim 69, wherein the sample comprises blood.

95. The method of claim 92, wherein the tissue sample comprises biopsy or necropsy tissue.

96. The method of claim 92, 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.

97. The method of claim 69, further comprising using the pharmacological property common to the drug candidates to generate additional drug candidates having the pharmacological property.

98. The method of claim 69, wherein the subject is an animal.

99. The method of claim 69, wherein the subject is a human.

100. A method of generating mass-defined drug candidate libraries, the method comprising:

producing a plurality of drug candidates, at least some of the drug candidates each having a unique mass signature or digest fragment mass signature;

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; and

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.

101. The method of claim 100, 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, or a combination thereof.

102. The method of claim 100, wherein the pharmacological property comprises oral bioavailability.

103. The method of claim 100, wherein the drug candidates are selected from the group consisting of small chemical molecules, biologics, and peptides.

104. The method of claim 100, wherein the plurality of drug candidates comprises greater than five drug candidates.

105. The method of claim 100, wherein the plurality of drug candidates comprises greater than 10 drug candidates.

106. The method of claim 100, wherein the plurality of drug candidates comprises greater than 100 drug candidates.

107. The method of claim 100, wherein the plurality of drug candidates comprises greater than 1000 drug candidates.

108. The method of claim 100, wherein the plurality of drug candidates comprises greater than 10000 drug candidates.

109. The method of claim 100, wherein the plurality of drug candidates comprise knotted-peptides.

110. The method of claim 109, wherein the unique mass signature or digest fragment mass signature of the knotted-peptides are defined by the natural amino acid sequence of the knotted peptides.

111. The method of claim 100, wherein at least one of the drug candidates in the plurality comprises a pre-defined number of a heavy isotope atom to modify the unique mass signature or digest fragment mass signature of the at least one drug candidate.

112. The method of claim 101, wherein the heavy isotope atom comprises ^{13}C or deuterium.

113. The method of claim 109, wherein the unique mass signature or digest fragment mass signature of at least one of the knotted-peptides in the plurality is defined by a moiety conjugated to the at least one knotted-peptide.

114. The method of claim 113, wherein the moiety is conjugated to the N-terminus of the at least one knotted-peptide.

115. The method of claim 113, wherein the moiety comprises a pre-defined number of a heavy isotope atom to modify the unique mass signature or digest fragment mass signature of the at least one knotted-peptide.

116. The method of claim 115, wherein the heavy isotope atom comprises ^{13}C or deuterium.

117. The method of claim 109, wherein the plurality of knotted peptides comprises greater than 100 peptides.

118. The method of claim 109, wherein the plurality of knotted peptides comprises greater than 1000 peptides.

119. The method of claim 109, wherein the plurality of knotted peptides comprises greater than 10000 peptides.

120. A method of determining a distribution profile of knotted-peptides administered to a subject by different administration pathways, the method comprising:

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;

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; and

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.

121. The method of claim 120, wherein the light knotted peptide comprises fewer heavier isotopes than the light knotted peptide.

122. The method of claim 120, wherein the heavy knotted-peptide comprises at least one more ^{13}C atom or deuterium atom than the light knotted-peptide.

123. The method of claim 120, wherein the first knotted-peptide is conjugated to a first moiety and the heavy knotted-peptide is conjugated to a second moiety.

124. The method of claim 123, wherein the first moiety, second moiety, or both is conjugated to the N-terminus of the light and heavy knotted peptide, respectively.

125. The method of claim 123, wherein the first moiety, second moiety, or both comprises a hydrophobic moiety.

126. The method of claim 120, wherein the first route and second route of delivery are independently selected from an oral route, a topical route, a transmucosal route, an intravenous route, an intramuscular route, and an inhalation route.

127. The method of claim 120, wherein the either the first route or the second route of delivery comprises an oral route.

128. The method of claim 120, wherein at least one ^{13}C atom is present in the first moiety, the second moiety, or both.

129. The method of claim 120, wherein the light knotted-peptide, the heavy knotted-peptide, or both have unique mass signature or digest mass signature when analyzed by mass spectrometry.

130. A peptide for imaging a tumor in a subject, the peptide comprising a chlorotoxin comprising 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 comprises a detectable label.

131. The peptide of claim 130, wherein 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 peptide are D-amino acids.

132. The peptide of claim 130 or 131, having an amino acid sequence at least 80%, 83%, 86%, 89%, 90% or 92 % identical to the following sequence of MCMPCFTTDHQMARXCDDCCGGXGRGXCYGPQCLCR, wherein X is selected from K, A and R.

133. A peptide comprising an amino acid sequence having at least 80% identical to the following sequence of MCMPCFTTDHQMARXCDDCCGGXGRGXCYGPQCLCR, wherein at least three of the amino acids in the amino acid sequence are D-amino acids, wherein X is selected from K, A and R.

134. The peptide of claim 133, wherein at least 50%, at least 60%, or at least 70% of the amino acids in the peptide are D-amino acids.

135. The peptide of any of claims 133-134, further comprising a detectable label.

136. The peptide of any of claim 130-135, wherein all the amino acids in the peptide are D-amino acids.

137. The peptide any of claims 130-132 or 135-137, wherein the detectable label is conjugated to the N-terminus of the peptide or conjugated to a lysine residue in the peptide.

138. The peptide of claim 137, wherein the detectable label comprises a near-infrared dye.

139. The peptide of claim 137, wherein the detectable label comprises a cyanine dye.

140. A composition comprising the peptide of any one of claims 130-139, or a combination thereof.

141. A method of detecting a peptide in a subject, the method comprising:

administering to the subject an effective amount of the peptide according to any of claims 130-132 or the peptide according to any of claims 135-139 or the composition according to claim 126, and detecting detectable label.

142. The method of claim 141, wherein the detecting comprises obtaining an image of a region in the subject by detecting the detectable label.

143. The method of claim 142, wherein the detecting comprises intra-operative visualization of cancerous tissue.

144. The method of claim 142, wherein visualization of the detectable label guides surgical removal of a tumor in the subject.

145. A method of treating a disease associated with cells expressing a chlorotoxin target, the method comprising:

administering, to a subject in need thereof, a therapeutically effective amount of a pharmaceutical composition comprising the peptide according to claims 130-139 or the composition of claim 140, thereby treating the disease.

146. The method of claim 145, wherein the peptide of the composition further comprises a cytotoxic agent, a toxin, an antisense nucleotide, a cancer drug, a nucleotide drug, a metabolic modulator, a radiosensitizer, a peptide therapeutic, a peptide-drug conjugate, or a combination thereof.

147. The method of claim 145 or of claim 146, wherein the disease comprises cancerous tissue associated with glioma, skin cancer, lung cancer, lymphoma, medulloblastoma, prostate cancer, pancreatic cancer, breast cancer, mammary cancer, colon cancer, sarcoma, oral squamous cell carcinoma, hemangiopericytoma, or a combination thereof.