LIPOCALIN FUSION PARTNERS

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

[0001] 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

[0002] Efforts toward drug discovery continue to use vast technical and financial resources to identify and develop new and useful drugs. Unfortunately, finding new drugs has continued to be difficult. For example, development of less damaging, more precisely targeted cancer therapies is essential. But even after decades of research, scientists still struggle to identify therapeutic compounds with the right mix of medicinal and cancer-targeting properties which not only reduce the likelihood that a therapeutic compound could serve as a treatment strategy but creates a need for successful methods of surgical resection.

[0003] A wide variety of types of compounds have been studied and pursued for a large breadth of 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.

[0004] A lack of methods for rapid and efficient production of peptides and proteins for clinical applications has limited the discovery of peptides and proteins that might serve as therapeutic compounds. Creating fusions of peptides, protein domains, or proteins, to a different protein to enhance production of peptides and proteins has been explored in bacterial expression systems, e.g., *E. coli* gene expression. However, bacterial protein expression systems are generally limited due to errors in protein folding efficiency. Thus, there is still a need for methods enabling the rapid and efficient production of peptides, protein domains, and proteins.

SUMMARY OF THE INVENTION

[0005] The present invention relates to methods for the production of siderocalin fusion proteins. In various aspects, the present invention relates to the fusion proteins produced according to those methods. In certain apsects, the fusion protein is cleaved, thereby producing a peptide according to the present disclosure.The present invention further relates to methods for producing a fusion protein, the methods comprising expressing, in a cell, a fusion protein, the fusion protein comprising a peptide or protein domain and a lipocalin protein, thereby producing the fusion

protein. In some embodiments, the present invention further relates to a method of producing a peptide, the method comprising: expressing, in a cell, a fusion protein comprising a peptide and a lipocalin protein; and separating the peptide from the lipocalin protein, thereby producing the peptide.

[0006] In some embodiments, the present invention relates to a composition of a fusion protein, the composition comprising; a peptide or protein domain, and; a lipocalin protein. In some embodiments, the present invention further relates to a composition comprising a peptide library, the peptide library further comprising a plurality of peptides lacking at least one native lysine residue. The present invention further relates to a composition comprising a fusion protein comprising a peptide and a lipocalin protein.

INCORPORATION BY REFERENCE

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

BRIEF DESCRIPTION OF THE DRAWINGS

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

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

[0010] FIG. 2 shows an example method of producing peptide libraries, in accordance with an embodiment of the present invention.

[0011] FIG. 3 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.

[0012] FIG. 4 shows analysis of a large peptide library using mass spectrometry, in accordance with an embodiment of the present invention.

[0013] FIG. 5 shows an example method of using siderocalin fusions to express knottin variants, in accordance with an embodiment of the present invention.

[0014] FIG. 6 depicts SDS-PAGE analysis of expressing knottin scaffolds, in accordance with embodiments of the present invention.

[0015] FIG. 7 provides a schematic of pooled library production, in accordance with an embodiment of the present invention. FIG. 7B shows an expanded view of FIG. 7A.

[0016] FIG. 8 describes representative sequencing data from a cloned knottin library, in accordance with an embodiment of the present invention.

[0017] FIG. 9 shows SDS-PAGE analysis of 3000-member knottin libraries, in accordance with embodiments of the present invention.

[0018] FIG. 10 shows a generalized process for manufacturing knottins, in accordance with an aspect of the present disclosure.

[0019] FIG. 11 shows a generalized process for manufacturing knottins, in accordance with an aspect of the present disclosure.

[0020] FIG. 12 a generalized process for manufacturing knottins, in accordance with an aspect of the present disclosure. According to this aspect, cleavage occurs at the furin-knottin interface.

[0021] FIG. 13 shows examples of knottins made with the Daedalus system and corresponding SDS PAGE analyses in accordance with an aspect of the present disclosure. The Daedalus system is desribed in Bandaranayake A.D., et al., *Nucleic Acids Res.* (2011) 39(21):e143.

[0022] FIG. 14 shows the alignment of Scn sequences from 18 species. The alignment shows a high level of sequence conservation between the 18 species with recognizable orthologs. Positions with an asterisk are possible sites for ligand binding. Correnti, C. & Strong, R. K. (2013) 'Iron sequestration in immunity' In Metals in Cells; Encyclopedia of Inorganic and Bioinorganic Chemistry. (Culotta, V. & Scott, R. A., eds.) John Wiley & Sons, pp. 349-59

[0023] FIG. 15 depicts a 3D model of super-stable Scn in accordance with an aspect of the present disclosure. In this form of Scn, a second disulfide bond was engineered in order to secure the N-terminus and increase thermal stability.

[0024] FIG. 16 shows a schematic of a generic Scn fusion that can be used in various aspects of the present disclosure. The native signal peptide was been removed and an exogenous sFLAG and HIS tag were added to facilitate purification. These modifications are optionally present in various aspects of the disclosure.

[0025] FIG. 17 depicts a schematic of a light chain antibody fusion with Scn and corresponding SDS PAGE analysis according to an aspect of the present disclosure.

[0026] FIG. 18 shows the SDS PAGE analysis comparing the expression of a Scn fusion with a cytoplasmic enzyme, HMOX1, both before and after cleavage according to an aspect of the present dislcosure.

[0027] FIG. 19 depicts the expression of a Scn fusion with a cytoplasmic viral protein Adv2 and the corresponding SDS PAGE analysis according to one aspect of the present disclosure.

[0028] FIG. 20 depicts the expression of a Scn fusion with an extracellular viral glycoprotein HIV gp120 and the corresponding SDS PAGE analysis according to one aspect of the present disclosure.

[0029] FIG. 21 depicts the expression of a Scn fusion with a knottin protein, Imperatoxin, and the corresponding SDS PAGE analysis according to one aspect of the present disclosure.

[0030] FIG. 22 depicts the expression of a Scn fusion with a small subdomain (i.e., Kringle domain) of the extracellular tyrosine kinase receptor ROR1 and corresponding SDS PAGE analysis according to one aspect of the present disclosure.

[0031] FIG. 23 depicts the expression of a Scn fusion with heptameric and trimeric subdomains and corresponding SDS PAGE analysis according to one aspect of the present disclosure.

[0032] FIG. 24 depicts the expression of an ExFABP fusion with a knottin and corresponding SDS PAGE analysis according to one aspect of the present disclosure. ExFABP is another functional Scn. According to the present disclosure, this construct can be used in a periplasmic bacterial system to secret a variety of client proteins.

[0033] FIG. 25 depicts the crystal structure of Scn with a Th ligand.

[0034] FIG. 26 depicts a captured fluorescent siderophore. The left frame depicts a size-exclusion purification of the protein siderophore complex. The right frame is a schematic.

DETAILED DESCRIPTION OF THE INVENTION

[0035] The methods and systems of the present disclosure relate to fusion proteins and methods of producing them. According to certain aspects, the peptides produced using the present methods can be used as components of drug discovery platforms. In some aspects, the methods relate to creating a fusion of a first protein, or of a peptide, that has potential therapeutic value, with a second protein such that the second protein enhances production and folding of the first protein by an expression system. In some aspects, after expression and purification of the fusion protein, the fusion protein is cleaved such that the first protein and the second protein are separate proteins. In some aspects, the second protein may be a lipocalin. For example, the second protein may be a specific lipocalin such as siderocalin.

Conjugates

[0036] In some aspects, the present invention includes fusion proteins, peptides, or conjugates thereof as described herein. For example, some or all of the fusion proteins or peptides can be conjugated to a moiety selected to modify a property of the peptides.

[0037] In certain aspects, the present invention includes fusion proteins or peptides conjugated at the N-terminus to hydrophobic (e.g., lipophilic) moieties. All or some of the fusion proteins or peptides of the present disclosure 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 fusion protein or peptide of the present disclosure. In some embodiments, simple carbon chains (e.g., by myristoylation and/or palmitylation) can be conjugated to the fusion proteins or peptides. In some aspects, the simple carbon chains may render the fusion proteins or peptides easily separable from the unconjugated material. For example, methods that may be used to separate the fusion proteins or peptides from the unconjugated material include, but are not limited to, solvent extraction and reverse phase chromatography. 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 fusion protein or peptide can also be performed to allow for tracing of the conjugated fusion protein or 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 fusion proteins or 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.

[0038] Other modifications can be used. For example, the fusion proteins, peptides, or conjugates thereof can include post-translational modifications (e.g., methylation and/or amidation). In some embodiments, the fusion proteins or peptides of the present disclosure 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.

[0039] In some embodiments, the fusion proteins or peptides of the present disclosure 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.

[0040] 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.

[0041] 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.

[0042] The fusion proteins or peptides 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.

[0043] In some aspects, the fusion proteins and peptides 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.

[0044] 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.

[0045] 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.

[0046] 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.

[0047] Other modifications to fusion proteins, peptides, or conjugates thereof of the present disclosure can be used. For example, the fusion proteins or peptides of the present disclosure can include post-translational modifications (e.g., methylation and/or amidation), which can affect, e.g., serum half-life. In some embodiments, the fusion proteins or peptides 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.

[0048] The fusion proteins or peptides of the present disclosure 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 of the present invention can also be conjugated to biotin. In addition to 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.

Methods of Making Fusion Proteins and Peptides

[0049] In yet another aspect, the present invention includes methods for making fusion proteins or peptides according to the present disclosure.

[0050] In some embodiments, the present invention includes methods of making fusion proteins or peptides according to the present disclosure. As described further herein, the present invention includes scaffolds that can be used as a starting point for generating fusion proteins or peptides according to the present disclosure. These scaffolds as well as a large diversity of scaffold variants can be made using several different approaches. In some aspects, the fusion proteins or peptides according to the present disclosure 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.

[0051] 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.

[0052] 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.The present invention relates to methods for producing a fusion protein, the methods comprising expressing, in a cell, a fusion protein, the fusion protein comprising a peptide or protein domain and a lipocalin protein, thereby producing the fusion protein. In some embodiments, the methods further comprise separating the peptide or protein domain from the lipocalin protein, thereby producing a peptide or protein domain. In certain embodiments, the peptide or protein domain is an antibody fragment. In certain embodiments, the antibody is trastuzumab, infliximab, adalimumab, OKT3, or Fc.

[0053] In some embodiments, the peptide or protein domain is human heme oxygenase 1 or murine heme oxygenase 1. In some embodiments, the fusion protein further comprises a cleavage site. In certain embodiments, the cleavage site is a furin cleavage site, a trypsin cleavage site or a TEV cleavage site. In some embodiments, the separating of the peptide or protein domain from the lipocalin protein results from cleavage at the cleavage site in the fusion protein. In certain embodiments, the separating of the peptide or protein domain from the lipocalin protein occurs following secretion of the fusion protein from a cell. In certain embodiments, the cell is a mammalian cell.

[0054] The present invention relates to methods of producing a peptide, the method comprising: expressing, in a cell, a fusion protein comprising a peptide and a lipocalin protein; and separating the peptide from the lipocalin protein, thereby producing the peptide. In some embodiments, the peptide is separated from the lipocalin protein by proteolysis or by cleavage of a furin cleavage site in the fusion protein such that the peptide is cleaved from the lipocalin protein upon secretion from the cell. In certain embodiments, the cell is a mammalian cell. In some embodiments, the peptide is produced at a concentration less than about 200 mg/liter.

[0055] In some embodiments, the present invention relates to compositions comprising a fusion protein comprising a peptide and a lipocalin protein. In some embodiments, the peptide comprises a knotted peptide. In some embodiments, the knotted-peptide is selected from the group consisting of chymotrypsin inhibitor, hefutoxin, bubble protein, the C-terminal domain of midkine, potato carboxypeptidase inhibitor, and epiregulin. In certain embodiments, the lipocalin

protein is siderocalin. In certain embodiments, the knotted-peptide includes at least two disulfide bonds

[0056] The peptides described herein (e.g., knotted peptides) can be expressed using a variety of lipocalin proteins. As used herein, the term "lipocalin" refers to a protein as defined in "Structure and sequence relationships in the lipocalins and related proteins", Darren R. Flower, Anthony C.T. North, Teresa K. Attwood, *Protein Science* (1993) 2:5, 753–761. Lipocalins may include, but are not limited to, the numbered lipocalins (e.g., Lcn2 (also NGAL, Siderocalin, 24p3), and the like, chicken Ex-FABP and quail Q86.

[0057] In various apsects of the present disclosure, siderocalin is used as a secretion partner. As used herein, the term "siderocalin" refers to a lipocalin that is capable of binding a small chelator. In some aspects, a chelator may be natural or engineered. In some aspects, siderocalin binds to siderophores and ferric siderophore complexes. For example, a siderocalin can be, but is not limited to, a siderocalin or Ex-FABP. As used herein, the term "Siderocalin," (as a proper noun), refers to the orthologous family of proteins related to the human archetype Siderocalin.

[0058] Siderocalin advantegeously can be used for the secretion of a variety of peptides, proteins, and protein domains, including intracellular peptides, proteins, and peptide domains when used as a secretion partner.

[0059] Siderocalin is useful as a fusion partner for 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 (*see* Goetz, D. H., et al. 'The Neutrophil Lipocalin NGAL is a Bacteriostatic Agent that Interferes with Siderophore-mediated Iron Acquisition' *Molecular Cell* (2002) 10: 1033-43). 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 aspects, the peptides can also be expressed as fusion peptides with Murine SCN (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.In some embodiments, the peptide comprises a disulfide knotted-peptide. In certain embodiments, the knotted-peptide is PMP-D2, potato carboxypeptidase, huwentoxin, imperatoxin, epiregulin, midkine, bubble protein or conotoxin CVIC. In other embodiments, the peptide comprises a knottin. In some embodiments, the lipocalin protein comprises siderocalin. In some embodiments, the lipocalin

protein is siderocalin. In certain embodiments, the lipocalin protein is human siderocalin, murine siderocalin, chicken Ex-FABP, or quail Q86.

[0060] In some embodiments, the expression of peptides (e.g., knotted peptides) as fusions with SCN can be utilized with an endogenously cleaving SCN, 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 SCN and knotted peptide can be secreted into surrounding media. ENLYFQ is a tobacco etch virus (TEV) 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 SCN 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.

[0061] In some embodiments, the peptide comprises a knotted-peptide. In certain embodiments, the knotted-peptide is selected from the group consisting of chymotrypsin inhibitor, hefutoxin, bubble protein, the C-terminal domain of midkine, and epiregulin. In some embodiments, the lipocalin protein comprises siderocalin.

[0062] 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. SCN also has unique ligand specificity and tightly binds catecholate siderophores (bacterial iron chelators). This opens the possibility of loading the SCN fusion with specific ligands, such as a chemotherapeutic or radioactive reagent or some type or a compound that has beneficial properties. SCN, when loaded with siderophores and iron, has a deep red color that can aid in chromatography or other purification steps.

[0063] The lipocalin fusions of the present disclosure (e.g., SCN fused with HO-1) can be used in a variety of ways. For example, fusions can be used to increase the size of the target protein (e.g., a potential therapeutic) in order to increase its biological half-life. Fusions can be used to secrete the target protein where the target protein is naturally expressed in the cytoplasm. Fusions can also be used to target the fusion partner protein to specific locations to maximize therapeutic effects. SCN also has unique ligand specificity, tightly binding catecholate

siderophores (bacterial iron chelators), and has kidney protective effects. Combining SCN with other kidney protective agents, like HO-1, which could generate synergistic functional effects. This also opens the possibility of loading the SCN fusion with specific ligands, such as radioactive metal atoms (e.g.. ^{55}Fe or Th) to allow *in vivo* tracking or specific cytotoxic activities. 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/LIn some embodiments, the peptide is produced at a concentration less than about 500 mg/liter, less than about 400 mg/liter, less than about 300 mg/liter, less than about 200 mg/liter, less than about 100 mg/liter, or less than about 50 mg/liter.

[0064] 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. The peptides can also be expressed in insect cell systems and eukaryotic systems including mammalian systems.

[0065] In some aspects, the peptides disclosed herein can be introduced 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.

[0066] 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.

[0067] 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. 2 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.

[0068] In an example embodiment, the present invention includes fusion proteins of a knotted peptide fused to siderocalin via a cleavable linker. FIG. 3 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. In some embodiments, the fusion protein further comprises at least one of an IgK starter sequence, a sFLAG, a HIS, and a TEV. In certain embodiments, the fusion protein comprises the following

construct: IgK SP-sFLAG-HIS-siderocalin-TEP-peptide. In some embodiments, the peptide comprises a knotted-peptide.

[0069] 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.

[0070] 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.

Methods of Making Knotted Peptides and Related Compositions

[0071] The fusion systems of the present disclosure can be used in various aspects for the production of peptides, knottins, and cytoplasmic and secreted proteins. In some aspects, the methods and compositions described herein include fusion of target proteins and/or peptides to lipocalin such that lipocalin facilitates the expression and secretion of the target protein by a cell. For example, lipocalins have a conserved fold characterized by an eight-stranded beta barrel with a flanking alpha helix and supports a versatile scaffold. In some aspects, lipocalin fusion protein systems result in greater fusion protein expression in mammalian cells compared to systems without the use of a lipocalin fusion protein. For example, lipocalin fusion protein systems result in less than 0.5 times greater, 1 times greater, 2 times greater, 3 times greater, 4 times greater, 5 times greater, 6 times greater, 7 times greater, 8 times greater, 9 times greater, 10 times greater, 11 times greater, 12 times greater, 13 times greater, 14 times greater, 15 times greater, 16 times greater, 17 times greater, 18 times greater, 19 times greater, 20 times greater, 25 times greater, 30 times greater, 35 times greater, 40 times greater, 45 times greater, 50 times greater, 55 times greater, 60 times greater, 65 times greater, 70 times greater, 75 times greater, 80 times greater, 85 times greater, 90 times greater, 95 times greater, 100 times greater, 200 times greater, 300 times greater, 400 times greater, 500 times greater, 600 times greater, 700 times greater, 800 times greater, 900 times greater, or 1000 times fusion protein expression in mammalian cells compared to systems lacking the lipocalin fusion protein.

[0072] In some aspects, lipocalin (e.g., SCN) may be used as a fusion partner to stabilize proteins or peptides of interest as immunogens. In aspects, the species from which the lipocalin sequence is derived is different than the recipient species. In some aspects, the species from which the lipocalin sequence is derived is the same as the recipient species.

[0073] SCN, and related proteins, when used as a secretion partner advantageously improve the production of secreted proteins and peptides. Moreover, SCN, and related proteins, advantageously are small, thereby improving their bioavailability. For example, the mature protein is 178 amino acids and has a molecular weight of 20547 Da. SCN has a single intramolecular disulfide bond, which increases its stability and a single N-linked glycosylation site.

[0074] In some aspects, at least one of the amino acids in the native sequence of lipocalin 2 (SCN) may be substituted for a non-native amino acid. In some aspects, the mutations may be generated to prevent SCN from dimerizing. For example, one SCN protein may dimerize with another SCN protein at cysteine residues or one SCN protein may dimerize with different proteins at cysteine residues. For example, generating a C87S mutation in SCN may prevent

dimerization at a cysteine residue (see Goetz, D. H., et al., 'The Neutrophil Lipocalin NGAL is a Bacteriostatic Agent that Interferes with Siderophore-mediated Iron Acquisition' *Molecular Cell* (2002) 10: 1033-43).

[0075] In some aspects, at least one of the amino acids in the native sequence of the non-human lipocalin protein orthologous to the human SCN may be substituted for a non-native amino acid. For example, non-human orthologs of the human lipocalin protein that may be used with the methods and compositions described herein include, but are not limited to, murine Lcn2 (e.g., 24p3), Lcn1, Lcn6, Lcn8, Lcn9, Lcn10, Lcn12 and Lcn15 (see FIG. 14). In some aspects, the mutations may be generated to prevent non-human lipocalin protein orthologous to the human SCN from dimerizing. For example, one non-human lipocalin protein orthologous to the human SCN protein may dimerize with another non-human lipocalin protein orthologous to the human SCN protein at cysteine residues or one non-human lipocalin protein orthologous to the human SCN protein may dimerize with a different protein at cysteine residues. For example, generating a mutation at a site orthologous to the C87S mutation in human lipocalin protein may prevent dimeraizaiton at a cysteine residue (see Goetz, D. H., et al., 'The Neutrophil Lipocalin NGAL is a Bacteriostatic Agent that Interferes with Siderophore-mediated Iron Acquisition' *Molecular Cell* (2002) 10: 1033-43).

[0076] In an exemplary aspect, a lipocalin fusion protein may contain the following protein sequence:

MPLGLLWLGLALLGALHAQAQDSTSDLIPAPPLSKVPLQQNFQDNQFQGKWYVVGLAG
NAILREDKDPQKMYATIYELKEDKSY**N**VTSVLFRKKK**C**DYWIRTFVPGCQPGEFTLGNI
KSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNREYFKITLYGRTKELTSELKENFIRFSK
SLGLPENHIVFPVPIDQ**C**IDG

In this sequence, the annotations indicate, a signal peptide, **glycosylation site**, **disulfide bond** and the C87S mutation.

[0077] In some aspects, fusion of a protein or peptide of interest to lipocalin may improve the biological properties of the target protein (e.g., a potential therapeutic). For example, fusion of a protein or peptide of interest to lipocalin may increase the half-life of the protein or peptide of interest by increasing the size of the overall protein. For example, the increased size of the protein may prevent glomerular filtration of the protein or peptide of interest. In some aspects, fusions of small proteins, such as antibody fragments, exhibit decreased glomerular filtration. These effects are observed with both enzymes and Fab antibody fragments. In some embodiments, the fusion protein further comprises a peptide selected from the group consisting of: an IgK starter sequence, a sFLAG, a HIS, siderocalin, a TEV and the knotted peptide

sequence of interest. In certain embodiments, the fusion protein comprises the following construct: IgK SP-sFLAG-HIS-siderocalin-TEV-peptide. In some embodiments, the knotted-peptide includes at least two disulfide bonds.

[0078] In some embodiments, the fusion protein further comprises at least one of an IgK starter sequence, a sFLAG, a HIS, and a TEV. In certain embodiments, the fusion protein comprises the following construct: IgK SP-sFLAG-HIS-siderocalin-TEV-peptide. In some embodiments, the fusion protein is generated by direct fusion of each subunit to the adjacent subunits. In certain embodiments, the composition further comprises a linker sequence between the peptide or protein domain and the lipocalin protein. In some embodiments, the peptide comprises a knotted-peptide.

[0079] In some embodiments, the fusion protein further comprises at least one of an IgK starter sequence, a sFLAG, a HIS, and a TEV. In certain embodiments, the fusion protein comprises the following construct: IgK SP-sFLAG-HIS-siderocalin-TEV-peptide. In other embodiments, the peptide comprises a knotted-peptide

[0080] In some aspects, fusion of the protein or peptide of interest to lipocalin may enhance the purification of the protein or peptide of interest after production. For example, the protein or peptide of interest fused to lipocalin may be produced from a protein expression system (e.g., fusion protein expression system). For example, proteins or peptides of interest may be retained in a compartment of a cell during production if proteins or peptides of interest are not fused to lipocalin.

[0081] In some aspects, SCN has protective properties *in vivo* that are imparted to a fusion partner as the basis of *in vivo* therapeutics fusion partners. In some aspects, these therapeautics can be stabilized for use as immunogens, matching the species donating the lipocalin sequence to that of the recipient species to focus elicited immune responses to the fusion partner. SCN also has unique ligand specificity and tightly binds siderophores (ferric iron chelators).

[0082] The methods and compositions described herein include fusion proteins of a protein or peptide of interest coupled to SCN and a SCN ligand. In some aspects, the ligands for SCN may include, but are not limited to, a siderophore or an organizing metal. In some aspects, fusion proteins including the SCN ligands may be coupled to a fluorphore. For example, the coupling may be covalently coupled. In some aspects, fusion proteins including the SCN ligands may be coupled to a luminescent siderophore. In some aspects, the luminescent siderophore may include a metal complex.

[0083] In some aspects, the addition of SCN ligands to the fusion proteins could be used to detect or localize the target protein, or target peptide, of the fusion protein. For example, the

detection and localization could be performed either *in vitro* or *in vivo*. In some aspects, the addition of SCN ligands to the fusion proteins could be used to purify fusion proteins from mixtures. For example, an SCN ligand (e.g., a siderophore/metal complex) could be contacted with at least one purification resin metal other than, including but not limited to, aluminum, gadolinium, indium, vanadium, plutonium or thorium, and any related isotopes.

[0084] In some apects, the addition of SCN ligands to the fusion proteins could be used for the delivery of radionuclides to a target tissue. In some aspects, the addition of SCN ligands to the fusion proteins could be used to deliver iron to a target. In addition, SCN bound to siderophores and iron can be a dark red color. For example, the dark red color can be combined with methods and compositions of chromatography or other steps in a method of purification.

[0085] In an exemplary apect, a self-cleaving SCN isoform may be added to a peptide or protein of interest to generate a fusion protein. For example, the self-cleaving SCN isoform may contain the RARYKR amino acid sequence immediately following the CIDG amino acid sequence. In this case, the RARYKR sequence may be cleaved by an endogenous enzyme to the cells of the protein expression system (e.g., the mammalian cells) during export of the fusion protein. For example, furin may cleave the RARYKR sequence. In this case, SCN and the peptide or protein of interest may be free and located in the extracellular space.

[0086] In an exemplary apect, an exogenously cleaved SCN isoform may be added to a peptide or protein of interest to generate a fusion protein. For example, the exogenously cleaved SCN isoform may contain the ENLYFQ amino acid sequence immediately following the CIDG amino acid sequence. In this case, the ENLYFQ sequence may be cleaved by an exogenous enzyme to the cells of the protein expression system (e.g., the mammalian cells) during export or after export of the fusion protein. For example, a tobacco etch protease may cleave the ENLYFQ site.

[0087] In some apects, the fusion proteins may be secreted from the cells of the protein expression system (e.g., the mammalian cells) as fusion proteins. In this case, the peptide or protein of interest may be cleaved from the SCN protein by TEV protease. For example, the TEV protease may be added to the cells of the protein expression system (e.g., the mammalian cells) or added after removal of the fusion protein from the cells.

[0088] In some apects, the SCN peptide may be modified to include compounds for purification or isolation. In some aspects, the compounds may be an amino acid or more than one amino acid. For example, the compounds may be poly-histidine or poly-arginine and may be located between Lcn2 and a signal peptide in the fusion protein. In some aspects, the compounds may be removed from Lcn2 using an enzyme or proteolysis.

[0089] The present invention relates to compositions comprising a peptide library, the peptide library further comprising a plurality of peptides lacking at least one native lysine residue. In some embodiments, the peptides are conjugated to an adaptor molecule. In certain embodiments, the adaptor molecule is a peptide. In some embodiments, the peptide has a unique signature determined by mass spectroscopy.

Methods of Use

[0090] The fusion proteins, peptides, or conjugates thereof of the present disclosure can be used for a variety of other applications, such as therapeutic and/or diagnostic applications. In some embodiments, the fusion proteins, peptides, or conjugates thereof of the present disclosure can be used for methods of treating diseases. In some embodiments, the fusion proteins, peptides, or conjugates thereof of the present disclosure can be used to deliver drugs to, e.g., tumors in the brain of a subject.

[0091] The present invention also provides compositions for administering the fusion proteins, peptides, or conjugates thereof described herein to a subject to facilitate diagnostic and/or therapeutic applications.

[0092] 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.

[0093] The fusion proteins, peptides, or conjugates thereof of the present disclosure may be administered by any suitable technique available in the art, e.g., as compositions. For example, they can be administered as frequently as necessary, including hourly, daily, weekly or monthly. The fusion proteins, peptides, or conjugates thereof, can be 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 fusion proteins, peptides, 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 fusion proteins, peptides, or conjugates thereof of the present disclosure, 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.

[0094] The oral agents comprising fusion proteins, peptides, or conjugates thereof of the present disclosure 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 fusion proteins, peptides, or conjugates thereof of the present disclosure 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.

[0095] 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.

[0096] "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).

[0097] A "therapeutically effective amount (or dose)" or "effective amount (or dose)" of a composition including fusion proteins, peptides, or conjugates thereof of the present disclosure,

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 fusion proteins, peptides, or conjugates thereof of the present disclosure). 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).

[0098] 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.

[0099] 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 fusion proteins, peptides, or conjugates thereof of the present disclosure described herein.

[00100] In some embodiments, the fusion proteins, peptides, or conjugates thereof of the present disclosure can further include other agents to facilitate treatment. For example, a fusion proteins, peptides, or conjugates thereof of the present disclosure 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.

[00101] 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.

[00102] Suitable metabolic modulators can include, but are not limited to, lonidamine, dichloroacetate, alpha-tocopheryl succinate, methyl jasmonate, betulinic acid, and resveratrol

[00103] 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.

[00104] In some embodiments, the fusion proteins, peptides, or conjugates thereof of the present disclosure 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.

[00105] 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.

[00106] The present invention further includes methods of administering a fusion proteins, peptides, or conjugates thereof of the present disclosure. For example, in one aspect, the present invention includes a method comprising a step of administering an effective dose of a fusion proteins, peptides, or conjugates thereof of the present disclosure or a composition including fusion proteins, peptides, or conjugates thereof of the present disclosure to a subject with a tumor such that the peptide selectively targets tumor tissue over normal tissue.

[00107] 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 fusion proteins, peptides, or conjugates thereof of the present disclosure or a composition including fusion proteins, peptides, or conjugates thereof of the present disclosure 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 fusion proteins, peptides, or conjugates thereof of the present disclosure.

[00108] The present invention also provides compositions for administering fusion proteins, peptides, or conjugates thereof of the present disclosure 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.

[00109] The fusion proteins, peptides, or conjugates thereof of the present disclosure can be administered as frequently as necessary, including hourly, daily, weekly or monthly. The fusion proteins, peptides, or conjugates thereof of the present disclosure 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 fusion proteins, peptides, or conjugates thereof of the present disclosure 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.

[00110] The oral agents comprising peptides or protein fusions 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 peptides or fusion proteins 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.

[00111] The present invention further includes functional assays of fusion proteins, peptides, or conjugates thereof of the present disclosure. The capacity of fusion proteins, peptides, or conjugates thereof of the present disclosure, 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.

[00112] One skilled in the art will be knowledgeable about animal models that are useful for measuring the *in vivo* activity of fusion proteins, peptides, or conjugates thereof of the present disclosure. 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.

[00113] The fusion proteins, peptides, and conjugates therof generated and produced by the methods and systems described herein can be used for a range of applications. For example, the proteins and peptides can be used for therapeutic and/or diagnostic purposes. Some example uses include, but are not limited to, conjugating the fusion proteins or 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.

[00114] As will be appreciated by one of ordinary skill in the art, the fusion proteins and peptides 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.

[00115] The present invention also provides compositions for administering the fusion peptides, peptides, or conujuates thereof according to the present disclosure 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.

[00116] The fusion peptides, peptides, or conujuates thereof according to the present disclosure can be administered as frequently as necessary, including hourly, daily, weekly or monthly. The fusion peptides, peptides, or conujuates thereof according to the present disclosure 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 fusion peptides, peptides, or conujuates thereof according to the present disclosure 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 fusion peptides, peptides, or

conujuates thereof according to the present disclosure 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.

[00117] The oral agents comprising a fusion peptides, peptides, or conujuates thereof according to the present disclosure 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 fusion peptides, peptides, or conujuates thereof according to the present disclosure 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.

EXEMPLARY ASPECTS

EXAMPLE 1

Expressing Peptide Constructs for Knottin Generation

[00118] This example describes a method for expressing peptide constructs in culture and greatly facilitating their development, particularly as drugs.

[00119] As shown in FIG. 5, 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. 6 shows gel data of a number of example knottins that were made according to the method described in FIG. 5. As

shown, chymotrypsin inhibitor (CTI), epiregulin (EPI), hefutoxin (HTX), bubble protein (BUB), potato carboxypeptidase inhibitor (PCI) were properly folded.

[00120] 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.

[00121] FIG. 8 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 knottin variants, and the errors in oligonucleotide synthesis can explain the truncated and extended peptide sequences.

[00122] Using the methods described in this example, variants of several knottin scaffolds were generated and analyzed. FIG. 9 shows an SDS-PAGE analysis of 3000 member knottin libraries for, e.g., hefutoxin, knottin 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.

[00123] Four scaffolds were selected for the generation of defined libraries: hefutoxin, CTI, knottin, 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 SCN 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 SCN 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.

[00124] 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.

[00125] 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 A.D., et al., *Nucleic Acids Res.* (2011) 39(21):e143). 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.

[00126] 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

Fusion Protein Systems

[00127] This example describes expression systems for the efficient production of various peptides, including knotted peptides. Advantageously, the peptides produced according to these methods are secreted and stable.

[00128] Siderocalin can be used according to the present disclosure for construction of a Scn-peptide fusion, which can be secreted and cleaved for the efficient production of excreted peptides. FIG. 24 depicts the crystal structure of Scn with a Th ligand.

[00129] FIG. 14 shows the alignment of Scn sequences from 18 species. The alignment shows a high level of sequence conservation between the 18 species with recognizable orthologs. Positions with an asterisk are possible cites for ligand binding.

[00130] FIG. 15 depicts a 3D model of super-stable Scn in accordance with an aspect of the present disclosure. In this form of Scn, a second disulfide bond was engineered in order to secure the N-terminus and increase thermal stability.

[00131] FIGS. 10-12 depict various manufacturing methods according to the present disclosure, which enable the efficient production of knottin peptides. FIG. 10 depicts the elements of the siderocalin fusion protein and use thereof for the production of knottin peptides. In this example, the knottin is secreted as a fusion protein to the siderocalin. The IgK signal peptide is derived from the mouse Light Chain IgG. The sFLAG is the short FLAG (DYKDE) for enhanced cleavage of the signal peptide relative to the long FLAG signal peptide. The His is a histidine tag (HHHHHH) and TEV is the tobacco etch virus protease site (ENLYFQ). According to this method, and as depicted in FIG 12., the fusion protein is secreted, followed by cleavage and isolation. IgK SP is Murine IgK light chain signal peptide, sFLAG: Shortened FLAG epitope, HIS: 6xhistidine tag, TEV: Tobacco Etch Virus Protease recognition site and Furin – furin cleavage site with BamHI site.

[00132] FIG. 10 depicts the elements of the siderocalin fusion protein and the use thereof for the production of knottin peptides. In this example, the knottin is secreted as a fusion protein to HIS-tagged siderocalin. The HIS tag is six histidines (HHHHHH) that will reversibly bind nickel and is therefore a useful purification tag for Immobilized Metal Chromatography (IMAC). Following isolation of the fusion protein by IMAC, the fusion is cleaved with TEV protease and the cleaved siderocalin removed by another round of IMAC, leaving pure knottin.

[00133] FIG. 11 depicts the elements of the siderocalin fusion protein and use thereof for the production of knottin peptides. This method includes the secretion of a siderocalin fusion with concomitant, intracellular cleavage at the Furin-His junction and purification by IMAC. The HIS tag can then be removed from the knottin with TEV protease. The secretion method depicted in FIG. 11 utilizes the same restriction sites as the method depicted in FIG. 10.

[00134] FIG. 12 depicts the elements of a system for the secretion of cleaved knottin peptides. The siderocalin in this case is present during translation but is cleaved off by

intracellular furin during protein. The secretion method depicted in FIG. 12 utilizes the same restriction sites as the method depicted in FIG. 10.

[00135] FIG. 16 shows a schematic of a generic Scn fusion that can be used in various aspects of the present disclosure. The native signal peptide was removed and an exogenous sFLAG and HIS tag were added to facilitate purification. These modifications are optionally present in various aspects of the disclosure.

[00136] FIG. 13 depicts exemplary knottins made using the Daedalus system and the corresponding SDS PAGE analyses in accordance with an aspect of the present disclosure. The Dadalus system is desribed in Bandaranayake A.D., et al., *Nucleic Acids Res.* (2011) 39(21):e143. SDS PAGE analyses were performed under reducing (left) and non-reducing (right) conditions. As shown in FIG. 13, a number of challenging peptides can be produced according to the presenltly described methods.

[00137] FIG. 17 depicts a schematic of a light chain antibody fusion with Scn and a corresponding SDS PAGE analysis according to an aspect of the present disclosure. The Scn fusion enables the generation of Fab fragments that have an increased molecular weight (~75kDa) and improved serum half-life. The addition of the Scn moiety also enables the delivery of an exogenous ligand through the Scn ligand binding site. Also depicted is the SDS PAGE analysis under non-reducing and reducing conditions of a construct according to one aspect of the present disclosure.

[00138] FIG. 18 shows the SDS PAGE analysis comparing the expression of a Scn fusion with a cytoplasmic enzyme, HMOX1, both before and after cleavage according to an aspect of the present dislcosure. By cleaving the fusion protein as shown in FIG. 18, it is demonstrated that the HMOX1 protein is stable even after separation from Scn. This result further suggests that, using the presently described methods, it is possible to express enzymes in a mammalian cell culture system where they are secreted out of the cell.

[00139] FIG. 19 depicts the expression of a Scn fusion with a cytoplasmic viral protein Adv2 and the corresponding SDS PAGE analysis according to one aspect of the present disclosure. By cleaving the fusion protein as shown in FIG. 19, it is demonstrated that the Adv2 protein is stable even after separation from Scn. This result further suggests that, using the presently described methods, it is possible to express difficult viral proteins in a mammalian cell culture system where they are secreted out of the cell. Additionally this fusion protein has further utility because it can be used to raise antibodies against the viral antigen in rats, mice or rabbits using the corresponding Scn ortholog.

[00140] FIG. 20 depicts the expression of a Scn fusion with an extracellular viral glycoprotein HIV gp120 and the corresponding SDS PAGE analysis according to one aspect of the present disclosure. The results in FIG. 20 demonstrate that stabilized glycoproteins can be expressed according to one aspect of the present disclosure. Additionally this fusion protein has further utility because it can be used to raise antibodies against the viral antigen in rats, mice or rabbits using the corresponding Scn ortholog.

[00141] FIG. 21 depicts the expression of a Scn fusion with a knottin protein, Imperatoxin, and the corresponding SDS PAGE analysis according to one aspect of the present disclosure. The results in FIG. 21 show that the knottin protein, Imperatoxin, is stable even after separation from Scn. These results show that knottins can be expressed in a mammalian cell culture system where they are secreted out of the cell according to an aspect of the present disclosure. Although knottins are known to be incredibly difficult to secrete in a properly folded state, the results of FIG. 21 demonstrates that it is possible to do so using the presently described methods. Additionally, this fusion protein has further utility because it can be used to raise antibodies against the viral antigen in rats, mice or rabbits using the corresponding Scn ortholog.

[00142] FIG. 22 depicts the expression of a Scn fusion with a small subdomain (i.e., Kringle domain) of the extracellular tyrosine kinase receptor ROR1 and corresponding SDS PAGE analysis according to one aspect of the present disclosure. The results in FIG. 22 show that the Kringle domain failed to express when alone, while it was successfully experssed as part of a Scn fusion. These results show that the Kringle domain of the extracellular tyrosine kinase receptor ROR1 can be efficiently prepared according to one apsect of the present dislcosure. Additionally, this fusion protein has further utility because it can be used to raise antibodies against the viral antigen in rats, mice or rabbits using the corresponding Scn ortholog.

[00143] FIG. 23 depicts the expression of a Scn fusion with heptameric and trimeric subdomains and corresponding SDS PAGE analysis according to one aspect of the present disclosure. As demonstrated in FIG. 23, the presently described methods enable the expression of multimeric protein constructs in a mammalian cell culture system, where the proteins are secreted out of the cell.

[00144] FIG. 24 depicts the expression of an ExFABP fusion with a knottin and corresponding SDS PAGE analysis according to one aspect of the present disclosure. ExFABP is another functional Scn. According to the present disclosure, this construct can be used in a periplasmic bacterial system to secret a variety of client proteins.

[00145] 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

[00146] 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**GGCGATACCTGCGGCAGCGGCTATAATGTGGATCAGCGTCGTACCA**
ATAGCGGCTGCAAAGCGGGCAATGGCGATCGTCATTTTTGCGGCTGCGATCGT
ACCGGCGTGGTGGAATGCAAAGGCGGCAAATGGACCGAAGTGCAGGATTGCG
GCAGCAGCAGCTGCAAAGGCACCAGCAATGGCGGCGCGACCTGCTAATGCTAA
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
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**GGC<u>GATCAGAATTGCGATATTGGCAATATTACCAGCCAGTGCCAGA</u>**
<u>TGCAGCATAAAAATTGCGAAGATGCGAATGGCTGCGATACCATTATTGAAGAAT</u>
<u>GCAAAACCAGCATGGTGGAACGTTGCCAGAATCAGGAATTTGAAAGCGCGGCG</u>
<u>GGCAGCACCACCCTGGGCCCGCAG</u>TAATGCTAAGGATCCCGGA

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
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*ACCGCC*

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
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**GGCGAAATTAGCTGCGAACCGGGCAAAACCTTTAAAGATAAATGCA**
ATACCTGCCGTTGCGGCGCGGATGGCAAAAGCGCGGCGTGCACCCTGAAAGCG
TGCCCGAATCAGTAATGCTAAGGATCCCGGA

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
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
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**GGCGTGTGCCGTGATTGGTTTAAAGAAACCGCGTGCCGTCATGCGA**
AAAGCCTGGGCAATTGCCGTACCAGCCAGAAATATCGTGCGAATTGCGCGAAA
ACCTGCGAACTGTGCTAATGCTAAGGATCCCGGA

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**
GCGCGGCGCTGGGCTGCAGCTGCAAAAATAAAGTGTGCTATCGTAATTAATGC</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
ATTGCGAATAT</u>**TAATGCTAAGGAT**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 -

```
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: 21 - 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: 22 - 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: 23- 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 NO: 24 -

```
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: 25 - 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 NO: 26 -

```
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: 27 -
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 NO: 28 -

```
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
```

cagggaggcgtgtgctgcggctacaagctgtgccacccctgc
Q G G V C C G Y K L C H P C

SEQ ID NO: 29 - 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: 30 -
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: 31 - 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: 32 -

```
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
```

Anti-CD3-STa bispecific
The anti-CD3 is an OKT3 variant from the C-terminus of patent 7635462. It is supposed to work as either N- or C- terminal.

SEQ ID NO: 33-
GSDIKLQQSGAELARPGASVKMSCKTSGYTFTRYTMHWVKQRPGQGLEWIGYINPSRG
YTNYNQKFKDKATLTTDKSSSTAYMQLSSLTSEDSAVYYCARYYDDHYCLDYWGQGT
TLTVSSVEGGSGGSGGSGGSGGVDDIQLTQSPAIMSASPGEKVTMTCRASSSVSYMNWY
QQKSGTSPKRWIYDTSKVASGVPYRFSGSGSGTSYSLTISSMEAEDAATYYCQQWSSNP
LTFGAGTKLELKGGGGSNSSNYCCELCCNPACTGCY

This is designed to be dropped into the Bam/Not cut library vector as a TEV-cleavable siderocalin fusion:

SEQ ID NO: 34 -
ACCTGTATTTTCAGGGATCCgatattaaactgcagcagagcggcgcggaactggcgcgcccgggcgcgagcgtgaaa
atgagctgcaaaaccagcggctatacctttacccgctataccatgcattgggtgaaacagcgcccgggccagggcctggaatggattggct
atattaacccgagccgcggctataccaactataaccagaaatttaaagataaagcgaccctgaccaccgataaaagcagcagcaccgcgta
tatgcagctgagcagcctgaccagcgaagatagcgcggtgtattattgcgcgcgctattatgatgatcattattgcctggattattggggccag
ggcaccaccctgaccgtgagcagcgtggaaggcggcagcggcggcagcggcggcagcggcggcgtggatgatattca
gctgacccagagcccggcgattatgagcgcgagcccgggcgaaaaagtgaccatgacctgccgcgcgagcagcagcgtgagctatatg
aactggtatcagcagaaaagcggcaccagcccgaaacgctggatttatgataccagcaaagtggcgagcggcgtgccgtatcgctttagc
ggcagcggcagcggcaccagctatagcctgaccattagcagcatggaagcggaagatgcggcgacctattattgccagcagtggagcag
caacccgctgacctttggcgcgggcaccaaactggaactgaaaggcggcggcggcagcaacagcagcaactattgctgcgaactgtgct
gcaacccggcgtgcaccggctgctatTAATGCGGCCGCTCATCACCATTAATC

Parental cloning construct 1 for downstream fusions:

IgK-sFLAG-H6-GGS-**humanScn**-GGS-ENLYFQ-GG-NotI
SEQ ID NO: 35 -
CTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGCTGCTCTGGGTTCCA
GGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACCATGGTGGAAGC**CA**
GGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAGGTCCCTCTGCA
GCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTAGGCCTGG
CAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGCCACC
ATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAGG
AAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCAGCCCGG
CGAGTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGT
CCGAGTGGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGT
TTCTCAAAACAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGC
TGACTTCGGAACTAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCC
CTGAAAACCACATCGTCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAG
GTAGCGAAAACCTGTATTTTCAGGGAGGCGGCCGC

Parental cloning construct 2 for downstream fusions:

IgK-sFLAG-H6-GGS-**humanScnC87S**-GGS-ENLYFQ-GS-STUFFER
SEQ ID NO: 36 -
CTCGAGACCATGGAGACAGACACACTCCTGCTATGGGTACTGCTGCTCTGGGTTCCA
GGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACCATGGTGGAAGC**CA**
GGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAGGTCCCTCTGCA
GCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTAGGCCTGG
CAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTATGCCACC
ATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCTGTTTAGG
AAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTCCCAGCCCGGC
GAGTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTTACCTCGTC
CGAGTGGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCAAGAAAGTT
TCTCAAAACAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCAAGGAGCT
GACTTCGGAACTAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGGGCCTCCC
TGAAAACCACATCGTCTTCCCTGTCCCAATCGACCAGTGTATCGACGGCGGAGG
TAGCGAAAACCTGTATTTTCAGGGATCCATGTACGGTCTTAAGGGACCCGACATTTA
CAAAGGAGTTTACCAATTTAAGTCAGTGGAGTTTGATATGTCACATCTGAACCTGAC
CATGCCCAACGCATGTTCAGCCAACAACTCCCACCATTACATCAGTATGGGGACTTC
TGGACTAGAATTGACCTTCACCAATGATTCCATCATCAGTCACAACTTTTGCAATCTG
ACCTCTGCCTTCAACAAAAAGACCTTTGACCACACACTCATGAGTATAGTTTCGAGC
CTACACCTCAGTATCAGAGGGAACTCCAACTATAAGGCAGTATCCTGCGACTTCAAC
AATGGCATAACCATCCAATACAACTTGACATTCTCAGATCGACAAAGTGCTCAGAGC
CAGTGTAGAACCTTCAGAGGTAGAGTCCTAGATATGTTTAGAACTGCCTTCGGGGGG
AAATACATGAGGAGTGGCTGGGGCTGGACAGGCTCAGATGGCAAGACCACCTGGTG
TAGCCAGACGAGTTACCAATACCTGATTATACAAAATAGAACCTGGGAAAACCACT
GCACATATGCAGGTCCTTTTGGGATGTCCAGGATTCTCCTTTCCCAAGAGAAGACTA
AGTTCTTCACTAGGAGACTGGTGCCCAGGGGCAGCGGCCTGAACGACATCTTCGAGG
CCCAGAAGATCGAGTGGCACGAGTAATGCGGCCGCTCATCACCATTAATCATCACCA
TTAATCGGACCG

Parental cloning construct 3 for light chain fusions:
LightChain-GRGGSGGS-**humanScnC87S**

SEQ ID NO: 37 -
CTCGAGACCATGGATTTCCAGGTGCAGATTTTTAGCTTTCTGCTGATTTCCGCTTCCG
TGATTATGAGCCGAGGCGACATTGTGATGACCCAGGCAGCTCCTAGCGTGCCAGTCA
CCCCAGGAGAGTCAGTGAGCATCTCCTGCAGAAGTACTAAGTCACTGCTGCACAGCA
ACGGCAATACCTACCTGTATTGGTTCCTGCAGAGACCTGGGCAGTCCCCACAGAGGC
TGATCTACTATATGAGTAACCTGGCATCAGGAGTGCCTGACAGGTTCAGCGGACGAG
GCAGCGGCACTGATTTTACCCTGCGGATTTCTAGAGTGGAGGCAGAAGACGCCGGC

GTCTACTATTGCATGCAGAGTCTGGAGTACCCTTATACTTTCGGCGGGGGAACCAAA
CTGGAAATCAAGAGGGCCGATGCCGCTCCAACCGTGTCCATTTTTCCCCCTAGCTCC
GAGCAGCTGACATCTGGCGGGGCTAGTGTGGTCTGTTTCCTGAACAATTTTTACCCA
AAGGACATCAACGTGAAATGGAAGATTGATGGAAGTGAAAGGCAGAACGGCGTCCT
GAATTCATGGACAGACCAGGATAGCAAAGACTCCACTTATTCTATGTCTAGTACCCT
GACACTGACTAAGGATGAGTACGAACGCCACAATTCTTATACATGCGAGGCAACTC
ATAAAACCTCTACAAGTCCCATCGTGAAGAGCTTTAACCGAAATGAATGCGGCCGCG
GAGGCTCCGGAGGCTCC**CAGGACTCAACAAGCGATCTGATTCCAGCCCCACCCC**
TGAGCAAAGTGCCCCTGCAGCAGAACTTCCAGGACAATCAGTTTCAGGGCAAG
TGGTACGTGGTCGGGCTGGCTGGAAACGCAATCCTGCGGGAGGACAAAGATCC
CCAGAAGATGTACGCCACTATCTACGAGCTGAAAGAAGACAAGTCATACAATGT
GACCAGCGTCCTGTTCCGCAAGAAAAAGTGTGATTATTGGATCAGAACATTCGT
GCCCGGCTCCCAGCCTGGGGAGTTTACTCTGGGGAATATTAAGTCCTACCCTG
GACTGACCTCTTATCTGGTGCGAGTGGTCTCTACAAACTACAATCAGCATGCTA
TGGTGTTCTTTAAAAAGGTCAGCCAGAACCGGGAGTACTTTAAAATCACCCTGT
ATGGCAGAACCAAAGAACTGACAAGCGAGCTGAAGGAAAATTTCATTCGCTTTT
CCAAGTCTCTGGGGCTGCCAGAGAATCATATTGTGTTCCCAGTCCCCATTGACC
AGTGTATTGACGGGTGAGGATCC

Super stable ScnC87S
humanScn-I8C-N39C-C87S

SEQ ID NO: 38 -
CTCGAG**ATGCCCCTGGGCCTGCTGTGGCTGGGCCTGGCCCTGCTGGGCGCCCT**
GCACGCCCAGGCCCAGGACTCCACCTCAGACCTGTGTCCAGCCCCACCTCTGA
GCAAGGTCCCTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGG
TATGTGGTAGGCCTGGCAGGGTGTGCAATTCTCAGAGAAGACAAAGACCCGCA
AAAGATGTATGCCACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCAC
CTCCGTCCTGTTTAGGAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCC
AGGTTCCCAGCCGGGCGAGTTCACGCTGGGCAACATTAAGAGTTACCCTGGAT
TAACGAGTTACCTCGTCCGAGTGGTGAGCACCAACTACAACCAGCATGCTATG
GTGTTCTTCAAGAAAGTTTCTCAAAACAGGGAGTACTTCAAGATCACCCTCTAC
GGGAGAACCAAGGAGCTGACTTCGGAACTAAAGGAGAACTTCATCCGCTTCTC
CAAATCTCTGGGCCTCCCTGAAAACCACATCGTCTTCCCTGTCCCAATCGACCA
GTGTATCGACGGCGGATCC

SEQ ID NO: 39 -
Homosapiens
QDSTSDLIPAPPLSKVPLQQNFQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELK
EDKSYNVTSVLFRKKKCDYWIRTFVPGCQPGEFTLGNIKSYPGLTSYLVRVVSTNYNQH
AMVFFKKVSQNREYFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDG

SEQ ID NO: 40 -
Pantroglodytes
QDSTSDLIPAPPLSKVPLQQNFQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELK
EDKSYNVTSVLFRKKKCDYWIRTFVPGRQPGEFTLGNIKSYPGLTSYLVRVVSTNYNQH
AMVFFKKVSQNREYFKITLYGRTKELTSELQENFIRFSKSLGLPENHIVFPVPIDQCIDG

SEQ ID NO: 41 -
Canisfamiliaris
QDSTPSLIPAPPPLKVPLQPDFQHDQFQGKWYVIGIAGNILKKEGHGQLKMYTTTYELKD
DQSYNVTSTLLRNERCDYWNRDFVPSFQPGQFSLGDIQLYPGVQSYLVQVVATNYNQY
ALVYFRKVYKSQEYFKITLYGRTKELPLELKKEFIRFAKSIGLTEDHIIFPVPIDQCIDE

SEQ ID NO: 42 -
Bostaurus
RSSSSRLLRAPPLSRIPLQPNFQADQFQGKWYTVGVAGNAIKKEEQDPLKMYSSNYELK
EDGSYNVTSILLKDDLCDYWIRTFVPSSQPGQFTLGNIKSYRGIRSYTVRVVNTDYNQFA
IVYFKKVQRKKTYFKITLYGRTKELTPEVRENFINFAKSLGLTDDHIVFTVPIDRCIDDQ

SEQ ID NO: 43 -
Musmusculus
QDSTQNLIPAPSLLTVPLQPDFRSDQFRGRWYVVGLAGNAVQKKTEGSFTMYSTIYELQ
ENNSYNVTSILVRDQDQGCRYWIRTFVPSSRAGQFTLGNMHRYPQVQSYNVQVATTDY
NQFAMVFFRKTSENKQYFKITLYGRTKELSPELKERFTRFAKSLGLKDDNIIFSVPTDQCI
DN

SEQ ID NO: 44 -
Rattusnorvegicus
QDSTQNLIPAPPLISVPLQPGFWTERFQGRWFVVGLAANAVQKERQSRFTMYSTIYELQE
DNSYNVTSILVRGQGCRYWIRTFVPSSRPGQFTLGNIHSYPQIQSYDVQVADTDYDQFA
MVFFQKTSENKQYFKVTLYGRTKGLSDELKERFVSFAKSLGLKDNNIVFSVPTDQCIDN

SEQ ID NO: 45 -
Macacamulatta
QDSSSDLIPAPPLSKVPLQQNFQDNQFQGKWYVVGLSGNAVGRKDEAPLKMYATIYEL
KEDKSYNVTSILFRKEKCDYWIRTFVPGSQPGEFTLGNIQNHPGLTSYVVRVVSTNYKQ
YAMVFFKKVSQNKEYFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFSVPIDQCING

SEQ ID NO: 46 -
Tursiopstruncatus
QDSTPNLIPAPPLFRVPLQPNFQPDQFQGKWYIVGLAGNAFKKEKQGQFKMYATTYELK
EDRSYNVTSALLRGKTQRCDHWIRTFVPSSRPGQFTLGNIKGFPGVQSYTVRVATTNYN
QFAIVYFKKVYKNQEYFKTTLYGRTKELTPQLKENFIHFAKSLGLTDEYILFPVPIDKCID
DQ

SEQ ID NO: 47 -
Gorillagorilla
QDSTSDLIPAPPLSKVPLQQNFQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELK
EDKSYNVTSVLFREKAQKCDYWIRTFVPGSQPGEFTLGNIKSYPGLTSYLVRVVSTNYN
QHAMVFFKKVSQNREYFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCID
G

SEQ ID NO: 48 -
Procaviacapensis
QEPTPTLIPAPPLSSIPLKPNFHNDKFQGKWYVVGVAGNAITKEKDPSLMYTTTYELRDD
GSYNVTSTQFREKINCTHWTRTFVPTSQPGQFSLGNIDKYPHLSSYTVRVTATNYNYFAI
VYFKKVSKNQEYFKTTLYKRIKKLTHGLKKHFIQFAKSLGLPDNHITFLVPTDRCIDDA

SEQ ID NO: 49 -
Callithrixjacchus
QDSPSPLIPAPPLSKVPLQQNFQDNQFQGKWYVVGLAGNAIRREDQDSLKMYATIYELK
EDKSYNVTSVLFRKAQKCDYWIRTFVPSSRPGEFKLGNIESHPGLTSYIVRVVNTDYKQH
AMVFFMKASHNRKYFKVTLYGRTKELTSDLKENFTSFSKSLGLTENHIIFPVPIDQCIDG

SEQ ID NO: 50 -
Microcebusmurinus
QDSKEKLIPAPPLLRVPLQPDFQDDQFRETSWPRGSKMKETPAGSRDAGTGWATTYELK
DHSYNVTSTLLRQNGKCDYWIRTFVLTSQPGQFALGNINRYPGIQSYTVRVVTTNYNQF
AIVFFKKVSENKEYFKTTLYGRTKELPPELKENFIRFAKSLGLTEDHIIYPVPIDQCIDD

SEQ ID NO: 51 -
QDSTSDLIPAPPLSKVPLQQNFQDNQFQGKWYVVGLAGNAIRREDKDSQKMYATIYELK
EDKSYNVTSVLFRKKKCDYWIRTFVPGSQPGEFTLGNTKGYPGLTSYLVRVVSTNYNQY
AMVFFKKVSQNREYFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPAPIDQCIDG

SEQ ID NO: 52 -
Ochotonaprinceps

QELTTDLIPVPSLRKIHVQKNFQSDQFQGKWYVVGLAGNNIHNSDQEHQQMYSTTYELK
EDGSYNVTSTLLRQRNQQCDHWIRTFVPGSKLGHFNLGNIKSYPTLKSYLIRVVTTDYN
QFAIVFFRKVYKNNKKFFKIVLYGRTKELSPELRGRFTSFAKTLGLTDNHIVFPAPIGQCI
DD

SEQ ID NO: 53 -
Loxodontaafricana
QTHSPTLIPAPPLLRVPLQPDFQDDKFQGKWYVIGLAGNAVEKKEQGQFKMYTTTYELK
EDGSYNVTSTLLQEDGKCSYWIRTFVPSFQPGQFNLGNIKNFPGLQSYTVRVTATNYNQ
FAIVFFKKVSKNGEYFKTTLYGRTKELTPELKERFIRFAKSLGLSDHIIFPVPIDRCIDD

SEQ ID NO: 54 -
Oryctolaguscuniculus
QDPTPKLIPAPSLRRVPLQRNFQDEQFQGKWYVVGLAGNAVQKREEGQEPMYSTTYEL
NEDRSFNVTSTLLRDQRCDHWIRTFVPTSRPGQYNLGNIKSYPGVKNYIVRVVATDYSQ
YAMMFFRKGSRNKQFFKTTLYGRTKELSPELRERFTRFAKSLGLPDDRIVFPTPIDQCIDD

SEQ ID NO: 55 -
Murine Scn construct for downstream viral fusions:
IgK-H6-**murineScn**-StrepII-GGGGS-E7.16
gi|29468134|gb|AAO85409.1|AF402678_2 E7 [HPV16]
CTCGAGACC**ATGGAGACCGACACGCTCTTGTTGTGGGTTCTCTTGTTGTGGGTG**
CCTGGGTCTACAGGCGACCACCACCATCATCACCACCTCGTTCCTAGAGGCAG
CCAGGATAGTACCCAGAATCTTATCCCAGCACCATCTTTGCTCACAGTACCATT
GCAACCCGACTTTCGGTCTGATCAATTTCGGGGACGCTGGTACGTGGTTGGAC
TGGCCGGCAATGCTGTACAGAAAAAAACAGAGGGCAGTTTCACCATGTACTCA
ACAATCTATGAGCTCCAAGAGAATAATAGTTACAACGTTACCTCCATCTTGGTG
AGGGACCAGGATCAGGGATGTCGCTACTGGATTCGGACATTCGTACCAAGTTC
TCGGGCCGGTCAGTTTACTCTGGGCAACATGCACAGGTATCCCCAAGTTCAATC
TTACAACGTGCAGGTGGCGACTACCGACTACAACCAATTCGCTATGGTGTTCTT
CCGCAAAACAAGCGAGAACAAGCAGTATTTTAAAATCACTCTGTACGGTAGAAC
TAAGGAGCTGAGCCCTGAACTTAAGGAGCGGTTCACCAGATTCGCTAAGTCCC
TGGGACTGAAGGATGATAATATAATCTTTTCCGTCCCCACCGATCAGTGTATCG
ATAATTCAGCTTGGTCACATCCCCAGTTCGAGAAAGGAGGCGGTGGATCCATGCATG
GCGACACTCCGACCCTGCACGAATACATGCTGGACCTGCAGCCCGAAACCACTGAC
CTGTATTGTTACGAGCAACTCAACGATTCTAGCGAGGAGGAGGACGAGATCGATGG
ACCGGCAGGCCAGGCCGAACCAGACCGCGCACATTATAACATTGTTACCTTCTGTTG
CAAATGTGATTCAACTCTTAGACTTTGTGTCCAGAGTACACACGTGGACATCCGCAC
CCTGGAAGATCTGCTGATGGGAACTCTGGGTATCGTGTGTCCTATATGTAGCCAGAA
ACCCTGACGGACCG

SEQ ID NO: 56 -
METDTLLLWVLLLWVPGSTGDHHHHHHLVPRGSQDSTQNLIPAPSLLTVPLQPDF
RSDQFRGRWYVVGLAGNAVQKKTEGSFTMYSTIYELQENNSYNVTSILVRDQDQG
CRYWIRTFVPSSRAGQFTLGNMHRYPQVQSYNVQVATTDYNQFAMVFFRKTSENK
QYFKITLYGRTKELSPELKERFTRFAKSLGLKDDNIIFSVPTDQCIDNSAWSHPQFEK
GGGGSMHGDTPTLHEYMLDLQPETTDLYCYEQLNDSSEEEDEIDGPAGQAEPDRAHYNI
VTFCCKCDSTLRLCVQSTHVDIRTLEDLLMGTLGIVCPICSQKP

IgK-H6-**murineScn**-StrepII-GGGGS-E6.16
gi|4927720|gb|AAD33252.1|AF125673_1 E6 [HPV16]
SEQ ID NO: 57 -
CTCGAGACC**ATGGAGACCGACACGCTCTTGTTGTGGGTTCTCTTGTTGTGGGTG**
CCTGGGTCTACAGGCGACCACCACCATCATCACCACCTCGTTCCTAGAGGCAG
CCAGGATAGTACCCAGAATCTTATCCCAGCACCATCTTTGCTCACAGTACCATT
GCAACCCGACTTTCGGTCTGATCAATTTCGGGGACGCTGGTACGTGGTTGGAC
TGGCCGGCAATGCTGTACAGAAAAAAACAGAGGGCAGTTTCACCATGTACTCA
ACAATCTATGAGCTCCAAGAGAATAATAGTTACAACGTTACCTCCATCTTGGTG
AGGGACCAGGATCAGGGATGTCGCTACTGGATTCGGACATTCGTACCAAGTTC

TCGGGCCGGTCAGTTTACTCTGGGCAACATGCACAGGTATCCCCAAGTTCAATC
TTACAACGTGCAGGTGGCGACTACCGACTACAACCAATTCGCTATGGTGTTCTT
CCGCAAAACAAGCGAGAACAAGCAGTATTTTAAAATCACTCTGTACGGTAGAAC
TAAGGAGCTGAGCCCTGAACTTAAGGAGCGGTTCACCAGATTCGCTAAGTCCC
TGGGACTGAAGGATGATAATATAATCTTTTCCGTCCCCACCGATCAGTGTATCG
ATAATTCAGCTTGGTCACATCCCCAGTTCGAGAAAGGAGGCGGTGGATCCATGCACC
AGAAGAGAACCGCCATGTTCCAGGAcCCaCAAGAGCGGCCCCGGAAACTGCCCCAAC
TGTGCACTGAATTGCAGACCACCATCCACGACATCATTTTGGAATGTGTCTACTGTA
AGCAGCAGCTCCTCAGGCGAGAGGTGTATGACTTCGCCTTCCGGGATTTGTGTATTG
TCTACAGGGATGGTAATCCCTATGCCGTTTGTGATAAGTGCCTGAAATTTTATAGCA
AGATCAGCGAGTACCGACATTACTGTTACAGCGTTTATGGAACAACATTGGAGCAGC
AGTACAACAAACCTCTTTGCGACCTCCTGATTCGCTGCATCAACTGCCAGAAGCCCC
TGTGCCCCGAAGAGAAACAAAGGCATTTGGATAAGAAGCAGAGGTTCCACAACATC
CGCGGTCGCTGGACGGGGCGCTGCATGAGTTGCTGCAGGAGTTCCCGCACTCGGCGC
GAGACCCAACTCTGACGGACCGCCTCTCCCTCCC

IgK-H6-**murineScn**-StrepII-GGGGS-E6.18
gi|30172005|gb|AAP20594.1| E6 protein [HPV18]
SEQ ID NO: 58 -
CTCGAGACC**ATGGAGACCGACACGCTCTTGTTGTGGGTTCTCTTGTTGTGGGTG**
CCTGGGTCTACAGGCGACCACCACCATCATCACCACCTCGTTCCTAGAGGCAG
CCAGGATAGTACCCAGAATCTTATCCCAGCACCATCTTTGCTCACAGTACCATT
GCAACCCGACTTTCGGTCTGATCAATTTCGGGGACGCTGGTACGTGGTTGGAC
TGGCCGGCAATGCTGTACAGAAAAAAACAGAGGGCAGTTTCACCATGTACTCA
ACAATCTATGAGCTCCAAGAGAATAATAGTTACAACGTTACCTCCATCTTGGTG
AGGGACCAGGATCAGGGATGTCGCTACTGGATTCGGACATTCGTACCAAGTTC
TCGGGCCGGTCAGTTTACTCTGGGCAACATGCACAGGTATCCCCAAGTTCAATC
TTACAACGTGCAGGTGGCGACTACCGACTACAACCAATTCGCTATGGTGTTCTT
CCGCAAAACAAGCGAGAACAAGCAGTATTTTAAAATCACTCTGTACGGTAGAAC
TAAGGAGCTGAGCCCTGAACTTAAGGAGCGGTTCACCAGATTCGCTAAGTCCC
TGGGACTGAAGGATGATAATATAATCTTTTCCGTCCCCACCGATCAGTGTATCG
ATAATTCAGCTTGGTCACATCCCCAGTTCGAGAAAGGAGGCGGTGGATCCATGGCC
AGATTTGAAGACCCCACAAGGCGCCCCTATAAACTGCCGGATCTTTGCACCGAACTG
AATACTAGCCTGCAAGATATTGAGATTACCTGCGTGTACTGTAAAACGGTGCTCGAA
TTGACCGAGGTTTTTGAGTTCGCATTCAAGGACCTGTTTGTTGTATATCGCGATTCCA
TCCCGCACGCAGCTTGCCATAAATGCATTGACTTTTACTCCCGGATACGCGAGCTGC
GACACTATAGTGATAGCGTGTACGGCGATACACTTGAGAAGCTTACCAACACCGGTC
TGTACAATCTTCTGATTCGGTGTTTGAGGTGCCAGAAGCCGCTCAACCCAGCTGAGA
AACTGCGGCATCTGAACGAAAAAAGAAGATTCCACAACATTGCTGGCCACTACAGG
GGCCAGTGCCATTCTTGTTGTAATAGAGCAAGGCAGGAGCGGCTGCAACGGCGGCG
CGAGACCCAGGTATGACGGACCGCCTCTCCCTCCC

IgK-H6-**murineScn**-StrepII-GGGGS-E7.18
gi|285804409|gb|ADC35717.1| E7 [HPV 18]
SEQ ID NO: 59 -
CTCGAGACC**ATGGAGACCGACACGCTCTTGTTGTGGGTTCTCTTGTTGTGGGTG**
CCTGGGTCTACAGGCGACCACCACCATCATCACCACCTCGTTCCTAGAGGCAG
CCAGGATAGTACCCAGAATCTTATCCCAGCACCATCTTTGCTCACAGTACCATT
GCAACCCGACTTTCGGTCTGATCAATTTCGGGGACGCTGGTACGTGGTTGGAC
TGGCCGGCAATGCTGTACAGAAAAAAACAGAGGGCAGTTTCACCATGTACTCA
ACAATCTATGAGCTCCAAGAGAATAATAGTTACAACGTTACCTCCATCTTGGTG
AGGGACCAGGATCAGGGATGTCGCTACTGGATTCGGACATTCGTACCAAGTTC
TCGGGCCGGTCAGTTTACTCTGGGCAACATGCACAGGTATCCCCAAGTTCAATC
TTACAACGTGCAGGTGGCGACTACCGACTACAACCAATTCGCTATGGTGTTCTT
CCGCAAAACAAGCGAGAACAAGCAGTATTTTAAAATCACTCTGTACGGTAGAAC
TAAGGAGCTGAGCCCTGAACTTAAGGAGCGGTTCACCAGATTCGCTAAGTCCC
TGGGACTGAAGGATGATAATATAATCTTTTCCGTCCCCACCGATCAGTGTATCG
ATAATTCAGCTTGGTCACATCCCCAGTTCGAGAAAGGAGGCGGTGGATCCATGCAC
GGACCTAAAGCAACACTCCAGGACATCGTCCTGCATTTGGAACCACAAAACGAAAT

ACCCGTGGACCTTTTGTGTCACGAACAGCTTTCAGATTCTGAGGAAGAGAATGATGA
AATCGACGGTGTCAACCACCAGCATCTCCCCGCTAGGCGGGCAGAACCCCAGCGCC
ACACAATGCTGTGCATGTGTTGCAAATGCGAAGCTCGAATTGAACTCGTGGTTGAGT
CCTCCGCGGACGACTTGAGGGCATTCCAGCAACTGTTCCTCAACACACTGAGCTTTG
TCTGTCCTTGGTGCGCTAGTCAGCAGTGACGGACCGCCTCTCCCTCCC

IgK-H6-**murineScn**-StrepII-GGGGS-E6.33
gi|218931423|gb|ACL12326.1| E6 [HPV 33]
SEQ ID NO: 60 -
CTCGAGACC**ATGGAGACCGACACGCTCTTGTTGTGGGTTCTCTTGTTGTGGGTG**
CCTGGGTCTACAGGCGACCACCACCATCATCACCACCTCGTTCCTAGAGGCAG
CCAGGATAGTACCCAGAATCTTATCCCAGCACCATCTTTGCTCACAGTACCATT
GCAACCCGACTTTCGGTCTGATCAATTTCGGGGACGCTGGTACGTGGTTGGAC
TGGCCGGCAATGCTGTACAGAAAAAAACAGAGGGCAGTTTCACCATGTACTCA
ACAATCTATGAGCTCCAAGAGAATAATAGTTACAACGTTACCTCCATCTTGGTG
AGGGACCAGGATCAGGGATGTCGCTACTGGATTCGGACATTCGTACCAAGTTC
TCGGGCCGGTCAGTTTACTCTGGGCAACATGCACAGGTATCCCCAAGTTCAATC
TTACAACGTGCAGGTGGCGACTACCGACTACAACCAATTCGCTATGGTGTTCTT
CCGCAAAACAAGCGAGAACAAGCAGTATTTTAAAATCACTCTGTACGGTAGAAC
TAAGGAGCTGAGCCCTGAACTTAAGGAGCGGTTCACCAGATTCGCTAAGTCCC
TGGGACTGAAGGATGATAATATAATCTTTTCCGTCCCCACCGATCAGTGTATCG
ATAATTCAGCTTGGTCACATCCCCAGTTCGAGAAAGGAGGCGGTGGATCCATGTTCC
AAGACACTGAGGAGAAGCCACGCACGCTGCACGATCTGTGCCAGGCCCTTGAGACT
ACCATCCATAACATCGAGCTCCAGTGTGTCGAATGCAGGAATCCTCTTCAGCGGAGC
GAGGTGTACGATTTTGCCTTCGCGGACCTGACGGTGGTCTACCGGGAAGGTAACCCA
TTCGGGATTTGCAAGCTGTGTCTCAGATTTCTTAGTAAGATAAGTGAATACCGGCAC
TACAACTATTCAGTTTACGGTCACACTCTGGAACAGACCGTGAACAAACCCCTGAAC
GAGATCCTCATTCGATGTATCATCTGTCAGAGACCTCTCTGTCCGCGCGAAAAGAAG
AGGCACGTCGACCTGAATAAGCGATTTCATAATATCTCTGGACGGTGGGCGGGGCGC
TGTGCAGCCTGTTGGAGATCCCGGAGACGGGAAACAGCTCTTTGACGGACCGCCTCT
CCCTCCC

IgK-H6-**murineScn**-StrepII-GGGGS-E7.33
gi|218931424|gb|ACL12327.1| E7 [HPV 33]
SEQ ID NO: 61 -
CTCGAGACC**ATGGAGACCGACACGCTCTTGTTGTGGGTTCTCTTGTTGTGGGTG**
CCTGGGTCTACAGGCGACCACCACCATCATCACCACCTCGTTCCTAGAGGCAG
CCAGGATAGTACCCAGAATCTTATCCCAGCACCATCTTTGCTCACAGTACCATT
GCAACCCGACTTTCGGTCTGATCAATTTCGGGGACGCTGGTACGTGGTTGGAC
TGGCCGGCAATGCTGTACAGAAAAAAACAGAGGGCAGTTTCACCATGTACTCA
ACAATCTATGAGCTCCAAGAGAATAATAGTTACAACGTTACCTCCATCTTGGTG
AGGGACCAGGATCAGGGATGTCGCTACTGGATTCGGACATTCGTACCAAGTTC
TCGGGCCGGTCAGTTTACTCTGGGCAACATGCACAGGTATCCCCAAGTTCAATC
TTACAACGTGCAGGTGGCGACTACCGACTACAACCAATTCGCTATGGTGTTCTT
CCGCAAAACAAGCGAGAACAAGCAGTATTTTAAAATCACTCTGTACGGTAGAAC
TAAGGAGCTGAGCCCTGAACTTAAGGAGCGGTTCACCAGATTCGCTAAGTCCC
TGGGACTGAAGGATGATAATATAATCTTTTCCGTCCCCACCGATCAGTGTATCG
ATAATTCAGCTTGGTCACATCCCCAGTTCGAGAAAGGAGGCGGTGGATCCATGCGG
GGACATGAACCTACTCTGAAGGAGTACGTCCTGGACCTTTACCCGGAGCCGACAGAT
CTTTACTGTTACGAGCAATTGTCTGACTCCAGCGACGAGGATGAGGGCCTTGACAGA
CCTGATGGCCAGGCTCAGCCAGCTACTGCCGATTATTATATCGTTACGTGTTGTCACA
CCTGCAACACAACCGTAAGGTTGTGTGTGAACTCCACCGCCAGTGACTTGAGAACGA
TACAACAACTCCTCATGGGCACTGTCAATATCGTCTGTCCTACATGTGCTCAGCTGCT
GACGGACCGCCTCTCCCTCCC

IgK-H6-**murineScn**-StrepII-GGGGS-E6.45
gi|145968371|gb|ABP99896.1| E6 [HPV45]
SEQ ID NO: 62 -
CTCGAGACC**ATGGAGACCGACACGCTCTTGTTGTGGGTTCTCTTGTTGTGGGTG**

CCTGGGTCTACAGGCGACCACCACCATCATCACCACCTCGTTCCTAGAGGCAG
CCAGGATAGTACCCAGAATCTTATCCCAGCACCATCTTTGCTCACAGTACCATT
GCAACCCGACTTTCGGTCTGATCAATTTCGGGGACGCTGGTACGTGGTTGGAC
TGGCCGGCAATGCTGTACAGAAAAAAACAGAGGGCAGTTTCACCATGTACTCA
ACAATCTATGAGCTCCAAGAGAATAATAGTTACAACGTTACCTCCATCTTGGTG
AGGGACCAGGATCAGGGATGTCGCTACTGGATTCGGACATTCGTACCAAGTTC
TCGGGCCGGTCAGTTTACTCTGGGCAACATGCACAGGTATCCCCAAGTTCAATC
TTACAACGTGCAGGTGGCGACTACCGACTACAACCAATTCGCTATGGTGTTCTT
CCGCAAAACAAGCGAGAACAAGCAGTATTTTAAAATCACTCTGTACGGTAGAAC
TAAGGAGCTGAGCCCTGAACTTAAGGAGCGGTTCACCAGATTCGCTAAGTCCC
TGGGACTGAAGGATGATAATATAATCTTTTCCGTCCCCACCGATCAGTGTATCG
ATAATTCAGCTTGGTCACATCCCCAGTTCGAGAAAGGAGGCGGTGGATCCATGGCC
AGGTTCGATGATCCCACCCAGCGACCCTATAAGTTGCCCGATCTCTGCACAGAACTT
AACACTAGCTTGCAGGACGTAAGCATTGCATGTGTTTACTGTAAAGCTACGCTGGAG
CGAACCGAGGTGTACCAATTCGCCTTCAAAGACTTGTTCATCGTGTATAGAGACTGT
ATCGCTTATGCCGCCTGCCACAAATGCATAGACTTTTACAGCAGGATCAGGGAATTG
AGGTACTATTCCAACTCAGTCTATGGAGAAACGCTGGAGAAGATAACTAACACTGA
GCTTTATAACCTCCTGATTCGCTGCCTCCGGTGTCAGAAGCCACTGAATCCTGCCGA
AAAGAGACGCCATCTGAAGGACAAGCGGCGCTTTCATAGCATTGCAGGACAGTACA
GAGGCCAATGTAATACTTGCTGTGACCAAGCACGCCAAGAAAGGCTCAGGAGAAGG
AGAGAGACACAGGTGTGACGGACCGCCTCTCCCTCCC

IgK-H6-**murineScn**-StrepII-GGGGS-E7.45
gi|145968372|gb|ABP99897.1| E7 [HPV 45]
SEQ ID NO: 63 -
CTCGAGACC**ATGGAGACCGACACGCTCTTGTTGTGGGTTCTCTTGTTGTGGGTG**
CCTGGGTCTACAGGCGACCACCACCATCATCACCACCTCGTTCCTAGAGGCAG
CCAGGATAGTACCCAGAATCTTATCCCAGCACCATCTTTGCTCACAGTACCATT
GCAACCCGACTTTCGGTCTGATCAATTTCGGGGACGCTGGTACGTGGTTGGAC
TGGCCGGCAATGCTGTACAGAAAAAAACAGAGGGCAGTTTCACCATGTACTCA
ACAATCTATGAGCTCCAAGAGAATAATAGTTACAACGTTACCTCCATCTTGGTG
AGGGACCAGGATCAGGGATGTCGCTACTGGATTCGGACATTCGTACCAAGTTC
TCGGGCCGGTCAGTTTACTCTGGGCAACATGCACAGGTATCCCCAAGTTCAATC
TTACAACGTGCAGGTGGCGACTACCGACTACAACCAATTCGCTATGGTGTTCTT
CCGCAAAACAAGCGAGAACAAGCAGTATTTTAAAATCACTCTGTACGGTAGAAC
TAAGGAGCTGAGCCCTGAACTTAAGGAGCGGTTCACCAGATTCGCTAAGTCCC
TGGGACTGAAGGATGATAATATAATCTTTTCCGTCCCCACCGATCAGTGTATCG
ATAATTCAGCTTGGTCACATCCCCAGTTCGAGAAAGGAGGCGGTGGATCCATGCAC
GGCCCACAGGCAACCCTGCAAGAGATCGTGCTGCATCTCGAACCACAGAATGAATT
GGACCCTGTGGATCTGCTGTGTTACGAGCAGCTCTCTGAAAGCGAAGAGGAGAATG
ACGAGGCCGACGGCGTGTCTCATGCACAGCTGCCTGCTCGCCGGGCCGAACCTCAGC
GACACAAAATTCTGTGCGTGTGCTGCAAATGCGACGGCCGCATAGAGCTGACGGTA
GAATCATCAGCCGACGATCTGCGAACTCTTCAACAACTCTTCCTGAGCACGCTCAGC
TTCGTGTGTCCTTGGTGTGCTACAAATCAGTGACGGACCGCCTCTCCCTCCC

IgK-H6-**murineScn**-StrepII-GGGGS-E6.31
gi|148727550|gb|ABR08438.1| E6 protein [HPV31]
SEQ ID NO: 64 -
CTCGAGACC**ATGGAGACCGACACGCTCTTGTTGTGGGTTCTCTTGTTGTGGGTG**
CCTGGGTCTACAGGCGACCACCACCATCATCACCACCTCGTTCCTAGAGGCAG
CCAGGATAGTACCCAGAATCTTATCCCAGCACCATCTTTGCTCACAGTACCATT
GCAACCCGACTTTCGGTCTGATCAATTTCGGGGACGCTGGTACGTGGTTGGAC
TGGCCGGCAATGCTGTACAGAAAAAAACAGAGGGCAGTTTCACCATGTACTCA
ACAATCTATGAGCTCCAAGAGAATAATAGTTACAACGTTACCTCCATCTTGGTG
AGGGACCAGGATCAGGGATGTCGCTACTGGATTCGGACATTCGTACCAAGTTC
TCGGGCCGGTCAGTTTACTCTGGGCAACATGCACAGGTATCCCCAAGTTCAATC
TTACAACGTGCAGGTGGCGACTACCGACTACAACCAATTCGCTATGGTGTTCTT
CCGCAAAACAAGCGAGAACAAGCAGTATTTTAAAATCACTCTGTACGGTAGAAC
TAAGGAGCTGAGCCCTGAACTTAAGGAGCGGTTCACCAGATTCGCTAAGTCCC

TGGGACTGAAGGATGATAATATAATCTTTTCCGTCCCCACCGATCAGTGTATCG
ATAATTCAGCTTGGTCACATCCCCAGTTCGAGAAAGGAGGCGGTGGATCCATGTTCA
AAAACCCGGCTGAGAGACCGCGGAAGTTGCACGAGCTCTCATCCGCGCTGGAAATA
CCTTATGATGAGCTTCGCTTGAATTGTGTGTACTGCAAAGGCCAGCTCACTGAGACC
GAAGTACTTGATTTTGCCTTTACTGACCTGACAATCGTCTATAGAGACGACACTCCA
CACGGGGTCTGTACAAAATGTCTGCGGTTTTATAGTAAAGTGAGCGAATTCCGGTGG
TATCGCTATTCAGTGTATGGAACCACATTGGAGAAACTCACTAACAAAGGTATCTGT
GACCTGCTGATCAGGTGCATAACTTGTCAGAGGCCGCTCTGCCCCGAGGAGAAGCA
GCGCCACCTGGATAAGAAGAAGAGATTCCACAACATTGGAGGCAGATGGACAGGCC
GGTGCATTGCTTGTTGGCGCAGGCCAAGAACCGAGACCCAAGTTTGACGGACCGCCT
CTCCCTCCC

IgK-H6-**murineScn**-StrepII-GGGGS-E7.31
gi|338969947|gb|AEJ33624.1| E7, partial [HPV31]
SEQ ID NO: 65 -
CTCGAGACC**ATGGAGACCGACACGCTCTTGTTGTGGGTTCTCTTGTTGTGGGTG**
CCTGGGTCTACAGGCGACCACCACCATCATCACCACCTCGTTCCTAGAGGCAG
CCAGGATAGTACCCAGAATCTTATCCCAGCACCATCTTTGCTCACAGTACCATT
GCAACCCGACTTTCGGTCTGATCAATTTCGGGGACGCTGGTACGTGGTTGGAC
TGGCCGGCAATGCTGTACAGAAAAAAACAGAGGGCAGTTTCACCATGTACTCA
ACAATCTATGAGCTCCAAGAGAATAATAGTTACAACGTTACCTCCATCTTGGTG
AGGGACCAGGATCAGGGATGTCGCTACTGGATTCGGACATTCGTACCAAGTTC
TCGGGCCGGTCAGTTTACTCTGGGCAACATGCACAGGTATCCCCAAGTTCAATC
TTACAACGTGCAGGTGGCGACTACCGACTACAACCAATTCGCTATGGTGTTCTT
CCGCAAAACAAGCGAGAACAAGCAGTATTTTAAAATCACTCTGTACGGTAGAAC
TAAGGAGCTGAGCCCTGAACTTAAGGAGCGGTTCACCAGATTCGCTAAGTCCC
TGGGACTGAAGGATGATAATATAATCTTTTCCGTCCCCACCGATCAGTGTATCG
ATAATTCAGCTTGGTCACATCCCCAGTTCGAGAAAGGAGGCGGTGGATCCATGCGG
GGTGAGACACCAACTCTTCAGGATTATGTTCTGGATCTGCAGCCAGAGGCCACAGAT
CTGCACTGTTACGAGCAATTGCCTGATTCCAGCGACGAGGAGGATGTCATCGATAGC
CCTGCTGGGCAGGCCAAGCCAGACACTTCAAATTACAACATTGTAACGTTTTGTTGT
CAGTGCGAATCCACCCTCAGGCTTTGCGTCCAGAGCACTCAGGTTGACATTCGAATA
CTCCAGGAGCTGTTGATGGGGAGCTTTGGAATCGTGTGCCCAAATTGTAGTACACGA
CTGTGACGGACCGCCTCTCCCTCCC

IgK-H6-**murineScn**-StrepII-GGGGS-E6.58
gi|425892429|gb|AFY09749.1| E6 [HPV 58]
SEQ ID NO: 66 -
CTCGAGACC**ATGGAGACCGACACGCTCTTGTTGTGGGTTCTCTTGTTGTGGGTG**
CCTGGGTCTACAGGCGACCACCACCATCATCACCACCTCGTTCCTAGAGGCAG
CCAGGATAGTACCCAGAATCTTATCCCAGCACCATCTTTGCTCACAGTACCATT
GCAACCCGACTTTCGGTCTGATCAATTTCGGGGACGCTGGTACGTGGTTGGAC
TGGCCGGCAATGCTGTACAGAAAAAAACAGAGGGCAGTTTCACCATGTACTCA
ACAATCTATGAGCTCCAAGAGAATAATAGTTACAACGTTACCTCCATCTTGGTG
AGGGACCAGGATCAGGGATGTCGCTACTGGATTCGGACATTCGTACCAAGTTC
TCGGGCCGGTCAGTTTACTCTGGGCAACATGCACAGGTATCCCCAAGTTCAATC
TTACAACGTGCAGGTGGCGACTACCGACTACAACCAATTCGCTATGGTGTTCTT
CCGCAAAACAAGCGAGAACAAGCAGTATTTTAAAATCACTCTGTACGGTAGAAC
TAAGGAGCTGAGCCCTGAACTTAAGGAGCGGTTCACCAGATTCGCTAAGTCCC
TGGGACTGAAGGATGATAATATAATCTTTTCCGTCCCCACCGATCAGTGTATCG
ATAATTCAGCTTGGTCACATCCCCAGTTCGAGAAAGGAGGCGGTGGATCCATGTTTC
AGGACGCTGAGGAGAAGCCCAGAACTCTGCACGATCTGTGTCAGGCCTTGGAGACG
TCTGTGCATAAAATTGAGCTTAAATGTGTCGAATGTAAGAAGACACTCCAGCGCAGC
GAAGTTTATGACTTCGTGTTCGCGGATCTGAGAATCGTGTATCGGGACGGCAACCCT
TTTGCTGTTTGCAAGGTTTGCCTTAGGCTCCTGTCCAAAATTAGCGAGTACCGCCACT
ATAACTACTCTCTCTACGGTGATACTCTCGAGCAAACACTGAAGAAGTGCTTGAACG
AGATCCTGATTAGATGCATCATTTGTCAAAGGCCACTTTGTCCACAGGAGAAGAAGA
GGCACGTGGACCTGAATAAGCGCTTTCATAACATCTCTGGCAGATGGACAGGCCGAT
GCGCTGTATGTTGGCGCCCACGGAGAAGGCAAACCCAGGTGTGACGGACCGCCTCT

CCCTCCC

IgK-H6-**murineScn**-StrepII-GGGGS-E7.58
gi|414090989|gb|AFW98384.1| E7 [HPV 58]
SEQ ID NO: 67 -
CTCGAGACC**ATGGAGACCGACACGCTCTTGTTGTGGGTTCTCTTGTTGTGGGTG**
CCTGGGTCTACAGGCGACCACCACCATCATCACCACCTCGTTCCTAGAGGCAG
CCAGGATAGTACCCAGAATCTTATCCCAGCACCATCTTTGCTCACAGTACCATT
GCAACCCGACTTTCGGTCTGATCAATTTCGGGGACGCTGGTACGTGGTTGGAC
TGGCCGGCAATGCTGTACAGAAAAAAACAGAGGGCAGTTTCACCATGTACTCA
ACAATCTATGAGCTCCAAGAGAATAATAGTTACAACGTTACCTCCATCTTGGTG
AGGGACCAGGATCAGGGATGTCGCTACTGGATTCGGACATTCGTACCAAGTTC
TCGGGCCGGTCAGTTTACTCTGGGCAACATGCACAGGTATCCCCAAGTTCAATC
TTACAACGTGCAGGTGGCGACTACCGACTACAACCAATTCGCTATGGTGTTCTT
CCGCAAAACAAGCGAGAACAAGCAGTATTTTAAAATCACTCTGTACGGTAGAAC
TAAGGAGCTGAGCCCTGAACTTAAGGAGCGGTTCACCAGATTCGCTAAGTCCC
TGGGACTGAAGGATGATAATATAATCTTTTCCGTCCCCACCGATCAGTGTATCG
ATAATTCAGCTTGGTCACATCCCCAGTTCGAGAAAGGAGGCGGTGGATCCATGCGG
GGGAATAACCCCACCCTGCGCGAGTACATTCTTGACCTGCACCCAGAGCCTACGGAT
CTGTTTTGTTACGAACAACTGTGCGACTCCTCCGACGAGGATGAGATCGGGCTGGAT
GGCCCAGACGGGCAGGCACAGCCTGCTACAGCTAACTACTATATTGTGACATGTTGC
TACACATGCGGAACGACGGTCAGACTGTGCATTAATAGCACTGCCACAGACGTGCG
GACCCTGCAGCAACTGCTCATGGGGACCTGCACTATTGTGTGTCCTTCATGTGCGCA
GCAATGACGGACCGCCTCTCCCTCCC

Human Scn construct for downstream viral fusions:
IgK-sFLAG-H6-GGS-**humanScn**-GGS-ENLYFQ-GG-Adv2E3/19K
SEQ ID NO: 68 -
METDTLLLWVLLLWVPGSTGDYKDEHHHHHHGGSQDSTSDLIPAPPLSKVPLQQN
FQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKK
CDYWIRTFVPGSQPGEFTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNRE
YFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGGSENLYFQGGA
KKVEFKEPACNVTFKSEANECTTLIKCTTEHEKLIIRHKDKIGKYAVYAIWQPGDTNDYN
VTVFQGENRKTFMYKFPFYEMCDITMYMSKQYKLW

IgK-sFLAG-H6-GGS-**humanScn**-GGS-ENLYFQ-GG-SF162gp120
SEQ ID NO: 69 -
METDTLLLWVLLLWVPGSTGDYKDEHHHHHHGGSQDSTSDLIPAPPLSKVPLQQN
FQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKK
CDYWIRTFVPGSQPGEFTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNRE
YFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGGSENLYFQGG
WVTVYYGVPVWKEATTTLFCASDAKAYDTEVHNVWATHACVPTDPNPQEIVLENVTE
NFNMWKNNMVEQMHEDIISLWDQSLKPCVKLTPLCVTLHCTNLKNATNTKSSNWKEM
DRGEIKNCSFKVTTSIRNKMQKEYALFYKLDVVPIDNDNTSYKLINCNTSVITQACPKVS
FEPIPIHYCAPAGFAILKCNDKKFNGSGPCTNVSTVQCTHGIRPVVSTQLLLNGSLAEEGV
VIRSENFTDNAKTIIVQLKESVEINCTRPNNNTRKSITIGPGRAFYATGDIIGDIRQAHCNIS
GEKWNNTLKQIVTKLQAQFGNKTIVFKQSSGGDPEIVMHSFNCGGEFFYCNSTQLFNST
WNNTIGPNNTNGTITLPCRIKQIINRWQEVGKAMYAPPIRGQIRCSSNITGLLLTRDGGKE
ISNTTEIFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTKAKRRVVQQGLNDIFEAQKIE
WHE

IgK-sFLAG-H6-GGS-**humanScn**-GGS-ENLYFQ-GG-QH0692gp120
SEQ ID NO: 70 -
METDTLLLWVLLLWVPGSTGDYKDEHHHHHHGGSQDSTSDLIPAPPLSKVPLQQN
FQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKK
CDYWIRTFVPGSQPGEFTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNRE
YFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGGSENLYFQGG
WVTVYYGVPVWKEATTTLFCASDAKAYETEKHNVWATHACVPTDPNPQEVVLGNVTE
NFNMWKNNMVEQMHEDIISLWDESLKPCVKLTPLCVTLNCTDEVKTSYANKTSNETYK

TSNETFGEIKNCSFSVPTGIKDKVQNVYALFYKLDVIPIDDNNNSSKNNNGSYSSYRLINC
NTSVITQACPKVSFEPIPIHYCAPAGFAILKCNNKTFNGTGPCTNVSTVQCTHGIRPVVST
QLLLNGSLAEEEVVIRSENFTNNAKTIIVHLKKSVEINCTRPGNNTRKSIHIGPGRAFYAT
GDIIGDIRQAHCNLSSVQWNDTLKQIVIKLGEQFGTNKTIAFNQSSGGDPEIVMHSFNCG
GEFFYCNTTQLFNSTWEFHGNWTRSNFTESNSTTITLPCRIKQIVNMWQEVGKAMYAPPI
RGQIRCSSNITGLLLTRDGGVNGTRETFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPT
KAKRRVVQGLNDIFEAQKIEWHE

Human Scn construct for downstream peptide fusions:

IgK-sFLAG-H6-GGS-**humanScn**-GGS-ENLYFQ-GG-ITPR-1
SEQ ID NO: 71 -
METDTLLLWVLLLWVPGSTGDYKDEHHHHHHGGSQDSTSDLIPAPPLSKVPLQQN
FQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKK
CDYWIRTFVPGSQPGEFTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNRE
YFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGGSENLYFQGGS
KCRVFNTTERDEQGSKVNDFFQQTEDLYNEMKWQK

IgK-sFLAG-H6-GGS-**humanScn**-GGS-ENLYFQ-GG-ITPR-2
SEQ ID NO: 72 -
METDTLLLWVLLLWVPGSTGDYKDEHHHHHHGGSQDSTSDLIPAPPLSKVPLQQN
FQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKK
CDYWIRTFVPGSQPGEFTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNRE
YFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGGSENLYFQGGP
PHELTEEEKQQILHSEEFLSFFDHSTRIVERALSE

IgK-sFLAG-H6-GGS-**humanScn**-GGS-ENLYFQ-GG-ITPR-3
SEQ ID NO: 73 -
METDTLLLWVLLLWVPGSTGDYKDEHHHHHHGGSQDSTSDLIPAPPLSKVPLQQN
FQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKK
CDYWIRTFVPGSQPGEFTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNRE
YFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGGSENLYFQGGP
PPRCISTNKCTAPEVENAIRVPGNRSFFSLTEIVR

IgK-sFLAG-H6-GGS-**humanScn**-GGS-ENLYFQ-GG-ITPR-4
SEQ ID NO: 74 -
METDTLLLWVLLLWVPGSTGDYKDEHHHHHHGGSQDSTSDLIPAPPLSKVPLQQN
FQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKK
CDYWIRTFVPGSQPGEFTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNRE
YFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGGSENLYFQGGT
ERDEQGSKINDFFLRSEDLFNEMNWQKKLRAQPVL

IgK-sFLAG-H6-GGS-**humanScn**-GGS-ENLYFQ-GG-ITPR-5
SEQ ID NO: 75 -
METDTLLLWVLLLWVPGSTGDYKDEHHHHHHGGSQDSTSDLIPAPPLSKVPLQQN
FQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKK
CDYWIRTFVPGSQPGEFTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNRE
YFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGGSENLYFQGGL
TEETKHRLFTTTEQDEQGSKVSDFFDQSSFLHNEM

IgK-sFLAG-H6-GGS-**humanScn**-GGS-ENLYFQ-GG-ITPR-8
SEQ ID NO: 76 -
METDTLLLWVLLLWVPGSTGDYKDEHHHHHHGGSQDSTSDLIPAPPLSKVPLQQN
FQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKK
CDYWIRTFVPGSQPGEFTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNRE
YFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGGSENLYFQGGG
AQPPFDAQSPLDSQPQPSGQPWNFHASTSWYWRQS

Human Scn construct for downstream HMOX1 fusions:

IgK-sFLAG-H6-GGS-**humanScn**-GGS-ENLYFQ-GG-human HMOX1
SEQ ID NO: 77 -
METDTLLLWVLLLWVPGSTGDYKDEHHHHHHGGSQDSTSDLIPAPPLSKVPLQQN FQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKK CDYWIRTFVPGSQPGEFTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNRE YFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGGSENLYFQGG
MERPQPDSMPQDLSEALKEATKEVHTQAENAEFMRNFQKGQVTRDGFKLVMASLYHIY
VALEEEIERNKESPVFAPVYFPEELHRKAALEQDLAFWYGPRWQEVIPYTPAMQRYVKR
LHEVGRTEPELLVAHAYTRYLGDLSGGQVLKKIAQKALDLPSSGEGLAFFTFPNIASATK
FKQLYRSRMNSLEMTPAVRQRVIEEAKTAFLLNIQLFEELQELLTHDTKDQSPSRAPGLR
QRASNKVQDSAPVETPRGKPPLNTRSQA

IgK-sFLAG-H6-GGS-**humanScn**-GGS-ENLYFQ-GG-murineHMOX1
SEQ ID NO: 78 -
METDTLLLWVLLLWVPGSTGDYKDEHHHHHHGGSQDSTSDLIPAPPLSKVPLQQN FQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKK CDYWIRTFVPGSQPGEFTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNRE YFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGGSENLYFQGG
MERPQPDSMPQDLSEALKEATKEVHIQAENAEFMKNFQKGQVSREGFKLVMASLYHIY
TALEEEIERNKQNPVYAPLYFPEELHRRAALEQDMAFWYGPHWQEIIPCTPATQHYVKR
LHEVGRTHPELLVAHAYTRYLGDLSGGQVLKKIAQKAMALPSSGEGLAFFTFPNIDSPTK
FKQLYRARMNTLEMTPEVKHRVTEEAKTAFLLNIELFEELQVMLTEEHKDQSPSQMASL
RQRPASLVQDTAPAETPRGKPQISTSSSQ*

humanScn-humanHMOX1 (Second Generation)
SEQ ID NO: 79 -
MPLGLLWLGLALLGALHAQAQDSTSDLIPAPPLSKVPLQQNFQDNQFQGKWYVVG LAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKKCDYWIRTFVPGSQPGE FTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNREYFKITLYGRTKELTSE LKENFIRFSKSLGLPENHIVFPVPIDQCIDGGGSENLYFQGGGMERPQPDSMPQDLSEA
LKEATKEVHT
QAENAEFMRNFQKGQVTRDGFKLVMASLYHIYVALEEEIERNKESPVFAPVYFPEELHR
KAALEQDLAFWYGPRWQEVIPYTPAMQRYVKRLHEVGRTEPELLVAHAYTRYLGDLS
GGQVLKKIAQKALDLPSSGEGLAFFTFPNIASATKFKQLYRSRMNSLEMTPAVRQRVIEE
AKTAFLLNIQLFEELQELLTHDTKDQSPSRAPGLRQRASNKVQDSAPVETPRGKPPLNTR
SQAGGLVPRGSHHHHHH

humanScn-cTHAP4
SEQ ID NO: 80 -
MPLGLLWLGLALLGALHAQAQDSTSDLIPAPPLSKVPLQQNFQDNQFQGKWYVVG LAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKKCDYWIRTFVPGSQPGE FTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNREYFKITLYGRTKELTSE LKENFIRFSKSLGLPENHIVFPVPIDQCIDGGGSENLYFQGGGPPKMNPVVEPLSWML
GTWLSDPPGAG
TYPTLQPFQYLEEVHISHVGQPMLNFSFNSFHPDTRKPMHRECGFIRLKPDTNKVAFVSA
QNTGVVEVEEGEVNGQELCIASHSIARISFAKEPHVEQITRKFRLNSEGKLEQTVSMATTT
QPMTQHLHVTYKKVTPGGLVPRGSHHHHHH

Human Scn construct for downstream multimer fusions:
SEQ ID NO: 81 -
IgK-sFLAG-H6-GGS-**humanScn**-GGS-ENLYFQ-GG-heptamer
METDTLLLWVLLLWVPGSTGDYKDEHHHHHHGGSQDSTSDLIPAPPLSKVPLQQN FQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKK CDYWIRTFVPGSQPGEFTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNRE YFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGGSENLYFQGGG
RSAGAHAGWETPEGCEQVLTGKRLMQCLPNPEDVKMALEVYKLSLEIEQLELQRDSAR
QSTLDKELVPRGS

IgK-sFLAG-H6-GGS-**humanScn**-GGS-ENLYFQ-GG-CD80heptamer

SEQ ID NO: 82 -

METDTLLLWVLLLWVPGSTGDYKDEHHHHHHGGSQDSTSDLIPAPPLSKVPLQQN FQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKK CDYWIRTFVPGSQPGEFTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNRE YFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGGSENLYFQGGG IIQVNKTVKEVAVLSCDYNISTTELMKVRIYWQKDDEVVLAVTSGQTKVWSKYENRTF ADFTNNLSIVIMALRLSDNGKYTCIVQKTEKRSYKVKHMTSVMLLVRADFPVPSITDLG NPSHDIKRIMCSTSGGFPKPHLSWWENEEELNAANTTVSQDPDTELYTISSELDFNITSNH SFVCLVKYGDLTVSQIFNWQKSVEPHPPNNSAWSHPQFEKGGSLVPRGSGSAGAHAGW ETPEGCEQVLTGKRLMQCLPNPEDVKMALXVYKLSLEIEQLELQRDSARQSTLDKELVP RGS

IgK-sFLAG-H6-GGS-**humanScn**-GGS-ENLYFQ-GG-trimer

SEQ ID NO: 83 -

METDTLLLWVLLLWVPGSTGDYKDEHHHHHHGGSQDSTSDLIPAPPLSKVPLQQN FQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKK CDYWIRTFVPGSQPGEFTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNRE YFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGGSENLYFQGGG RNLVTAFSNMDDMLQKAHLVIEGTFIYLRDSTEFFIRVRDGWKKLQLGELIPIPA

IgK-sFLAG-H6-GGS-**humanScn**-GGS-ENLYFQ-GG-CD80trimer

SEQ ID NO: 84 -

METDTLLLWVLLLWVPGSTGDYKDEHHHHHHGGSQDSTSDLIPAPPLSKVPLQQN FQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKK CDYWIRTFVPGSQPGEFTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNRE YFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGGSENLYFQGGG IIQVNKTVKEVAVLSCDYNISTTELMKVRIYWQKDDEVVLAVTSGQTKVWSKYENRTF ADFTNNLSIVIMALRLSDNGKYTCIVQKTEKRSYKVKHMTSVMLLVRADFPVPSITDLG NPSHDIKRIMCSTSGGFPKPHLSWWENEEELNAANTTVSQDPDTELYTISSELDFNITSNH SFVCLVKYGDLTVSQIFNWQKSVEPHPPNNSAWSHPQFEKGGSLVPRGSGNLVTAFSNM DDMLQKAHLVIEGTFIYLRDSTEFFIRVRDGWKKLQLGELIPIPA

IgK-sFLAG-H6-GGS-**humanScn**-GGS-ENLYFQ-GG-pentamer

SEQ ID NO: 85 -

METDTLLLWVLLLWVPGSTGDYKDEHHHHHHGGSQDSTSDLIPAPPLSKVPLQQN FQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKK CDYWIRTFVPGSQPGEFTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNRE YFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGGSENLYFQGGG RSSNAKWDQWSSDWQTWNAKWDQWSNDWNAWRSDWQAWKDDWARWNQRWDN WAT

Human Scn construct for downstream subdomain fusions:

SEQ ID NO: 86 -

IgK-sFLAG-H6-GGS-**humanScn**-GGS-ENLYFQ-GG-ROR1Kringle

METDTLLLWVLLLWVPGSTGDYKDEHHHHHHGGSQDSTSDLIPAPPLSKVPLQQN FQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKK CDYWIRTFVPGSQPGEFTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNRE YFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGGSENLYFQGGC YNSTGVDYRGTVSVTKSGRQCQPWNSQYPHTHTFTALRFPELNGGHSYCRNPGNQKEA PWCFTLDENFKSDLCDIPAC

IgK-sFLAG-H6-GGS-**humanScn**-GGS-ENLYFQ-GG-CTLA4

SEQ ID NO: 87 -

METDTLLLWVLLLWVPGSTGDYKDEHHHHHHGGSQDSTSDLIPAPPLSKVPLQQN FQDNQFQGKWYVVGLAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKK CDYWIRTFVPGSQPGEFTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNRE YFKITLYGRTKELTSELKENFIRFSKSLGLPENHIVFPVPIDQCIDGGGSENLYFQGG MHVAQPAVVLASSRGVASFVCEYGSSGNAAEVRVTVLRQAGSQMTEVCAATYTVEDE LAFLDDSTCTGTSSGNKVNLTIQGLRAMDTGLYICKVELMYPPPYYVGMGNGTQIYVID

PEPC

Human Scn construct for downstream knottin fusions:
IgK-sFLAG-H6-GGS-**humanScn**-GGS-ENLYFQ-GG-Imperatoxin
SEQ ID NO: 88 -
CTCGAGACC**ATGGAGACAGACACACTCCTGCTATGGGTACTGCTGCTCTGGGTT**
CCAGGTTCCACTGGTGACTACAAGGACGAGCATCACCATCATCACCATGGTGG
AAGCCAGGACTCCACCTCAGACCTGATCCCAGCCCCACCTCTGAGCAAGGTCC
CTCTGCAGCAGAACTTCCAGGACAACCAATTCCAGGGGAAGTGGTATGTGGTA
GGCCTGGCAGGGAATGCAATTCTCAGAGAAGACAAAGACCCGCAAAAGATGTA
TGCCACCATCTATGAGCTGAAAGAAGACAAGAGCTACAATGTCACCTCCGTCCT
GTTTAGGAAAAAGAAGTGTGACTACTGGATCAGGACTTTTGTTCCAGGTTGCCA
GCCCGGCGAGTTCACGCTGGGCAACATTAAGAGTTACCCTGGATTAACGAGTT
ACCTCGTCCGAGTGGTGAGCACCAACTACAACCAGCATGCTATGGTGTTCTTCA
AGAAAGTTTCTCAAAACAGGGAGTACTTCAAGATCACCCTCTACGGGAGAACCA
AGGAGCTGACTTCGGAACTAAAGGAGAACTTCATCCGCTTCTCCAAATCTCTGG
GCCTCCCTGAAAACCACATCGTCTTCCCTGTCCCAATCGACCAGTGTATCGACG
GCGGAGGTAGCGAAAACCTGTATTTTCAGGGAGGCGACTGCCTGCCCCACCTGAGG
AGGTGCAGGGCCGACAACGACTGCTGCGGCAGGAGGTGCAGGAGGAGGGGCACCA
ACGCCGAGAGGAGGTGCAGGTAAGCTAAGGATCC

SEQ ID NO: 89 -
IgK SP: Murine IgK light chain signal peptide:
METDTLLLWVLLLWVPGSTG

SEQ ID NO: 90 -
ATGGAGACAGACACACTCCTGCTATGGGTACTGCTGCTCTGGGTTCCAGGTTCCACT
GGT

SEQ ID NO: 91 -
sFLAG: Shortened FLAG epitope:
DYKDE

SEQ ID NO: 92 -
GACTACAAGGACGAG

SEQ ID NO: 93 -
HIS: 6xhistidine tag
HHHHHH

SEQ ID NO: 94 -
CATCATCATCATCATCAT

SEQ ID NO: 95 -
TEV: Tobacco Etch Virus Protease recognition site:
ENLYFQ

SEQ ID NO: 96 -
GAGAATTTATATTTTCAG

SEQ ID NO: 97 –
Furin – furin cleavage site with BamHI site (GGATCC):

SEQ ID NO: 99 –
RARYKRGS

SEQ ID NO: 100
CGGGCCCGGTATAAACGGGGATCC

CLAIMS

WHAT IS CLAIMED IS:

1. A method for producing a fusion protein, the method comprising expressing, in a cell, a fusion protein, the fusion protein comprising a peptide or protein domain and a lipocalin protein, thereby producing the fusion protein.

2. The method of claim 1, further comprising separating the peptide or protein domain from the lipocalin protein, thereby producing a peptide or protein domain.

3. The method of claim 1 or 2, wherein the peptide or protein domain is an antibody fragment.

4. The method of claim 3, wherein the antibody is trastuzumab, infliximab, adalimumab, OKT3, or Fc.

5. The method of any of the preceding claims, wherein the peptide or protein domain is human heme oxygenase 1 or murine heme oxygenase 1.

6. The method of any one of the preceding claims, wherein the fusion protein further comprises a cleavage site.

7. The method of claim 6, wherein the cleavage site is a furin cleavage site, a trypsin cleavage site or a TEV cleavage site.

8. The method of claim 7, wherein the separating of the peptide or protein domain from the lipocalin protein results from cleavage at the cleavage site in the fusion protein.

9. The method of any one of claims 2 to 7, wherein the separating of the peptide or protein domain from the lipocalin protein occurs following secretion of the fusion protein from a cell.

10. The method of claim 9, wherein the cell is a mammalian cell.

11. The method of any one of the preceeding claims, wherein the peptide is produced at a concentration less than about 500 mg/liter, less than about 400 mg/liter, less than about 300 mg/liter, less than about 200 mg/liter, less than about 100 mg/liter, or less than about 50 mg/liter.

12. The method of any one of the preceding claims, wherein the peptide comprises a disulfide knotted-peptide.

13. The method of claim 12, wherein the knotted-peptide is PMP-D2, potato carboxypeptidase, huwentoxin, imperatoxin, epiregulin, midkine, bubble protein or conotoxin CVIC.

14. The method of any one of the preceeding claims, wherein the lipocalin protein comprises siderocalin.

15. The method of any one of the preceeding claims, wherein the fusion protein further comprises a peptide selected from the group consisting of: an IgK starter sequence, a sFLAG, a HIS, siderocalin, a TEV, and the knotted peptide sequence of interest.

16. The method of claim 15, wherein the fusion protein comprises the following construct: IgK SP-sFLAG-HIS-siderocalin-TEV-peptide.

17. The method of any one of the preceeding claims, wherein the peptide comprises a knottin.

18. A composition of a fusion protein, the composition comprising;

a peptide or protein domain, and;

a lipocalin protein.

19. The composition of claim 18, wherein the peptide comprises a knotted peptide.

20. The composition of claim 19, wherein the knotted-peptide is selected from the group consisting of PMP-D2, potato carboxypeptidase, huwentoxin, imperatoxin, epiregulin, midkine, bubble protein and conotoxin CIVC.

21. The composition of any one of claims 18 to 20, wherein the lipocalin protein is siderocalin.

22. The composition of any one of claims 18 to 21, wherein the lipocalin protein is human siderocalin, murine siderocalin, chicken Ex-FABP, or quail Q86.

23. The composition of claim 19 or 20, wherein the knotted-peptide includes at least two disulfide bonds.

24. The composition of any one of claims 18 to 23, wherein the fusion protein further comprises at least one of an IgK starter sequence, a sFLAG, a HIS, and a TEV.

25. The composition of claim 24, wherein the fusion protein comprises the following construct: IgK SP-sFLAG-HIS-siderocalin-TEV-peptide.

26. The composition of any one of claims 18 to 25, wherein the fusion protein is generated by direct fusion of each subunit to the adjacent subunits.

27. The composition of any one of claims 18 to 26, further comprising a linker sequence between the peptide or protein domain and the lipocalin protein.

28. The composition of any one of claims 18 to 27, wherein the peptide comprises a knotted-peptide.

29. A composition comprising a peptide library, the peptide library further comprising a plurality of peptides lacking at least one native lysine residue.

30. The composition of claim 29, wherein the peptides are conjugated to an adaptor molecule.

31. The composition of claim 30, wherein the adaptor molecule is a peptide.

32. The composition of claim 29 or 30, wherein the peptide has a unique signature determined by mass spectroscopy.

33. A method of producing a peptide, the method comprising:

expressing, in a cell, a fusion protein comprising a peptide and a lipocalin protein; and

separating the peptide from the lipocalin protein, thereby producing the peptide.

34. The method of claim 33, wherein the peptide is separated from the lipocalin protein by proteolysis or by cleavage of a furin cleavage site in the fusion protein such that the peptide is cleaved from the lipocalin protein upon secretion from the cell.

35. The method of claim 34, wherein the cell is a mammalian cell.

36. The method of claim 34, wherein the peptide is produced at a concentration less than about 200 mg/liter.

37. The method of claim 33, wherein the peptide comprises a knotted-peptide.

38. The method of claim 37, wherein the knotted-peptide is selected from the group consisting of chymotrypsin inhibitor, hefutoxin, bubble protein, the C-terminal domain of midkine, and epiregulin.

39. The method of claim 33, wherein the lipocalin protein comprises siderocalin.

40. The method of claim 33, wherein the fusion protein further comprises at least one of an IgK starter sequence, a sFLAG, a HIS, and a TEV.

41. The method of claim 40, wherein the fusion protein comprises the following construct: IgK SP-sFLAG-HIS-siderocalin-TEP-peptide.

42. The method of claim 41, wherein the peptide comprises a knotted-peptide.

43. A composition comprising a fusion protein comprising a peptide and a lipocalin protein.

44. The composition of claim 43, wherein the peptide comprises a knotted peptide.

45. The composition of claim 44, wherein the knotted-peptide is selected from the group consisting of chymotrypsin inhibitor, hefutoxin, bubble protein, the C-terminal domain of midkine, potato carboxypeptidase inhibitor, and epiregulin.

46. The composition of claim 44, wherein the lipocalin protein is siderocalin.

47. The composition of claim 43, wherein the knotted-peptide includes at least two disulfide bonds.

48. The composition of claim 43, wherein the fusion protein further comprises at least one of an IgK starter sequence, a sFLAG, a HIS, and a TEV.

49. The composition of claim 43, wherein the fusion protein comprises the following construct: IgK SP-sFLAG-HIS-siderocalin-TEV-peptide.

50. The composition of claim 49, wherein the peptide comprises a knotted-peptide.