(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 May 2014 (22.05.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/078734 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/US2013/070424

(22) International Filing Date:
15 November 2013 (15.11.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/727,094 15 November 2012 (15.11.2012) US
61/727,093 15 November 2012 (15.11.2012) US

(71) Applicant: **APELLIS PHARMACEUTICALS, INC.** [US/US]; 6400 Westwind Way, Suite A, Crestwood, Kentucky 40014 (US).

(72) Inventors: **FRANCOIS, Cedric**; 1434 Goddard Avenue, Louisville, Kentucky 40204 (US). **DESCHATELETS, Pascal**; 11031 Indian Legends Drive, Apt. 202, Louisville, Kentucky 40241 (US).

(74) Agents: **HUDDLESON, Justin P.** et al.; Choate, Hall & Stewart LLP, Two International Place, Boston, Massachusetts 02110 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

WO 2014/078734 A2

(54) Title: CELL-REACTIVE, LONG-ACTING, OR TARGETED COMPSTATIN ANALOGS AND RELATED COMPOSITIONS AND METHODS

(57) Abstract: In some aspects, the present invention provides cell-reactive compstatin analogs and compositions comprising cell-reactive compstatin analogs. In some aspects, the invention further provides methods of using cell-reactive compstatin analogs, e.g., treat a complement-mediated disorder, e.g., to inhibit complement-mediated damage to a cell, tissue, or organ. In some aspects, the invention provides long-acting compstatin analogs and compositions comprising long-acting compstatin analogs. In some aspects, the invention further provides methods of using long-acting compstatin analogs, e.g., to treat a complement-mediated disorder, e.g., to inhibit complement-mediated damage to a cell, tissue, or organ. In some aspects, the invention provides targeted compstatin analogs and compositions comprising targeted compstatin analogs. In some aspects, the invention further provides methods of using targeted compstatin analogs, e.g., to treat a complement-mediated disorder, e.g., to inhibit complement-mediated damage to a cell, tissue, or organ.

CELL-REACTIVE, LONG-ACTING, OR TARGETED COMPSTATIN ANALOGS AND RELATED COMPOSITIONS AND METHODS

Cross Reference to Related Applications

[0001] The present application claims priority to United States provisional patent applications nos. 61/727,094 and 61/727,093, filed November 15, 2012, the entire contents of both of which are hereby incorporated by reference.

Background of the Invention

[0002] Complement is a system consisting of more than 30 plasma and cell-bound proteins that plays a significant role in both innate and adaptive immunity. The proteins of the complement system act in a series of enzymatic cascades through a variety of protein interactions and cleavage events. Complement activation occurs via three main pathways: the antibody-dependent classical pathway, the alternative pathway, and the mannose-binding lectin (MBL) pathway. Inappropriate or excessive complement activation is an underlying cause or contributing factor to a number of serious diseases and conditions, and considerable effort has been devoted over the past several decades to exploring various complement inhibitors as therapeutic agents. However, there remains a need for innovative approaches to inhibiting complement activation for a variety of therapeutic purposes.

Summary of the Invention

[0003] In some aspects, the invention provides cell-reactive compstatin analogs. For example, the invention provides cell-reactive compstatin analogs, compositions comprising cell-reactive compstatin analogs, and methods of making, identifying, characterizing, and/or using cell-reactive compstatin analogs. In some aspects, the invention provides a physiologically acceptable composition comprising a cell-reactive compstatin analog. In some aspects, the invention provides a pharmaceutical grade composition comprising a cell-reactive compstatin analog.

[0004] In some aspects, the invention provides long-acting compstatin analogs. For example, the invention provides long-acting compstatin analogs, compositions comprising long-acting compstatin analogs, and methods of making, identifying, characterizing, and/or using long-acting compstatin analogs. In some aspects, the invention provides a physiologically acceptable composition comprising a long-acting compstatin analog. In

some aspects, the invention provides a pharmaceutical grade composition comprising a long-acting compstatin analog.

[0005] In some aspects, the invention provides targeted compstatin analogs. For example, the invention provides targeted compstatin analogs, compositions comprising targeted compstatin analogs, and methods of making, identifying, characterizing, and/or using targeted compstatin analogs. In some aspects, the invention provides a physiologically acceptable composition comprising a targeted compstatin analog. In some aspects, the invention provides a pharmaceutical grade composition comprising a targeted compstatin analog.

[0006] The invention further provides methods of protecting a cell from complement-mediated damage. In some embodiments, the methods comprise contacting the cell with a cell-reactive compstatin analog. The cell can be any type of cell in various embodiments. For example, in some embodiments, the cell is a blood cell. In some embodiments the blood cell is a red blood cell (RBC), also referred to as an erythrocyte. In some embodiments, the cell has abnormally low expression, surface density, and/or activity of one or more complement regulatory proteins. For example, the cell may have a mutation in a gene encoding such protein, wherein the mutation results in reduced or absent expression and/or reduced activity of the encoded protein. The cell can be of any animal type or species in various embodiments. For example, the cell can be mammalian, e.g., primate (human or a non-human primate), rodent (e.g., mouse, rat, rabbit), ungulate (e.g., pig, sheep, cow), canine, or feline. In many embodiments, the protection is from primate complement, e.g., human complement. In some embodiments, the cell is contacted ex vivo (outside the body of a subject). In some embodiments the cell is contacted in vivo (in a subject, e.g., a human). In some embodiments the cell is to be transplanted into a subject or has been transplanted into a subject. In some aspects, the invention provides an isolated cell having a compstatin analog covalently attached thereto. In some aspects, the invention provides an isolated tissue or organ having a compstatin analog attached to at least some of its cells.

[0007] The invention provides methods of treating a subject in need of treatment for a complement-mediated disorder. In some embodiments the method comprises administering a cell-reactive compstatin analog to the subject. In some embodiments the method comprises administering a long-acting compstatin analog to the subject. In some embodiments, a long-acting compstatin analog is a cell-reactive compstatin analog. In some embodiments, the complement-mediated disorder is paroxysmal nocturnal hemoglobinuria (PNH), atypical hemoloytic uremic syndrome (aHUS), or another disorder associated with complement-

mediated hemolysis. In some embodiments the disorder is ischemia/reperfusion (I/R) injury (e.g., due to myocardial infarction, thromboembolic stroke, or surgery. In some embodiments, the disorder is trauma. In some embodiments, the disorder is transplant rejection. In some embodiments the disorder is a chronic respiratory disorder, e.g., asthma or COPD.

[0008] All articles, books, patent applications, patents, other publications, websites, and databases mentioned in this application are incorporated herein by reference. In the event of a conflict between the specification and any of the incorporated references the specification (including any amendments thereto) shall control. Unless otherwise indicated, art-accepted meanings of terms and abbreviations are used herein. The practice of certain aspects described herein may employ conventional techniques of molecular biology, cell culture, recombinant nucleic acid (e.g., DNA) technology, immunology, and/or nucleic acid and polypeptide synthesis, detection, manipulation, and quantification, etc., that are within the ordinary skill of the art. See, e.g., Ausubel, F., et al., (eds.), Current Protocols in Molecular Biology, Current Protocols in Immunology, Current Protocols in Protein Science, and Current Protocols in Cell Biology, all John Wiley & Sons, N.Y., e.g., edition current as of January 2010 or later; Sambrook, Russell, and Sambrook, Molecular Cloning: A Laboratory Manual, 3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2001 or 4th ed, 2012.

Brief Description of the Drawing

[0009] **Figure 1** is a plot that shows percent complement activation inhibiting activity of compstatin analog CA28 (SEQ ID NO: 28) and three long-acting compstatin analogs (CA28-1, CA28-2, CA28-3), as a function of peptide concentration (μM). Inhibition of complement activation was tested in vitro using a classical complement inhibition assay. The plot shows values obtained by averaging the results of two sets of measurements. CA28 (circles; red), CA28-1 (crosses (x); blue); CA28-2 (triangles, green), CA28-3 (squares (purple).

[0010] **Figure 2** is a plot that shows percent complement activation inhibiting activity of CA28 and long-acting compstatin analogs CA28-2 and CA28-3, as a function of compound concentration (μM). CA28 (squares, light gray), CA28-2 (diamonds, black), CA28-3 (circles, dark gray). CA28-3 is a compound that contains multiple peptide moieties. Although the activity per peptide moiety is less than the activity of an individual CA28 molecule, the total activity of CA28-3 exceeds the activity of CA28 on a molar basis.

[0011] **Figure 3** is a plot that shows plasma concentrations versus time of CA28 and long-acting compstatin analogs CA28-2 and CA28-3 in Cynomolgus monkeys following a single intravenous injection. CA28 was administered at 200 mg/kg. CA28-2 and CA28-3 were each administered at 50 mg/kg. In calculating the doses for these experiments the administered CA28-2 and CA28-3 substance was assumed to consist 80% of active compound w/w based on dry weight. However, during the sample analysis, the standard curve assumed a 100% of active compound w/w based on dry weight, by an estimated 30%. Thus, the values for Cmax overestimate the actual Cmax. CA28 (squares, light gray), CA28-2 (triangles, black), CA28-3 (circles, dark gray).

[0012] **Figure 4** is a plot that shows percent complement activation inhibiting activity of CA28 and long-acting compstatin analogs CA28-4, as a function of compound concentration (μM). Inhibition of complement activation was tested in vitro using a classical complement inhibition assay. The plot shows values obtained by averaging the results of four sets of measurements for CA28-4. CA28 (squares, light gray), CA28-4 (crosses, black).

[0013] **Figure 5** is a plot that shows concentrations versus time of CA28 and long-acting compstatin analogs CA28-2, CA28-3, and CA28-4 in Cynomolgus monkeys following a single intravenous injection. CA28 was administered at 200 mg/kg. CA28-2, CA28-3, and CA28-4 were each administered at 50 mg/kg. In calculating the doses for these experiments the administered CA28-2 and CA28-3 substance was assumed to consist 80% of active compound w/w based on dry weight. However, during the sample analysis, the standard curve assumed a 100% of active compound w/w based on dry weight. Thus, the values for Cmax overestimate the Cmax that would be achieved if these compounds had been administered at the indicated doses on a dry mass basis, by an estimated 30%. CA28 (squares, light gray), CA28-2 (triangles, black), CA28-3 (circles, dark gray), CA28-4 (inverted triangles, black).

[0014] **Figure 6** is a representative chromatogram showing ultraviolet (UV) detection of a PEG-based long-acting compstatin analog using reverse phase HPLC. The peak with a retetion time (RT) of 33.68 minutes represents the PEGylated compstatin analog and had a relative area of 96%.

[0015] **Figure 7** is a plot that shows percent complement activation inhibiting activity of CA28 and long-acting compstatin analogs, as a function of compound concentration (μM). CA28-2CS (diamonds, red); CA28-2GS (crosses, blue); CA28-2HS (triangles, green); CA28-2TS (squares, black).

[0016] **Figure 8** is a plot that shows percent complement activation inhibiting activity of CA28 and bifunctionalized long-acting compstatin analog, CA28-2GS-BF, as a function of compound concentration (micromoles). CA28 (open circles, blue); CA28-2G-SBF (filled circles, red)

[0017] **Figure 9** is a plot that shows the plasma concentration versus time of CA28 and long-acting compstatin analog CA28-2GS-BF in Cynomolgus monkeys following either a single intravenous injection (CA28 (squares, red) and CA28-2GS-BF (circles, purple) or when administered by subcutaneous injection once daily for 7 days (CA28-2GS-BF only, asterisks, blue). CA28-2GSBF was administered at 25 mg/ml. Dosing volume was 2 ml/kg for IV and 0.28 ml/kg/day for the subcutaneous administration. Data for CA28 was from a different experiment in which the compound was also in 5% dextrose and was formulated as 20 mg/ml with a 10 ml/kg dosing volume. The vehicle in each case was 5% dextrose in water.

[0018] **Figures 10(A) and 10(B)** presents plots that show percent complement activation inhibiting activity of CA28 and bifunctionalized long-acting compstatin analog, CA28-2TS-BF, as a function of compound concentration (micromoles). (A) Classical pathway inhibition by CA28 (circles, red) and CA28-2TS-BF (crosses, blue). (B) Alternative pathway inhibition. CA28 (circles, red) and CA28-2TS-BF (crosses, blue).

[0019] **Figure 10(C)** (assuming a PEG moiety of 40 kD) shows the structure of CA28-2TS-BF.

[0020] **Figure 11** is a plot that shows the plasma concentration versus time of CA28 and long-acting compstatin analog CA28-2TS-BF in Cynomolgus monkeys following a single intravenous injection of CA28 at 200 mg/kg (squares, red), a single intravenous injection of CA28-2TS-BF at 7 mg/kg (asterisks, purple), subcutaneous injection of CA28-2TS-BF at 7 mg/kg once only (circles, blue) or subcutaneous injection of CA28-2TS-BF at 7 mg/kg once daily for 7 consecutive days (inverted triangles, green). The vehicle in each case was 5% dextrose in water.

[0021] **Figure 12** shows flow cytometric analysis of C3 deposition on red blood cells from a patient with PNH, which were exposed to activated complement in a modified Ham's test. (A) Results of a dilution experiment demonstrating the effect of CA28 on C3 deposition are shown. (B) Results of a dilution experiment demonstrating the effect of CA28-2GS-BF on C3 deposition are shown. Compound concentrations used are shown on and above each panel.

[0022] **Figure 13** shows flow cytometric analysis of C3 deposition on red blood cells from a patient with PNH, which were exposed to activated complement in a modified Ham's test in the absence of complement inhibitors (left panel), in the presence of anti-C5 monoclonal antibody eculizumab (middle panel) and in the presence of CA28-2GS-BF (right panel).

[0023] **Figure 14** is a schematic diagram showing examples of XTEN polypeptides with various reactive functional groups to which a compstatin analog can be conjugated or that may be used to link multiple XTEN polypeptides together to form a longer XTEN polypeptide.

[0024] **Figure 15** is a schematic representation of XTEN polypeptides functionalized with various reactive functional groups and the conjugation of moieties comprising appropriate partner groups thereto. The lower left panel shows installation of an alkyne at the N-terminus, which then reacts with an azide-functionalized moiety.

[0025] **Figure 16** shows conjugation of two XTEN polypeptides bearing different active agents (denoted as peptide and drug) via click chemistry.

[0026] **Figure 17** shows schematic diagrams and product names of various available XTEN polypeptides useful for making certain long-acting compstatin analogs of the present invention.

Detailed Description of Certain Embodiments of the Invention

[0027] **I. Definitions**

[0028] The terms "approximately" or "about" in reference to a number generally include numbers that fall within ±10%, in some embodiments ±5%, in some embodiments ±1%, in some embodiments ±0.5% of the number unless otherwise stated or otherwise evident from the context (except where such number would impermissibly exceed 100% of a possible value).

[0029] A "complement component" or "complement protein" is a protein that is involved in activation of the complement system or participates in one or more complement-mediated activities. Components of the classical complement pathway include, e.g., C1q, C1r, C1s, C2, C3, C4, C5, C6, C7, C8, C9, and the C5b-9 complex, also referred to as the membrane attack complex (MAC) and active fragments or enzymatic cleavage products of any of the foregoing (e.g., C3a, C3b, C4a, C4b, C5a, etc.). Components of the alternative pathway include, e.g., factors B, D, and properdin. Components of the lectin pathway include, e.g.,

MBL2, MASP-1, and MASP-2. Complement components also include cell-bound receptors for soluble complement components, wherein such receptor mediates one or more biological activities of such soluble complement component following binding of the soluble complement component. Such receptors include, e.g., C5a receptor (C5aR), C3a receptor (C3aR), Complement Receptor 1 (CR1), Complement Receptor 2 (CR2), Complement Receptor 3 (CR3, also known as CD45), etc. It will be appreciated that the term "complement component" is not intended to include those molecules and molecular structures that serve as "triggers" for complement activation, e.g., antigen-antibody complexes, foreign structures found on microbial or articifial surfaces, etc.

[0030] A "complement-mediated disorder" is any disorder in which complement activation is known or suspected of being a contributing and/or at least partially causative factor in at least some subjects suffering from the disorder, e.g., disorders in which complement activation results in tissue damage. Non-limiting examples of complement-mediated disorders include, but are not limited to, (i) various disorders characterized by hemolysis or hemolytic anemia such as atypical hemolytic uremic syndrome, cold agglutinin disease, paroxysmal nocturnal hemoglobinuria, transfusion reactions; (ii) transplant rejection (e.g., hyperacute or acute transplant rejection) or transplant dysfunction; (iii) disorders involving ischemia/reperfusion injury such as trauma, surgery (e.g., aneurysm repair), myocardial infarction, ischemic stroke; (iv) disorders of the respiratory system such as asthma and chronic obstructive pulmonary disease (COPD); (v) arthritis, e.g., rheumatoid arthritis; (vi) ocular disorders such as age-related macular degeneration (AMD), diabetic retinopathy, glaucoma, and uveitis. "Disorder" is used interchangeably herein with "disease", "condition", and similar words to refer to any impairment of health or state of abnormal functioning of an organism, e.g., any state in which medical and/or surgical management is indicated or for which a subject appropriately seeks medical and/or surgical attention. It should also be understood that the listing of a particular disorder within a particular category is for convenience and is not intended to limit the invention. It will be understood that certain disorders could appropriately be listed in multiple categories.

[0031] A "complement regulatory protein" is a protein involved in regulating complement activity. A complement regulatory protein may down-regulate complement activity by, e.g., inhibiting complement activation or by inactivating or accelerating decay of one or more activated complement proteins. Examples of complement regulatory proteins include C1 inhibitor, C4 binding protein, clusterin, vitronectin, CFH, factor I, and the cell-bound proteins CD46, CD55, CD59, CR1, CR2, and CR3.

[0032] "Isolated", as used herein, means 1) separated from at least some of the components with which it is usually associated in nature; 2) prepared or purified by a process that involves the hand of man; and/or 3) not occurring in nature, e.g., present in an artificial environment. In general, unless otherwise indicated or clearly evident, any entity, product, agent, composition, etc., may be deemed "isolated", if desired.

[0033] "Linked", as used herein with respect to two or more moieties, means that the moieities are physically associated or connected with one another to form a molecular structure that is sufficiently stable so that the moieties remain associated under the conditions in which the linkage is formed and, preferably, under the conditions in which the new molecular structure is used, e.g., physiological conditions. In certain preferred embodiments of the invention the linkage is a covalent linkage. In other embodiments the linkage is noncovalent. Moieties may be linked either directly or indirectly. When two moieties are directly linked, they are either covalently bonded to one another or are in sufficiently close proximity such that intermolecular forces between the two moieties maintain their association. When two moieties are indirectly linked, they are each linked either covalently or noncovalently to a third moiety, which maintains the association between the two moieties. In general, when two moieties are referred to as being linked by a "linking moiety" or "linking portion", the linkage between the two linked moieties is indirect, and typically each of the linked moieties is covalently bonded to the linking moiety. Two moieties may be linked using a "linker". A linker can be any suitable moiety that reacts with the entities to be linked within a reasonable period of time, under conditions consistent with stability of the entities (portions of which may be protected as appropriate, depending upon the conditions), and in sufficient amount, to produce a reasonable yield. Typically the linker will contain at least two functional groups, one of which reacts with a first entity and the other of which reacts with a second entity. It will be appreciated that after the linker has reacted with the entities to be linked, the term "linker" may refer to the part of the resulting structure that originated from the linker, or at least the portion that does not include the reacted functional groups. A linking moiety may comprise a portion that does not participate in a bond with the entities being linked, and whose main purpose may be to spatially separate the entities from each other. Such portion may be referred to as a "spacer".

[0034] As used herein, "physiological conditions" refers to a set of conditions such as temperature, salt concentration, pH that at least in part mimic those conditions as typically found in a living subject, e.g., a mammalian subject. In some aspects, physiological conditions refer to conditions in an aqueous medium, e.g., a medium comprising at least 90%,

95%, 96%, 97%, 97%, 99%, or about 100% water on a volume/volume basis. In some embodiments other liquids, if present, do not substantially affect protein secondary or tertiary structure. In some embodiments physiological conditions at least in part mimic those found in a body fluid such as blood or extracellular fluid, e.g., interstitial fluid, e.g., of a mammalian subject. A variety of physiological conditions useful for, e.g., in vitro assays, are known in the art. Generally, a medium under physiological conditions contains a physiological concentration of salt, e.g., sodium chloride. In some embodiments a physiological concentration of salt refers to a concentration ranging from about 250 mOsm/L to about 350 mOsm/L, e.g., about 275 mOsm/L to about 325 mOsm/L, e.g., about 300 mOsm/L. In some embodiments physiological conditions are approximately isotonic to a body fluid, e.g., blood or extracellular fluid, e.g., interstitial fluid. In some embodiments physiological conditions include a pH ranging from about 6.5 to about 7.8, e.g., about 7.0 to about 7.5. In some embodiments a physiological medium comprises a buffer substance that helps maintain the pH of the medium within a physiological range. In some embodiments physiological conditions comprise conditions such that a typical mammalian protein, e.g., a protein typically found in a body fluid, such as blood or extracellular fluid, substantially retains the secondary and, if applicable, tertiary structure that such protein has in the body fluid in which it is normally found. In some embodiments components of a physiological medium are typically substantially non-toxic to mammalian cells at the concentration at which they are present in the physiological medium. A variety of physiological media (sometimes termed "buffers") are listed in various standard references, such as those cited above (e.g., Sambrook, et al., Protocols series). In some embodiments a physiological temperature ranges from about 25 degrees C to about 38 degrees C, e.g., from about 30 degrees C to about 37 degrees C, e.g., 35 degrees C to 37 degrees C.

[0035] "Polypeptide", as used herein, refers to a polymer of amino acids, optionally including one or more amino acid analogs. A protein is a molecule composed of one or more polypeptides. A peptide is a relatively short polypeptide, typically between about 2 and 60 amino acids in length, e.g., between 8 and 40 amino acids in length. The terms "protein", "polypeptide", and "peptide" may be used interchangeably. Polypeptides used herein may contain amino acids such as those that are naturally found in proteins, amino acids that are not naturally found in proteins, and/or amino acid analogs that are not amino acids. As used herein, an "analog" of an amino acid may be a different amino acid that structurally resembles the amino acid or a compound other than an amino acid that structurally resembles the amino acid. A large number of art-recognized analogs of the 20 amino acids commonly

found in proteins (the "standard" amino acids) are known. One or more of the amino acids in a polypeptide may be modified, for example, by the addition of a chemical entity such as a carbohydrate group, a phosphate group, a farnesyl group, an isofarnesyl group, a fatty acid group, a linker for conjugation, functionalization, or other modification, etc. Certain non-limiting suitable analogs and modifications are described in WO2004026328 and/or below. The polypeptide may be acetylated, e.g., at the N-terminus and/or amidated, e.g., at the C-terminus.

[0036] The term "purified", as used herein, refers to substances that have been separated from at least some or most of the components with which they are associated in nature or when originally generated or with which they were associated prior to purification. In general, such purification involves action of the hand of man. Purified agents may be partially purified, substantially purified, or pure. Such agents may be, for example, at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or more than 99% pure. In some embodiments, a nucleic acid, polypeptide, or small molecule is purified such that it constitutes at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or more, of the total nucleic acid, polypeptide, or small molecule material, respectively, present in a preparation. In some embodiments, an organic substance, e.g., a nucleic acid, polypeptide, or small molecule, is purified such that it constitutes at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or more, of the total organic material present in a preparation. Purity may be based on, e.g., dry weight, size of peaks on a chromatography tracing (GC, HPLC, etc.), molecular abundance, electrophoretic methods, intensity of bands on a gel, spectroscopic data (e.g., NMR), elemental analysis, high throughput sequencing, mass spectrometry, or any art-accepted quantification method. In some embodiments, water, buffer substances, ions, and/or small molecules (e.g., synthetic precursors such as nucleotides or amino acids), can optionally be present in a purified preparation. A purified agent may be prepared by separating it from other substances (e.g., other cellular materials), or by producing it in such a manner to achieve a desired degree of purity. In some embodiments "partially purified" with respect to a molecule produced by a cell means that a molecule produced by a cell is no longer present within the cell, e.g., the cell has been lysed and, optionally, at least some of the cellular material (e.g., cell wall, cell membrane(s), cell organelle(s)) has been removed and/or the molecule has been separated or segregated from at least some molecules of the same type (protein, RNA, DNA, etc.) that were present in the lysate.

[0037] "Recombinant host cells", "host cells", and other such terms, denote prokaryotic or eukaryotic cells or cell lines that contain an exogenous nucleic acid (typically DNA) such

as an expression vector comprising a nucleic acid that encodes a polypeptide of interest. It will be understood that such terms include the descendants of the original cell(s) into which the vector or other nucleic acid has been introduced. Appropriate host cells include any of those routinely used in the art for expressing polynucleotides (e.g., for purposes of producing polypeptide(s) encoded by such polynucleotides) including, for example, prokaryotes, such as E. coli or other bacteria such as species of Escherichia; Lactobacillus, Bacillus (e.g., B subtilis), Salmonella Pseudomonas, Streptomyces ,Staphylococcus, etc; and eukaryotes, including for example, fungi, such as yeast (e.g., Pichia (e.g., Pichia pastoris), Kluyveromyces, such as K. lactis, Hansenula, e.g. H. polymorpha). Examples of other fungal cells are cells of filamentous fungi, e.g. Aspergillus spp., Neurospora spp., Fusarium spp. or Trichoderma spp., e.g., strains of A. oryzae, A. nidulans or A. niger; insect cells (e.g., Sf9), plant cells, and animal cells, e.g., mammalian cells such as CHO, R1.1, B-W, L-M, African Green Monkey Kidney cells (e.g. COS-1, COS-7, BSC-1, BSC-40 and BMT-10), and cultured human cells. Also encompassed are genetically modified cells in genetically modified (e.g., transgenic) plants or animals, wherein a recombinant polypeptide is produced by at least some such cells. A polypeptide may be secreted in milk, harvested from plant material, etc. The exogenous nucleic acid may be stably maintained as an episome such as a plasmid or may at least in part be integrated into the host cell's genome, optionally after being copied or reverse transcribed. Terms such as "host cells", etc., are also used to refer to cells or cell lines that can be used as recipients for an exogenous nucleic acid, prior to introduction of the nucleic acid. A "recombinant polynucleotide" generally is a polynucleotide that contains nucleic acid sequences that are not found joined directly to one another in nature. For example, the nucleic acid sequences may occur in different genes or different species or one or more of the sequence(s) may be a variant of a naturally occurring sequence or may at least in part be an artificial sequence that is not homologous to a naturally occurring sequence. A "recombinant polypeptide" generally is a polypeptide that is at least in part produced by transcription and translation of an exogenous nucleic acid by a recombinant host cell or by a cell-free in vitro expression system and/or that contains amino acid sequences that are not found joined directly to one another in nature. In the latter case, the recombinant polypeptide may be referred to as a "chimeric polypeptide". The amino acid sequences in a chimeric polypeptide may, for example, occur in different genes or in different species or one or more of the sequence(s) may be a variant of a naturally occurring sequence or may at least in part be an artificial sequence that is not identical or in some embodiments is not homologous to a naturally occurring sequence over a substantial portion

of the length. It will be understood that a chimeric polypeptide may comprise two or more polypeptides. For example, first and second polypeptides A and B of a chimeric polypeptide may be directly linked (A-B or B-A) or may be separated by a third polypeptide portion C (A-C-B or B-C-A). In some embodiments, portion C represents a polypeptide linker which may, for example, comprise multiple glycine and/or serine residues or any of a variety of other amino acids. In some embodiments, two or more polypeptides may be linked by non-polypeptide linker(s). "Recombinant" as used herein encompasses in certain embodiments polypeptides produced by joining (e.g., chemically conjugating, enzymatically conjugating), shorter recombinant polypeptides that may be produced in recombinant host cells. In some embodiments a recombinant polypeptide may comprise a signal sequence that directs secretion of the polypeptide or a sequence that directs the expressed polypetpide to a specific compartment or organelle. Suitable sequences are known in the art. Appropriate sequences for a host cell type of interest (e.g., bacterial, fungal, mammalian, plant, etc.) may be selected. A signal sequence may be located at or near (e.g., within up to 10 - 50 amino acids of) the N-terminus or C-terminus in some embodiments. In some embodiments a polypeptide comprises a tag. A tag may be useful to facilitate detection and/or purification of a protein that contains it. Examples of tags include polyhistidine-tag (e.g., 6X-His tag), glutathione-S-transferase, maltose binding protein, NUS tag, SNUT tag, Strep tag, epitope tags such as V5, HA, Myc, or FLAG. In some embodiments a protease cleavage site is located in the region between the tag and the polypeptide, allowing the polypeptide to be separated from the tag by exposure to the protease. In some embodiments a polynucleotide that encodes a recombinant polypeptide is at least in part codon optimized for expression in a host cell of interest (e.g., bacterial, fungal, mammalian, plant, etc.). A tag may be located at or near (e.g., within up to 10 – 50 amino acids of) the N- or C-terminus of a polypeptide in various embodiments. A recombinant polypeptide may be isolated, purified, etc., using any of a variety of methods. See, e.g., Sambrook, Protocols series, or other standard references. Methods of use may include, e.g., dialysis (e.g., using membranes having defined pore size), chromatography, precipitation, gel purification, or affinity-based methods that may, in some embodiments, utilize a tag or a specific binding reagent such as an antibody.

[0038] "Reactive functional groups" as used herein refers to groups including, but not limited to, olefins, acetylenes, alcohols, phenols, ethers, oxides, halides, aldehydes, ketones, carboxylic acids, esters, amides, cyanates, isocyanates, thiocyanates, isothiocyanates, amines, hydrazines, hydrazones, hydrazides, diazo, diazonium, nitro, nitriles, mercaptans, sulfides, disulfides, sulfoxides, sulfones, sulfonic acids, sulfinic acids, acetals, ketals, anhydrides,

sulfates, sulfenic acids isonitriles, amidines, imides, imidates, nitrones, hydroxylamines, oximes, hydroxamic acids thiohydroxamic acids, allenes, ortho esters, sulfites, enamines, ynamines, ureas, pseudoureas, semicarbazides, carbodiimides, carbamates, imines, azides, azo compounds, azoxy compounds, and nitroso compounds, N-hydroxysuccinimide esters, maleimides, sulfhydryls, and the like. Methods to prepare each of these functional groups are well known in the art and their application to or modification for a particular purpose is within the ability of one of skill in the art (see, for example, Sandler and Karo, eds. ORGANIC FUNCTIONAL GROUP PREPARATIONS, Academic Press, San Diego, 1989, and Hermanson, G., *Bioconjugate Techniques*, 2nd ed., Academic Press, San Diego, 2008).

[0039] "Specific binding" generally refers to a physical association between a target polypeptide (or, more generally, a target molecule) and a binding molecule such as an antibody or ligand. The association is typically dependent upon the presence of a particular structural feature of the target such as an antigenic determinant, epitope, binding pocket or cleft, recognized by the binding molecule. For example, if an antibody is specific for epitope A, the presence of a polypeptide containing epitope A or the presence of free unlabeled A in a reaction containing both free labeled A and the binding molecule that binds thereto, will reduce the amount of labeled A that binds to the binding molecule. It is to be understood that specificity need not be absolute but generally refers to the context in which the binding occurs. For example, it is well known in the art that numerous antibodies cross-react with other epitopes in addition to those present in the target molecule. Such cross-reactivity may be acceptable depending upon the application for which the antibody is to be used. One of ordinary skill in the art will be able to select antibodies or ligands having a sufficient degree of specificity to perform appropriately in any given application (e.g., for detection of a target molecule, for therapeutic purposes, etc). It is also to be understood that specificity may be evaluated in the context of additional factors such as the affinity of the binding molecule for the target versus the affinity of the binding molecule for other targets, e.g., competitors. If a binding molecule exhibits a high affinity for a target molecule that it is desired to detect and low affinity for nontarget molecules, the antibody will likely be an acceptable reagent. Once the specificity of a binding molecule is established in one or more contexts, it may be employed in other, preferably similar, contexts without necessarily re-evaluating its specificity. In some embodiments, the affinity (as measured by the equilibrium dissociation constant, Kd) of two molecules that exhibit specific binding is 10^{-3} M or less, e.g., 10^{-4} M or less, e.g., 10^{-5} M or less, e.g., 10^{-6}M or less, 10^{-7}M or less, 10^{-8}M or less, or 10^{-9} M or less under the conditions tested, e.g., under physiological conditions.

[0040] A "subject" treated according to the instant invention is typically a human, a non-human primate, or a lower animal (e.g., a mouse or rat), which expresses or contains at least some primate (e.g., human) complement component C3 and, optionally, one or more additional primate complement component(s). In some embodiments the subject is male. In some embodiments the subject is female. In some embodiments the subject is an adult, e.g., a human at least 18 years of age, e.g., between 18 and 100 years of age. In some embodiments, a human subject is at least 12 years of age. In some embodiments a subject is an adult, e.g., a human at least 18 years of age, e.g., between 18 and 100 years of age. In some embodiments a subject is at least 40, 45, 50, 55, 60, 65, 70, 75, or 80 years of age. In some embodiments the subject is a child, e.g., a human between 0 and 4 years of age, or between 5 and 11 years of age.

[0041] "Treating", as used herein in regard to treating a subject, refers to providing treatment, i.e, providing any type of medical or surgical management of a subject. The treatment can be provided in order to reverse, alleviate, inhibit the progression of, prevent or reduce the likelihood of a disease, or in order to reverse, alleviate, inhibit or prevent the progression of, prevent or reduce the likelihood of one or more symptoms or manifestations of a disease. "Prevent" refers to causing a disease or symptom or manifestation of a disease not to occur for at least a period of time in at least some individuals. Treating can include administering a compound or composition to the subject following the development of one or more symptoms or manifestations indicative of a disease, e.g., in order to reverse, alleviate, reduce the severity of, and/or inhibit or prevent the progression of the disease and/or to reverse, alleviate, reduce the severity of, and/or inhibit or one or more symptoms or manifestations of the disease. A compound or composition can be administered to a subject who has developed a disease, or is at increased risk of developing the disease relative to a member of the general population. A compound or composition can be administered to a subject who has developed a disease and is at increased risk of developing one or more particular symptoms or manifestations of the disease or an exacerbation of the disease relative to other individuals diagnosed with the disease, or relative to the subject's typical or average risk for such symptom or manifestation or exacerbation. For example, the subject may have been exposed to a "trigger" that places the subject at increased risk (e.g., temporarily increased risk) of experiencing an exacerbation. A compound or composition can be administered prophylactically, i.e., before development of any symptom or manifestation of the disease. Typically in this case the subject will be at risk of developing

the disease, e.g., relative to a member of the general population, optionally matched in terms of age, sex, and/or other demographic variable(s).

[0042] A "vector" may be any of a variety of nucleic acid molecules, viruses, or portions thereof that are capable of mediating entry of, e.g., transferring, transporting, etc., a nucleic acid of interest between different genetic environments or into a cell. The nucleic acid of interest may be linked to, e.g., inserted into, the vector using, e.g., restriction and ligation. Vectors include, for example, DNA or RNA plasmids, cosmids, naturally occurring or modified viral genomes or portions thereof, nucleic acids that can be packaged into viral capsids, mini-chromosomes, artificial chromosomes, etc. Plasmid vectors typically include an origin of replication (e.g., for replication in prokaryotic cells). A plasmid may include part or all of a viral genome (e.g., a viral promoter, enhancer, processing or packaging signals, and/or sequences sufficient to give rise to a nucleic acid that can be integrated into the host cell genome and/or to give rise to infectious virus). Viruses or portions thereof that can be used to introduce nucleic acids into cells may be referred to as viral vectors. Viral vectors include, e.g., adenoviruses, adeno-associated viruses, retroviruses (e.g., lentiviruses, vaccinia virus and other poxviruses, herpesviruses (e.g., herpes simplex virus), and others. Baculovirus are of use, e.g., in insect cells. A wide range of plant viral vectors are known and include, e.g., those based on or comprising Cauliflower Mosaic Virus, Tobacco Mosaic Virus, or one or more genetic elements thereof (e.g., Cauliflower Mosaic Virus 35S promoter). Viral vectors may or may not contain sufficient viral genetic information for production of infectious virus when introduced into host cells, i.e., viral vectors may be replication-competent or replication-defective. In some embodiments, e.g., where sufficient information for production of infectious virus is lacking, it may be supplied by a host cell or by another vector introduced into the cell, e.g., if production of virus is desired. In some embodiments such information is not supplied, e.g., if production of virus is not desired. A nucleic acid to be transferred may be incorporated into a naturally occurring or modified viral genome or a portion thereof or may be present within a viral capsid as a separate nucleic acid molecule. A vector may contain one or more nucleic acids encoding a marker suitable for identifying and/or selecting cells that have taken up the vector. Markers include, for example, various proteins that increase or decrease either resistance or sensitivity to antibiotics or other agents (e.g., a protein that confers resistance to an antibiotic such as puromycin, hygromycin or blasticidin), enzymes whose activities are detectable by assays known in the art (e.g., β-galactosidase or alkaline phosphatase), and proteins or RNAs that detectably affect the phenotype of cells that express them (e.g., fluorescent proteins).

Vectors often include one or more appropriately positioned sites for restriction enzymes, which may be used to facilitate insertion into the vector of a nucleic acid, e.g., a nucleic acid to be expressed. An expression vector is a vector into which a desired nucleic acid has been inserted or may be inserted such that it is operably linked to regulatory elements (also termed "regulatory sequences", "expression control elements", or "expression control sequences") and may be expressed as an RNA transcript (e.g., an mRNA that can be translated into protein or a noncoding RNA). Expression vectors include regulatory sequence(s), e.g., expression control sequences, sufficient to direct transcription of an operably linked nucleic acid under at least some conditions; other elements required or helpful for expression may be supplied by, e.g., the host cell or by an in vitro expression system. Such regulatory sequences typically include a promoter and may include enhancer sequences or upstream activator sequences. In some embodiments a vector may include sequences that encode a 5' untranslated region and/or a 3' untranslated region, which may comprise a cleavage and/or polyadenylation signal. In general, regulatory elements may be contained in a vector prior to insertion of a nucleic acid whose expression is desired or may be contained in an inserted nucleic acid or may be inserted into a vector following insertion of a nucleic acid whose expression is desired. As used herein, a nucleic acid and regulatory element(s) are said to be "operably linked" when they are covalently linked so as to place the expression or transcription of the nucleic acid under the influence or control of the regulatory element(s). For example, a promoter region would be operably linked to a nucleic acid if the promoter region were capable of effecting transcription of that nucleic acid. One of ordinary skill in the art will be aware that the precise nature of the regulatory sequences useful for gene expression may vary between species or cell types, but may in general include, as appropriate, sequences involved with the initiation of transcription, RNA processing, or initiation of translation. The choice and design of an appropriate vector and regulatory element(s) is within the ability and discretion of one of ordinary skill in the art. For example, one of skill in the art will select an appropriate promoter (or other expression control sequences) for expression in a desired species (e.g., a prokaryotic (bacterial) or eukaryotic (e.g., fungal, plant, mammalian species) or cell type. A vector may contain a promoter capable of directing expression in mammalian cells, such as a suitable viral promoter, e.g., from a cytomegalovirus (CMV), retrovirus, simian virus (e.g., SV40), papilloma virus, herpes virus or other virus that infects mammalian cells, or a mammalian promoter from, e.g., a gene such as EF1alpha, ubiquitin (e.g., ubiquitin B or C), globin, actin, phosphoglycerate kinase (PGK), etc., or a composite promoter such as a CAG promoter (combination of the CMV

early enhancer element and chicken beta-actin promoter). In some embodiments a human promoter may be used. In some embodiments, a promoter that ordinarily directs transcription by a eukaryotic RNA polymerase I (a "pol I promoter"), e.g., (a U6, H1, 7SK or tRNA promoter or a functional variant thereof) may be used. In some embodiments, a promoter that ordinarily directs transcription by a eukaryotic RNA polymerase II (a "pol II promoter") or a functional variant thereof is used. In some embodiments, a promoter that ordinarily directs transcription by a eukaryotic RNA polymerase III (a "pol III promoter"), e.g., a promoter for transcription of ribosomal RNA (other than 5S rRNA) or a functional variant thereof is used. One of ordinary skill in the art will select an appropriate promoter for directing transcription of a sequence of interest. Examples of expression vectors that may be used in mammalian cells include, e.g., the pcDNA vector series, pSV2 vector series, pCMV vector series, pRSV vector series, pEF1 vector series, Gateway® vectors, etc. In some embodiments, regulatable (e.g., inducible or repressible) expression control element(s), e.g., a regulatable promoter, is/are used so that expression can be regulated, e.g., turned on or increased or turned off or decreased. In some embodiments a vector may comprise a polynucleotide sequence that encodes a polypeptide, wherein the polynucleotide sequence is positioned in frame with a nucleic acid inserted into the vector so that an N- or C-terminal fusion is created. In some embodiments a polypeptide encoded by the polynucleotide sequence may comprise a signal sequence (which directs secretion of a protein) or a sequence that directs the expressed protein to a specific organelle or location in the cell such as the nucleus or mitochondria. In some embodiments a polypeptide comprises a tag. A tag may be useful to facilitate detection and/or purification of a protein that contains it. Examples of tags include polyhistidine-tag (e.g., 6X-His tag), glutathione-S-transferase, maltose binding protein, NUS tag, SNUT tag, Strep tag, epitope tags such as V5, HA, Myc, or FLAG. In some embodiments a protease cleavage site is located in the region between the protein encoded by the inserted nucleic acid and the polypeptide, allowing the polypeptide to be removed by exposure to the protease. Vectors may be introduced into host cells using methods known in the art. One of ordinary skill will select an appropriate method based, e.g., on the vector, cell type, etc. Examples of suitable methods include, e.g., calcium phosphate-mediated transfection, transfection with any of a variety of commercially available reagents, e.g., lipid-based or non-lipid based, such as FuGENE, Lipofectamine, TurboFect; electroporation; microparticle bombardment, etc. Such methods are explained in detail in standard references such as Sambrook, Protocols series, and others.

[0043] As used herein the term "aliphatic" denotes a hydrocarbon moiety that may be straight-chain (i.e., unbranched), branched, or cyclic (including fused, bridging, and spiro-fused polycyclic) and may be completely saturated or may contain one or more units of unsaturation, but which is not aromatic. Unless otherwise specified, aliphatic groups contain 1–30 carbon atoms. In some embodiments, aliphatic groups contain 1–10 carbon atoms. In other embodiments, aliphatic groups contain 1–8 carbon atoms. In still other embodiments, aliphatic groups contain 1–6 carbon atoms, and in yet other embodiments aliphatic groups contain 1–4 carbon atoms. Suitable aliphatic groups include, but are not limited to, linear or branched, alkyl, alkenyl, and alkynyl groups, and hybrids thereof such as (cycloalkyl)alkyl, (cycloalkenyl)alkyl or (cycloalkyl)alkenyl.

[0044] As used herein, "alkyl" refers to a saturated straight, branched, or cyclic hydrocarbon having from about 1 to about 22 carbon atoms (and all combinations and subcombinations of ranges and specific numbers of carbon atoms therein), with from about 1 to about 12, or about 1 to about 7 carbon atoms being preferred in certain embodiments of the invention. Alkyl groups include, but are not limited to, methyl, ethyl, n-propyl, isopropyl, n-butyl, isobutyl, t-butyl, n- pentyl, cyclopentyl, isopentyl, neopentyl, n-hexyl, isohexyl, cyclohexyl, cyclooctyl, adamantyl, 3- methylpentyl, 2,2-dimethylbutyl, and 2,3-dimethylbutyl.

[0045] As used herein, "halo" refers to F, Cl, Br or I.

[0046] As used herein, "alkanoyl" refers to an optionally substituted straight or branched aliphatic acyclic residue having about 1 to 10 carbon atoms (and all combinations and subcombinations of ranges and specific number of carbon atoms) therein, e.g., from about 1 to 7 carbon atoms which, as will be appreciated, is attached to a terminal C=O group with a single bond (and may also be referred to as an "acyl group"). Alkanoyl groups include, but are not limited to, formyl, acetyl, propionyl, butyryl, isobutyryl, pentanoyl, isopentanoyl, 2-methyl-butyryl, 2,2-dimethoxypropionyl, hexanoyl, heptanoyl, octanoyl, and the like, and for purposes of the present invention a formyl group is considered an alkanoyl group. "Lower alkanoyl" refers to an optionally substituted straight or branched aliphatic acyclic residue having about 1 to about 5 carbon atoms (and all combinations and subcombinations of ranges and specific number of carbon atoms). Such groups include, but are not limited to, formyl, acetyl, propionyl, butyryl, isobutyryl, pentanoyl, isopentanoyl, etc.

[0047] As used herein, "aryl" refers to an optionally substituted, mono- or bicyclic aromatic ring system having from about 5 to about 14 carbon atoms (and all combinations and subcombinations of ranges and specific numbers of carbon atoms therein), with from

about 6 to about 10 carbons being preferred. Non-limiting examples include, for example, phenyl and naphthyl.

[0048] As used herein, "aralkyl" refers to alkyl radicals bearing an aryl substituent and having from about 6 to about 22 carbon atoms (and all combinations and subcombinations of ranges and specific numbers of carbon atoms therein), with from about 6 to about 12 carbon atoms being preferred in certain embodiments. Aralkyl groups can be optionally substituted. Non-limiting examples include, for example, benzyl, naphthylmethyl, diphenylmethyl, triphenylmethyl, phenylethyl, and diphenylethyl.

[0049] As used herein, the terms "alkoxy" and "alkoxyl" refer to an optionally substituted alkyl-O- group wherein alkyl is as previously defined. Exemplary alkoxy and alkoxyl groups include methoxy, ethoxy, n-propoxy, i-propoxy, n-butoxy, and heptoxy.

[0050] As used herein, "carboxy" refers to a -C(=O)OH group.

[0051] As used herein, "alkoxycarbonyl" refers to a -C(=O)O-alkyl group, where alkyl is as previously defined.

[0052] As used herein, "aroyl" refers to a -C(=O)-aryl group, wherein aryl is as previously defined. Exemplary aroyl groups include benzoyl and naphthoyl.

[0053] The term "cyclic ring system" refers to an aromatic or non-aromatic, partially unsaturated or fully saturated, 3- to 10-membered ring system, which includes single rings of 3 to 8 atoms in size and bi- and tri-cyclic ring systems which may include aromatic 5- or 6-membered aryl or aromatic heterocyclic groups fused to a non-aromatic ring. These heterocyclic rings include those having from 1 to 3 heteroatoms independently selected from the group consisting of oxygen, sulfur, and nitrogen. In certain embodiments, the term heterocyclic refers to a non-aromatic 5-, 6-, or 7-membered ring or a polycyclic group wherein at least one ring atom is a heteroatom selected from the group consisting of O, S, and N, including, but not limited to, a bi- or tri-cyclic group, comprising fused six-membered rings having between one and three heteroatoms independently selected from the group consisting of the oxygen, sulfur, and nitrogen. In some embodiments, "cyclic ring system" refers to a cycloalkyl group which, as used herein, refers to groups having 3 to 10, e.g., 4 to 7 carbon atoms. Cycloalkyls include, but are not limited to cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, cycloheptyl and the like, which, is optionally substituted. In some embodiments, "cyclic ring system" refers to a cycloalkenyl or cycloalkynyl moiety, which is optionally substituted.

[0054] Typically, substituted chemical moieties include one or more substituents that replace hydrogen. Exemplary substituents include, for example, halo, alkyl, cycloalkyl,

aralkyl, aryl, sulfhydryl, hydroxyl (-OH), alkoxyl, cyano (-CN), carboxyl (-COOH), -C(=O)O-alkyl, aminocarbonyl (-C(=O)NH_2), -N-substituted aminocarbonyl (-C(=O)NHR"), CF_3, CF_2CF_3, and the like. In relation to the aforementioned substituents, each moiety R'' can be, independently, any of H, alkyl, cycloalkyl, aryl, or aralkyl, for example.

[0055] As used herein, "L-amino acid" refers to any of the naturally occurring levorotatory alpha-amino acids normally present in proteins or the alkyl esters of those alpha-amino acids. The term "D-amino acid" refers to dextrorotatory alpha-amino acids. Unless specified otherwise, all amino acids referred to herein are L-amino acids.

[0056] As used herein, an "aromatic amino acid" is an amino acid that comprises at least one aromatic ring, e.g., it comprises an aryl group.

[0057] As used herein, an "aromatic amino acid analog" is an amino acid analog that comprises at least one aromatic ring, e.g., it comprises an aryl group.

[0058] **II. Overview**

[0059] The present invention provides cell-reactive compstatin analogs and methods of relating thereto, e.g., methods of use thereof. Cell-reactive compstatin analogs are compounds that comprise a compstatin analog moiety and a cell-reactive functional group that is capable of reacting with a functional group exposed at the surface of a cell, e.g., under physiological conditions, to form a covalent bond. The cell-reactive compstatin analog thus becomes covalently attached to the cell. Without wishing to be bound by any particular theory, a cell-tethered compstatin analog protects the cell from complement-mediated damage by, for example, binding to C3 (which may be in the form of C3(H_2O)) at the cell surface and/or in the vicinity of the cell and inhibiting C3 cleavage and activation, and/or by binding to C3b and inhibiting its deposition on the cell or participation in the complement activation cascade. In some aspects of the invention, isolated cells are contacted with a cell-reactive compstatin analog ex vivo (outside the body). In some aspects of the invention, the cells are present in an isolated tissue or organ, e.g., a tissue or organ to be transplanted into a subject. In some aspects of the invention, cells are contacted with a cell-reactive compstatin analog in vivo, by administering the cell-reactive compstatin analog to a subject. The cell-reactive compstatin analog becomes covalently attached to cells in vivo. In some aspects, the inventive approach protects cells, tissues, and/or organs from the deleterious effects of complement activation for at least two weeks, without need for retreatment during that time.

[0060] In some aspects, the invention provides compstatin analogs comprising a targeting moiety that binds non-covalently to a target molecule present at the surface of cells or tissues or to an extracellular substance not attached to cells or tissues. Such compstatin analogs are

referred to herein as "targeted compstatin analogs"). Often the target molecule is a protein or carbohydrate attached to the cell membrane and exposed at the cell surface. The targeting moiety targets the compstatin analog to a cell, tissue, or location susceptible to complement activation. In some aspects of the invention, isolated cells are contacted with a targeted compstatin analog ex vivo (outside the body). In some aspects of the invention, the cells are present in an isolated tissue or organ, e.g., a tissue or organ to be transplanted into a subject. In some aspects of the invention, a targeted compstatin analog is administered to a subject and becomes non-covalently attached to a cell, tissue, or extracellular substance in vivo. In some aspects, the inventive approach protects cells, tissues, and/or organs from the deleterious effects of complement activation for at least two weeks, without need for retreatment during that time. In some embodiments, a targeted compstatin analog comprises both a targeting moiety and a cell-reactive moiety. The targeting moiety targets the compstatin analog, e.g., to a particular cell type, by binding non-covalently to a molecule on such cells. The cell-reactive moiety then binds covalently to the cell or extracellular substance. In other embodiments, a targeted compstatin analog does not comprise a cell-reactive moiety.

[0061] In some aspects, the invention provides long-acting compstatin analogs, wherein the long-acting compstatin analogs comprise a moiety such as polyethylene glycol (PEG) that prolongs the lifetime of the compound in the body (e.g., by reducing its clearance from the blood). In some embodiments, a long-acting compstatin analog does not comprise a targeting moiety or a cell-reactive moiety. In some embodiments, a long-acting compstatin analog comprises a targeting moiety and/or a cell-reactive moiety.

[0062] **III. Complement System**

[0063] In order to facilitate understanding of the invention, and without intending to limit the invention in any way, this section provides an overview of complement and its pathways of activation. Further details are found, e.g., in *Kuby Immunology*, 6th ed., 2006; Paul, W.E., *Fundamental Immunology*, Lippincott Williams & Wilkins; 6th ed., 2008; and Walport MJ., Complement. First of two parts. *N Engl J Med.*, 344(14):1058-66, 2001.

[0064] Complement is an arm of the innate immune system that plays an important role in defending the body against infectious agents. The complement system comprises more than 30 serum and cellular proteins that are involved in three major pathways, known as the classical, alternative, and lectin pathways. The classical pathway is usually triggered by binding of a complex of antigen and IgM or IgG antibody to C1 (though certain other activators can also initiate the pathway). Activated C1 cleaves C4 and C2 to produce C4a

and C4b, in addition to C2a and C2b. C4b and C2a combine to form C3 convertase, which cleaves C3 to form C3a and C3b. Binding of C3b to C3 convertase produces C5 convertase, which cleaves C5 into C5a and C5b. C3a, C4a, and C5a are anaphylotoxins and mediate multiple reactions in the acute inflammatory response. C3a and C5a are also chemotactic factors that attract immune system cells such as neutrophils.

[0065] The alternative pathway is initiated by and amplified at, e.g., microbial surfaces and various complex polysaccharides. In this pathway, hydrolysis of C3 to C3(H2O), which occurs spontaneously at a low level, leads to binding of factor B, which is cleaved by factor D, generating a fluid phase C3 convertase that activates complement by cleaving C3 into C3a and C3b. C3b binds to targets such as cell surfaces and forms a complex with factor B, which is later cleaved by factor D, resulting in a C3 convertase. Surface-bound C3 convertases cleave and activate additional C3 molecules, resulting in rapid C3b deposition in close proximity to the site of activation and leading to formation of additional C3 convertase, which in turn generates additional C3b. This process results in a cycle of C3 cleavage and C3 convertase formation that signicantly amplifies the response. Cleavage of C3 and binding of another molecule of C3b to the C3 convertase gives rise to a C5 convertase. C3 and C5 convertases of this pathway are regulated by host cell molecules CR1, DAF, MCP, CD59, and fH. The mode of action of these proteins involves either decay accelerating activity (i.e., ability to dissociate convertases), ability to serve as cofactors in the degradation of C3b or C4b by factor I, or both. Normally the presence of complement regulatory proteins on host cell surfaces prevents significant complement activation from occurring thereon.

[0066] The C5 convertases produced in both pathways cleave C5 to produce C5a and C5b. C5b then binds to C6, C7, and C8 to form C5b-8, which catalyzes polymerization of C9 to form the C5b-9 membrane attack complex (MAC). The MAC inserts itself into target cell membranes and causes cell lysis. Small amounts of MAC on the membrane of cells may have a variety of consequences other than cell death.

[0067] The lectin complement pathway is initiated by binding of mannose-binding lectin (MBL) and MBL-associated serine protease (MASP) to carbohydrates. The MB1-1 gene (known as LMAN-1 in humans) encodes a type I integral membrane protein localized in the intermediate region between the endoplasmic reticulum and the Golgi. The MBL-2 gene encodes the soluble mannose-binding protein found in serum. In the human lectin pathway, MASP-1 and MASP-2 are involved in the proteolysis of C4 and C2, leading to a C3 convertase described above.

[0068] Complement activity is regulated by various mammalian proteins referred to as complement control proteins (CCPs) or regulators of complement activation (RCA) proteins (U.S. Pat. No. 6,897,290). These proteins differ with respect to ligand specificity and mechanism(s) of complement inhibition. They may accelerate the normal decay of convertases and/or function as cofactors for factor I, to enzymatically cleave C3b and/or C4b into smaller fragments. CCPs are characterized by the presence of multiple (typically 4-56) homologous motifs known as short consensus repeats (SCR), complement control protein (CCP) modules, or SUSHI domains, about 50-70 amino acids in length that contain a conserved motif including four disulfide-bonded cysteines (two disulfide bonds), proline, tryptophan, and many hydrophobic residues. The CCP family includes complement receptor type 1 (CR1; C3b:C4b receptor), complement receptor type 2 (CR2), membrane cofactor protein (MCP; CD46), decay-accelerating factor (DAF), complement factor H (fH), and C4b-binding protein (C4bp). CD59 is a membrane-bound complement regulatory protein unrelated structurally to the CCPs. Complement regulatory proteins normally serve to limit complement activation that might otherwise occur on cells and tissues of the mammalian, e.g., human host. Thus, "self" cells are normally protected from the deleterious effects that would otherwise ensue were complement activation to proceed on these cells. Defiencies or defects in complement regulatory protein(s) are involved in the pathogenesis of a variety of complement-mediated disorders, e.g., as discussed herein.

[0069] **IV. Compstatin Analogs**

[0070] Compstatin is a cyclic peptide that binds to C3 and inhibits complement activation. U.S. Pat. No. 6,319,897 describes a peptide having the sequence Ile- [Cys-Val-Val-Gln-Asp-Trp-Gly-His-His-Arg-Cys]-Thr (SEQ ID NO: 1), with the disulfide bond between the two cysteines denoted by brackets. It will be understood that the name "compstatin" was not used in U.S. Pat. No. 6,319,897 but was subsequently adopted in the scientific and patent literature (see, e.g., Morikis, et al., *Protein Sci.*, 7(3):619-27, 1998) to refer to a peptide having the same sequence as SEQ ID NO: 2 disclosed in U.S. Pat. No. 6,319,897, but amidated at the C terminus as shown in Table 1 (SEQ ID NO: 8). The term "compstatin" is used herein consistently with such usage (i.e., to refer to SEQ ID NO: 8). Compstatin analogs that have higher complement inhibiting activity than compstatin have been developed. See, e.g., WO2004/026328 (PCT/US2003/029653), Morikis, D., et al., *Biochem Soc Trans.* 32(Pt 1):28-32, 2004, Mallik, B., et al., *J. Med. Chem.*, 274-286, 2005; Katragadda, M., et al. *J. Med. Chem.*, 49: 4616-4622, 2006; WO2007062249

(PCT/US2006/045539); WO2007044668 (PCT/US2006/039397), WO/2009/046198 (PCT/US2008/078593); WO/2010/127336 (PCT/US2010/033345) and discussion below.

[0071] Compstatin analogs may be acetylated or amidated, e.g., at the N-terminus and/or C-terminus. For example, compstatin analogs may be acetylated at the N-terminus and amidated at the C-terminus. Consistent with usage in the art, "compstatin" as used herein, and the activities of compstatin analogs described herein relative to that of compstatin, refer to compstatin amidated at the C-terminus (Mallik, 2005, *supra*).

[0072] Concatamers or multimers of compstatin or a complement inhibiting analog thereof are also of use in the present invention.

[0073] As used herein, the term "compstatin analog" includes compstatin and any complement inhibiting analog thereof. The term "compstatin analog" encompasses compstatin and other compounds designed or identified based on compstatin and whose complement inhibiting activity is at least 50% as great as that of compstatin as measured, e.g., using any complement activation assay accepted in the art or substantially similar or equivalent assays. Certain suitable assays are described in U.S. Pat. No. 6,319,897, WO2004/026328, Morikis, *supra*, Mallik, *supra*, Katragadda 2006, *supra*,WO2007062249 (PCT/US2006/045539); WO2007044668 (PCT/US2006/039397), WO/2009/046198 (PCT/US2008/078593); and/or WO/2010/127336 (PCT/US2010/033345). The assay may, for example, measure alternative or classical pathway-mediated erythrocyte lysis or be an ELISA assay. In some embodiments, an assay described in WO/2010/135717 (PCT/US2010/035871) is used.

[0074] The activity of a compstatin analog may be expressed in terms of its IC_{50} (the concentration of the compound that inhibits complement activation by 50%), with a lower IC_{50} indicating a higher activity as recognized in the art. The activity of a preferred compstatin analog for use in the present invention is at least as great as that of compstatin. It is noted that certain modifications known to reduce or eliminate complement inhibiting activity and may be explicitly excluded from any embodiment of the invention. The IC_{50} of compstatin has been measured as 12 μM using an alternative pathway-mediated erythrocyte lysis assay (WO2004/026328). It will be appreciated that the precise IC_{50} value measured for a given compstatin analog will vary with experimental conditions (e.g., the serum concentration used in the assay). Comparative values, e.g., obtained from experiments in which IC_{50} is determined for multiple different compounds under substantially identical conditions, are of use. In one embodiment, the IC_{50} of the compstatin analog is no more than

the IC_{50} of compstatin. In certain embodiments of the invention the activity of the compstatin analog is between 2 and 99 times that of compstatin (i.e., the analog has an IC_{50} that is less than the IC_{50} of compstatin by a factor of between 2 and 99). For example, the activity may be between 10 and 50 times as great as that of compstatin, or between 50 and 99 times as great as that of compstatin. In certain embodiments of the invention the activity of the compstatin analog is between 99 and 264 times that of compstatin. For example, the activity may be 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, or 264 times as great as that of compstatin. In certain embodiments the activity is between 250 and 300, 300 and 350, 350 and 400, or 400 and 500 times as great as that of compstatin. The invention further contemplates compstatin analogs having activities between 500 and 1000 times that of compstatin, or more. In certain embodiments the IC_{50} of the compstatin analog is between about 0.2 μM and about 0.5 μM. In certain embodiments the IC_{50} of the compstatin analog is between about 0.1 μM and about 0.2 μM. In certain embodiments the IC_{50} of the compstatin analog is between about 0.05 μM and about 0.1 μM. In certain embodiments the IC_{50} of the compstatin analog is between about 0.001 μM and about 0.05 μM.

[0075] The K_d of compstatin binding to C3 can be measured using isothermal titration calorimetry (Katragadda, et al., *J. Biol. Chem.*, 279(53), 54987-54995, 2004). Binding affinity of a variety of compstatin analogs for C3 has been correlated with their activity, with a lower K_d indicating a higher binding affinity, as recognized in the art. A linear correlation between binding affinity and activity was shown for certain analogs tested (Katragadda, 2004, *supra*; Katragadda 2006, *supra*). In certain embodiments of the invention the compstatin analog binds to C3 with a K_d of between 0.1 μM and 1.0 μM, between 0.05 μM and 0.1 μM, between 0.025 μM and 0.05 μM, between 0.015 μM and 0.025 μM, between 0.01 μM and 0.015 μM, or between 0.001 μM and 0.01μM.

[0076] Compounds "designed or identified based on compstatin" include, but are not limited to, compounds that comprise an amino acid chain whose sequence is obtained by (i) modifying the sequence of compstatin (e.g., replacing one or more amino acids of the sequence of compstatin with a different amino acid or amino acid analog, inserting one or more amino acids or amino acid analogs into the sequence of compstatin, or deleting one or more amino acids from the sequence of compstatin); (ii) selection from a phage display peptide library in which one or more amino acids of compstatin is randomized, and optionally further modified according to method (i); or (iii) identified by screening for compounds that

compete with compstatin or any analog thereof obtained by methods (i) or (ii) for binding to C3 or a fragment thereof. Many useful compstatin analogs comprise a hydrophobic cluster, a β-turn, and a disulfide bridge.

[0077] In certain embodiments of the invention the sequence of the compstatin analog comprises or consists essentially of a sequence that is obtained by making 1, 2, 3, or 4 substitutions in the sequence of compstatin, i.e., 1, 2, 3, or 4 amino acids in the sequence of compstatin is replaced by a different standard amino acid or by a non-standard amino acid. In certain embodiments of the invention the amino acid at position 4 is altered. In certain embodiments of the invention the amino acid at position 9 is altered. In certain embodiments of the invention the amino acids at positions 4 and 9 are altered. In certain embodiments of the invention only the amino acids at positions 4 and 9 are altered. In certain embodiments of the invention the amino acid at position 4 or 9 is altered, or in certain embodiments both amino acids 4 and 9 are altered, and in addition up to 2 amino acids located at positions selected from 1, 7, 10, 11, and 13 are altered. In certain embodiments of the invention the amino acids at positions 4, 7, and 9 are altered. In certain embodiments of the invention amino acids at position 2, 12, or both are altered, provided that the alteration preserves the ability of the compound to be cyclized. Such alteration(s) at positions 2 and/or 12 may be in addition to the alteration(s) at position 1, 4, 7, 9, 10, 11, and/or 13. Optionally the sequence of any of the compstatin analogs whose sequence is obtained by replacing one or more amino acids of compstatin sequence further includes up to 1, 2, or 3 additional amino acids at the C-terminus. In one embodiment, the additional amino acid is Gly. Optionally the sequence of any of the compstatin analogs whose sequence is obtained by replacing one or more amino acids of compstatin sequence further includes up to 5, or up to 10 additional amino acids at the C-terminus. It should be understood that compstatin analogs may have any one or more of the characteristics or features of the various embodiments described herein, and characteristics or features of any embodiment may additionally characterize any other embodiment described herein, unless otherwise stated or evident from the context. In certain embodiments of the invention the sequence of the compstatin analog comprises or consists essentially of a sequence identical to that of compstatin except at positions corresponding to positions 4 and 9 in the sequence of compstatin.

[0078] Compstatin and certain compstatin analogs having somewhat greater activity than compstatin contain only standard amino acids ("standard amino acids" are glycine, leucine, isoleucine, valine, alanine, phenylalanine, tyrosine, tryptophan, aspartic acid, asparagine,

glutamic acid, glutamine, cysteine, methionine, arginine, lysine, proline, serine, threonine and histidine). Certain compstatin analogs having improved activity incorporate one or more non-standard amino acids. Useful non-standard amino acids include singly and multiply halogenated (e.g., fluorinated) amino acids, D-amino acids, homo-amino acids, N-alkyl amino acids, dehydroamino acids, aromatic amino acids (other than phenylalanine, tyrosine and tryptophan), ortho-, meta- or para-aminobenzoic acid, phospho-amino acids, methoxylated amino acids, and α,α-disubstituted amino acids. In certain embodiments of the invention, a compstatin analog is designed by replacing one or more L-amino acids in a compstatin analog described elsewhere herein with the corresponding D-amino acid. Such compounds and methods of use thereof are an aspect of the invention. Exemplary non-standard amino acids of use include 2-naphthylalanine (2-NaI), 1-naphthylalanine (1-NaI), 2-indanylglycine carboxylic acid (2Ig1), dihydrotrpytophan (Dht), 4-benzoyl-L-phenylalanine (Bpa), 2-α-aminobutyric acid (2-Abu), 3-α-aminobutyric acid (3-Abu), 4-α-aminobutyric acid (4-Abu), cyclohexylalanine (Cha), homocyclohexylalanine (hCha), 4-fluoro-L-tryptophan (4fW), 5-fluoro-L-tryptophan (5fW), 6-fluoro-L-tryptophan (6fW), 4-hydroxy-L-tryptophan (4OH-W), 5-hydroxy-L-tryptophan (5OH-W), 6-hydroxy-L-tryptophan (6OH-W), 1-methyl-L-tryptophan (1MeW), 4-methyl-L-tryptophan (4MeW), 5-methyl-L-tryptophan (5MeW), 7-aza-L-tryptophan (7aW), α-methyl-L-tryptophan (αMeW), β-methyl-L-tryptophan (βMeW), N-methyl-L-tryptophan (NMeW), ornithine (orn), citrulline, norleucine, γ-glutamic acid, etc.

[0079] In certain embodiments of the invention the compstatin analog comprises one or more Trp analogs (e.g., at position 4 and/or 7 relative to the sequence of compstatin). Exemplary Trp analogs are mentioned above. See also Beene, et. al. *Biochemistry* 41: 10262-10269, 2002 (describing, *inter alia*, singly- and multiply-halogenated Trp analogs); Babitzke & Yanofsky, *J. Biol. Chem.* 270: 12452-12456, 1995 (describing, *inter alia*, methylated and halogenated Trp and other Trp and indole analogs); and U.S. Patents 6,214,790, 6,169,057, 5,776,970, 4,870,097, 4,576,750 and 4,299,838. Other Trp analogs include variants that are substituted (e.g., by a methyl group) at the α or β carbon and, optionally, also at one or more positions of the indole ring. Amino acids comprising two or more aromatic rings, including substituted, unsubstituted, or alternatively substituted variants thereof, are of interest as Trp analogs. In certain embodiments of the invention the Trp analog, e.g., at position 4, is 5-methoxy, 5-methyl-, 1-methyl-, or 1-formyl-tryptophan. In certain embodiments of the invention a Trp analog (e.g., at position 4) comprising a 1-alkyl substituent, e.g., a lower

alkyl (e.g., C_1-C_5) substituent is used. In certain embodiments, N(α) methyl tryptophan or 5-methyltryptophan is used. In some embodiments, an analog comprising a 1-alkanyol substituent, e.g., a lower alkanoyl (e.g., C_1-C_5) is used. Examples include 1-acetyl-L-tryptophan and L-β-tryptophan.

[0080] In certain embodiments the Trp analog has increased hydrophobic character relative to Trp. For example, the indole ring may be substituted by one or more alkyl (e.g., methyl) groups. In certain embodiments the Trp analog participates in a hydrophobic interaction with C3. Such a Trp analog may be located, e.g., at position 4 relative to the sequence of compstatin. In certain embodiments the Trp analog comprises a substituted or unsubstituted bicyclic aromatic ring component or two or more substituted or unsubstituted monocyclic aromatic ring components.

[0081] In certain embodiments the Trp analog has increased propensity to form hydrogen bonds with C3 relative to Trp but does not have increased hydrophobic character relative to Trp. The Trp analog may have increased polarity relative to Trp and/or an increased ability to participate in an electrostatic interaction with a hydrogen bond donor on C3. Certain exemplary Trp analogs with an increased hydrogen bond forming character comprise an electronegative substituent on the indole ring. Such a Trp analog may be located, e.g., at position 7 relative to the sequence of compstatin.

[0082] In certain embodiments of the invention the compstatin analog comprises one or more Ala analogs (e.g., at position 9 relative to the sequence of compstatin), e.g., Ala analogs that are identical to Ala except that they include one or more CH_2 groups in the side chain. In certain embodiments the Ala analog is an unbranched single methyl amino acid such as 2-Abu. In certain embodiments of the invention the compstatin analog comprises one or more Trp analogs (e.g., at position 4 and/or 7 relative to the sequence of compstatin) and an Ala analog (e.g., at position 9 relative to the sequence of compstatin).

[0083] In certain embodiments of the invention the compstatin analog is a compound that comprises a peptide that has a sequence of $(X'aa)_n$- Gln - Asp – Xaa – Gly-$(X"aa)_m$, (SEQ ID NO: 2) wherein each X'aa and each X"aa is an independently selected amino acid or amino acid analog, wherein Xaa is Trp or an analog of Trp, and wherein n>1 and m>1 and n+m is between 5 and 21. The peptide has a core sequence of Gln - Asp – Xaa – Gly, where Xaa is Trp or an analog of Trp, e.g., an analog of Trp having increased propensity to form hydrogen bonds with an H-bond donor relative to Trp but, in certain embodiments, not having increased hydrophobic character relative to Trp. For example, the analog may be one in

which the indole ring of Trp is substituted with an electronegative moiety, e.g., a halogen such as fluorine. In one embodiment Xaa is 5-fluorotryptophan. Absent evidence to the contrary, one of skill in the art would recognize that any non-naturally occurring peptide whose sequence comprises this core sequence and that inhibits complement activation and/or binds to C3 will have been designed based on the sequence of compstatin. In an alternative embodiment Xaa is an amino acid or amino acid analog other than a Trp analog that allows the Gln - Asp – Xaa – Gly peptide to form a β-turn.

[0084] In certain embodiments of the invention the peptide has a core sequence of X'aa-Gln - Asp – Xaa – Gly (SEQ ID NO: 3), where X'aa and Xaa are selected from Trp and analogs of Trp. In certain embodiments of the invention the peptide has a core sequence of X'aa-Gln - Asp – Xaa – Gly (SEQ ID NO: 3), where X'aa and Xaa are selected from Trp, analogs of Trp, and other amino acids or amino acid analogs comprising at least one aromatic ring. In certain embodiments of the invention the core sequence forms a β-turn in the context of the peptide. The β–turn may be flexible, allowing the peptide to assume two or more conformations as assessed for example, using nuclear magnetic resonance (NMR). In certain embodiments X'aa is an analog of Trp that comprises a substituted or unsubstituted bicyclic aromatic ring component or two or more substituted or unsubstituted monocyclic aromatic ring components. In certain embodiments of the invention X'aa is selected from the group consisting of 2-napthylalanine, 1-napthylalanine, 2-indanylglycine carboxylic acid, dihydrotryptophan, and benzoylphenylalanine. In certain embodiments of the invention X'aa is an analog of Trp that has increased hydrophobic character relative to Trp. For example, X'aa may be 1-methyltryptophan. In certain embodiments of the invention Xaa is an analog of Trp that has increased propensity to form hydrogen bonds relative to Trp but, in certain embodiments, not having increased hydrophobic character relative to Trp. In certain embodiments of the invention the analog of Trp that has increased propensity to form hydrogen bonds relative to Trp comprises a modification on the indole ring of Trp, e.g., at position 5, such as a substitution of a halogen atom for an H atom at position 5. For example, Xaa may be 5-fluorotryptophan.

[0085] In certain embodiments of the invention the peptide has a core sequence of X'aa-Gln - Asp – Xaa – Gly-X"aa (SEQ ID NO: 4), where X'aa and Xaa are each independently selected from Trp and analogs of Trp and X"aa is selected from His, Ala, analogs of Ala, Phe, and Trp. In certain embodiments of the invention X'aa is an analog of Trp that has increased hydrophobic character relative to Trp, such as 1-methyltryptophan or another Trp

analog having an alkyl substituent on the indole ring (e.g., at position 1, 4, 5, or 6). In certain embodiments X'aa is an analog of Trp that comprises a substituted or unsubstituted bicyclic aromatic ring component or two or more substituted or unsubstituted monocyclic aromatic ring components. In certain embodiments of the invention X'aa is selected from the group consisting of 2-napthylalanine, 1-napthylalanine, 2-indanylglycine carboxylic acid, dihydrotryptophan, and benzoylphenylalanine. In certain embodiments of the invention Xaa is an analog of Trp that has increased propensity to form hydrogen bonds with C3 relative to Trp but, in certain embodiments, not having increased hydrophobic character relative to Trp. In certain embodiments of the invention the analog of Trp that has increased propensity to form hydrogen bonds relative to Trp comprises a modification on the indole ring of Trp, e.g., at position 5, such as a substitution of a halogen atom for an H atom at position 5. For example, Xaa may be 5-fluorotryptophan. In certain embodiments X"aa is Ala or an analog of Ala such as Abu or another unbranched single methyl amino acid. In certain embodiments of the invention the peptide has a core sequence of X'aa-Gln - Asp – Xaa – Gly-X"aa (SEQ ID NO: 4), where X'aa and Xaa are each independently selected from Trp, analogs of Trp, and amino acids or amino acid analogs comprising at least one aromatic side chain, and X"aa is selected from His, Ala, analogs of Ala, Phe, and Trp. In certain embodiments X"aa is selected from analogs of Trp, aromatic amino acids, and aromatic amino acid analogs.

[0086] In certain preferred embodiments of the invention the peptide is cyclic. The peptide may be cyclized via a bond between any two amino acids, one of which is $(X'aa)_n$ and the other of which is located within $(X"aa)_m$. In certain embodiments the cyclic portion of the peptide is between 9 and 15 amino acids in length, e.g., 10-12 amino acids in length. In certain embodiments the cyclic portion of the peptide is 11 amino acids in length, with a bond (e.g., a disulfide bond) between amino acids at positions 2 and 12. For example, the peptide may be 13 amino acids long, with a bond between amino acids at positions 2 and 12 resulting in a cyclic portion 11 amino acids in length.

[0087] In certain embodiments the peptide comprises or consists of the sequence X'aa1 - X'aa2 - X'aa3 - X'aa4 -Gln-Asp-Xaa-Gly- X"aa1- X"aa2- X"aa3- X"aa4- X"aa5 (SEQ ID NO: 5). In certain embodiments X'aa4 and Xaa are selected from Trp and analogs of Trp, and X'aa1, X'aa2, X'aa3, X"aa1, X"aa2, X"aa3, X"aa4, and X"aa5 are independently selected from among amino acids and amino acid analogs. In certain embodiments X'aa4 and Xaa are selected from aromatic amino acids and aromatic amino acid analogs. Any one or more of X'aa1, X'aa2, X'aa3, X"aa1, X"aa2, X"aa3, X"aa4, and X"aa5 may be identical to the amino acid at the corresponding position in compstatin. In one embodiment, X"aa1 is Ala or a

single methyl unbranched amino acid. The peptide may be cyclized via a covalent bond between (i) X'aa1, X'aa2, or X'aa3; and (ii) X"aa2, X"aa3, X"aa4 or X"aa5. In one embodiment the peptide is cyclized via a covalent bond between X'aa2 and X"aa4. In one embodiment the covalently bound amino acid are each Cys and the covalent bond is a disulfide (S-S) bond. In other embodiments the covalent bond is a C-C, C-O, C-S, or C-N bond. In certain embodiments one of the covalently bound residues is an amino acid or amino acid analog having a side chain that comprises a primary or secondary amine, the other covalently bound residue is an amino acid or amino acid analog having a side chain that comprises a carboxylic acid group, and the covalent bond is an amide bond. Amino acids or amino acid analogs having a side chain that comprises a primary or secondary amine include lysine and diaminocarboxylic acids of general structure $NH_2(CH_2)_nCH(NH_2)COOH$ such as 2,3-diaminopropionic acid (dapa), 2,4-diaminobutyric acid (daba), and ornithine (orn), wherein n = 1 (dapa), 2 (daba), and 3 (orn), respectively. Examples of amino acids having a side chain that comprises a carboxylic acid group include dicarboxylic amino acids such as glutamic acid and aspartic acid. Analogs such as beta-hydroxy-L-glutamic acid may also be used. In some embodiments a peptide is cyclized with a thioether bond, e.g., as described in PCT/US2011/052442 (WO/2012/040259). For example, in some embodiments a disulfide bond in any of the peptides is replaced with a thioether bond. In some embodiments, a cystathionine is formed. In some embodiments the cystathionine is a delta-cystathionine or a gamma-cystathionine. In some embodiments a modification comprises replacement of a Cys-Cys disulfide bond between cysteines at X'aa2 and X"aa4 in SEQ ID NO: 5 (or corresponding positions in other sequences) with addition of a CH_2, to form a homocysteine at X'aa2 or X"aa4, and introduction of a thioether bond, to form a cystathionine. In one embodiment, the cystathionine is a gamma-cystathionine. In another embodiment, the cystathionine is a delta-cystathionine. Another modification in accordance with the present invention comprises replacement of the disulfide bond with a thioether bond without the addition of a CH_2, thereby forming a lantithionine. In some embodiments a compstatin analog having a thioether in place of a disulfide bond has increased stability, at least under some conditions, as compared with the compstatin analog having the disulfide bond.

[0088] In certain embodiments, the compstatin analog is a compound that comprises a peptide having a sequence:

[0089] Xaa1 – Cys – Val – Xaa2 - Gln - Asp – Xaa2* - Gly – Xaa3 - His - Arg – Cys – Xaa4 (SEQ ID NO: 6); wherein:

Xaa1 is Ile, Val, Leu, B^1-Ile, B^1-Val, B^1-Leu or a dipeptide comprising Gly-Ile or B^1-Gly-Ile, and B^1 represents a first blocking moiety;
Xaa2 and Xaa2* are independently selected from Trp and analogs of Trp;
Xaa3 is His, Ala or an analog of Ala, Phe, Trp, or an analog of Trp;
Xaa4 is L-Thr, D-Thr, Ile, Val, Gly, a dipeptide selected from Thr-Ala and Thr-Asn, or a tripeptide comprising Thr-Ala-Asn, wherein a carboxy terminal –OH of any of the L-Thr, D-Thr, Ile, Val, Gly, Ala, or Asn optionally is replaced by a second blocking moiety B^2; and the two Cys residues are joined by a disulfide bond. In some embodiments, Xaa4 is Leu, Nle, His, or Phe or a depeptide selected from Xaa5-Ala and Xaa5-Asn, or a tripeptide Xaa5-Ala-Asn, wherein Xaa5 is selected from Leu, Nle, His or Phe, and wherein a carboxy terminal –OH of any of the L-Thr, D-Thr, Ile, Val, Gly, Leu, Nle, His, Phe, Ala, or Asn optionally is replaced by a second blocking moiety B^2; and the two Cys residues are joined by a disulfide bond.

[0090] In other embodiments Xaa1 is absent or is any amino acid or amino acid analog, and Xaa2, Xaa2*, Xaa3, and Xaa4 are as defined above. If Xaa1 is absent, the N-terminal Cys residue may have a blocking moiety B^1 attached thereto.

[0091] In another embodiment, Xaa4 is any amino acid or amino acid analog and Xaa1, Xaa2, Xaa2*, and Xaa3 are as defined above. In another embodiment Xaa4 is a dipeptide selected from the group consisting of: Thr-Ala and Thr-Asn, wherein the carboxy terminal –OH or the Ala or Asn is optionally replaced by a second blocking moiety B^2.

[0092] In any of the embodiments of the compstatin analog of SEQ ID NO: 6, Xaa2 may be Trp.

[0093] In any of the embodiments of the compstatin analog of SEQ ID NO: 6, Xaa2 may be an analog of Trp comprising a substituted or unsubstituted bicyclic aromatic ring component or two or more substituted or unsubstituted monocyclic aromatic ring components. For example, the analog of Trp may be selected from 2-naphthylalanine (2-NaI), 1-naphthylalanine (1-NaI), 2-indanylglycine carboxylic acid (Ig1), dihydrotrpytophan (Dht), and 4-benzoyl-L-phenylalanine.

[0094] In any of the embodiments of the compstatin analog of SEQ ID NO: 6, Xaa2 may be an analog of Trp having increased hydrophobic character relative to Trp. For example, the analog of Trp may be selected from 1-methyltryptophan, 4-methyltryptophan, 5-methyltryptophan, and 6-methyltryptophan. In one embodiment, the analog of Trp is 1-

methyltryptophan. In one embodiment, Xaa2 is 1-methyltryptophan, Xaa2* is Trp, Xaa3 is Ala, and the other amino acids are identical to those of compstatin.

[0095] In any of the embodiments of the compstatin analog of SEQ ID NO: 6, Xaa2* may be an analog of Trp such as an analog of Trp having increased hydrogen bond forming propensity with C3 relative to Trp, which, in certain embodiments, does not have increased hydrophobic character relative to Trp. In certain embodiments the analog of Trp comprises an electronegative substituent on the indole ring. For example, the analog of Trp may be selected from 5-fluorotryptophan and 6-fluorotryptophan.

[0096] In certain embodiments of the invention Xaa2 is Trp and Xaa2* is an analog of Trp having increased hydrogen bond forming propensity with C3 relative to Trp which, in certain embodiments, does not have increased hydrophobic character relative to Trp. In certain embodiments of the compstatin analog of SEQ ID NO: 6, Xaa2 is analog of Trp having increased hydrophobic character relative to Trp such as an analog of Trp selected from 1-methyltryptophan, 4-methyltryptophan, 5-methyltryptophan, and 6-methyltryptophan, and and Xaa2* is an analog of Trp having increased hydrogen bond forming propensity with C3 relative to Trp which, in certain embodiments, does not have increased hydrophobic character relative to Trp. For example, in one embodiment Xaa2 is methyltryptophan and Xaa2* is 5-fluorotryptophan.

[0097] In certain of the afore-mentioned embodiments, Xaa3 is Ala. In certain of the afore-mentioned embodiments Xaa3 is a single methyl unbranched amino acid, e.g., Abu.

[0098] The invention further provides compstatin analogs of SEQ ID NO: 6, as described above, wherein Xaa2 and Xaa2* are independently selected from Trp, analogs of Trp, and other amino acids or amino acid analogs that comprise at least one aromatic ring, and Xaa3 is His, Ala or an analog of Ala, Phe, Trp, an analog of Trp, or another aromatic amino acid or aromatic amino acid analog.

[0099] In certain embodiments of the invention the blocking moiety present at the N- or C-terminus of any of the compstatin analogs described herein is any moiety that stabilizes a peptide against degradation that would otherwise occur in mammalian (e.g., human or non-human primate) blood or interstitial fluid. For example, blocking moiety B^1 could be any moiety that alters the structure of the N-terminus of a peptide so as to inhibit cleavage of a peptide bond between the N-terminal amino acid of the peptide and the adjacent amino acid. Blocking moiety B^2 could be any moiety that alters the structure of the C-terminus of a peptide so as to inhibit cleavage of a peptide bond between the C-terminal amino acid of the

peptide and the adjacent amino acid. Any suitable blocking moieties known in the art could be used. In certain embodiments of the invention blocking moiety B^1 comprises an acyl group (i.e., the portion of a carboxylic acid that remains following removal of the –OH group). The acyl group typically comprises between 1 and 12 carbons, e.g., between 1 and 6 carbons. For example, in certain embodiments of the invention blocking moiety B^1 is selected from the group consisting of: formyl, acetyl, proprionyl, butyryl, isobutyryl, valeryl, isovaleryl, etc. In one embodiment, the blocking moiety B^1 is an acetyl group, i.e., Xaa1 is Ac-Ile, Ac-Val, Ac-Leu, or Ac-Gly-Ile.

[00100] In certain embodiments of the invention blocking moiety B^2 is a primary or secondary amine ($–NH_2$ or $–NHR^1$, wherein R is an organic moiety such as an alkyl group).

[00101] In certain embodiments of the invention blocking moiety B^1 is any moiety that neutralizes or reduces the positive charge that may otherwise be present at the N-terminus at physiological pH. In certain embodiments of the invention blocking moiety B^2 is any moiety that neutralizes or reduces the negative charge that may otherwise be present at the C-terminus at physiological pH.

[00102] In certain embodiments of the invention, the compstatin analog is acetylated or amidated at the N-terminus and/or C-terminus, respectively. A compstatin analog may be acetylated at the N-terminus, amidated at the C-terminus, and or both acetylated at the N-terminus and amidated at the C-terminus. In certain embodiments of the invention a compstatin analog comprises an alkyl or aryl group at the N-terminus rather than an acetyl group.

[00103] In certain embodiments, the compstatin analog is a compound that comprises a peptide having a sequence:

[00104] Xaa1 – Cys – Val – Xaa2 - Gln - Asp – Xaa2* - Gly – Xaa3 - His - Arg – Cys – Xaa4 (SEQ ID NO: 7); wherein:

Xaa1 is Ile, Val, Leu, Ac-Ile, Ac-Val, Ac-Leu or a dipeptide comprising Gly-Ile or Ac-Gly-Ile;

Xaa2 and Xaa2* are independently selected from Trp and analogs of Trp;

Xaa3 is His, Ala or an analog of Ala, Phe, Trp, or an analog of Trp;

Xaa4 is L-Thr, D-Thr, Ile, Val, Gly, a dipeptide selected from Thr-Ala and Thr-Asn, or a tripeptide comprising Thr-Ala-Asn, wherein a carboxy terminal –OH of any of L-Thr, D-Thr, Ile, Val, Gly, Ala, or Asn optionally is replaced by $–NH_2$; and the two Cys residues are joined by a disulfide bond. In some embodiments, Xaa4 is Leu, Nle, His, or Phe or a depeptide

selected from Xaa5-Ala and Xaa5-Asn, or a tripeptide Xaa5-Ala-Asn, wherein Xaa5 is selected from Leu, Nle, His or Phe, and wherein a carboxy terminal –OH of any of the L-Thr, D-Thr, Ile, Val, Gly, Leu, Nle, His, Phe, Ala, or Asn optionally is replaced by a second blocking moiety B2; and the two Cys residues are joined by a disulfide bond.

[00105] In some embodiments, Xaa1, Xaa2, Xaa2*, Xaa3, and Xaa4 are as described above for the various embodiments of SEQ ID NO: 6. For example, in certain embodiments Xaa2* is Trp. In certain embodiments Xaa2 is an analog of Trp having increased hydrophobic character relative to Trp, e.g., 1-methyltryptophan. In certain embodiments Xaa3 is Ala. In certain embodiments Xaa3 is a single methyl unbranched amino acid.

[00106] In certain embodiments of the invention Xaa1 is Ile and Xaa4 is L-Thr.

[00107] In certain embodiments of the invention Xaa1 is Ile, Xaa2* is Trp, and Xaa4 is L-Thr.

[00108] The invention further provides compstatin analogs of SEQ ID NO: 7, as described above, wherein Xaa2 and Xaa2* are independently selected from Trp, analogs of Trp, other amino acids or aromatic amino acid analogs, and Xaa3 is His, Ala or an analog of Ala, Phe, Trp, an analog of Trp, or another aromatic amino acid or aromatic amino acid analog.

[00109] In certain embodiments of any of the compstatin analogs described herein, an analog of Phe is used rather than Phe.

[00110] Table 1 provides a non-limiting list of compstatin analogs useful in the present invention. The analogs are referred to in abbreviated form in the left column by indicating specific modifications at designated positions (1-13) as compared to the parent peptide, compstatin. Consistent with usage in the art, "compstatin" as used herein, and the activities of compstatin analogs described herein relative to that of compstatin, refer to the compstatin peptide amidated at the C-terminus. Unless otherwise indicated, peptides in Table 1 are amidated at the C-terminus. Bold text is used to indicate certain modifications. Activity relative to compstatin is based on published data and assays described therein (WO2004/026328, WO2007044668, Mallik, 2005; Katragadda, 2006). Where multiple publications reporting an activity were consulted, the more recently published value is used, and it will be recognized that values may be adjusted in the case of differences between assays. It will also be appreciated that in certain embodiments of the invention the peptides listed in Table 1 are cyclized via a disulfide bond between the two Cys residues when used in the therapeutic compositions and methods of the invention. Alternate means for cyclizing the peptides are also within the scope of the invention. As noted above, in various embodiments

of the invention one or more amino acid(s) of a compstatin analog (e.g., any of the compstatin analogs disclosed herein) can be an N-alkyl amino acid (e.g., an N-methyl amino acid). For example, and without limitation, at least one amino acid within the cyclic portion of the peptide, at least one amino acid N-terminal to the cyclic portion, and/or at least one amino acid C-terminal to the cyclic portion may be an N-alkyl amino acid, e.g., an N-methyl amino acid. In some embodiments of the invention, for example, a compstatin analog comprises an N-methyl glycine, e.g., at the position corresponding to position 8 of compstatin and/or at the position corresponding to position 13 of compstatin. In some embodiments, one or more of the compstatin analogs in Table 1 contains at least one N-methyl glycine, e.g., at the position corresponding to position 8 of compstatin and/or at the position corresponding to position 13 of compstatin. . In some embodiments, one or more of the compstatin analogs in Table 1 contains at least one N-methyl isoleucine, e.g., at the position corresponding to position 13 of compstatin. For example, a Thr at or near the C-terminal end of a peptide whose sequence is listed in Table 1 or any other compstatin analog sequence may be replaced by N-methyl Ile. As will be appreciated, in some embodiments the N-methylated amino acids comprise N-methyl Gly at position 8 and N-methyl Ile at position 13. In some embodiments the N-methylated amino acids comprise N-methyl Gly in a core sequence such as SEQ ID NO: 3 or SEQ ID NO: 4. In some embodiments the N-methylated amino acids comprise N-methyl Gly in a core sequence such as SEQ ID NO: 5, SEQ ID NO: 6, or SEQ ID NO: 7.

[00111] Table 1

Peptide	Sequence	SEQ ID NO:	Activity over compstatin
Compstatin	*H*-ICVVQDWGHHRCT-*CONH2*	8	*
Ac-compstatin	*Ac*-ICVVQDWGHHRCT-*CONH2*	9	3xmore
Ac-V4Y/H9A	*Ac*-ICV**Y**QDWG**A**HRCT-*CONH2*	10	14xmore
Ac-V4W/H9A -OH	*Ac*-ICV**W**QDWG**A**HRCT-*COOH*	11	27xmore
Ac-V4W/H9A	*Ac*-ICV**W**QDWG**A**HRCT-*CONH2*	12	45xmore
Ac-V4W/H9A/T13dT -OH	*Ac*-ICV**W**QDWG**A**HRC**dT**-*COOH*	13	55xmore
Ac-V4(2-Nal)/H9A	*Ac*-ICV**(2-Nal)**QDWG**A**HRCT-*CONH2*	14	99xmore
Ac V4(2-Nal)/H9A -OH	*Ac*-ICV**(2-Nal)**QDWG**A**HRCT-*COOH*	15	38xmore
Ac V4(1-Nal)/H9A -OH	*Ac*-ICV**(1-Nal)**QDWG**A**HRCT-*COOH*	16	30xmore
Ac-V42Igl/H9A	*Ac*-ICV**(2-Igl)**QDWG**A**HRCT-*CONH2*	17	39xmore
Ac-V42Igl/H9A -OH	*Ac*-ICV**(2-Igl)Q**DWG**A**HRCT-*COOH*	18	37xmore
Ac-V4Dht/H9A -OH	*Ac*-ICV**Dht**QDWG**A**HRCT-*COOH*	19	5xmore
Ac-V4(Bpa)/H9A -OH	*Ac*-ICV**(Bpa)**QDWG**A**HRCT-*COOH*	20	49xmore
Ac-V4(Bpa)/H9A	*Ac*-ICV**(Bpa)**QDWG**A**HRCT-*CONH2*	21	86xmore
Ac-V4(Bta)/H9A -OH	*Ac*-ICV**(Bta)**QDWG**A**HRCT-*COOH*	22	65xmore
Ac-V4(Bta)/H9A	*Ac*-ICV**(Bta)**QDWG**A**HRCT-*CONH2*	23	64xmore
Ac-V4W/H9(2-Abu)	*Ac*-ICV**W**QDWG(2-**Abu)**HRCT-*CONH2*	24	64xmore
+G/V4W/H9A +AN -OH	*H*-**G**ICV**W**QDWG**A**HRCT**AN**-*COOH*	25	38xmore
Ac-V4(5fW)/H9A	*Ac*-ICV**(5fW)**QDWG**A**HRCT- *$CONH_2$*	26	31xmore
Ac-V4(5-MeW)/H9A	*Ac*-ICV**(5-methyl-W)**QDWG**A**HRCT- *$CONH_2$*	27	67xmore
Ac-V4(1-MeW)/H9A	*Ac*-ICV**(1-methyl-W)**QDWG**A**HRCT- *$CONH_2$*	28	264xmore
Ac-V4W/W7(5fW)/H9A	*Ac*-ICV**W**QD**(5fW)**G**A**HRCT-*$CONH_2$*	29	121xmore
Ac-V4(5fW)/W7(5fW)/H9A	*Ac*-ICV**(5fW)**QD**(5fW)**G**A**HRCT- *$CONH_2$*	30	NA
Ac-V4(5-MeW)/W7(5fW)H9A	*Ac*-ICV**(5-methyl-W)**QD**(5fW)**G**A**HRCT-*$CONH_2$*	31	NA
Ac-V4(1MeW)/W7(5fW)/H9A	*Ac*-ICV**(1-methyl-W)**QD**(5fW)**G**A**HRCT-*$CONH_2$*	32	264xmore
+G/V4(6fW)/W7(6fW)H9A+N-OH	*H*-GICV**(6fW)**QD(6fW)G**A**HRCT**N**-*COOH*	33	126xmore
Ac-V4(1-formyl-W)/H9A	*Ac*-ICV**(1-formyl-W)**QDWG**A**HRCT-*$CONH_2$*	34	264xmore
Ac-V4(5-methoxy-W)/H9A	*Ac*-ICV**(1-methyoxy-W)**QDWG**A**HRCT-*$CONH_2$*	35	76xmore
G/V4(5f-W)/W7(5fW)/H9A+N-OH	*H*-GICV(**5fW**)QD(**5fW**)G**A**HRCT**N**-*COOH*	36	112xmore

NA = not available

[00112] In certain embodiments of the compositions and methods of the invention the compstatin analog has a sequence selected from sequences 9-36. In certain embodiments of the compositions and methods of the invention the compstatin analog has a sequence selected from SEQ ID NOs: 14, 21, 28, 29, 32, 33, 34, and 36. In certain embodiments of the compositions and/or methods of the invention the compstatin analog has a sequence selected from SEQ ID NOs: 30 and 31. In one embodiment of the compositions and methods of the invention the compstatin analog has a sequence of SEQ ID NO: 28. In one embodiment of the compositions and methods of the invention the compstatin analog has a sequence of SEQ ID

NO: 32. In one embodiment of the compositions and methods of the invention the compstatin analog has a sequence of SEQ ID NO: 34. In one embodiment of the compositions and methods of the invention the compstatin analog has a sequence of SEQ ID NO: 36.

[00113] In some embodiments a blocking moiety B^1 comprises an amino acid, which may be represented as Xaa0. In some embodiments blocking moiety B^2 comprises an amino acid, which may be represented as XaaN. In some embodiments blocking moiety B^1 and/or B^2 comprises a non-standard amino acid, such as a D-amino acid, N-alkyl amino acid (e.g., N-methyl amino acid). In some embodiments a blocking moiety B^1 and/or B^2 comprises a non-standard amino acid that is an analog of a standard amino acid. In some embodiments an amino acid analog comprises a lower alkyl, lower alkoxy, or halogen substituent, as compared with a standard amino acid of which it is an analog. In some embodiments a substituent is on a side chain. In some embodiments a substituent is on an alpha carbon atom. In some embodiments, a blocking moiety B^1 comprising an amino acid, e.g., a non-standard amino acid, further comprises a moiety B^{1a}. For example, blocking moiety B^1 may be represented as B^{1a}-Xaa0. In some embodiments B^{1a} neutralizes or reduces a positive charge that may otherwise be present at the N-terminus at physiological pH. In some embodiments B^{1a} comprises or consists of, e.g., an acyl group that, e.g., comprises between 1 and 12 carbons, e.g., between 1 and 6 carbons. In certain embodiments blocking moiety B^{1a} is selected from the group consisting of: formyl, acetyl, proprionyl, butyryl, isobutyryl, valeryl, isovaleryl, etc. In some embodiments, a blocking moiety B^2 comprising an amino acid, e.g., a non-standard amino acid, may further comprise a moiety B^{2a} For example, blocking moiety B^2 may be represented as XaaN-B^{2a}, where N represents the appropriate number for the amino acid (which will depend on the numbering used in the rest of the peptide). In some embodiments B^{2a} neutralizes or reduces a negative charge that may otherwise be present at the C-terminus at physiological pH. In some embodiments B^{2a} comprises or consistsof a primary or secondary amine (e.g., NH_2). It will be understood that a blocking activity of moiety B^{1a}-Xaa0 and/or XaaN-B^{2a} may be provided by either or both components of the moiety in various embodiments. In some embodiments a blocking moiety or portion thereof, e.g., an amino acid residue, may contribute to increasing affinity of the compound for C3 or C3b and/or improve the activity of the compound. In some embodiments a contribution to affinity or activity of an amino acid residue may be at least as important as a contribution to blocking activity. For example, in some embodiments Xaa0 and/or XaaN in B^{1a}-Xaa0 and/or XaaN-B^{2a} may function mainly to increase affinity or activity of the compound, while B^{1a}

and/or B^{2a} may inhibit digestion of and/or neutralize a charge of the peptide. In some embodiments a compstatin analog comprises the amino acid sequence of any of SEQ ID NOs: 5-36, wherein SEQ ID NOs: 5-36 is further extended at the N- and/or C-terminus. In some embodiments, the sequence may be represented as B^{1a}-Xaa0 -SEQUENCE - XaaN-B^{2a}, where SEQUENCE represents any of SEQ ID NOs: 5-36, wherein B^{1a} and B^{2a} may independently be present or absent. For example, in some embodiments a compstatin analog comprises B^{1a}-Xaa0 - X'aa1 - X'aa2 - X'aa3 - X'aa4 -Gln-Asp-Xaa-Gly- X"aa1- X"aa2- X"aa3- X"aa4- X"aa5 - XaaN-B^{2a} (SEQ ID NO: 37A) ,where X'aa1 - X'aa2 - X'aa3 - X'aa4 , Xaa, X"aa1, X"aa2, X"aa3, X"aa4, and X"aa5 are as set forth above for SEQ ID NO: 5.

[00114] In some embodiments a compstatin analog comprises B^{1a}-Xaa0 – Xaa1 – Cys – Val -Xaa2 -Gln - Asp – Xaa2*-Gly – Xaa3 -His -Arg – Cys-Xaa4 - XaaN-B^{2a} (SEQ ID NO: 38A), where Xaa1, Xaa2, Xaa2*, Xaa3, and Xaa4 are as set forth above for SEQ ID NO: 6 or wherein Xaa1, Xaa2, Xaa2*, Xaa3, and Xaa4 are as set forth for SEQ ID NO: 6 or SEQ ID NO: 7.

[00115] In some embodiments a compstatin analog comprises B^{1a}-Xaa0 – Xaa1 – Xaa2– Xaa3 – Xaa4 – Xaa5 – Xaa6 – Xaa7 – Xaa8 – Xaa9 - Xaa10- Xaa11- Xaa12-Xaa13-XaaN-B^{2a} (SEQ ID NO: 39A) wherein Xaa1, Xaa2, Xaa3, Xaa4 , Xaa5, Xaa6, Xaa7, Xaa8, Xaa9, Xaa10, Xaa11, Xaa12, and Xaa13 are identical to amino acids at positions 1-13 of any of SEQ ID NOs: 9-36.

[00116] In some embodiments Xaa0 and/or XaaN in any compstatin analog sequence comprises an amino acid that comprises an aromatic ring having an alkyl substituent at one or more positions. In some embodiments an alkyl substituent is a lower alkyl substituent. For example, in some embodiments an alkyl substituent is a methyl or ethyl group. In some embodiments a substituent is located at any position that does not destroy the aromatic character of the compound. In some embodiments a substituent is located at any position that does not destroy the aromatic character of a ring to which the substituent is attached. In some embodiments a substituent is located at position 1, 2, 3, 4, or 5. In some embodiments Xaa0 comprises an O-methyl analog of tyrosine, 2-hydroxyphenylalanine or 3-hydroxyphenylalanine. For purposes of the present disclosure, a lower case "m" followed by a three letter amino acid abbreviation may be used to specifically indicate that the amino acid is an N-methyl amino acid. For example, where the abbreviation "mGly" appears herein, it denotes N-methyl glycine (also sometimes referred to as sarcosine or Sar). In some embodiments Xaa0 is or comprises mGly, Tyr, Phe, Arg, Trp, Thr, Tyr(Me), Cha, mPhe, mVal, mIle, mAla, DTyr, DPhe, DArg, DTrp, DThr, DTyr(Me), mPhe, mVal, mIle, DAla, or

DCha. For example, in some embodiments a compstatin analog comprises a peptide having a sequence B^1-Ile-[Cys-Val-Trp(Me)-Gln-Asp-Trp-mGly-Ala-His-Arg-Cys]-mIle-B^2 (SEQ ID NO: 40A) or B^1-Ile-[Cys-Val-Trp(Me)-Gln-Asp-Trp-mGly-Ala-His-Arg-Cys]-mIle-B^2 (SEQ ID NO: 41A). The two Cys residues are joined by a disulfide bond in the active compounds. In some embodiments the peptide is acetylated at the N-terminus and/or amidated at the C-terminus. In some embodiments B^1 comprises B^{1a}-Xaa0 and/or B^2 comprises XaaN-B^{2a}, as described above. For example, in some embodiments B^1 comprises or consists of Gly, mGly, Tyr, Phe, Arg, Trp, Thr, Tyr(Me), mPhe, mVal, mIle, mAla, DTyr, DPhe, DTrp, DCha, DAla and B^2 comprises NH_2, e.g., a carboxy terminal –OH of mIle is replaced by NH_2.. In some embodiments B^1 comprises or consists of mGly, Tyr, DTyr, or Tyr(Me) and B^2 comprises NH_2, e.g., a carboxy terminal –OH of mIle is replaced by NH_2. .In some embodiments an Ile at position Xaa1 is replaced by Gly. Complement inhibition potency and/or C3b binding parameters of selected compstatin analogs are described in WO/2010/127336 (PCT/US2010/033345) and/or in Qu, et al., Immunobiology (2012), doi:10.1016/j.imbio.2012.06.003.

[00117] In some embodiments a blocking moiety or portion thereof, e.g., an amino acid residue, may contribute to increasing affinity of the compound for C3 or C3b and/or improve the activity of the compound. In some embodiments a contribution to affinity or activity of an amino acid or amino acid analog may be more significant than a blocking activity.

[00118] In certain embodiments of the compositions and methods of the invention the compstatin analog has a sequence as set forth in Table 1, but where the Ac- group is replaced by an alternate blocking moiety B^1, as described herein. In some embodiments the $–NH_2$ group is replaced by an alternate blocking moiety B^2, as described herein.

[00119] In one embodiment, the compstatin analog binds to substantially the same region of the β chain of human C3 as does compstatin. In one embodiment the compstatin analog is a compound that binds to a fragment of the C-terminal portion of the β chain of human C3 having a molecular weight of about 40 kDa to which compstatin binds (Soulika, A.M., et al., *Mol. Immunol.*, 35:160, 1998; Soulika, A.M., et al., *Mol. Immunol.* 43(12):2023-9, 2006). In certain embodiments the compstatin analog is a compound that binds to the binding site of compstatin as determined in a compstatin-C3 structure, e.g., a crystal structure or NMR-derived 3D structure. In certain embodiments the compstatin analog is a compound that could substitute for compstatin in a compstatin-C3 structure and would form substantially the same intermolecular contacts with C3 as compstatin. In certain embodiments the compstatin

analog is a compound that binds to the binding site of a peptide having a sequence set forth in Table 1, e.g., SEQ ID NO: 14, 21, 28, 29, 32, 33, 34, or 36, 37, 37A, 38A, 39A, 40A, or 41A or another compstatin analog sequence disclosed herein in a peptide-C3 structure, e.g., a crystal structure. In certain embodiments the compstatin analog is a compound that binds to the binding site of a peptide having SEQ ID NO: 30 or 31 in a peptide-C3 structure, e.g., a crystal structure. In certain embodiments the compstatin analog is a compound that could substitute for the peptide of SEQ ID NO: 9-36, e.g., a compound that could substitute for the peptide of SEQ ID NO: 14, 21, 28, 29, 32, 33, 34, or 36, 37 ,37A, 38A, 39A, 40A, or 41A or another compstatin analog sequence disclosed herein in a peptide-C3 structure and would form substantially the same intermolecular contacts with C3 as the peptide. In certain embodiments the compstatin analog is a compound that could substitute for the peptide of SEQ ID NO: 30 or 31 in a peptide-C3 structure and would form substantially the same intermolecular contacts with C3 as the peptide.

[00120] One of ordinary skill in the art will readily be able to determine whether a compstatin analog binds to a fragment of the C-terminal portion of the β chain of C3 using routine experimental methods. For example, one of skill in the art could synthesize a photocrosslinkable version of the compstatin analog by including a photo-crosslinking amino acid such as *p*-benzoyl-L-phenylalanine (Bpa) in the compound, e.g., at the C-terminus of the sequence (Soulika, A.M., et al, *supra*). Optionally additional amino acids, e.g., an epitope tag such as a FLAG tag or an HA tag could be included to facilitate detection of the compound, e.g., by Western blotting. The compstatin analog is incubated with the fragment and crosslinking is initiated. Colocalization of the compstatin analog and the C3 fragment indicates binding. Surface plasmon resonance may also be used to determine whether a compstatin analog binds to the compstatin binding site on C3 or a fragment thereof. One of skill in the art would be able to use molecular modeling software programs to predict whether a compound would form substantially the same intermolecular contacts with C3 as would compstatin or a peptide having the sequence of any of the peptides in Table 1, e.g., SEQ ID NO: 14, 21, 28, 29, 32, 33, 34, or 36, or in some embodiments SEQ ID NO: 30 or 31, 37, 37A, 38A, 39A, 40A, or 41A or another compstatin analog sequence disclosed herein.

[00121] Compstatin analogs may be prepared by various synthetic methods of peptide synthesis known in the art *via* condensation of amino acid residues, e.g., in accordance with conventional peptide synthesis methods, may be prepared by expression *in vitro* or in living cells from appropriate nucleic acid sequences encoding them using methods known in the art.

For example, peptides may be synthesized using standard solid-phase methodologies as described in Malik, *supra*, Katragadda, *supra*, WO2004026328, and/or WO2007062249. Potentially reactive moieties such as amino and carboxyl groups, reactive functional groups, etc., may be protected and subsequently deprotected using various protecting groups and methodologies known in the art. See, e.g., "Protective Groups in Organic Synthesis", 3rd ed. Greene, T. W. and Wuts, P. G., Eds., John Wiley & Sons, New York: 1999. Peptides may be purified using standard approaches such as reversed-phase HPLC. Separation of diasteriomeric peptides, if desired, may be performed using known methods such as reversed-phase HPLC. Preparations may be lyophilized, if desired, and subsequently dissolved in a suitable solvent, e.g., water. The pH of the resulting solution may be adjusted, e.g. to physiological pH, using a base such as NaOH. Peptide preparations may be characterized by mass spectrometry if desired, e.g., to confirm mass and/or disulfide bond formation. See, e.g., Mallik, 2005, and Katragadda, 2006.

[00122] A compstatin analog, optionally linked to a cell-reactive moiety or targeting moiety, can be modified by addition of a molecule such as polyethylene glycol (PEG) or similar molecules to stabilize the compound, reduce its immunogenicity, increase its lifetime in the body, increase or decrease its solubility, and/or increase its resistance to degradation. Methods for pegylation are well known in the art (Veronese, F.M. & Harris, Adv. Drug Deliv. Rev. 54, 453-456, 2002; Davis, F.F., Adv. Drug Deliv. Rev. 54, 457-458, 2002); Hinds, K.D. & Kim, S.W. *Adv. Drug Deliv. Rev.* 54, 505-530 (2002; Roberts, M.J., Bentley, M.D. & Harris, J.M. *Adv. Drug Deliv. Rev.* 54, 459-476; 2002); Wang, Y.S. et al. *Adv. Drug Deliv. Rev.* 54, 547-570, 2002). A wide variety of polymers such as PEGs and modified PEGs, including derivatized PEGs to which polypeptides can conveniently be attached are described in Nektar Advanced Pegylation 2005-2006 Product Catalog, Nektar Therapeutics, San Carlos, CA, which also provides details of appropriate conjugation procedures. In another embodiment a compstatin analog is fused to the Fc domain of an immunoglobulin or a portion thereof. In some other embodiments a compstatin analog is conjugated to an albumin moiety or to an albumin binding peptide. Thus in some embodiments a compstatin analog is modified with one or more polypeptide or non-polypeptide components, e.g., the compstatin analog is pegylated or conjugated to another moiety. In some embodiments the component is not the Fc domain of an immunoglobulin or a portion thereof. A compstatin analog can be provided as a multimer or as part of a supramolecular complex, which can include either a single molecular species or multiple different species (e.g., multiple different analogs).

[00123] In some embodiments, a compstatin analog is a multivalent compound comprising a plurality of compstatin analog moieties covalently or noncovalently linked to a polymeric backbone or scaffold. The compstatin analog moieties can be identical or different. In certain embodiments of the invention the multivalent compound comprises multiple instances, or copies, of a single compstatin analog moiety. In other embodiments of the invention the multivalent compound comprises one or more instances of each of two of more non-identical compstatin analog moieties, e.g., 3, 4, 5, or more different compstatin analog moieties. In certain embodiments of the invention the number of compstatin analog moieties ("n") is between 2 and 6. In other embodiments of the invention n is between 7 and 20. In other embodiments of the invention n is between 20 and 100. In other embodiments n is between 100 and 1,000. In other embodiments of the invention n is between 1,000 and 10,000. In other embodiments n is between 10,000 and 50,000. In other embodiments n is between 50,000 and 100,000. In other embodiments n is between 100,000 and 1,000,000.

[00124] The compstatin analog moieties may be attached directly to the polymeric scaffold or may be attached via a linking moiety that connects the compstatin analog moiety to the polymeric scaffold. The linking moiety may be attached to a single compstatin analog moiety and to the polymeric scaffold. Alternately, a linking moiety may have multiple compstatin analog moieties joined thereto so that the linking moiety attaches multiple compstatin analog moieties to the polymeric scaffold.

[00125] In some embodiments, the compstatin analog comprises an amino acid having a side chain comprising a primary or secondary amine, e.g., a Lys residue. For example, a Lys residue, or a sequence comprising a Lys residue, is added at the N-terminus and/or C-terminus of the compstatin analog. In some embodiments, the Lys residue is separated from the cyclic portion of the compstatin analog by a rigid or flexible spacer. The spacer may, for example, comprise a substituted or unsubstitued, saturated or unsaturated alkyl chain, oligo(ethylene glycol) chain, and/or other moieties, e.g., as described in Section VI with regard to linkers. The length of the chain may be, e.g., between 2 and 20 carbon atoms. In other embodiments the spacer is a peptide. The peptide spacer may be, e.g., between 1 and 20 amino acids in length, e.g., between 4 and 20 amino acids in length. Suitable spacers can comprise or consist of multiple Gly residues, Ser residues, or both, for example. Optionally, the amino acid having a side chain comprising a primary or secondary amine and/or at least one amino acid in a spacer is a D-amino acid. Any of a variety of polymeric backbones or scaffolds could be used. For example, the polymeric backbone or scaffold may be a polyamide, polysaccharide, polyanhydride, polyacrylamide, polymethacrylate, polypeptide,

polyethylene oxide, or dendrimer. Suitable methods and polymeric backbones are described, e.g., in WO98/46270 (PCT/US98/07171) or WO98/47002 (PCT/US98/06963). In one embodiment, the polymeric backbone or scaffold comprises multiple reactive functional groups, such as carboxylic acids, anhydride, or succinimide groups. The polymeric backbone or scaffold is reacted with the compstatin analogs. In one embodiment, the compstatin analog comprises any of a number of different reactive functional groups, such as carboxylic acids, anhydride, or succinimide groups, which are reacted with appropriate groups on the polymeric backbone. Alternately, monomeric units that could be joined to one another to form a polymeric backbone or scaffold are first reacted with the compstatin analogs and the resulting monomers are polymerized. In another embodiment, short chains are prepolymerized, functionalized, and then a mixture of short chains of different composition are assembled into longer polymers.

[00126] V. Compstatin Mimetics

[00127] The structure of compstatin is known in the art, and NMR structures for a number of compstatin analogs having higher activity than compstatin are also known (Malik, *supra*). Structural information may be used to design compstatin mimetics.

[00128] In one embodiment, the compstatin mimetic is any compound that competes with compstatin or any compstatin analog (e.g., a compstatin analog whose sequence is set forth in Table 1) for binding to C3 or a fragment thereof (such as a 40 kD fragment of the β chain to which compstatin binds). In some embodiments, the compstatin mimetic has an activity equal to or greater than that of compstatin. In some embodiments, the compstatin mimetic is more stable, orally available, or has a better bioavailability than compstatin. The compstatin mimetic may be a peptide, nucleic acid, or small molecule. In certain embodiments the compstatin mimetic is a compound that binds to the binding site of compstatin as determined in a compstatin-C3 structure, e.g., a crystal structure or a 3-D structure derived from NMR experiments. In certain embodiments the compstatin mimetic is a compound that could substitute for compstatin in a compstatin-C3 structure and would form substantially the same intermolecular contacts with C3 as compstatin. In certain embodiments the compstatin mimetic is a compound that binds to the binding site of a peptide having a sequence set forth in Table 1, e.g., SEQ ID NO: 14, 21, 28, 29, 32, 33, 34, or 36 or in certain embodiments SEQ ID NO: 30 or 31 or other compstatin analog sequence, in a peptide-C3 structure. In certain embodiments the compstatin mimetic is a compound that could substitute for a peptide having a sequence set forth in Table 1, e.g., SEQ ID NO: 14, 21, 28, 29, 32, 33, 34, or 36 or

in certain embodiments SEQ ID NO: 30 or 31 or other compstatin analog sequence, in a peptide-C3 structure and would form substantially the same intermolecular contacts with C3 as the peptide. In certain embodiments the compstatin mimetic has a non-peptide backbone but has side chains arranged in a sequence designed based on the sequence of compstatin.

[00129] One of skill in the art will appreciate that once a particular desired conformation of a short peptide has been ascertained, methods for designing a peptide or peptidomimetic to fit that conformation are well known. See, *e.g.*, G.R. Marshall (1993), Tetrahedron, 49: 3547-3558; Hruby and Nikiforovich (1991), in Molecular Conformation and Biological Interactions, P. Balaram & S. Ramasehan, eds., Indian Acad. of Sci., Bangalore, PP. 429-455), Eguchi M, Kahn M., Mini Rev Med Chem., 2(5):447-62, 2002. Of particular relevance to the present invention, the design of peptide analogs may be further refined by considering the contribution of various side chains of amino acid residues, e.g., for the effect of functional groups or for steric considerations as described in the art for compstatin and analogs thereof, among others.

[00130] It will be appreciated by those of skill in the art that a peptide mimic may serve equally well as a peptide for the purpose of providing the specific backbone conformation and side chain functionalities required for binding to C3 and inhibiting complement activation. Accordingly, it is contemplated as being within the scope of the present invention to produce and utilize C3-binding, complement-inhibiting compounds through the use of either naturally-occurring amino acids, amino acid derivatives, analogs or non-amino acid molecules capable of being joined to form the appropriate backbone conformation. A non-peptide analog, or an analog comprising peptide and non-peptide components, is sometimes referred to herein as a "peptidomimetic" or “isosteric mimetic,” to designate substitutions or derivations of a peptide that possesses much the same backbone conformational features and/or other functionalities, so as to be sufficiently similar to the exemplified peptides to inhibit complement activation. More generally, a compstatin mimetic is any compound that would position pharmacophores similarly to their positioning in compstatin, even if the backbone differs.

[00131] The use of peptidomimetics for the development of high-affinity peptide analogs is well known in the art. Assuming rotational constraints similar to those of amino acid residues within a peptide, analogs comprising non-amino acid moieties may be analyzed, and their conformational motifs verified, by means of the Ramachandran plot (Hruby & Nikiforovich 1991), among other known techniques.

[00132] One of skill in the art will readily be able to establish suitable screening assays to identify additional compstatin mimetics and to select those having desired inhibitory activities. For example, compstatin or an analog thereof could be labeled (e.g., with a radioactive or fluorescent label) and contacted with C3 in the presence of different concentrations of a test compound. The ability of the test compound to diminish binding of the compstatin analog to C3 is evaluated. A test compound that significantly diminishes binding of the compstatin analog to C3 is a candidate compstatin mimetic. For example, a test compound that diminishes steady-state concentration of a compstatin analog-C3 complex, or that diminishes the rate of formation of a compstatin analog-C3 complex by at least 25%, or by at least 50%, is a candidate compstatin mimetic. One of skill in the art will recognize that a number of variations of this screening assay may be employed. Compounds to be screened include natural products, libraries of aptamers, phage display libraries, compound libraries synthesized using combinatorial chemistry, etc. The invention encompasses synthesizing a combinatorial library of compounds based upon the core sequence described above and screening the library to identify compstatin mimetics. Any of these methods could also be used to identify new compstatin analogs having higher inhibitory activity than compstatin analogs tested thus far. It will be appreciated that compstatin mimetics could be used in the cell-reactive compounds of the invention, and the invention provides such cell-reactive compstatin mimetics.

[00133] VI. Cell-reactive or Long-Acting Compstatin Analogs

[00134] The invention provides a variety of cell-reactive compstatin analogs. In some aspects, a cell-reactive compstatin analog comprises a compound of formula A–L–M, wherein A is a moiety that comprises a cell-reactive functional group J, L is an optionally present linking portion, and M comprises a compstatin analog moiety. The compstatin analog moiety can comprise any compstatin analog, e.g., any compstatin analog described above, in various embodiments. Formula A–L–M encompasses embodiments in which A–L is present at the N-terminus of the compstatin analog moiety, embodiments in which A–L is present at the C-terminus of the compstatin analog moiety, embodiments in which A–L is attached to a side chain of an amino acid of the compstatin analog moiety, and embodiments where the same or different A-Ls are present at both ends of M. It will be appreciated that when certain compstatin analog(s) are present as a compstatin analog moiety in a compound of formula A–L–M, a functional group of the compstatin analog will have reacted with a functional group of L to form a covalent bond to A or L. For example, a cell-reactive compstatin analog in which the compstatin analog moiety comprises a compstatin analog that

contains an amino acid with a side chain containing a primary amine (NH_2) group (which compstatin analog can be represented by formula R^1— (NH_2)), can have a formula R^1—NH—L– A in which a new covalent bond to L (e.g., N—C) has been formed and a hydrogen lost. Thus the term "compstatin analog moiety" includes molecular structures in which at least one atom of a compstatin analog participates in a covalent bond with a second moiety, which may, e.g., modification of a side chain. Similar considerations apply to compstatin analog moieties present in multivalent compound described above. In some embodiments, a blocking moiety at the N-terminus or C-terminus of a compstatin analog, e.g., a compstatin analog described in Section IV above, is replaced by A–L in the structure of a cell-reactive compstatin analog. In some embodiments, A or L comprises a blocking moiety. In some embodiments, a cell-reactive compstatin analog has a molar activity of at least about 10%, 20%, or 30%, e.g., between 30% and 40%, between 30% and 50%, between 30% and 60%, between 30% and 70%, between 30% and 80%, between 30% and 90%, or more, of the activity of a corresponding compstatin analog having the same amino acid sequence (and, if applicable, one or more blocking moiet(ies)) but not comprising a cell-reactive moiety. In some embodiments in which a cell-reactive compstatin analog comprises multiple compstatin analog moieties, the molar activity of the cell-reactive compstatin analog is at least about 10%, 20%, or 30%, e.g., between 30% and 40%, between 30% and 50%, between 30% and 60%, between 30% and 70%, between 30% and 80%, between 30% and 90%, or more, of the sum of the activities of said compstatin analog moieties.

[00135] Cell-reactive moiety A can comprise any of a variety of different cell-reactive functional groups J, in various embodiments. In general, a cell-reactive functional group may be selected based at least in part on factors such as (a) the particular functional group to be targeted; (b) the ability of the reactive functional group to react with the target functional group under physiologically acceptable ex vivo conditions (e.g., physiologically acceptable pH and osmolarity) and/or in vivo conditions (e.g., in blood); (c) the specificity of the reaction between the reactive functional group and the target functional group under physiologically acceptable ex vivo conditions and/or in vivo; (d) the stability (e.g., under in vivo conditions) of the covalent bond that would result from reaction of the reactive functional group with its target functional group; (e) the ease of synthesizing a cell-reactive compstatin analog comprising the reactive functional group, etc. In some embodiments, a reactive functional group that reacts with its target chemical group without releasing a leaving group is selected. In some embodiments, a reactive functional group that results in release of a leaving group upon reaction with a target is selected. Compounds containing

such groups may be useful, e.g., to monitor progress and/or extent of a reaction. In some embodiments, a leaving group is physiologically acceptable to cells, tissues, or organs in the amount generated (e.g., based on concentration and/or absolute amount generated) and/or is medically acceptable to a subject in the amount generated in vivo (e.g., based on concentration in a relevant body fluid such as blood and/or based on the absolute amount generated). In some embodiments, a leaving group generated ex vivo is at least in part removed, e.g., by washing cells or by washing or perfusing a tissue or organ, e.g., with saline.

[00136] In many embodiments, a cell-reactive functional group of use in the invention reacts with a side chain of an amino acid residue and/or with an N-terminal amino group or C-terminal carboxyl group of a protein. In some embodiments, the cell-reactive functional group is reactive with sulfhydryl (-SH) groups, which are found in the side chains of cysteine residues. In some embodiments, a maleimide group is used. Maleimide groups react with sulfhydryl groups of cysteine residues of proteins at physiologic pH and form a stable thioether linkage. In some embodiments, a haloacetyl group, such as an iodoacetyl or a bromoacetyl group, is used. Haloacetyls react with sulfhydryl groups at physiologic pH. The reaction of the iodoacetyl group proceeds by nucleophilic substitution of iodine with a sulfur atom from a sulfhydryl group resulting in a stable thioether linkage. In other embodiments, an iodoacetamide group is used. In some embodiments, the cell-reactive functional group reacts with amino ($-NH_2$) groups, which are present at the N-termini of proteins and in the side chain of lysine residues (ε-amino group). In some embodiments an activated ester, e.g., a succinimidyl ester (i.e., NHS ester) is used. For example, N-hydroxysuccinimide (NHS) or its water-soluble analog (sulfo-NHS) can be used in the synthesis, whereby the resulting cell-reactive compstatin analog comprises an NHS ester. In some embodiments, the cell-reactive functional group reacts with carboxyl (-COOH) groups, which are present at the C-termini of proteins and in the side chains of various amino acid residues. In some embodiments, the cell-reactive compstatin analog is reactive with hydroxyl (-OH) groups, which are present in the side chains of various amino acids and in carbohydrate moieties of glycosylated proteins.

[00137] In general, linking portion L can comprise any one or more aliphatic and/or aromatic moieties consistent with the formation of a stable compound joining the linked moieties. The term "stable", as used herein, preferably refers to compounds which possess stability sufficient to allow manufacture and which maintain the integrity of the compound for a sufficient period of time, e.g., to be useful for one or more purposes described herein. In some embodiments, L comprises a saturated or unsaturated, substituted or unsubstituted,

branched or unbranched, aliphatic chain having a length of between 1 and 30, between 1 and 20, between 1 and 10, between 1 and 6, or 5 or less carbon atoms, where length refers to the number of C atoms in the main (longest) chain. In some embodiments, the aliphatic chain comprises one or more heteroatoms (O, N, S), which may be independently selected. In some embodiments, at least 50% of the atoms in the main chain of L are carbon atoms. In some embodiments, L comprises a saturated alkyl moiety $(CH_2)_n$, wherein n is between 1 and 30.

[00138] In some embodiments, L comprises one or more heteroatoms and has a length of between 1 and 1000, between 1 and 800, between 1 and 600, between 1 and 400, between 1 and 300, between 1 and 200, between 1 and 100, between 1 and 50, between 1 and 30, or between 1 and 10 total carbon atoms in a chain. In some embodiments, L comprises an oligo(ethylene glycol) moiety $(-(O-CH_2-CH_2-)_n)$ wherein n is between 1 and 500, between 1 and 400, between 1 and 300, between 1 and 200, between 1 and 100, between 10 and 200, between 200 and 300, between 100 and 200, between 40 and 500, between 30 and 500, between 20 and 500, between 10 and 500, between 1 and 40, between 1 and 30, between 1 and 20, or between 1 and 10.

[00139] In some embodiments, L comprises an unsaturated moiety such as -CH=CH- or $-CH_2-CH=CH-$; a moiety comprising a non-aromatic cyclic ring system (e.g., a cyclohexyl moiety), an aromatic moiety (e.g., an aromatic cyclic ring system such as a phenyl moiety); an ether moiety (-C-O-C-); an amide moiety (-C(=O)-N-); an ester moiety (-CO-O-); a carbonyl moiety (-C(=O)-); an imine moiety (-C=N-); a thioether moiety (-C-S-C-); an amino acid residue; and/or any moiety that can be formed by the reaction of two compatible reactive functional groups. In certain embodiments, one or more moieties of a linking portion or cell-reactive moiety is/are substituted by independent replacement of one or more of the hydrogen (or other) atoms thereon with one or more moieties including, but not limited to aliphatic; aromatic, aryl; alkyl, aralkyl, alkanoyl, aroyl, alkoxy; thio; F; C1; Br; I; -NO2; -CN; -CF3; -CH2CF3; -CHC12; -CH2OH; -CH2CH2OH; -CH2NH2; -CH2SO2CH3; - or –GRG1 wherein G is –O-, -S-, -NRG2-, -C(=O)-, -S(=O)-, -SO2-, -C(=O)O-, -C(=O)NRG2-, -OC(=O)-, -NRG2C(=O)-, -OC(=O)O-, -OC(=O)NRG2-, -NRG2C(=O)O-, -NRG2C(=O)NRG2-, -C(=S)-, -C(=S)S-, -SC(=S)-, -SC(=S)S-, -C(=NRG2)-, -C(=NRG2)O-, -C(=NRG2)NRG3-, -OC(=NRG2)-, -NRG2C(=NRG3)-, -NRG2SO2-, -NRG2SO2NRG3-, or -SO2NRG2-, wherein each occurrence of RG1, RG2 and RG3 independently includes, but is not limited to, hydrogen, halogen, or an optionally substituted aliphatic, aromatic, or aryl moiety. It will be appreciated that cyclic ring systems when present as substituents may

optionally be attached via a linear moiety. Combinations of substituents and variables envisioned by this invention are preferably those that result in the formation of stable compounds useful in any one or more of the methods described herein, e.g., useful for the treatment of one or more disorders and/or for contacting a cell, tissue, or organ, as described herein, and/or useful as intermediates in the manufacture of one or more such compounds.

[00140] L can comprise one or more of any of the moieties described in the preceding paragraph, in various embodiments. In some embodiments, L comprises two or more different moieties linked to one another to form a structure typically having a length of between 1 to about 60 atoms, between 1 to about 50 atoms, e.g., between 1 and 40, between 1 and 30, between 1 and 20, between 1 and 10, or between 1 and 6 atoms, where length refers to the number of atoms in the main (longest) chain. In some embodiments, L comprises two or more different moieties linked to one another to form a structure typically having between 1 to about 40, e.g., between 1 and 30, e.g., between 1 and 20, between 1 and 10, or between 1 and 6 carbon atoms in the main (longest) chain. In general, the structure of such a cell-reactive compstatin analog can be represented by formula A-(L^{Pj})j-M, wherein j is typically between 1 and 10, and each L^{Pj} is independently selected from among the moieties described in the preceding paragraph. In many embodiments, L comprises one or morecarbon-containing chains such as -(CH_2)n- and/or -(O-CH_2-CH_2-)n, which are joined covalently to each other and/or to a cell-reactive functional group or compstatin analog, e.g., by moieties (e.g., amide, ester, or ether moieties) that result from the reaction of two compatible reactive functional groups. In some embodiments, L comprises an oligo(ethylene glycol) moiety and/or a saturated alkyl chain. In some embodiments, L comprises -$(CH_2)_m$-C(=O)-NH-$(CH_2CH_2O)_n(CH_2)_p$C(=O)- or -$(CH_2)_m$-C(=O)-NH-$(CH_2)_p(OCH_2CH_2)_n$C(=O)-. In some embodiments, m, n, and p are selected so that the number of carbons in the chain is between 1 and 500, e.g., between 2 and 400, between 2 and 300, between 2 and 200, between 2 and 100, between 2 and 50, between 4 and 40, between 6 and 30, or between 8 and 20. In some embodiments, m is between 2 and 10, n is between 1 and 500, and/or p is between 2 and 10. In some embodiments, m is between 2 and 10, n is between 1 and 400, and/or p is between 2 and 10. In some embodiments, m is between 2 and 10, n is between 1 and 300, and/or p is between 2 and 10. In some embodiments, m is between 2 and 10, n is between 1 and 200, and/or p is between 2 and 10. In some embodiments, m is between 2 and 10, n is between 1 and 100, and/or p is between 2 and 10. In some embodiments, m is between 2 and 10, n is between 1 and 50, and/or p is between 2 and 10. In some embodiments, m is between 2 and 10, n is between 1 and 25, and/or p is between 2 and 10. In some embodiments, m is

between 2 and 10, n is between 1 and 8, and/or p is between 2 and 10. Optionally, at least one $-CH_2-$ is replaced by CH-R, wherein R can be any substituent. Optionally, at least one $-CH_2-$ is replaced by a heteroatom, cyclic ring system, amide, ester, or ether moiety. In some embodiments, L does not comprise an alkyl group having more than 3 carbon atoms in the longest chain. In some embodiments, L does not comprise an alkyl group having more than 4, 5, 6, 7, 8, 9, 10, or 11 carbon atoms in the longest chain.

[00141] In some embodiments of the invention, A comprises a cell-reactive functional group J and a linker L^1 comprising a linking portion L^{P1} and a reactive functional group that reacts with the compstatin analog to generate A-M In some embodiments, a bifunctional linker L^2 comprising two reactive functional groups and a linking portion L^{P2} is used. The reactive functional groups of L react with appropriate reactive functional groups of A and M to produce a cell-reactive compstatin analog A-L-M. In some embodiments, the compstatin analog comprises a linker L^3 comprising a linking portion L^{P3}. For example, as discussed below, a linker comprising a reactive functional group may be present at the N- or C-terminus or a moiety comprising a reactive functional group may be attached to the N- or C-terminus via a linker. Thus L may contain multiple linking portions L^P contributed, e.g., by A, by linker(s) used to join A and M, and/or by the compstatin analog. It will be understood that, when present in the structure A-L-M, certain reactive functional group(s) present prior to reaction in L^1, L^2, L^3, etc., will have undergone reaction, so that only a portion of said reactive functional group(s) will be present in the final structure A-L-M, and the compound will contain moieties formed by reaction of said functional groups. In general, if a compound contains two or more linking portions, the linking portions can be the same or different, and can be independently selected in various embodiments. Multiple linking portions L^P can be attached to one another to form a larger linking portion L, and at least some of such linking portions can have one or more compstatin analog(s) and/or cell-reactive functional group(s) attached thereto. In molecules comprising multiple compstatin analogs, the compstatin analogs can be the same or different and, if different, can be independently selected. The same applies to the linking portions and reactive functional groups. The invention encompasses the use of multivalent compstatin analogs comprising one or more cell-reactive functional group(s) and use of concatamers of compstatin analogs comprising one or more cell-reactive functional group(s). In some embodiments, at least one linkage is a stable non-covalent linkage such as a biotin/(strept)avidin linkage or other noncovalent linkage of approximately equivalent strength.

[00142] In some embodiments a cell-reactive compstatin analog comprises a compstatin analog in which any of SEQ ID NOs: 3-36, 37A, 38A, 39A, 40A, or 41A is extended by one or more amino acids at the N-terminus, C-terminus, or both, wherein at least one of the amino acids has a side chain that comprises a reactive functional group such as a primary or secondary amine, a sulfhydryl group, a carboxyl group (which may be present as a carboxylate group), a guanidino group, a phenol group, an indole ring, a thioether, or an imidazole ring. In some embodiments, the amino acid(s) is/are L-amino acids. In some embodiments, any one or more of the amino acid(s) is a D-amino acid. If multiple amino acids are added, the amino acids can be independently selected. In some embodiments, the reactive functional group (e.g., a primary or secondary amine) is used as a target for addition of a moiety comprising a cell-reactive functional group. Amino acids having a side chain that comprises a primary or secondary amine include lysine (Lys) and diaminocarboxylic acids of general structure $NH_2(CH_2)_nCH(NH_2)COOH$ such as 2,3-diaminopropionic acid (dapa), 2,4-diaminobutyric acid (daba), and ornithine (orn), wherein n = 1 (dapa), 2 (daba), and 3 (orn), respectively. In some embodiments at least one amino acid is cysteine, aspartic acid, glutamic acid, arginine, tyrosine, tryptophan, methionine, or histidine. Cysteine has a side chain comprising a sulfhydryl group. Aspartic acid and glutamic acid have a side chain comprising a carboxyl group (ionizable to a carboxylate group). Arginine has a side chain comprising a guanidino group. Tyrosine has a side chain comprising a phenol group (ionizable to a phenolate group). Tryptophan has a side chain comprising an indole ring include include, e.g., tryptophan. Methionine has a side chain comprising a thioether group include, e.g., methionine. Histidine has a side chain comprising an imidazole ring. A wide variety of non-standard amino acids having side chains that comprise one or more such reactive functional group(s) are available, including naturally occurring amino acids and amino acids not found in nature. See, e.g., Hughes, B. (ed.), *Amino Acids, Peptides and Proteins in Organic Chemistry*, Volumes 1-4, Wiley-VCH (2009-2011); Blaskovich, M., Handbook on Syntheses of Amino Acids General Routes to Amino Acids, Oxford University Press, 2010. The invention encompasses embodiments in which one or more non-standard amino acid(s) is/are used to provide a target for addition of a moiety comprising a cell-reactive functional group. Any one or more of the amino acid(s) may be protected as appropriate during synthesis of the compound. For example, one or more amino acid(s) may be protected during reaction(s) involving the target amino acid side chain. In some embodiments, wherein a sulfhydryl-containing amino acid is used as a target for addition of a moiety comprising a cell-reactive functional group, the sulfhydryl is protected while the

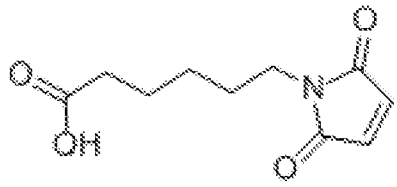
compound is being cyclized by formation of an intramolecular disulfide bond between other amino acids such as cysteines.

[00143] In the discussion in this paragraph, an amino acid having a side chain containing an amine group is used as an example. The invention encompasses analogous embodiments in which an amino acid having a side chain containing a different reactive functional group is used. In some embodiments, an amino acid having a side chain comprising a primary or secondary amine is attached directly to the N-terminus or C-terminus of any of SEQ ID NOs: 3-36, 37, 37A, 38A, 39A, 40A, or 41A or via a peptide bond. In some embodiments, an amino acid having a side chain comprising a primary or secondary amine is attached to the N- or C-terminus of any of SEQ ID NOs: 3-36, 37, 37A, 38A, 39A, 40A, or 41A or via a linking portion, which may contain any one or more of the linking moieties described above. In some embodiments, at least two amino acids are appended to either or both termini. The two or more appended amino acids may be joined to each other by peptide bonds or at least some of the appended amino acids may be joined to each other by a linking portion, which may contain any one or more of the linking moieties described herein. Thus in some embodiments, a cell-reactive compstatin analog comprises a compstatin analog moiety M of formula B1-R1- M_1 -R2-B2, wherein M_1 represents any of SEQ ID NOs: 3-36, 37, 37A, 38A, 39A, 40A, or 41A, either R1 or R2 may be absent, at least one of R1 and R2 comprises an amino acid having a side chain that contains a primary or secondary amine, and B1 and B2 are optionally present blocking moieties. R1 and/or R2 may be joined to M_1 by a peptide bond or a non-peptide bond. R1 and/or R2 may comprise a linking portion L^{P3}. For example, R1 can have formula M_2-$L^{P3}$and/or R2 can have formula L^{P3}- M_2 wherein L^{P3} is a linking portion, and M_2 comprises at least one amino acid having a side chain comprising a primary or secondary amine. For example, M_2 can be Lys or an amino acid chain comprising Lys. In some embodiments, L^{P3} comprises of consists of one or more amino acids. For example, L^{P3} can be between 1 and about 20 amino acids in length, e.g., between 4 and 20 amino acids in length. In some embodiments, L^{P3} comprises or consist of multiple Gly, Ser, and/or Ala residues. In some embodiments, L^{P3} does not comprise an amino acid that comprises a reactive SH group, such as Cys. In some embodiments, L^{P3} comprises an oligo(ethylene glycol) moiety and/or a saturated alkyl chain. In some embodiments, L^{P3} is attached to the N-terminal amino acid of M_1 via an amide bond. In some embodiments, L^{P3} is attached to the C-terminal amino acid of M_1 via an amide bond. The compound may be further extended at either or both termini by addition of further linking portion(s) and/or amino acid(s). The amino acids can the same or different and, if different, can be independently selected. In

some embodiments, two or more amino acids having side chains comprising reactive functional groups are used, wherein the reactive functional groups can be the same or different. The two or more reactive functional groups can be used as targets for addition of two or more moieties. In some embodiments, two or more cell-reactive moieties are added. In some embodiments, a cell-reactive moiety and a targeting moiety are added. In some embodiments, a linker and/or cell-reactive moiety is attached to an amino acid side chain after incorporation of the amino acid into a peptide chain. In some embodiments, a linker and/or cell-reactive moiety is already attached to the amino acid side chain prior to use of the amino acid in the synthesis of a cell-reactive compstatin analog. For example, a Lys derivative having a linker attached to its side chain can be used. The linker may comprise a cell-reactive functional group or may subsequently be modified to comprise a cell-reactive functional group.

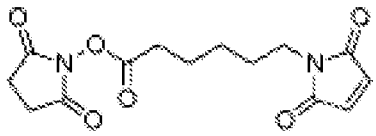
[00144] Certain cell-reactive compstatin analogs are described in further detail below In the following discussion, a peptide having the amino acid sequence Ile-Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr (SEQ ID NO: 37) (corresponding to the compstatin analog of SEQ ID NO: 28, wherein asterisks in SEQ ID NO: 37 represent cysteines joined by a disulfide bond in the active compound, and (1Me)Trp represents 1-methyl-tryptophan)), is used as an exemplary compstatin analog moiety; maleimide (abbreviated Mal) is used as an example of a cell-reactive functional group; $(CH2)_n$ and $(O\text{-}CH2\text{-}CH2)_n$ are used as examples of linking portions; lysine is used as an example of an amino acid comprising a reactive functional group (in some compounds), and acetylation and amidation of the N- and C-termini, respectively, are used as optionally present exemplary blocking moieties in some compounds and are represented in italics, i.e., as *Ac* and NH_2 respectively. It will be appreciated that the compounds can be prepared using a variety of synthetic approaches and using a variety of precursors. The discussion of various synthetic approaches and precursors below is not intended to limit the invention. In general, any of the features of any of the compounds described below or herein can be freely combined with feature(s) of other compounds described below or elsewhere herein, and the invention encompasses such embodiments.

[00145] In some embodiments, the cell-reactive moiety is provided by a cell-reactive compound comprising a maleimide group (as a cell-reactive functional group) and an alkanoic acid (RCOOH), where R is an alkyl group. For example, 6-malemeidocaproic acid (Mal-$(CH_2)_5$-COOH), depicted below, can be used.



Compound **I**

[00146] In some embodiments, the cell-reactive moiety is provided by a derivative of an alkanoic acid in which the carboxylic acid moiety has been activated, e.g., the OH moiety has been converted to a better leaving group. For example, the carboxyl group of compound **I** may be reacted with EDC, followed by reaction with NHS (which can optionally be provided as water-soluble sulfo-NHS), resulting in an N-hydroxysuccinimide ester derivative of 6-malemeidocaproic acid, i.e., 6-maleimidohexanoic acid N-hydroxysuccinimide (NHS) ester (depicted below).



Compound **II**

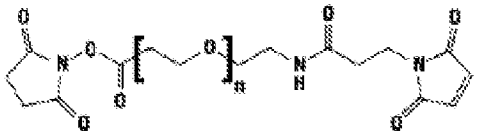
[00147] The compound of SEQ ID NO: 37 can be modified at the N- and/or C- terminus to generate a cell-reactive compstatin analog. For example, compound **II** can be used to generate the following cell-reactive compstatin analog by reaction with the N-terminal amino group of Ile.

[00148] Maleimide-$(CH_2)_5$-C(=O)-Ile-Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-NH_2 (SEQ ID NO: 38). It will be appreciated that in SEQ ID NO: 38 the -C(=O) moiety is attached to the immediately C-terminal amino acid (Ile), via a C-N bond, wherein the N is part of the amino acid and is not shown.

[00149] In other embodiments, a maleimide group is linked to Thr at the C-terminus, resulting in the following cell-reactive compstatin analog:

[00150] *Ac*-Ile-Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-(C=O)-$(CH_2)_5$-maleimide (SEQ ID NO: 39).

[00151] In some embodiments, a cell-reactive compstatin analog can be synthesized using bifunctional linker (e.g., a heterobifunctional linker). An exemplary heterobifunctional linker comprising $(CH_2\text{-}CH_2\text{-}O)_n$ and $(CH_2)_m$ (where m=2) moieties is shown below:



Compound **III**

[00152] Compound **III** comprises a maleimide group as a cell-reactive functional group and an NHS ester moiety that reacts readily with an amino group (e.g., an N-terminal amino group or an amino group of an amino acid side chain).

[00153] An embodiment of compound **III** in which n = 2 can be used to generate the following cell-reactive compstatin analog using the compstatin analog of SEQ ID NO: 37:

[00154] Maleimide-$(CH_2)_2$-C(=O)-NH-$CH_2CH_2OCH_2CH_2OCH_2CH_2$C(=O)-Ile-Cys*-Val-(1Me)Trp -Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-NH_2 (SEQ ID NO: 40)

[00155] It will be appreciated that in the compound of SEQ ID NO: 40 a -C(=O) moiety is attached to the N-terminal amino acid (Ile residue via a C-N bond, wherein the N is part of the amino acid and is not shown. In some embodiments a linker has the formula of Compound III wherein $n \geq 1$. Exemplary values for n in a $(CH_2\text{-}CH_2\text{-}O)_n$ moiety are provided herein.

[00156] In some embodiments, the alkyl chain that links the maleimide moiety to the rest of the molecule contains more or fewer methylene units, the oligo(ethylene glycol) moiety contains more or fewer ethylene glycol units, and/or there are more or fewer methylene units flanking either or both sides of the oligo(ethylene glycol) moiety, as compared with the compound of SEQ ID NO: 39 or SEQ ID NO: 40. Exemplary cell-reactive compstatin analogs illustrative of a few such variations are presented below (SEQ ID NOs: 41-46):

[00157] Maleimide-$(CH_2)_2$-C(=O)-NH-$CH_2CH_2OCH_2CH_2$C(=O)-Ile-Cys*-Val-(1Me)Trp -Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-NH_2 (SEQ ID NO: 41)

[00158] Maleimide-$(CH_2)_3$-C(=O)-NH-$CH_2CH_2OCH_2CH_2OCH_2$C(=O)-Ile-Cys*-Val-(1Me)Trp -Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-NH_2 (SEQ ID NO: 42)

[00159] Maleimide-$(CH_2)_5$-C(=O)-NH-$CH_2CH_2OCH_2CH_2OCH_2$C(=O)-Ile-Cys*-Val-(1Me)Trp -Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-NH_2 (SEQ ID NO:43)

[00160] Maleimide-$(CH_2)_4$-C(=O)-NH-$CH_2CH_2OCH_2CH_2OCH_2CH_2$C(=O)-Ile-Cys*-Val-(1Me)Trp -Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-*NH_2* (SEQ ID NO: 44)

[00161] Maleimide-$(CH_2)_2$-C(=O)-NH-$CH_2CH_2OCH_2CH_2OCH_2CH_2$C(=O)-Ile-Cys*-Val-(1Me)Trp -Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-*NH_2* (SEQ ID NO: 45)

[00162] Maleimide-$(CH_2)_5$-C(=O)-NH-$CH_2CH_2OCH_2CH_2OCH_2$C(=O)-Ile-Cys*-Val-(1Me)Trp -Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-*NH_2* (SEQ ID NO: 46)

[00163] In some embodiments, SEQ ID NO: 37 is extended to comprise a Lys residue at the N- or C- terminus of the peptide, e.g., as exemplified below for a C-terminal linkage:

[00164] *Ac*-Ile-Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-Lys-*NH_2* (SEQ ID NO: 47).

[00165] In some embodiments, a Lys residue is attached to the N- or C- terminus of SEQ ID NO: 37 via a peptide linker, e.g., as exemplified below for a C-terminal linkage:

[00166] *Ac*-Ile-Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-$(Gly)_5$-Lys-*NH_2* (SEQ ID NO: 48).

[00167] In some embodiments, a linker comprising a primary or secondary amine is added to the N- or C-terminus of a compstatin analog. In some embodiments, the linker comprises an alkyl chain and/or an oligo(ethylene glycol) moiety. For example, $NH_2(CH_2CH_2O)nCH_2C(=O)OH$ (e.g., 8-amino-3,6-dioxaoctanoic acid (AEEAc) or 11-amino-3,6,9-trioxaundecanoic acid) or an NHS ester thereof (e.g., an NHS ester of 8-amino-3,6-dioxaoctanoic acid or 11-amino-3,6,9-trioxaundecanoic acid), can be used. In some embodiments, the resulting compound is as follows (wherein the portion contributed by the linker is shown in bold):

[00168] **$NH_2(CH_2)_5C(=O)$**-Ile-Cys-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys-Thr-*NH_2* (SEQ ID NO: 49)

[00169] **$NH_2(CH_2CH_2O)_2CH_2C(=O)$**-Ile-Cys-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys-Thr-*NH_2* (SEQ ID NO: 50)

[00170] In some embodiments, a Lys residue is attached to the N- or C-terminus of SEQ ID NO: 37 via a linker comprising a non-peptide portion. For example, the linker can comprise an alkyl chain, oligo(ethylene glycol) chain, and/or cyclic ring system. In some embodiments, 8-AEEAc or an NHS ester thereof is used, resulting (in the case of attachment of Lys at the C-terminus) in the following compound (wherein the portion contributed by 8-AEEAc is shown in bold):

[00171] *Ac*-Ile-Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr -**NH-$CH_2CH_2OCH_2CH_2OCH_2$-C(=O)**-Lys-*NH_2* (SEQ ID NO: 51)

[00172] It will be appreciated that in SEQ ID NOs: 49 and 50, a -C(=O) moiety is attached to the adjacent Ile residue via a C-N bond, wherein the N is part of the amino acid and is not shown. Similarly, in SEQ ID NO: 51, a -C(=O) moiety is attached to the adjacent Lys residue via a C-N bond, wherein the N is part of the amino acid and is not shown. It will also be appreciated that that in SEQ ID NO: 51 the NH moiety is attached to the immediately N-terminal amino acid (Thr), via a C-N bond, wherein the C is the carbonyl carbon of the amino acid and is not shown.

[00173] The compounds of SEQ ID NOs: 47-51 can readily be modified at the primary amine group to produce a cell-reactive compstatin analog. For example, the compounds of SEQ ID NOs: 47-51 (or other compounds comprising a primary or secondary amine and a compstatin analog moiety) can be reacted with 6-maleimidocaproic acid N-succinimidyl ester to produce the following cell-reactive compstatin analogs:

[00174] *Ac*-Ile-Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-Lys-(C(=O)-$(CH_2)_5$-Mal)-*NH_2* (SEQ ID NO: 52).

[00175] *Ac*-Ile-Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-$(Gly)_5$-Lys--(C(=O)-$(CH_2)_5$-Mal)-*NH_2* (SEQ ID NO: 53).

[00176] Mal-$(CH_2)_5$-(C(=O)-**$NH(CH2)_5C(=O)$**-Ile-Cys-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys-Thr-*NH_2* (SEQ ID NO: 54)

[00177] Mal-$(CH_2)_5$-(C(=O)**$NH(CH_2CH_2O)_2CH_2C(=O)$**-Ile-Cys-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys-Thr-*NH_2* (SEQ ID NO: 55)

[00178] *Ac*-Ile-Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr -**$NH-CH_2CH_2OCH_2CH_2OCH_2-C(=O)$**-Lys-(C(=O)-$(CH_2)_5$-Mal)-*$NH_2$* (SEQ ID NO: 56)

[00179] In another embodiment, a cell-reactive compstatin analog is represented as: *Ac*-Ile-Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-Lys-C(=O)-$CH_2(OCH_2CH2)_2NH$(C(=O)-$(CH_2)_5$-Mal)-*NH_2* (SEQ ID NO: 57).

[00180] The invention provides variants of SEQ ID NOs: 38-57 in which -Ile-Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr- is replaced by an amino acid sequence comprising the amino acid sequence of any other compstatin analog, e.g., of any of SEQ ID NOs 3-27 or 29-36, 37, 37A, 38A, 39A, 40A, or 41A with the proviso that blocking moiet(ies) present at the N- and/or C-termini of a compstatin analog may be absent, replaced by a linker (which may comprise a blocking moiety), or attached to a different N- or C-terminal amino acid present in the corresponding variant(s).

[00181] Other bifunctional cross-linkers comprising a maleimide as a cell-reactive moiety and an NHS ester as an amine-reactive moiety of use in various embodiments of the

invention include, e.g., succinimidyl 4-(p-maleimidophenyl)butyrate (SMPB); succinimidyl 4-(N-maleimidomethyl)-cyclohexane-1-carboxylate (SMCC); N-γ-maleimidobutyryl-oxysuccinimide ester (GMBS). Addition of a sulfonate to the NHS ring results in water-soluble analogs such as sulfo-succinimidyl(4-iodoacetyl)-aminobenzoate (sulfo-SIAB), sulfo-succinimidyl 4-(N-maleimidomethyl)-cyclohexane-1-carboxylate (sulfo-SMCC), sulfo-succinimidyl 4-(p-maleimidophenyl)butyrate (sulfo-SMPB), sulfo-N-γ-maleimidobutyryl-oxysuccinimide ester (sulfo-GMBS) etc., which can avoid the need for an organic solvent. In some embodiments, a long chain version of any of the foregoing, comprising a spacer arm between the NHS ester moiety and the remainder of the molecule, is used. The spacer can comprise, e.g., an alkyl chain. An example is succinimidyl-4-[N-Maleimidomethyl]cyclohexane-1-carboxy-[6-amidocaproate].

[00182] In some embodiments, a bifunctional linker comprising an NHS ester (as an amine-reactive moiety) and an iodoacetyl group (reactive with sulfhydryl groups) is used. Such linkers include, e.g., *N*-succinimidyl(4-iodoacetyl)-aminobenzoate (SIAB); succinimidyl 6-[(iodoacetyl)-amino]hexanoate (SIAX); succinimidyl 6-[6-(((iodoacetyl)amino)-hexanoyl) amino]hexanoate (SIAXX); succinimidyl 4-((iodoacetyl)amino)methyl)-cyclohexane-1-carboxylate (SIAC); succinimidyl 6-((((4-(iodoacetyl)amino)methyl-cyclohexane-1-carbonyl)amino)hexanoate (SIACX);

[00183] In some embodiments, a bifunctional linker comprising an NHS ester (as an amine-reactive moiety) and a pyridy disulfide group (as a cell-reactive moiety reactive with sulfhydryl groups) is used. Examples include *N*-succinimidyl 3-(2-pyridyldithio)propionate (SPDP); succinimidyloxycarbonyl-α-methyl-α-(2-pyridyldithio)toluene (SMPT) and versions comprising a sulfonate on the NHS ring and/or a spacer compsing an alkyl chain between the NHS ester moiety and the rest of the molecule (e.g., succinimidyl 6-(3-[2-pyridyldithio]-propionamido)hexanoate) (LC-SPDP). Variations of such linkers that include additional or different moieties could be used. For example, a longer or shorter alkyl chain could be used in a spacer, or an oligo(ethylene glycol) moiety instead of an alkyl chain.

[00184] In general, a cell-reactive compstatin analog can be synthesized using a variety of approaches. Cell-reactive compounds that comprise a cell-reactive functional group and a linker can often be purchased as preformed building blocks. For example, 6-malemeidocaproic acid and 6-maleimidocaproic acid N-hydroxysuccinimide ester can be purchased from various suppliers. Alternately, such compounds can be synthesized using methods known in the art. See, e.g., Keller O, Rudinger J. Helv Chim Acta. 58(2):531-41, 1975 and Hashida S, et al., J Appl Biochem., 6(1-2):56-63, 1984. See also, Hermanson, G.

supra, and references therein, for discussion of methods and reagents of use for synthesizing conjugates. In general, the invention encompasses any method of producing a compound comprising a compstatin analog moiety and a cell-reactive functional group, and the resulting compounds.

[00185] In some embodiments, an amino acid having a linker attached to a side chain is used in the synthesis of a linear peptide. The linear peptide can be synthesized using standard methods for peptide synthesis known in the art, e.g., standard solid-phase peptide synthesis. The linear peptide is then cyclized (e.g., by oxidation of the Cys residues to form an intramolecular disulfide). The cyclic compound may then be reacted with a linker comprising a cell-reactive functional group. In other embodiments, a moiety comprising a cell-reactive functional group is reacted with a linear compound prior to cyclization thereof. In general, reactive functional groups can be appropriately protected to avoid undesired reaction with each other during synthesis of a cell-reactive compstatin analog. The cell-reactive functional group, any of the amino acid side chains, and/or either or both termini of the peptide may be protected during the reaction and subsequently deprotected. For example, SH groups of Cys residues and/or SH-reactive moieties such as maleimides can be protected until after cyclization to avoid reaction between them. The reaction conditions are selected based at least in part on the requirements of the particular reactive functional group(s) to achieve reasonable yield in a reasonable time period. Temperature, pH, and the concentration of the reagents can be adjusted to achieve the desired extent or rate of reaction. See, e.g., Hermanson, *supra*. The desired product can be purified, e.g., to remove unreacted compound comprising the cell-reactive functional group, unreacted compstatin analog, linker(s), products other than the desired cell-reactive compstatin analog that may have been generated in the reaction, other substances present in the reaction mixture, etc. Compositions and methods for making the cell-reactive compstatin analogs, and intermediates in the synthesis, are aspects of the invention.

[00186] In some aspects of the invention, linker(s) described above are used in the production of compstatin analogs comprising a moiety such as a polyethylene glycol (PEG) chain or other polymer(s) that, e.g., stabilize the compound, increase its lifetime in the body, increase its solubility, decrease its immunogenicity, and/or increase its resistance to degradation. Without limiting the invention in any way, such a moiety may be referred to herein as a "clearance reducing moiety" (CRM), and a compstatin analog comprising such a moiety may be referred to as a "long-acting compstatin analog". In some embodiments, a long-acting compstatin analog has an average plasma half-life of at least 1 day, e.g., 1 - 3

days, 3 - 7 days, 7 - 14 days, or 14 - 28 days, when administered IV at a dose of 10 mg/kg to humans or to non-human primates primates, or a dose of about 1-3 mg/kg, 3-5 mg/kg, 5-10 mg/kg, e.g., 7 mg/kg. In some embodiments, a long-acting compstatin analog has an average plasma half-life of at least 1 day, e.g., 1 - 3 days, 3 - 7 days, 7 - 14 days, or 14 - 28 days, when administered subcutaneously at, e.g., a dose of about 1-3 mg/kg, 3-5 mg/kg, 5 -10 mg/kg, e.g., 7 mg/kg to humans or to non-human primates. In some embodiments, a long-acting compstatin analog has an average plasma half-life (e.g., a terminal half-life) of between about 4-10, 5-9, 5-8, 6-9, 7-9, or 8-9 days, e.g., about 4,4.5, 5,5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5 or 10 days when administered intravenously at, e.g., a dose of about 1-3 mg/kg, 3-5 mg/kg, or 5-10 mg/kg, e.g., 7 mg/kg to humans or to non-human primates. In some embodiments, a long-acting compstatin analog has an average plasma half-life (e.g., a terminal half-life) of between about 4-10, 5-9, 5-8, 6-9, 7-9, or 8-9 days, e.g., about 4,4.5, 5,5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5 or 10 days, when administered subcutaneously at, e.g., a dose of about 1-3 mg/kg, 3-5 mg/kg, 5 -10 mg/kg, e.g., 7 mg/kg to humans or to non-human primates. In certain embodiments a long-acting compstatin analog is characterized in that it is extensively absorbed from the site of administration during the time period following subcutaneous injection and provides, e.g., at or after about 1-2 days following administration, a blood level comparable to that which would be achieved had the same amount of compound been administered intravenously instead. In some embodiments, the blood level at or after about 2, 3, 4, 5, 6, 7, 8, or more days following administration of a subcutaneous dose is within about 5%, 10%, 15%, 20%, or 25% of the blood level which would be achieved had the same amount of compound been administered intravenously instead. See, e.g., Figure 11, showing pharmacokinetics of an intravenously and subcutaneously administered dose of an exemplary compound described herein after about 1-2 days following administration. In some embodiments, average plasma half-life of a long-acting compstatin analog following administration IV at a dose of 10 mg/kg to humans or to non-human primates is increased by at least a factor of 2, e.g., by a factor of 2-5, 5-10, 10-50, or 50-100 -fold or 100-150-fold or 150-200 fold as compared with that of a corresponding compstatin analog having the same amino acid sequence (and, if applicable, one or more blocking moiet(ies)) but not comprising the CRM. It will be understood that in various embodiments such an increase in half-life may be observed following administration via other routes such as subcutaneous administration and/or using other doses, e.g., other doses described herein, e.g., 20 mg/kg. As noted above, in some embodiments a compstatin analog of any of SEQ ID NOs: 3-36 , 37, 37A. 38A, 39A, 40A, or 41A is extended by one or more amino acids at the N-terminus, C-

terminus, or both, wherein at least one of the amino acids has a side chain that comprises a reactive functional group such as a primary or secondary amine, a sulfhydryl group, a carboxyl group (which may be present as a carboxylate group), a guanidino group, a phenol group, an indole ring, a thioether, or an imidazole ring, which facilitate conjugation with a reactive functional group to attach a CRM to the compstatin analog. It will be understood that a corresponding compstatin analog not comprising the CRM may also lack one or more such amino acids which are present in the long-acting compstatin analog to which it corresponds. Thus, a corresponding compstatin analog comprising any of SEQ ID NOs: 3-36, 37, 37A. 38A, 39A, 40A, or 41A and lacking a CRM will be understood to "have the same amino acid sequence" as SEQ ID NO: 3-36, 37, 37A. 38A, 39A, 40A, or 41A, respectively. For example, a corresponding compstatin analog comprising the amino acid sequence of SEQ ID NO: 14, 21, 28, 29, 32, 33, 34, or 36 and lacking a CRM will be understood to "have the same amino acid sequence" as SEQ ID NO: 14, 21, 28, 29, 32, 33, 34, or 36, respectively. In some embodiments, a plasma half-life is a terminal half-life after administration of a single IV dose. In some embodiments, a plasma half-life is a terminal half-life after steady state has been reached following administration of multiple IV doses. In some embodiments, a long-acting compstatin analog achieves a Cmax in plasma at least 5-fold as great as that of a corresponding compstatin analog not comprising the CRM, e.g., between 5- and 50-fold as great, following administration of a single IV dose to a primate, or following administration of multiple IV doses. In some embodiments, a long-acting compstatin analog achieves a Cmax in plasma between 10- and 20-fold as great as that of a corresponding compstatin analog not comprising the CRM following administration of a single IV dose to a primate, or following administration of multiple IV doses. In some embodiments a primate is human. In some embodiments a primate is a non-human primate, e.g., a monkey, such as a Cynomolgus monkey or Rhesus monkey. In some embodiments, renal clearance of a long-acting compstatin analog during the first 24 hours following administration IV at a dose of 10 mg/kg or 20 mg/kg to humans or to non-human primates is reduced by at least a factor of 2, e.g., by a factor of 2-5, 5-10, 10-50, or 50-100 -fold or 100-150-fold or 150-200 fold as compared with renal clearance of a corresponding compstatin analog. It will be understood that in various embodiments such a reduction in renal clearance may be observed following administration via other routes such as subcutaneous administration and/or using other doses, e.g., other doses described herein, e.g., 20 mg/kg. The concentration of compstatin analog can be measured in blood and/or urine samples using, e.g., UV, HPLC, mass spectrometry (MS) or antibody to the CRM, or combinations of such

methods, such as LC/MS or LC/MS/MS. Pharmacokinetic parameters such as half-life and clearance can be determined using methods known to those of ordinary skill in the art. Pharmacokinetic analysis can be performed, e.g., with WinNonlin software v 5.2 (Pharsight Corporation, St. Louis, MO) or other suitable programs.

[00187] In certain embodiments a CRM is stable in physiological conditions for at least 24 hours or more. In certain embodiments a CRM is stable in mammalian, e.g., primate, e.g., human or non-human primate (e.g., monkey) blood, plasma, or serum for at least 24 hours. In various embodiments at least 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or more, of the CRM molecules remains intact upon incubation in physiological conditions for 24 hours, 48 hours, 72 hours, 96 hours, 120 hours, 144 hours, 168 hours, or more. In various embodiments at least 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or more, of the CRM molecules remains intact upon incubation in blood, plasma, or serum at 37 degrees C for 48 hours, 72 hours, 96 hours, 120 hours, 144 hours, 168 hours, or more. Incubation may be performed using a CRM at a concentration of between 1 microgram/ml to about 100 mg/ml in various embodiments. Samples may be analyzed at various time points. Size or intactness may be assessed using, e.g., chromatography (e.g., HPLC), mass spectrometry, Western blot, or any other suitable method. Such stability characteristics may be conferred on a moiety conjugated to the CRM. In various embodiments, a long-acting compstatin analog comprising a CRM may have any of the afore-mentioned stability characteristics. In some aspects intact with regard to a long-acting compstatin analog means that the compstatin analog moiety remains conjugated to the CRM and the CRM size remains about the same as at the start of incubation or administration.

[00188] In some embodiments, a long-acting compstatin analog has a molar activity of at least about 10%, 20%, 30%, e.g., between 30% and 40%, between 30% and 50%, between 30% and 60%, between 30% and 70%, between 30% and 80%, between 30% and 90%, or more, of the activity of a corresponding compstatin analog having the same amino acid sequence (and, if applicable, one or more blocking moiet(ies)) but not comprising a CRM. In some embodiments wherein a long-acting compstatin analog comprises multiple compstatin analog moieties, the molar activity of the long-acting compstatin analog is at least about 10%, 20%, or 30%, e.g., between 30% and 40%, between 30% and 50%, between 30% and 60%, between 30% and 70%, between 30% and 80%, between 30% and 90%, or more, of the sum of the activities of said compstatin analog moieties. In some embodiments, a polyethylene glycol (PEG) comprises a $(CH_2CH_2O)_n$ moiety having a molecular weight of at least 500 daltons. In some embodiments, a linker described above comprises an $(CH_2CH_2O)_n$ moiety

having an average molecular weight of between about 500; 1,000; 1,500; 2,000; 5,000; 10,000; 20,000; 30,000; 40,000; 50,000; 60,000; 70,000; 80,000; 90,000; and 100,000 daltons. In some embodiments the average molecular weight of a PEG is at least 20,000 daltons, up to about 100,000; 120,000; 140,000; 160,000; 180,000; or 200,000 daltons. "Average molecular weight" refers to the number average molecular weight. In some embodiments, the polydispersity D of a $(CH_2CH_2O)n$ moiety is between 1.0005 and 1.50, e.g., between 1.005 and 1.10, 1.15,1.20, 1.25, 1.30, 1.40, or 1.50, or any value between 1.0005 and 1.50.

[00189] In some embodiments, a $(CH_2CH_2O)n$ moiety is monodisperse and the polydispersity of a $(CH_2CH_2O)n$ moiety is 1.0. Such monodisperse $(CH_2CH_2O)n$ moieties are known in the art and are commercially available from Quanta BioDesign (Powell, OH), and include, by way of nonlimiting example, monodisperse moieties where n is 2, 4, 6, 8, 12, 16, 20, or 24.

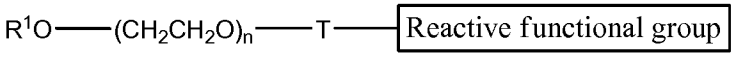
[00190] In some embodiments, a compound comprises multiple $(CH_2CH_2O)_n$ moieties wherein the total molecular weight of said $(CH_2CH_2O)_n$ moieties is between about 1,000; 5,000; 10,000; 20,000; 30,000; 40,000; 50,000; 60,000; 70,000; 80,000; 90,000; and 100,000 daltons. In some embodiments the average total molecular weight of the compound or $(CH_2CH_2O)_n$ moieties is at least 20,000 daltons, up to about 100,000; 120,000; 140,000; 160,000; 180,000; or 200,000 daltons. In some embodiments, the compound comprises multiple $(CH_2CH_2O)_n$ moieties having defined lengths, e.g., n = 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, or 30 or more. In some embodiments, the compound comprises a sufficient number of $(CH_2CH_2O)_n$ moieties having defined lengths to result in a total molecular weight of said $(CH_2CH_2O)_n$ moieties of between about 1,000; 5,000; 10,000; 20,000; 30,000; 40,000; 50,000; 60,000; 70,000; 80,000; 90,000; and 100,000 daltons. In some embodiments the average total molecular weight of the compound or $(CH_2CH_2O)_n$ moieties is at least 20,000 daltons, up to about 100,000; 120,000; 140,000; 160,000; 180,000; or 200,000 daltons. In some embodiments n is between about 30 and about 3000. In some embodiments a compstatin analog moiety is attached at each end of a linear PEG. A bifunctional PEG having a reactive functional group at each end of the chain may be used, e.g., as described above. In some embodiments the reactive functional groups are identical while in some embodiments different reactive functional groups are present at each end. In some embodiments, multiple $(CH_2CH_2O)_n$ moieties are provided as a branched structure. The branches may be attached to a linear polymer backbone (e.g., as a comb-shaped structure) or may emanate from one or more central core groups, e.g., as a star structure. In

some embodiments, a branched molecule has 3 to 10 $(CH_2CH_2O)_n$ chains. In some embodiments, a branched molecule has 4 to 8 $(CH_2CH_2O)_n$ chains. In some embodiments, a branched molecule has 10, 9, 8, 7, 6, 5, 4, or 3 $(CH_2CH_2O)_n$ chains. In some embodiments, a star -shaped molecule has 10–100, 10-50, 10-30, or 10-20 $(CH_2CH_2O)_n$ chains emanating from a central core group. In some embodiments a long-acting compstatin analog thus may comprise, e.g., 3-10 compstatin analog moieties, e.g., 4-8 compstatin analog moieties, each attached to a $(CH_2CH_2O)_n$ chain via a functional group at the end of the chain. In some embodiments a long-acting compstatin analog may comprise, e.g., 10-100 compstatin analog moieties, each attached to a $(CH_2CH_2O)_n$ chain via a functional group at the end of the chain. In some embodiments, branches (sometimes referred to as "arms") of a branched or star-shaped PEG contain about the same number of (CH_2CH_2O) moieties. In some embodiments, at least some of the branch lengths may differ. It will be understood that in some embodiments one or more $(CH_2CH_2O)_n$ chains does not have a comptatin analog moiety attached thereto. In some embodiments at least about 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 100% of the chains has a compstatin analog moiety attached thereto.

[00191] In genera and compounds depicted herein, a polyethylene glycol moiety is drawn with the oxygen atom on the right side of the repeating unit or the left side of the repeating unit. In cases where only one orientation is drawn, the present invention encompasses both orientations (i.e., $(CH_2CH_2O)_n$ and $(OCH_2CH_2)_n$) of polyethylene glycol moieties for a given compound or genus, or in cases where a compound or genus contains multiple polyethylene glycol moieties, all combinations of orientations are encompasses by the present disclosure.

[00192] Formulas of some exemplary monofunctional PEGs comprising a reactive functional group are illustrated below. For illustrative purposes, formulas in which the reactive functional group(s) comprise an NHS ester are depicted, but other reactive functional groups could be used, e.g., as described above. In some embodiments, the $(CH_2CH_2O)_n$ are depicted as terminating at the left end with a methoxy group (OCH_3) but it will be understood that the chains depicted below and elsewhere herein may terminate with a different OR moiety (e.g., an aliphatic group, an alkyl group, a lower alkyl group, or any other suitable PEG end group) or an OH group. It will also be appreciated that moieties other than those depicted may connect the $(CH_2CH_2O)_n$ moieties with the NHS group in various embodiments.

[00193] In some embodiments, a monofunctional PEG is of formula A:



Formula A

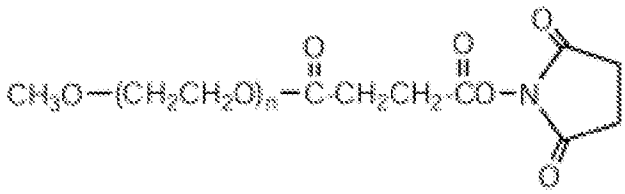
wherein "Reactive functional group" and n are as defined above and described in classes and subclasses herein;

R^1 is hydrogen, aliphatic, or any suitable end group; and

T is a covalent bond or a C_{1-12} straight or branched, hydrocarbon chain wherein one or more carbon units of T are optionally and independently replaced by -O-, -S-, -N(R^x)-, -C(O)-, -C(O)O-, -OC(O)-, -N(R^x)C(O)-, -C(O)N(R^x)-, -S(O)-, -S($O)_2$-, -N(R^x)SO_2-, or -SO_2N(R^x)-; and

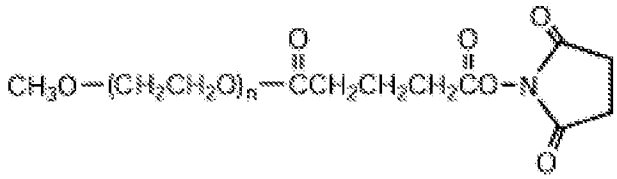
each R^x is independently hydrogen or C_{1-6} aliphatic.

[00194] Exemplary monofunctional PEGs of formula A include:

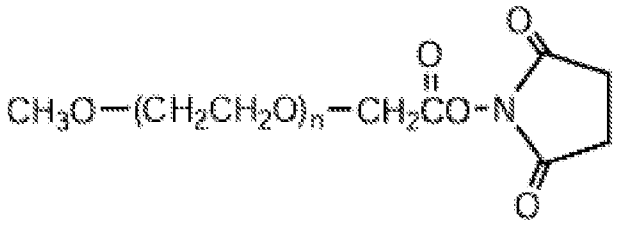


Formula I

[00195] In Formula I, the moiety comprising the reactive functional group has the general structure -CO-$(CH_2)_m$-COO-NHS, where m=2. In some embodiments, a monofunctional PEGs has the structure of Formula I, where m is between 1 and 10, e.g., between 1 and 5. For example, in some embodiments m is 3, as shown below:

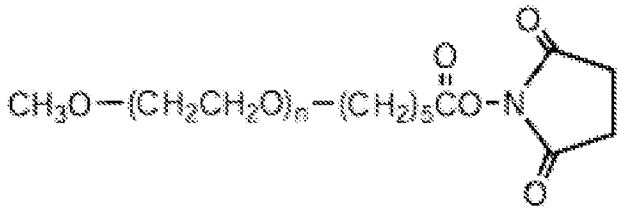


Formula Ia.

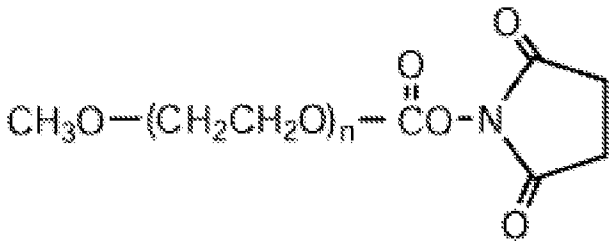


Formula II

[00196] In Formula II, the moiety comprising the reactive functional group has the general structure $-(CH_2)_m$-COO-NHS, where m=1. In some embodiments a monofunctional PEG has the structure of Formula II, where m is between 1 and 10 (e.g., wherein m is 5 as shown in Formula III below), or wherein m is 0 (as shown below in Formula IIIa).



Formula III

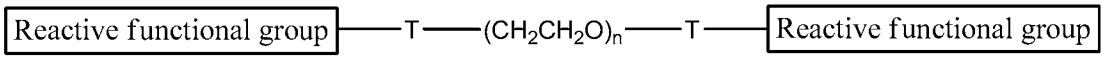


Formula IIIa

[00197] In some embodiments a bifunctional linear PEG comprises a moiety comprising a reactive functional group at each of its ends. The reactive functional groups may be the same (homobifunctional) or different (heterobifunctional). In some embodiments the structure of a bifunctional PEG may be symmetric, wherein the same moiety is used to connect the reactive functional group to oxygen atoms at each end of the $-(CH_2CH_2O)_n$ chain. In some

embodiments different moieties are used to connect the two reactive functional groups to the PEG portion of the molecule. The structures of exemplary bifunctional PEGs are depicted below. For illustrative purposes, formulas in which the reactive functional group(s) comprise an NHS ester are depicted, but other reactive functional groups could be used.

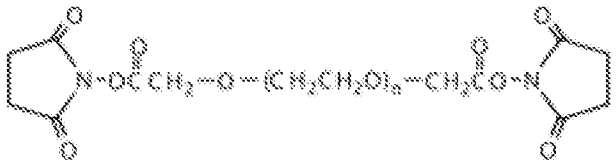
[00198] In some embodiments, a bifunctional linear PEG is of formula B:



Formula B

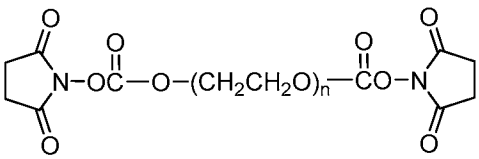
wherein each T and "Reactive functional group" is independently as defined above and described in classes and subclasses herein, and n is as defined above and described in classes and subclasses herein.

[00199] Exemplary bifunctional PEGs of formula B include:

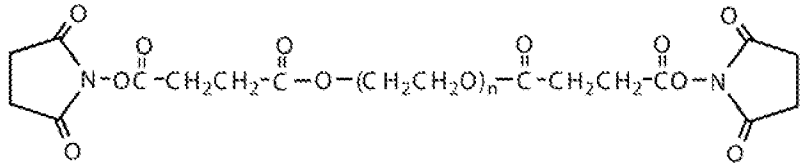


Formula IV

[00200] In Formula IV, the moiety comprising the reactive functional group has the general structure $-(CH_2)_m$-COO-NHS, where m=1. In some embodiments, a bifunctional PEG has the structure of Formula IV, where m is between 1 and 10, e.g., between 1 and 5. In certain embodiments m is 0, e.g., embodiments the moiety comprising the reactive functional group has the general structure –COO-NHS. For example, in some embodiments a bifunctional PEG has the structure of Formula IVa, as shown below:

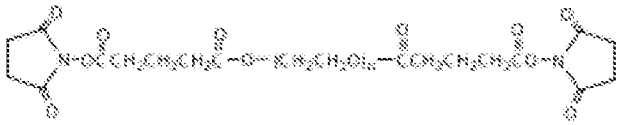


Formula IVa



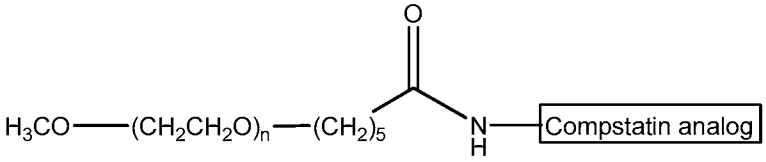
Formula V

[00201] In Formula V, the moiety comprising the reactive functional group has the general structure -CO-$(CH_2)_m$-COO-NHS, where m=2. In some embodiments, a bifunctional PEGs has the structure of Formula V, where m is between 1 and 10, e.g., between 1 and 5. In certain embodiments, for example, m is 2, as shown below:

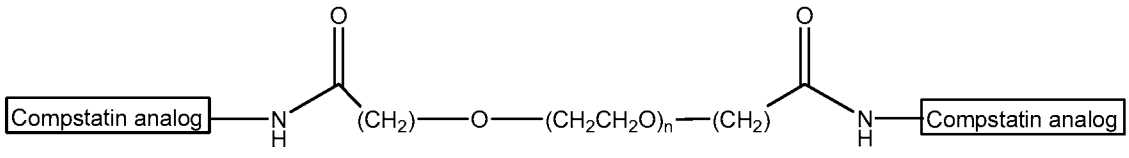


Formula Va

[00202] In some embodiments, the present invention provides a compstatin analog conjugated to a polymer. In certain embodiments, the present invention provides compstatin analog conjugates of PEG-containing compounds and genera depicted herein. In some embodiments, a functional group (for example, an amine, hydroxyl, or thiol group) on a compstatin analog is reacted with a PEG-containing compound having a "reactive functional group" as described herein, to generate such conjugates. By way of example, Formulae III and IV, respectively, can form compstatin analog conjugates having the structure:

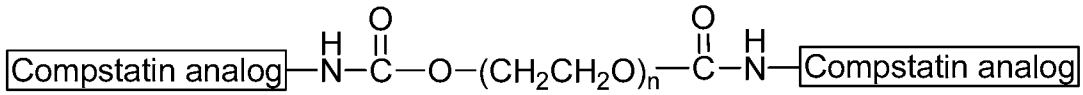


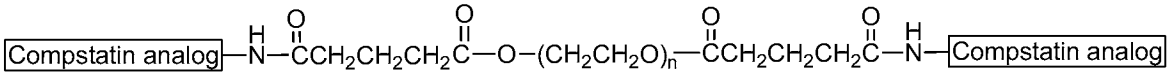
or



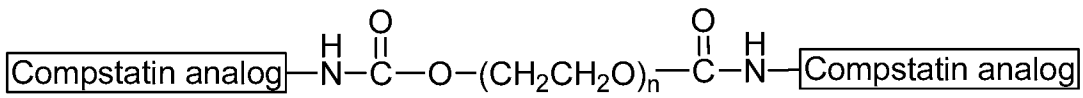
wherein,  represents the attachment point of an amine group on a compstatin analog. In certain embodiments, an amine group is a lysine side chain group. It will be appreciated that corresponding conjugates can be formed with any of the PEG-

containing compounds and genera depicted herein, depending on the choice of reactive functional group and/or compstatin functional group. For example, Formulae IVa and Va, respectively, can form compstatin analog conjugates having the following structures

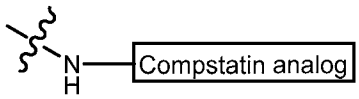




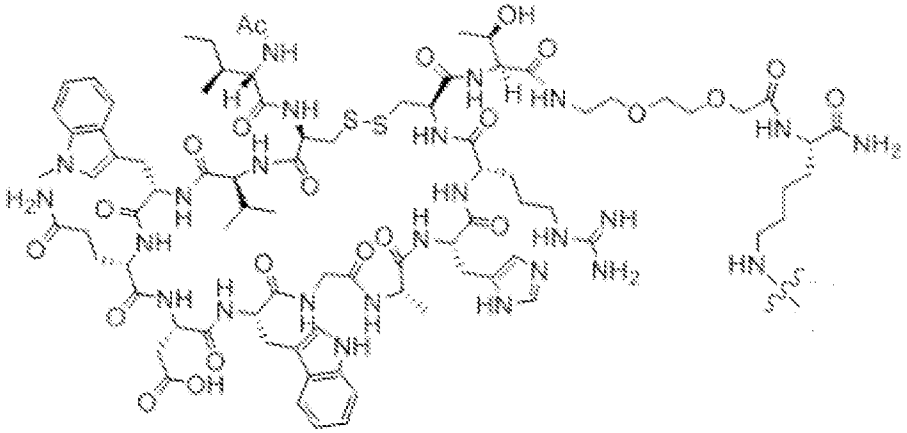
In certain embodiments, the PEG component of such conjugates has an average molecular weight of between about 20 kD -100 kD, about 20 kD-90 kD, about 20 kD-80 kD, about 20 kD-70 kD, about 20 kD-60 kD, about 20 kD-50 kD, about 30 kD-80 kD, about 30 kD-70 kD, about 30 kD-60 kD, about 30 kD-50 kD, about 30 kD-45 kD, about 35 kD-50 kD, about 35 kD-45 kD, about 36 kD-44 kD, about 37 kD-43 kD, about 38 kD-42 kD, or about 39 kD-41 kD. In certain embodiments, the PEG component of such conjugates has an average molecular weight of about 40 kD. The term "bifunctional" or "bifunctionalized" is sometimes used herein to refer to a compound comprising two compstatin analog moieties linked to a CRM. Such compounds may be designated with the letter "BF". In some embodiments a bifunctionalized compound is symmetrical. In some embodiments the linkages between the CRM and each of the compstatin analog moieties of a bifunctionalized compound are the same. In some embodiments, each linkage between a CRM and a compstatin analog of a bifunctionalized compound comprises a carbamate. In some embodiments, each linkage between a CRM and a compstatin analog of a bifunctionalized compound comprises a carbamate and does not comprise an ester. In some embodiments, each compstatin analog of a bifunctionalized compound is directly linked to a CRM via a carbamate. In some embodiments, each compstatin analog of a bifunctionalized compound is directly linked to a CRM via a carbamate, and the bifunctionalized compound has the structure:



[00203] In some embodiments of formulae and embodiments described herein,

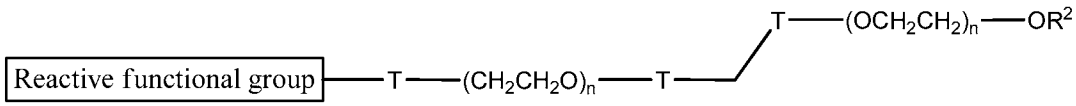


represents point of attachment of a lysine side chain group in a compstatin analog having the structure:

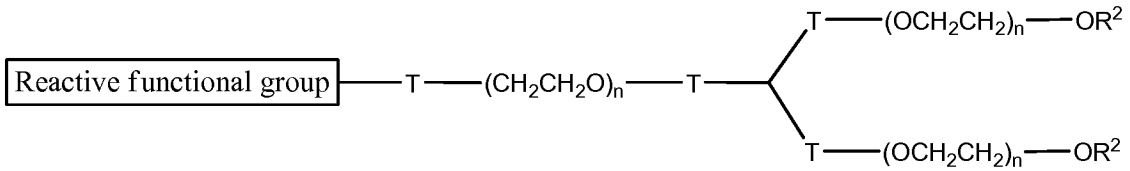


wherein the symbol "〰" denotes the point of attachment of a chemical moiety to the remainder of a molecule or chemical formula.

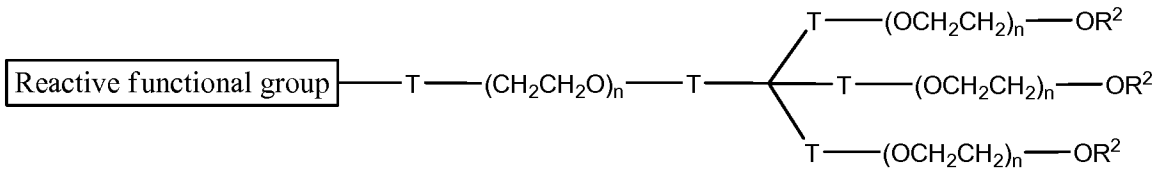
[00204] In some embodiments, a branched, comb, or star-shaped PEG comprises a moiety comprising a reactive functional group at the end of each of multiple -$(CH_2CH_2O)_n$ chains. The reactive functional groups may be the same or there may be at least two different groups. In some embodiments, a branched, comb, or star-shaped PEG is of the following formulae:



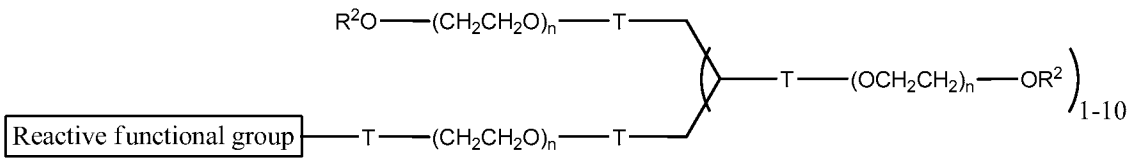
Formula C



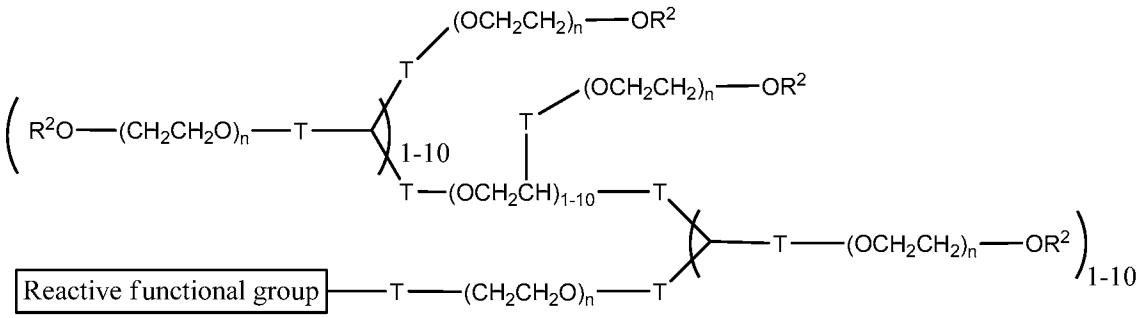
Formula D



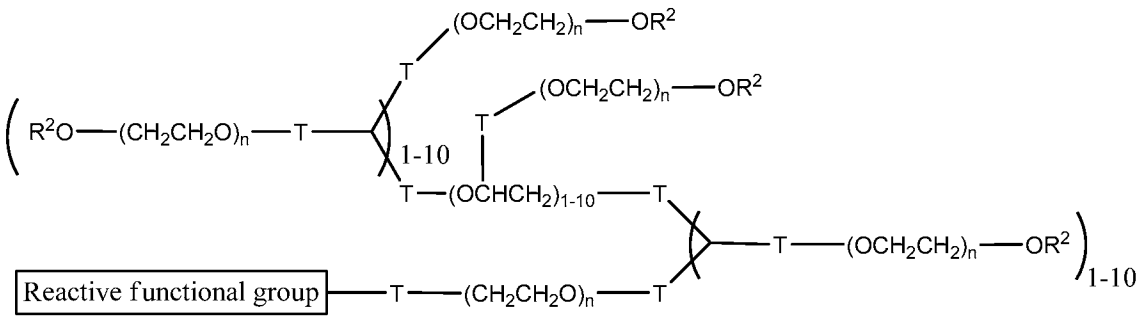
Formula E



Formula F

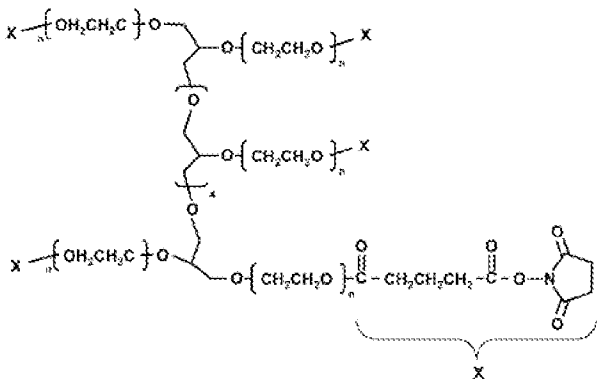


Formula G

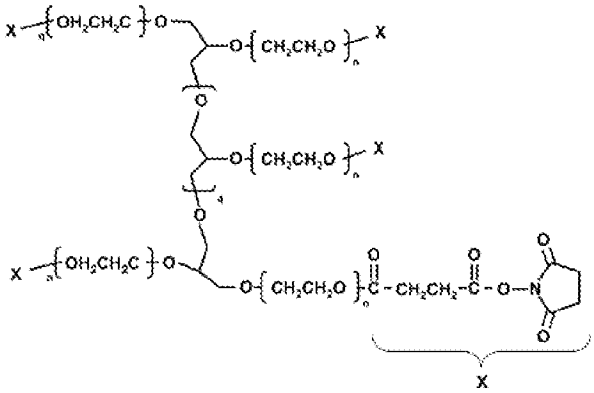


Formula H

[00205] wherein each R^2 is independently a "Reactive functional group" or R^1, and each T, n, and "Reactive functional group" is independently as defined above and described in classes and subclasses herein. The structure of exemplary branched PEGs (having 8 arms, or branches) comprising NHS moieties as reactive functional groups is depicted below:

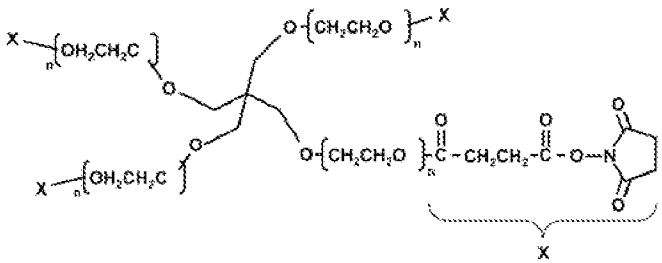


Formula VI

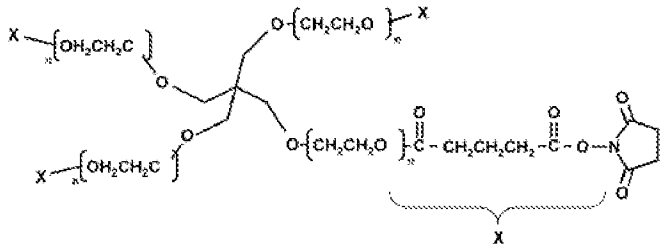


Formula VII

[00206] The structure of exemplary branched PEGs (having 4 arms, or branches) comprising NHS moieties as reactive functional groups is depicted below:



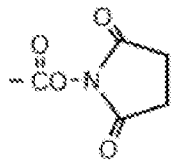
Formula VIII



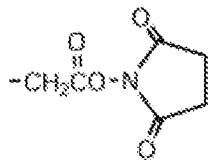
Formula IX

[00207] The number of branches emanating from the backbone may be varied. For example, the number 4 in the above formulae VI and VII may be changed to any other integer between 0 and 10 in various embodiments. In certain embodiments, one or more branches does not contain a reactive function group and the branch terminates with a $-CH_2CH_2OH$ or $-CH_2CH_2OR$ group, as described above.

[00208] In some embodiments a branched PEG has the structure of Formula VII, VIII, or IX (or variants thereof having different numbers of branches) with the proviso that x is

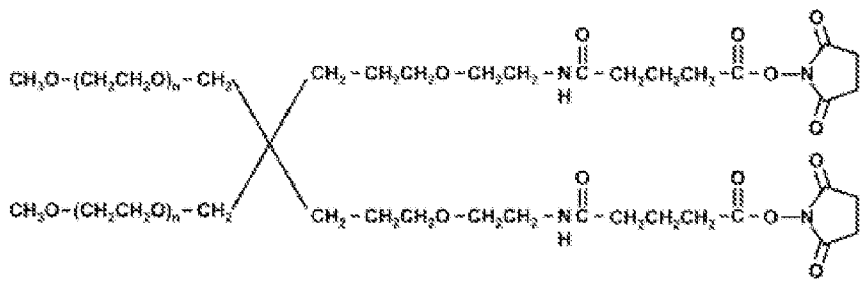


[00209] In some embodiments a branched PEG has the structure of Formula VII, VIII, or IX (or variants thereof having different numbers of branches) with the proviso that x is

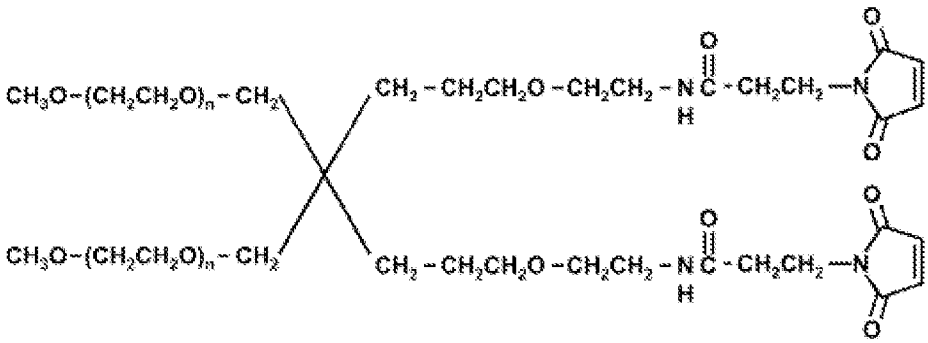


[00210] Of course the methylene (CH_2) group in the above x moiety may instead comprise a longer alkyl chain $(CH_2)_m$, where m is up to 2, 3, 4, 5, 6, 8, 10, 20, or 30, or may comprise one or more other moieties described herein.

[00211] In some embodiments, exemplary branched PEGs having NHS or maleimde reactive groups are depicted below:



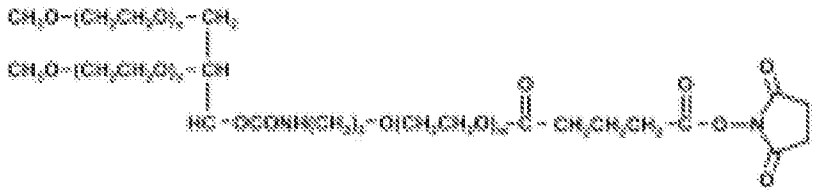
Formula X



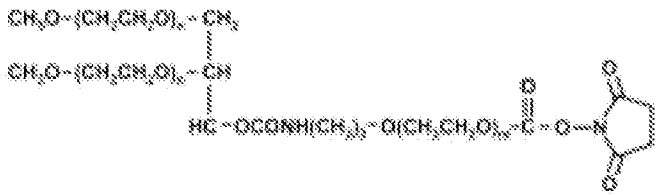
Formula XI

[00212] In some embodiments, a variant of Formula X or XI are used, wherein 3 or each of the 4 branches comprise a reactive functional group.

[00213] Still other examples of PEGs may be represented as follows:



Formula XII



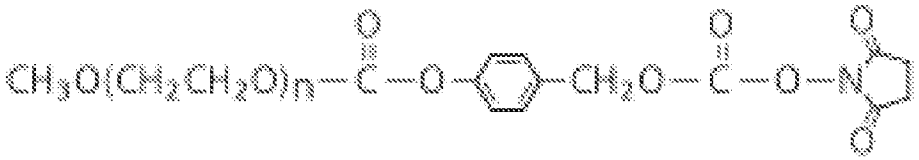
[00214]

Formula XIII

As noted above, it will be appreciated that, as described herein, in various embodiments any of a variety of moieties may be incorporated between the peptide component and $(CH_2CH_2O)_n$-R moiety of a long-acting compstatin analog, such as an linear alkyl, ester,

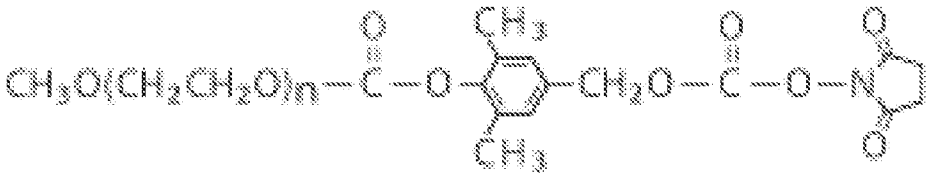
amide, aromatic ring (e.g., a substituted or unsubstituted phenyl), a substituted or unsubstituted cycloalkyl structure, or combinations thereof. In some embodiments such moiet(ies) may render the compound more susceptible to hydrolysis, which may release the peptide portion of the compound from the CRM. In some embodiments, such release may enhance the in vivo tissue penetration and/or activity of the compound. In some embodiments hydrolysis is general (e.g., acid-base) hydrolysis. In some embodiments hydrolysis is enzyme-catalyzed, e.g., esterase-catalyzed. Of course both types of hydrolysis may occur. Examples of PEGs comprising one or more such moieties and an NHS ester as a reactive functional group are as follows:

[00215]



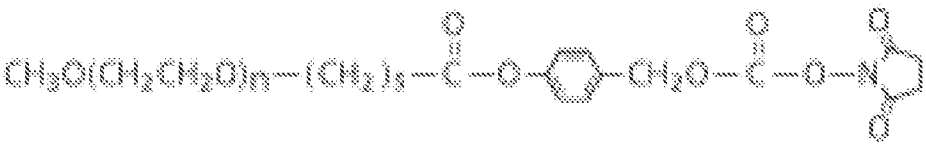
Formula XIV

[00216]



Formula XV

[00217]



Formula XVI

[00218] In some embodiments a branched (multi-arm) PEG or star-shaped PEG comprises a pentaerythritol core, hexaglycerin core, or tripentaerythritol core. It will be understood that the branches may not all emanate from a single point in certain embodiments.

[00219] Monofunctional, bifunctional, branched, and other PEGs comprising one or more reactive functional groups may, in some embodiments, be obtained from, e.g., NOF America Corp. White Plains, NY or BOC Sciences 45-16 Ramsey Road Shirley, NY 11967, USA, among others, or may be prepared using methods known in the art.

[00220] In some embodiments, a linkage between a CRM and a compstatin analog comprises a carbamate. In some embodiments, a compstatin analog is directly linked to a CRM via a carbamate. In some embodiments, a linkage between a CRM and a compstatin analog does not comprise an ester. In some embodiments, a linkage between a CRM and a compstatin analog comprises a carbamate and does not comprise an ester. In some embodiments, a linkage between a CRM and a compstatin analog comprises a carbamate and does not comprise a bond that is more susceptible to hydrolysis in aqueous medium than a carbamate. In some embodiments the CRM comprises or consists of a PEG moiety.

[00221] In some embodiments, a linkage between a CRM and a compstatin analog comprises an amide. In some embodiments, a compstatin analog is directly linked to a CRM via an amide. In some embodiments, a linkage between a CRM and a compstatin analog comprises an amide and does not comprise an ester. In some embodiments, a linkage between a CRM and a compstatin analog comprises an amide and does not comprise a bond that is more susceptible to hydrolysis in aqueous medium than an amide. In some embodiments the CRM comprises or consists of a PEG moiety.

[00222] In some embodiments, one or more compstatin analogs of a multifunctionalizedcompound (e.g., a bifunctionalized, trifunctionalized, or more extensively functionalized compound) is linked to a CRM by a linkage comprising a carbamate. In some embodiments, one or more compstatin analogs of a multifunctionalized compound (e.g., a bifunctionalized, trifunctionalized, or more extensively functionalized compound) is linked to a CRM by a linkage that does not comprise an ester. In some embodiments, one or more compstatin analogs of a multifunctionalizedcompound (e.g., a bifunctionalized, trifunctionalized, or more extensively functionalized compound) is linked to a CRM by a linkage that comprises a carbamate and does not comprise an ester. In some embodiments, one or more compstatin analogs of a multifunctionalizedcompound (e.g., a bifunctionalized, trifunctionalized, or more extensively functionalized compound) is linked to a CRM by a linkage that comprises a carbamate and does not comprise a bond that is more susceptible to hydrolysis in aqueous medium than a carbamate. In some embodiments, each compstatin analog of a multifunctionalized compound (e.g., a bifunctionalized, trifunctionalized, or more extensively functionalized compound) is directly linked to a CRM via a carbamate.

[00223] In some embodiments the CRM comprises or consists of a PEG moiety. In some embodiments, one or more compstatin analogs of a multifunctionalizedcompound (e.g., a bifunctionalized, trifunctionalized, or more extensively functionalized compound) is linked to a CRM by a linkage comprising an amide. In some embodiments, one or more compstatin

analogs of a multifunctionalizedcompound (e.g., a bifunctionalized, trifunctionalized, or more extensively functionalized compound) is linked to a CRM by a linkage that comprises an amide and does not comprise an ester. In some embodiments, one or more compstatin analogs of a multifunctionalizedcompound (e.g., a bifunctionalized, trifunctionalized, or more extensively functionalized compound) is linked to a CRM by a linkage that comprises an amide and does not comprise a bond that is more susceptible to hydrolysis in aqueous medium than an amide. In some embodiments, each compstatin analog of a multifunctionalized compound (e.g., a bifunctionalized, trifunctionalized, or more extensively functionalized compound) is directly linked to a CRM via an amide. In some embodiments the CRM comprises or consists of a PEG moiety.

[00224] In some embodiments, the present invention provides a compstatin analog conjugated with a polymer, wherein the polymer is other than PEG. In some embodiments, a polymer is a polyoxazoline (POZ). Exemplary mono- and poly-functionalized polyoxazoline derivatives for direct conjugation, or for conjugation via a linker, are depicted below:

$Z–T–[N(COR^x)CH_2CH_2]_n–T–R^1$;

$R^1–\{[N(CO–T–Z)CH_2CH_2]_m–[N(COR^x)CH_2CH_2]_n\}^a–T–R^1$;

$R^1–\{[N(CO–T–Z^1)CH_2CH_2]_p–[N(COR^x)CH_2CH_2]_n–[N(CO–T–Z^2)CH_2CH_2]_m\}^a–T–R^1$;

$R^1–\{[N(CO–T–Z^1)CH_2CH_2]_p–[N(COR^x)CH_2CH_2]_n–[N(CO–T–Z^2)CH_2CH_2]_m\}^a–T–Z$;

$R^1–[N(COR^x)CH_2CH_2]_n–T–B(–R^1)(–T–Z)–T–[N(COR^x)CH_2CH_2]_m–R^1$;

wherein:

each of Z, Z^1 and Z^2 is independently a reactive functional group as defined above and described in classes and subclasses herein;

each of T, R^x, and R^1 is independently as defined above and described in classes and subclasses herein;

each of m, n, and p is independently an integer 0–1000, with the limitation that the sum of m, n, and p for each formula is not 0;

a is "ran," which indicates a random copolymer, or "block," which indicates a block copolymer;

B is a branching moiety that is linked with or without a linker to the other parts of the polymer.

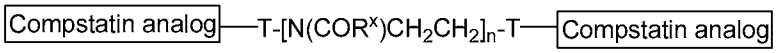
Other examples of functionalized polyoxazoline derivatives for conjugation are extensively described in the art, including but not limited to those described in PCT Patent Application

Publication Nos. WO/2010/006282, WO/2009/089542, WO/2009/043027 and WO/2008/106186, the entirety of each of which is hereby incorporated by reference.

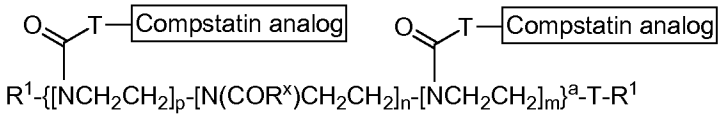
[00225] Exemplary compstatin analog conjugates with polyoxazoline polymers are depicted below:



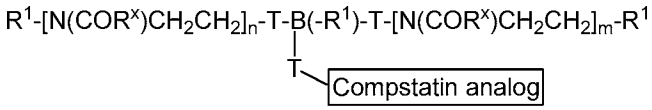
,



,



,



wherein each variable is independently as defined above and described in classes and subclasses herein.

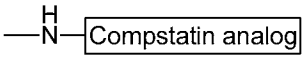
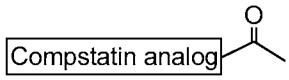
[00226] In some embodiments, the present invention provides a compstatin analog conjugated with a polymer, wherein the compstatin analog is connected to the polymer via one or more linkers. In some embodiments, a polymer is selected from PEG-containing compounds and genera described above and in classes and subclasses herein. In some embodiments, the present invention provides compstatin analog conjugates of PEG-containing compounds and genera depicted herein, wherein the compstatin analog is connected to the PEG-containing moieties via one or more linkers. Mono- and poly-functional PEGs that comprise one or more reactive functional groups for conjugation are defined above and described in classes and subclasses herein, including but not limited to those of formula A, I, Ia, II, III, IIIa, B, IV, IVa, V, Va, C, D, E, F, G, H, VI, VII, VIII, IX, X, XI, XII, XIII, XIV, XV, or XVI.

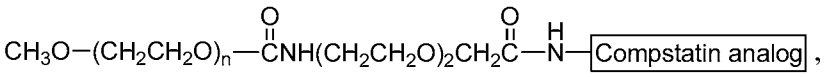
[00227] Suitable linkers for connecting a compstatin analog and a polymer moiety such as PEG or polyoxazoline are extensively described above and in classes and subclasses herein. In some embodiments, a linker has multiple functional groups, wherein one functional group is connected to a compstatin analog and another is connected to a polymer moiety. In some embodiments, a linker is a bifunctional compound. In some embodiments, a linker has the

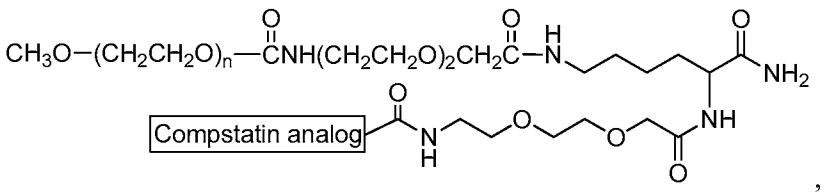
structure of $NH_2(CH_2CH_2O)nCH_2C(=O)OH$, wherein n is 1 to 1000. In some embodiments, a linker is 8-amino-3,6-dioxaoctanoic acid (AEEAc). In some embodiments, a linker is activated for conjugation with a polymer moiety or a functional group of a compstatin analog. For example, in some embodiments, the carboxyl group of AEEAc is activated before conjugation with the amine group of the side chain of a lysine group.

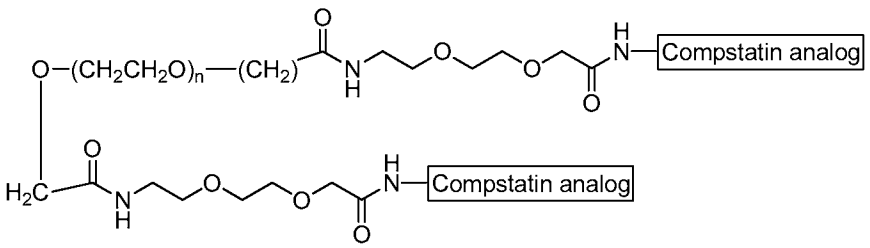
[00228] In some embodiments, a suitable functional group (for example, an amine, hydroxyl, thiol, or carboxylic acid group) on a compstatin analog is used for conjugation with a polymer moiety, either directly or via a linker. In some embodiments, a compstatin analog is conjugated through an amine group to a PEG moiety via a linker. In some embodiments, an amine group is the α-amino group of an amino acid residue. In some embodiments, an amine group is the amine group of the lysine side chain. In some embodiments, a compstatin analog is conjugated to a PEG moiety through the amino group of a lysine side chain (ε-amino group) via a linker having the structure of $NH_2(CH_2CH_2O)nCH_2C(=O)OH$, wherein n is 1 to 1000. In some embodiments, a compstatin analog is conjugated to the PEG moiety through the amino group of a lysine side chain via an AEEAc linker. In some embodiments, the $NH_2(CH_2CH_2O)nCH_2C(=O)OH$ linker introduces a $–NH(CH_2CH_2O)nCH_2C(=O)–$ moiety on a compstatin lysine side chain after conjugation. In some embodiments, the AEEAc linker introduces a $–NH(CH_2CH_2O)_2CH_2C(=O)–$ moiety on a compstatin lysine side chain after conjugation.

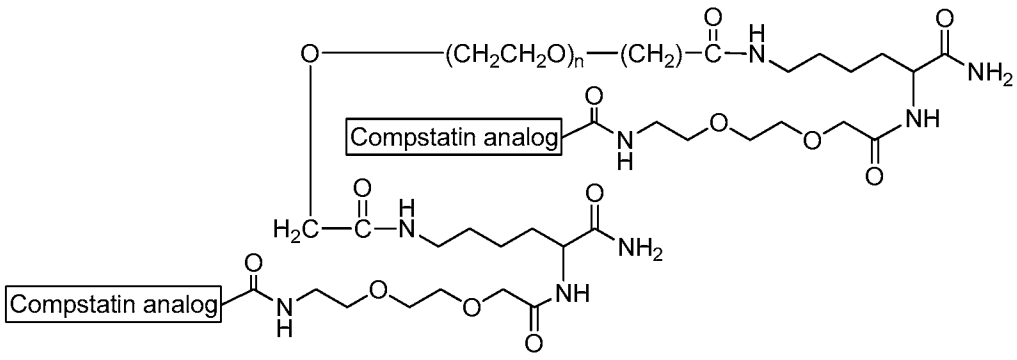
[00229] In some embodiments, a compstatin analog is conjugated to a polymer moiety via a linker, wherein the linker comprises an AEEAc moiety and an amino acid residue. In some embodiments, a compstatin analog is conjugated to a polymer moiety via a linker, wherein the linker comprises an AEEAc moiety and a lysine residue. In some embodiments, a polymer is PEG. In some embodiments, the C-terminus of a compstatin analog is connected to the amino group of AEEAc, and the C-terminus of AEEAc is connected to a lysine residue. In some embodiments, the C-terminus of a compstatin analog is connected to the amino group of AEEAc, and the C-terminus of AEEAc is connected to the α-amino group of a lysine residue. In some embodiments, the C-terminus of a compstatin analog is connected to the amino group of AEEAc, the C-terminus of AEEAc is connected to the α-amino group of the lysine residue, and a polymer moiety, such as a PEG moiety, is conjugated through the ε-amino group of said lysine residue. In some embodiments, the C-terminus of the lysine residue is modified. In some embodiments, the C-terminus of the lysine residue is modified by amidation. In some embodiments, the N-terminus of a compstatin analog is modified. In some embodiments, the N-terminus of a compstatin analog is acetylated.

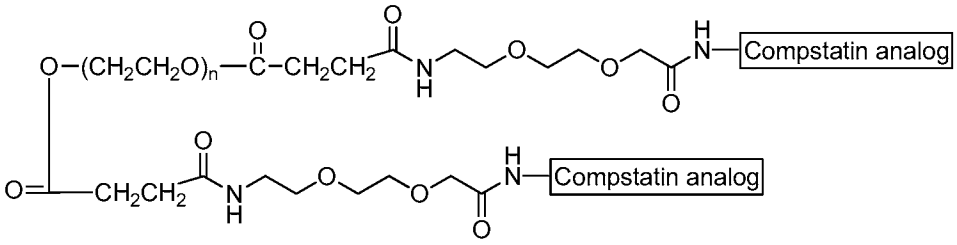
[00230] Exemplary conjugates comprising an AEEAc linker and a polymer are depicted below, wherein  represents the attachment point of an amine group on a compstatin analog,  represents a compstatin analog attaching through its C-terminus, and wherein each of the other variables is independently as defined above and described in classes and subclasses herewith. In some embodiments, an amine group is the amino group of a lysine side chain.



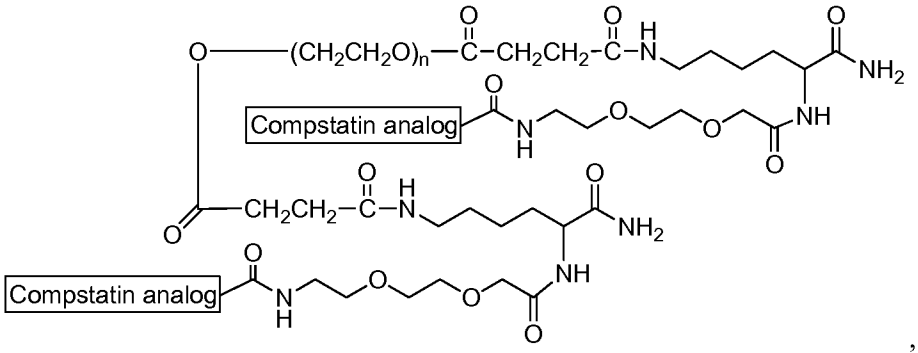




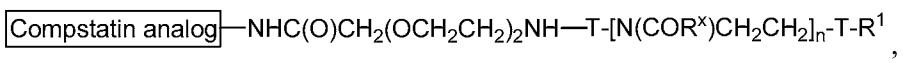



Compstatin analog
Compstatin analog

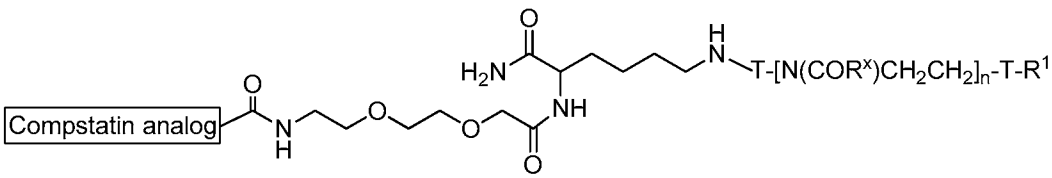
,


Compstatin analog
Compstatin analog

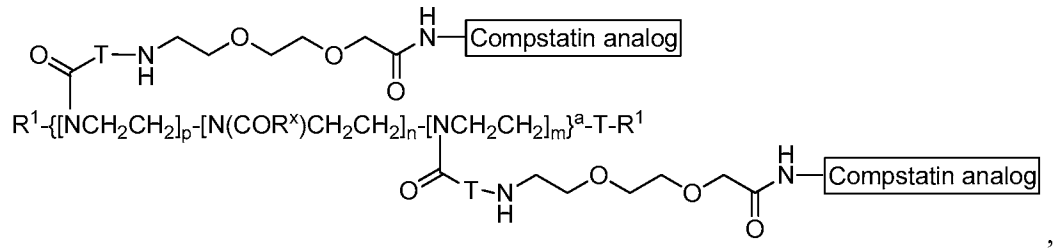
,


Compstatin analog
NHC(O)CH2(OCH2CH2)2NH—T-[N(CORx)CH2CH2]n-T-R1

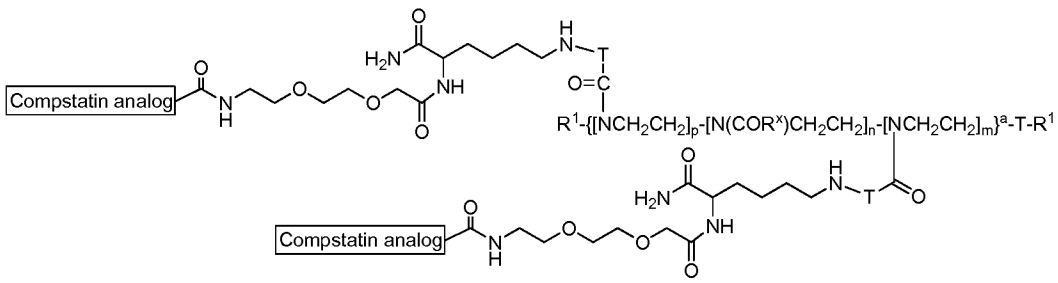
,


Compstatin analog
T-[N(CORx)CH2CH2]n-T-R1

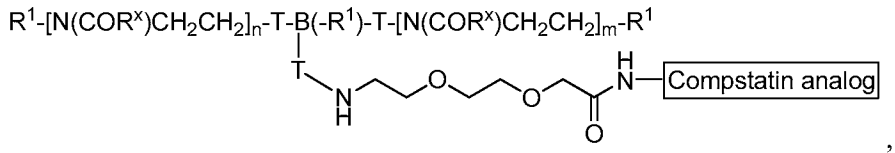
,


Compstatin analog
R1-{[NCH2CH2]p-[N(CORx)CH2CH2]n-[NCH2CH2]m}a-T-R1
Compstatin analog

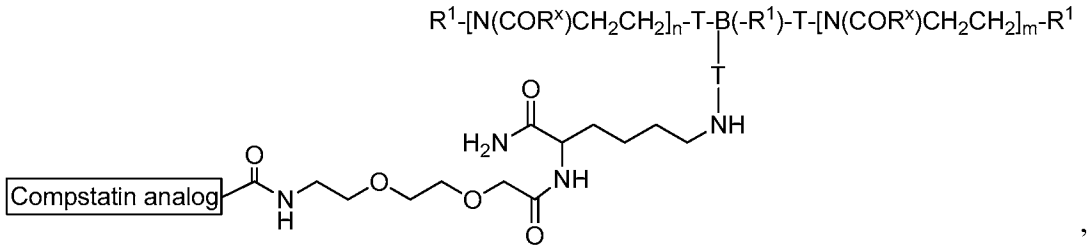
,


Compstatin analog
R1-{[NCH2CH2]p-[N(CORx)CH2CH2]n-[NCH2CH2]m}a-T-R1
Compstatin analog

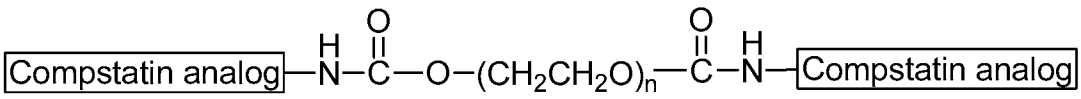
,


R1-[N(CORx)CH2CH2]n-T-B(-R1)-T-[N(CORx)CH2CH2]m-R1
T
N
H
O
O
H
N
Compstatin analog
O

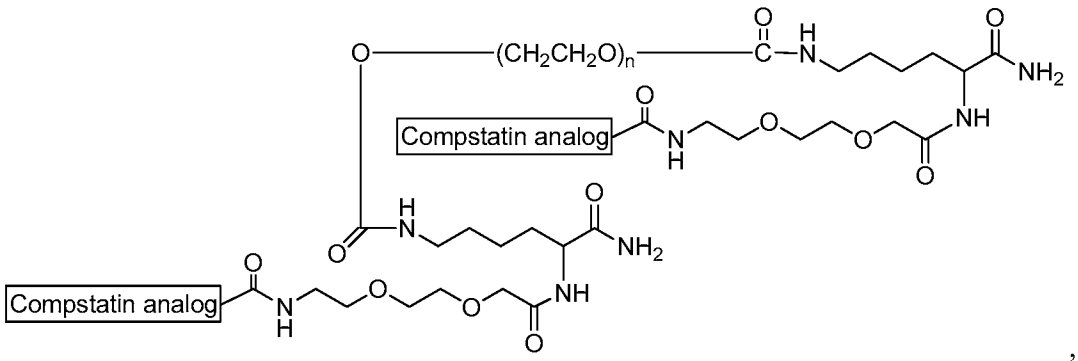
,


R1-[N(CORx)CH2CH2]n-T-B(-R1)-T-[N(CORx)CH2CH2]m-R1
T
NH
O
H2N
NH
Compstatin analog
O
N
H
O
O
O

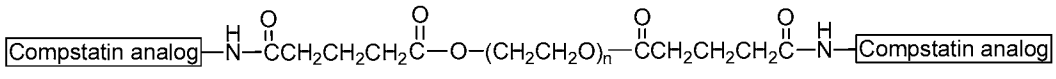
,


Compstatin analog
H
N
O
C
O
(CH2CH2O)n
O
C
H
N
Compstatin analog

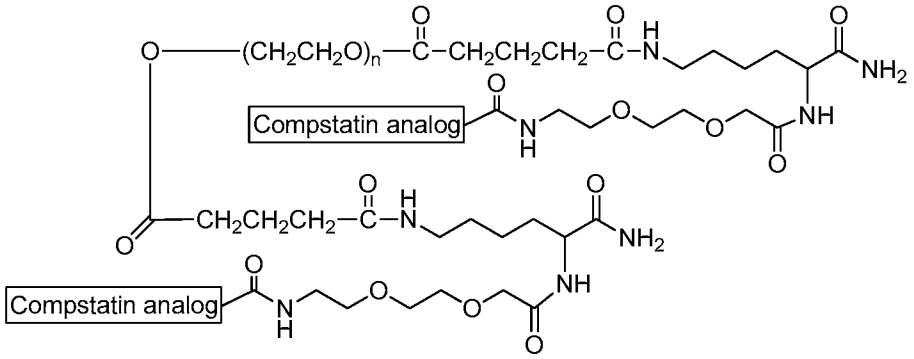
,


O
(CH2CH2O)n
O
C
H
N
O
NH2
Compstatin analog
O
N
H
O
O
NH
O
H
N
O
NH2
Compstatin analog
O
N
H
O
O
NH
O

,


Compstatin analog
H
N
O
CCH2CH2CH2C
O
O
(CH2CH2O)n
O
CCH2CH2CH2C
O
H
N
Compstatin analog

,



[00231] In certain embodiments a compstatin analog may be represented as M-AEEAc-Lys-B_2, wherein B_2 is a blocking moiety, e.g., NH_2, M represents any of SEQ ID NOs: 3-36, 37, 37A, 38A, 39A, 40A, or 41A, with the proviso that the C-terminal amino acid of any of SEQ ID NOs: 3-36, 37, 37A, 38A, 39A, 40A, or 41A is linked via a peptide bond to AEEAc-Lys-B_2. The NHS moiety of a monofunctional or multifunctional (e.g., bifunctional) PEG reacts with the free amine of the lysine side chain to generate a monofunctionalized (one compstatin analog moiety) or multifunctionalized (multiple compstatin analog moieties) long-acting compstatin analog. In various embodiments any amino acid comprising a side chain that comprises a reactive functional group may be used instead of Lys (or in addition to Lys). A monofunctional or multifunctional PEG comprising a suitable reactive functional group may be reacted with such side chain in a manner analogous to the reaction of NHS-ester activated PEGs with Lys.

[00232] With regard to any of the above formulae and structures, it is to be understood that embodiments in which the compstatin analog component comprises any compstatin analog described herein, e.g., any compstatin analog of SEQ ID NOs; 3-36, 37, 37A, 38A, 39A, 40A, 41A, are expressly disclosed. For example, and without limitation, a compstatin analog may comprise the amino acid sequence of SEQ ID NO: 28. An exemplary long-acting compstatin analog in which the compstatin analog component comprises the amino acid sequence of SEQ ID NO: 28 is depicted in Figure 10(C). It will be understood that the PEG moiety may have a variety of different molecular weights or average molecular weights in various embodiments, as described herein. For example, individual PEG chains within a preparation may vary in molecular weight and/or different preparations may have different average molecular weights and/or polydispersity, as described herein. In certain embodiments, the PEG moiety in the compound of Figure 10(C) has an average molecular weight of between about 20 kD -100 kD, about 20 kD-90 kD, about 20 kD-80 kD, about 20

kD-70 kD, about 20 kD-60 kD, about 20 kD-50 kD, about 30 kD-80 kD, about 30 kD-70 kD, about 30 kD-60 kD, about 30 kD-50 kD, about 30 kD-45 kD, about 35 kD-50 kD, about 35 kD-45 kD, about 36 kD-44 kD, about 37 kD-43 kD, about 38 kD-42 kD, or about 39 kD-41 kD. In some embodiments the PEG moiety in the compound of Figure 10(C) has an average molecular weight between about 30 kD and about 50 kD, e.g., between about 35 kD and about 45 kD, between about 37.5 kD and about 42.5 kD. In certain embodiments in which the PEG moiety has an average molecular weight of about 40 kD, e.g., 37.5 kD - 42.5 kD, 38 kD, 39 kD, 40 kD, 41 kD, 42 kD, the compound is sometimes referred to herein as CA28-2TS- BF. In certain embodiments a compound comprsing a CRM, e.g., a PEG moiety, thjat has an average molecular weight of about 40 kD, e.g., 37.5 kD - 42.5 kD, 38 kD, 39 kD, 40 kD, 41 kD, 42 kD, the compound has a terminal half-life of at least about 5 days, e.g., about 5-10 days, e.g., about 5, 6, 7, 8, 9 days, when administered IV or subcutaneously to non-human primates or humans, e.g., at a dose of about 1-3 mg/kg, 3-5 mg/kg, or 5-10 mg/kg.

[00233] In some aspects, the present invention relates to use of click chemistry in connection with compstatin analogs. "Click chemistry" is well known in the art and is useful in some aspects of the present invention. Click chemistry embodies, in certain embodiments, versatile cycloaddition reactions between azides and alkynes that enable a number of useful applications. Methods of carrying out click chemistry are known in the art, and are described by Kolb, H.C.; Sharpless, K.B., Drug Disc. Today, **2003**, 1128-1137; Moses, J.E.; Moorhouse, A.D.; *Chem. Soc. Rev.*, **2007**, 1249-1262; the entire contents of each are hereby incorporated by reference. Click chemistry is a popular method of bioconjugation due to its high reactivity and selectivity, even in biological media. See Kolb, H.C.; Finn, M.G.; Sharpless, K.B. *Angew. Chem. Int. Ed.* **2001**, *40*, 2004-2021; and Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G. *J. Am. Chem. Soc.* **2003**, *125*, 3192-3193. In addition, currently available recombinant techniques and synthetic methods permit the introduction of azides and alkyne-bearing non-canonical amino acids into peptides, proteins, cells, viruses, bacteria, and other biological entities that consist of or display proteins. See Link, A. J.; Vink, M. K. S.; Tirrell, D. A. *J. Am. Chem. Soc.* **2004**, *126*, 10598-10602; Deiters, A.; Cropp, T. A.; Mukherji, M.; Chin, J. W.; Anderson, C.; Schultz, P. G. *J. Am. Chem. Soc.* **2003**, *125*, 11782-11783.

[00234] As used herein, the term "click chemistry group" is sometimes used to refer\ to a reactive functional group capable of participating in a click chemistry reaction with an appropriate second reactive functional group, which second reactive functional group is also a click chemistry group. The first and second click chemistry groups, or entities (e.g.,

molecules) comprising such groups, may be referred to as complementary. First and second entities, e.g., molecules, that comprise complementary click chemistry groups may be referred to as click chemistry partners. An entity or molecule comprising a click chemistry group may be referred to as "click-functionalized". A bond formed by reaction of complementary click chemistry partners may be referred to as a "click chemistry bond".

[00235] In some embodiments, the present invention provides click-functionalized compstatin analogs for, e.g., conjugation to a complementary moiety on a partner molecule or biomolecule. In some embodiments, a complementary partner molecule or biomolecule is a polymer, peptide, protein, or a molecule that functions as a clearance-reducing moiety. In some embodiments, the "click-functionalized" moiety is an alkyne or an alkyne derivative which is capable of undergoing [3+2] cycloaddition reactions with complementary azide-bearing molecules and biomolecules. In another embodiment, the "click-functionalized" functionality is an azide or an azide derivative which is capable of undergoing [3+2] cycloaddition reactions with complementary alkyne-bearing molecules and biomolecules (*i.e.* click chemistry).

[00236] In some embodiments, a click-functionalized compstatin analog bears an azide group on any side chain group of the compstatin analog. In some embodiments, a click-functionalized compstatin analog bears an azide group on a lysine side chain group.

[00237] In some embodiments, a click-functionalized compstatin analog bears an alkyne group on any side chain group of the compstatin compstatin analog. In some embodiments, a click-functionalized compstatin analog bears an alkyne group on a lysine side chain group.

[00238] In some embodiments, the present invention provides compstatin conjugates comprising a compstatin analog, a molecule that functions as a clearance-reducing moiety, and a triazole linker. In some embodiments, a triazole linker is the result of click conjugation chemistry between a compstatin conjugate and a molecule that functions as a clearance-reducing moiety. In some embodiments the CRM may be any CRM disclosed herein. For example, the CRM may be a PEG, a polypeptide, or a POZ.

[00239] In some embodiments, the present invention provides compstatin conjugates comprising a compstatin analog, a PEG moiety, and a triazole linker. In some embodiments, a triazole linker is the result of click conjugation chemistry between a compstatin conjugate and a PEG moiety.

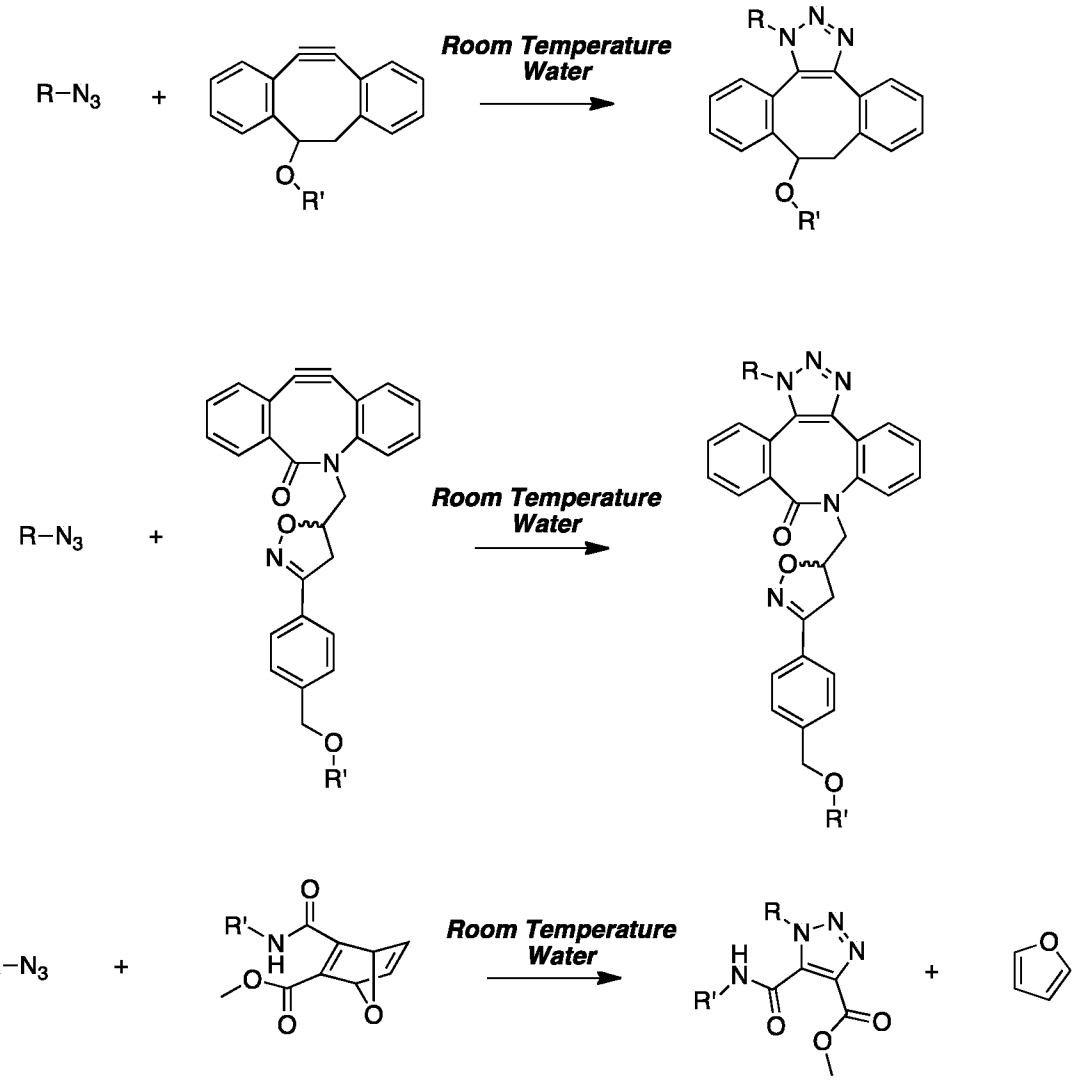
[00240] In some embodiments, the present invention provides compstatin conjugates comprising a compstatin analog, a polyoxazoline moiety, and a triazole linker. In some

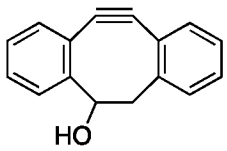
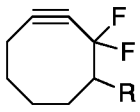
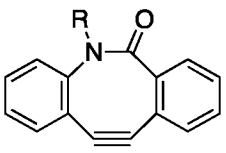
embodiments, a triazole linker is the result of click conjugation chemistry between a compstatin conjugate and a polyoxazoline moiety.

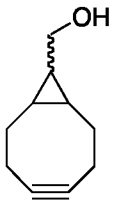
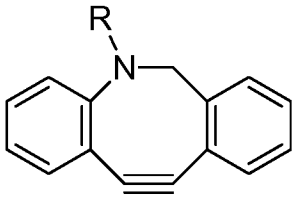
[00241] In some embodiments, click chemistry between a compstatin analog and another moiety is transition metal catalyzed. Copper-containing molecules which catalyze the "click" reaction include, but are not limited to, copper wire, copper bromide (CuBr), copper chloride (CuCl), copper sulfate ($CuSO_4$), copper sulfate pentahydrate ($CuSO_4 \bullet 5H_2O$), copper acetate ($Cu_2(AcO_4)$, copper iodide (CuI), $[Cu(MeCN)_4](OTf)$, $[Cu(MeCN)_4](PF_6)$, colloidal copper sources, and immobilized copper sources. In some embodiments other metals, such as ruthenium. Reducing agents as well as organic and inorganic metal-binding ligands can be used in conjunction with metal catalysts and include, but are not limited to, sodium ascorbate, tris(triazolyl)amine ligands, tris(carboxyethyl)phosphine (TCEP), sulfonated bathophenanthroline ligands, and benzimidazole-based ligands.

[00242] In some embodiments, compstatin analogs are conjugated to other moieties using metal free click chemistry (also known as copper free click chemistry) to give a metal free composition or conjugates. In contrast to standard click chemistry, also known as copper assisted click chemistry (CuACC), metal free click chemistry occurs between either a strained, cyclic alkyne or an alkyne precursor such as an oxanorbornadiene, and an azide group. As the name implies, no metal catalyst is necessary for the reaction to occur. Examples of such chemistries include reactions involving cyclooctyne derivatives (Codelli, et. al. *J. Am. Chem. Soc.*, **2008**, *130*, 11486-11493; Jewett, et. al. *J. Am. Chem. Soc.*, **2010**, *132*, 3688-3690; Ning, et. al. *Angew. Chem. Int. Ed.*, **2008**, *47*, 2253-2255), difluoro-oxanorbornene derivatives (van Berkel, et. al. *ChemBioChem*, **2007**, *8*, 1504-1508), or nitrile oxide derivatives (Lutz, et. al. *Macromolecules*, **2009**, *42*, 5411-5413). In certain embodiments a metal-free click chemistry reaction is a metal-free [3+2] cycloaddition reaction, Diels-Alder reaction, or thiol-alkene radical addition reaction. Exemplary click chemistry reactions and click chemistry groups are described in, e.g., Joerg Lahann, Click Chemistry for Biotechnology and Materials Science, 2009, John Wiley & Sons Ltd, ISBN 978-0-470-69970-6; Becer, Hoogenboom, and Schubert, Click Chemistry beyond Metal-Catalyzed Cycloaddition, Angewandte Chemie International Edition (2009) 48: 4900 – 4908. In certain embodiments a click chemistry group comprises a diarylcyclooctyne.

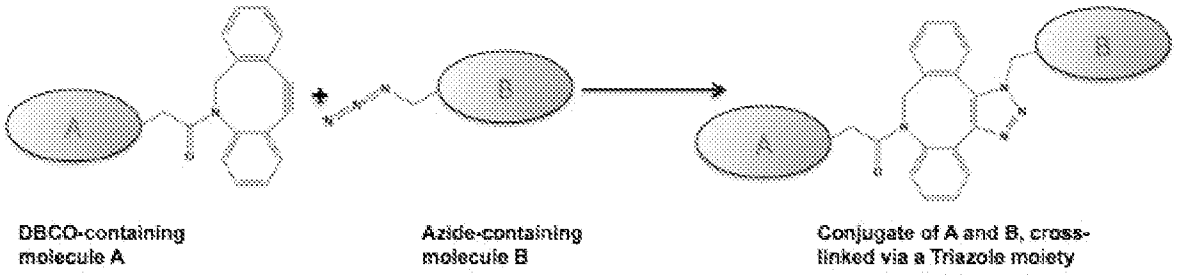
[00243] Certain examples of metal free click chemistry are shown in the scheme below.



[00244] Certain metal-free click moieties are known in the literature. Examples include 4-dibenzocyclooctynol (DIBO)  (from Ning et. al; Angew Chem Int Ed, 2008, 47, 2253); difluorinated cyclooctynes (DIFO or DFO)  (from Codelli, et. al.; J. Am. Chem. Soc. 2008, 130, 11486–11493.); biarylazacyclooctynone (BARAC)  (from Jewett et. al.; J. Am. Chem. Soc. 2010, 132, 3688); or

bicyclononyne (BCN)  (From Dommerholt, et. al.; *Angew Chem Int Ed*, **2010**, *49*, 9422-9425) or dibenzylcyclooctyne (DBCO)  .

[00245] A reaction scheme involving reaction of DBCO and an azide is shown below:

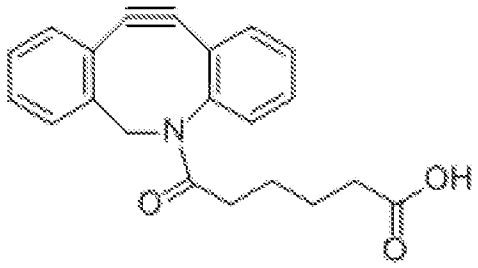


[00246] In the above scheme, in various embodiments, A may comprise or consist of a compstatin analog moiety and B may comprise or consist of a CRM, e.g., a polymer, such as a PEG or a POZ or a polypeptide, or B may comprise or consist of a compstatin analog moiety and A may comprise or consist of a CRM, e.g., a polymer, such as a PEG or a POZ or a polypeptide.

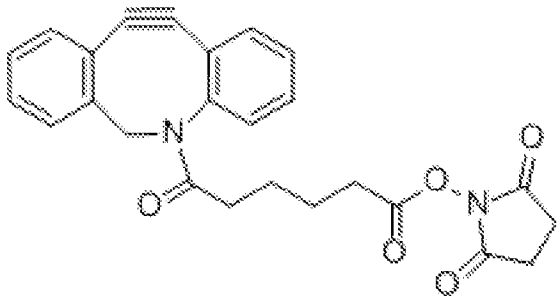
[00247] In some embodiments, the "metal free click-functionalized" moiety is an acetylene or an acetylene derivative which is capable of undergoing [3+2] cycloaddition reactions with complementary azide-bearing molecules and biomolecules without the use of a metal catalyst.

[00248] In some embodiments, the R and R' groups of the metal-free click chemistry reagents are a compstatin analog or any molecule described herein to which a compstatin analog may be conjugated. In some embodiments, such compstatin analogs bear a click-functionalized moiety on a lysine side chain. In some embodiments, such compstatin analogs are connected to a click-functionalized moiety via a linker. In some embodiments, such compstatin analogs are connected to a click-functionalized moiety via AEEAc.

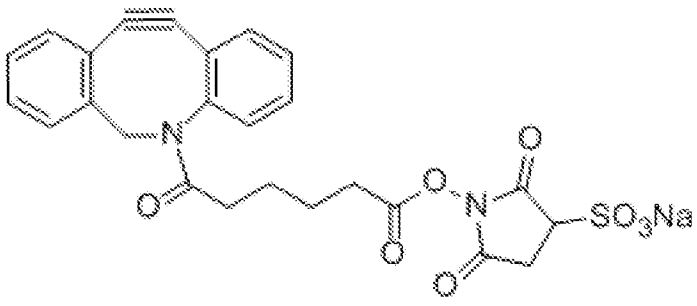
[00249] In some embodiments, a click chemistry reagent comprises DBCO. Exemplary reagents and exemplary uses thereof are set forth below:

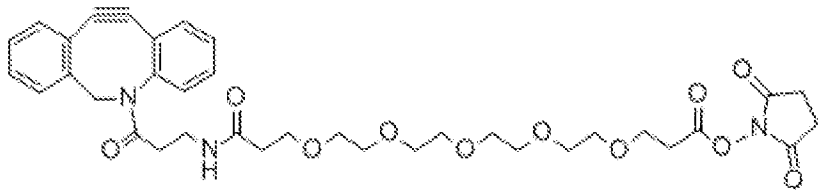


[00250] DBCO-Acid. In some embodiments a DBCO-Acid may be used to react with an amine-containing moiety.

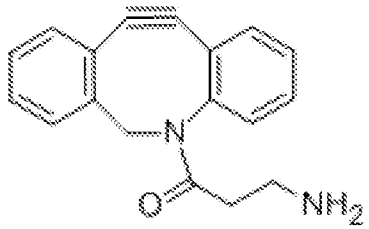


[00251] DBCO-NHS ester (above) or DBCO-*sulfo*-NHS ester (below) may be used to incorporate a DBCO functionality into an amine-containing molecule, such as a compstatin analog or a polypeptide comprising a lysine residue.



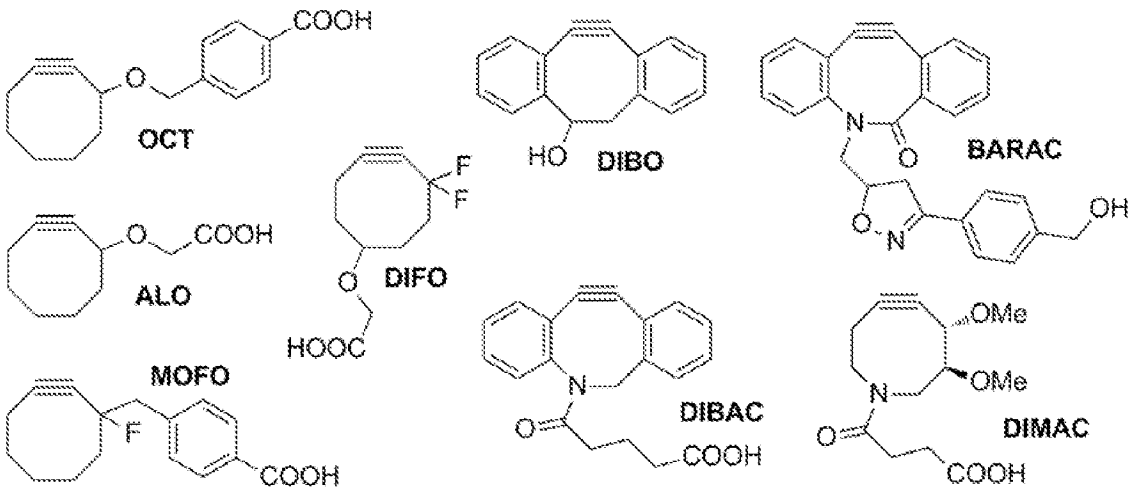


[00252] DBCO-PEG4-NHS ester. In some embodiments such reagent is useful for introducing a DBCO moiety by reaction with an available amine functionality. In some aspects, the presence of a PEG chain as a hydrophilic spacer may be useful to, e.g., increase solubility or provide flexibility.



[00253] DBCO-Amine. In some embodiments a click chemistry reagent comprises a carbonyl/carboxyl reactive dibenzylcyclooctyne, which may react with acids, active esters and/or aldehydes.

[00254] In certain embodiments a click chemisty reaction involves a cyclooctyne depicted below:



[00255] In certain embodiments click chemistry reactions comprise reactions between nitrones and cyclooctynes (see, e.g., Ning, Xinghai; Temming, Rinske P.; Dommerholt, Jan; Guo, Jun; Ania, Daniel B.; Debets, Marjoke F.; Wolfert, Margreet A.; Boons, Geert-Jan et al. (2010). "Protein Modification by Strain-Promoted Alkyne-Nitrone Cycloaddition". Angewandte Chemie International Edition 49 (17): 3065), oxime/hydrazone formation from aldehydes and ketones, tetrazine ligations (see, e.g., Blackman, Melissa L.; Royzen, Maksim; Fox, Joseph M. (2008). "The Tetrazine Ligation: Fast Bioconjugation based on Inverse-electron-demand Diels-Alder Reactivity". Journal of the American Chemical Society 130 (41): 13518–9), tetrazole ligations, the isonitrile-based click reaction (see, e.g., Stackmann, Henning; Neves, AndrÃ© A.; Stairs, Shaun; Brindle, Kevin M.; Leeper, Finian J. (2011). "Exploring isonitrile-based click chemistry for ligation with biomolecules". Organic & Biomolecular Chemistry 9 (21): 7303), and the quadricyclane ligation (see, e.g., Sletten, Ellen M.; Bertozzi, Carolyn R. (2011). "A Bioorthogonal Quadricyclane Ligation". Journal of the American Chemical Society 133 (44): 17570–3). In certain embodiments a click chemistry reaction is a Staudinger ligation (phosphine-azide).

[00256] Any compstatin analog may be modified to incorporate a click chemistry group in various embodiments. For example, a compstatin analog comprising the sequence of any of SEQ ID NOs: 3 – 36, 37, 37A, 38A, 39A, 40A, or 41A may be so modified. In some embodiments any such sequence further comprises a lysine residue or an AEEAc-Lys moiety, e.g., at the C-terminus. In some embodiments a click chemistry group is incorporated after peptide synthesis. For example, a Lys side chain may be reacted with azido acetic acid in order to introduce an azide moiety as a click chemistry group. In some embodiments a click

chemistry group is incorporated after cyclization and, in some embodiments, after addition of a blocking moiety at the N- and/or C-terminus. In some embodiments a click chemistry group is incorporated during peptide synthesis. For example, an amino acid comprising a side chain that comprises a click chemistry group may be used in the synthesis of a compstatin analog. A variety of such amino acids are commercially available from a number of sources, e.g., AAPPTec (Louisville, KY), Jena Bioscience GmBH (Jena, Germany). In some aspects, methods of making a click chemistry functionalized compstatin analog are provided herein.

[00257] In some embodiments compositions comprising a compstatin analog and a click chemistry reagent are provided. The click chemistry reagent may be any molecule capable of reacting with an amino acid side chain or terminus of a compound comprising a compstatin analog so as to install a click chemistry group, e.g., any click chemistry group known in the art. In some aspects, the composition may be incubated under suitable conditions (which may include providing a suitable catalyst, light (e.g., UV)) to functionalize the compstatin analog with a click chemistry functionality. In some embodiments, the invention provides compstatin analogs that comprise any click chemistry group including, but not limited to, those described herein. In some embodiments methods of making a long-acting compstatin analog are provided. In some embodiments the methods comprise mixing a compstatin analog comprising a first click chemistry group with a CRM comprising a complementary click chemistry group under conditions suitable for a click chemistry reaction to occur. Additional steps may comprise purifying the resulting conjugate. In some embodiments purifying comprises removing at least some unreacted components, e.g., with an appropriate scavenger.

[00258] In some embodiments a click chemistry reaction is used to join two or more CRMs, at least two of which have a compstatin analog moiety attached thereto. The compstatin analog moieties may be the same or different in various embodiments. The compstatin analog moieties may or may not be attached to the CRM via a click chemistry reaction. For example, in some embodiments a first heterobifunctional PEG comprising a first click chemistry group at a first terminus and an NHS ester at a second terminus is coupled to a compstatin analog moiety via the NHS ester. In a separate reaction, a second heterobifunctional PEG comprising a second click chemistry group at a first terminus and an NHS ester at a second terminus is coupled to a compstatin analog moiety via the NHS ester. The resulting two compounds are then reacted via a click chemistry reaction to form a larger molecule comprising two compstatin analog moieties. PEG is mentioned as an example of a

CRM but it should be understood that this approach may be used with any CRM. For example, in some embodiments it may be used with a CRM comprising a polypeptide, e.g., HSA or a portion thereof, or an albumin or albumin-binding peptide, or an antibody or portion thereof. In some embodiments this approach may be used with a POZ.

[00259] Comptatin analogs comprising a click chemistry group have a variety of uses. In some embodiments a compstatin analog comprising a first click chemistry group is reacted with any entity that comprses a complementary click chemistry group. The entity comprising the complementary click chemisty group may comprise, for example, a label (e.g., a flurophore, fluorescent protein, radioisotope, etc.), an affinity reagent, an antibody, a targeting moiety, a metal, a particle, etc. In some embodiments a click chemistry group is used to attach a compstatin analog moiety to a surface, wherein the surface comprises or is functionalized to comprise a complementary click chemistry group. In some embodiments a surface is for a sensor, e.g., a surface or sensor for capture/detection of C3. In some embodiments a surface forms part of a medical device, tubing, membrane, reservoir, implant, or other material that may come in contact with blood (e.g., extracorporeally) or be temporarily or indefinitely implanted into the body of a subject (e.g., a prosthetic device or drug delivery device). In some embodiments a surface is functionalized with compstatin analog to reduce complement activation thereon. In some embodiments a device or tubing is used for circulating blood, e.g., for dialysis, during surgery, etc. In some embodiments a device is a hemodialyzer or an extracorporeal circulatory support unit. Such compstatin analog functionalized devices and methods of making thereof are provided herein.

[00260] In some embodiments of the invention, a compstatin analog comprises both a cell-reactive functional group and a CRM. In some aspects, the invention provides variants of the molecules of any of the afore-mentioned cell-reactive compstatin analogs wherein a cell-reactive functional group or moiety is replaced by a $(CH_2CH_2O)_n$ moiety having a molecular weight of at least 500 daltons, e.g., at least 1,500 daltons up to about 100,000 daltons (e.g., an average molecular weight of about 20,000; 30,000; 40,000; 50,000; 60,000; 70,000; 80,000; 90,000; or 100,000 daltons). In some embodiments the average molecular weight of the compound or $(CH_2CH_2O)_n$ moieties is at least 20,000 daltons, up to about 100,000; 120,000; 140,000; 160,000; 180,000; or 200,000 daltons.

[00261] Exemplary long-acting compstatin analogs are set forth below, wherein n is sufficient to provide an average molecular weight of between about 500; 1,000; 1,500; 2,000; 5,000; 10,000; 20,000; 30,000; 40,000; 50,000; 60,000; 70,000; 80,000; 90,000; and 100,000 daltons. In some embodiments n is sufficient to provide an average molecular weight of

between about 20,000 daltons, up to about 100,000; 120,000; 140,000; 160,000; 180,000; or 200,000 daltons.

[00262] $(CH_2CH_2O)_nC(=O)$-Ile-Cys-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys-Thr- NH_2) (SEQ ID NO: 58)

[00263] *Ac*-Ile-Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr -NH-$CH_2CH_2OCH_2CH_2OCH_2$-C(=O)-Lys-C(=O)-(CH_2CH_2O)n-NH_2 (SEQ ID NO: 59)

[00264] *Ac*-Ile-Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-Lys-C(=O)-(CH_2CH_2O)n -NH_2 (SEQ ID NO: 60).

[00265] *Ac*-Ile-Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-$(Gly)_5$-Lys-C(=O)-(CH_2CH_2O)n-NH_2 (SEQ ID NO: 61)

[00266] *Ac*-(CH_2CH_2O)nC(=O)Lys-(Gly)5-Ile- Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr - NH_2) (SEQ ID NO: 62)

[00267] *Ac*-(CH_2CH_2O)nC(=O)Lys-Ile- Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr - NH_2) (SEQ ID NO: 63)

[00268] In SEQ ID NO: 58, the (CH_2CH_2O)n is coupled via an amide bond to the N-terminal amino acid. In SEQ ID NOs: 59-63, the (CH_2CH_2O)n moiety is coupled via an amide bond to a Lys side chain; thus it will be understood that the NH_2 at the C-terminus in SEQ ID NOs: 59, 60, and 61, represents amidation of the C-terminus of the peptide, and it will be understood that in SEQ ID NOs: 62 and 63, the *Ac* at the N-terminus represents acetylation of the N-terminus of the peptide, as described above. It will also be appreciated by those of ordinary skill in the art that a free end of a $(CH_2CH_2O)_n$ moiety typically terminates with an (OR) where the underlined O represents the O atom in the terminal (CH_2CH_2O) group. (OR) is often a moiety such as a hydroxyl (OH) or methoxy (-OCH_3) group though other groups (e.g., other alkoxy groups) could be used. Thus SEQ ID NO: 59, for example, may be represented as *Ac*-Ile-Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-NH-$CH_2CH_2OCH_2CH_2OCH_2$-C(=O)-Lys-(C(=O)-$(CH_2CH_2O)_n$-R)-*NH2* (SEQ ID NO: 64) wherein R is, e.g., either H or CH_3 in the case of a linear PEG. In the case of a bifunctional, branched or star-shaped PEG, R represents the remainder of the molecule. Further, it will be understood that the moiety comprising the reactive functional group may vary, as described herein (e.g., according to any of the formulas described herein). For example, long-acting compstatin analogs comprising the same peptide sequence as SEQ ID NO: 64, in which the moiety comprising the reactive functional group comprises an ester and/or alkyl chain may be represented as follows

Ac-Ile-Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-NH-$CH_2CH_2OCH_2CH_2OCH_2$-C(=O)-Lys-(C(=O)-$(CH2)_m$-$(CH_2CH_2O)_n$-R)-*NH2* (SEQ ID NO: 65);

Ac-Ile-Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-NH-$CH_2CH_2OCH_2CH_2OCH_2$-C(=O)-Lys-(C(=O)-$(CH_2)_m$-C(=O)-$(CH_2CH_2O)_n$-R)-*NH2* (SEQ ID NO: 66)

Ac-Ile-Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-NH-$CH_2CH_2OCH_2CH_2OCH_2$-C(=O)-Lys-(C(=O)-$(CH_2)_m$-C(=O)-(CH_2)j $(CH_2CH_2O)_n$-R)-*NH2* (SEQ ID NO: 67)

In SEQ ID NOs: 65-67 m may range from 1 up to about 2, 3, 4, 5, 6, 7, 8, 10, 15, 20, or 30 in various embodimetns, In SEQ ID NOs: 67 j may range from 1 up to about 2, 3, 4, 5, 6, 7, 8, 10, 15, 20, or 30 in various embodiments.

It will also be appreciated that, as described herein, in various embodiments other moieties may be incorporated between the Lys-(C(=O)- and $(CH_2CH_2O)_n$-R, such as an amide, aromatic ring (e.g., a substituted or unsubstituted phenyl), or a substituted or unsubstituted cycloalkyl structure.

[00269] The invention provides variants of SEQ ID NOs: 58-67 in which -Ile-Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr- is replaced by an amino acid sequence comprising the amino acid sequence of any other compstatin analog, e.g., of any of SEQ ID NOs 3-27 or 29-36, 37, 37A, 38A, 39A, 40A, or 41A with the proviso that blocking moiet(ies) present at the N- and/or C-termini of a compstatin analog may be absent, replaced by a linker (which may comprise a blocking moiety), or attached to a different N- or C-terminal amino acid present in the corresponding variant(s).

[00270] Any compstatin analog, e.g., any compound comprising any of SEQ ID NOs: 3-37, 37A, 38A, 39A, 40A, or 41A may, in various embodiments, can be attached via or near its N-terminal or C-terminal end (e.g., via a side chain of an amino acid at or near its N-terminal or C-terminal amino acid) directly or indirectly to any moiety comprising a reactive functional group, e.g., any compound of Formulae I – XVI or Formulae A-H.

[00271] In some embodiments the CRM comprises a polypeptide that occurs in human serum, or a fragment thereof or a substantially similar variant of the polypeptide or fragment thereof. In some embodiments the polypeptide, fragment, or variant has a molecular weight of between 5 kD and 150 kD, e.g., at least 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 kd, or more, e.g., between 100 and 120, or 120 and 150 kD. In some embodiments, producing a long-acting compstatin analog comprises reacting a compstatin analog comprising a reactive

functional group with one or more amino acid side chains of the polypeptide, wherein the side chain comprises a compatible functional group. In some embodiments, producing a long-acting compstatin analog comprises reacting a compstatin analog comprising a reactive functional group with the N-terminal amine and/or C-terminal carboxyl group of the polypeptide. In some embodiments, producing a long-acting compstatin analog comprises reacting a compstatin analog comprising an amine-reactive functional group with amino acids having a side chain comprising a primary amine (e.g., lysine) and/or with the N-terminal amine of the polypeptide. In some embodiments, producing a long-acting compstatin analog comprises reacting a compstatin analog comprising a carboxyl-reactive functional group with the C-terminal carboxyl group of the polypeptide. In some embodiments a compstatin analog moiety is attached at each terminus of the polypeptide and, optionally, to the side chain of one or more internal amino acids. In some embodiments, producing a long-acting compstatin analog comprises reacting a compstatin analog comprising a sulfhydryl-reactive functional group with one or more sulfhydryl groups of the polypeptide.

[00272] In some embodiments, at least one reactive functional group is introduced into the polypeptide. For example, in some embodiments at least one side chain of the polypeptide is modified to convert a first reactive functional group to a different reactive functional group prior to reaction with the compstatin analog. In some embodiments a thiol is introduced. Several methods are available for introducing thiols into biomolecules, including the reduction of intrinsic disulfides, as well as the conversion of amine, aldehyde or carboxylic acid groups to thiol groups. Disulfide crosslinks of cystines in proteins can be reduced to cysteine residues by dithiothreitol (DTT), tris-(2-carboxyethyl)phosphine (TCEP), or or tris-(2-cyanoethyl)phosphine. Amines can be indirectly thiolated by reaction with succinimidyl 3-(2-pyridyldithio)propionate (SPDP) followed by reduction of the 3-(2-pyridyldithio)propionyl conjugate with DTT or TCEP. Amines can be indirectly thiolated by reaction with succinimidyl acetylthioacetate followed by removal of the acetyl group with 50 mM hydroxylamine or hydrazine at near-neutral pH. Amines can be directly thiolated by reaction with 2-iminothiolane, which preserve the overall charge of the molecule and introduces a free thiol. Tryptophan residues in thiol-free proteins can be oxidized to mercaptotryptophan residues, which can then be modified by iodoacetamides or maleimides. A polypeptide comprising one or more thiols may be reacted with a compstatin analog comprising a maleimide group, such as *Ac*-Ile-Cys*-Val-Trp(1-Me)-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-AEEAc-Lys-(C(=O)-$(CH_2)_5$-Mal)-NH_2 (SEQ ID NO: 68) to generate a long-acting compstatin analog.

[00273] In some embodiments the polypeptide is recombinantly produced. In some embodiments the polypeptide is at least in part recombinantly produced (e.g., in bacteria or in eukaryotic host cells such as fungal, insect, plant, or vertebrate) and/or at least in part produced using chemical synthesis. In some embodiments the polypeptide is purified. For example, in some embodiments the polypeptide is purified from a host cell lysate or from culture medium into which it has been secreted by host cells. In some embodiments the polypeptide is glycosylated. In some embodiments the polypeptide is non-glycosylated. In some embodiments the polypeptide is human serum albumin (HSA). In some embodiments a substantially similar variant of the polypeptide is sufficiently similar to the polypeptide of which it is a variant so as to not be recognized as foreign by a normal immune system of a subject, e.g., a human subject. In some embodiments alterations in the sequence of substantially similar variant as compared with the polypeptide of which it is a variant are selected so as to avoid generating MHC Class I epitopes. Various methods known in the art can be used to predict whether a sequence comprises an MHC Class I epitope.

[00274] In some embodiments, one or more amino acids in a polypeptide or linker or composition may be selected to be hydrophobic or hydrophilic or selected to confer increased hydrophilicity or, in some embodiments, increased hydrophobicity, on a compound that contains it. As known in the art, the terms "hydrophilic" and "hydrophobic" are used to refer to the degree of affinity that a substance has with water. In some aspects a hydrophilic substance has a strong affinity for water, tending to dissolve in, mix with, or be wetted by water, while a hydrophobic substance substantially lacks affinity for water, tending to repel and not absorb water and tending not to dissolve in or mix with or be wetted by water. Amino acids can be classified based on their hydrophobicity as well known in the art. Examples of "hydrophilic amino acids" are arginine, lysine, threonine, alanine, asparagine, glutamine, aspartate, glutamate, serine, and glycine. Examples of "hydrophobic amino acids" are tryptophan, tyrosine, phenylalanine, methionine, leucine, isoleucine, and valine. In certain embodiments an analog of a standard amino acid is used, wherein the analog has increased or decreased hydrophilic or hydrophobic character as compared with the amino acid of which it is an analog.

[00275] In some aspects the disclosure provides the recognition that unstructured polypeptides may be used to produce a long-acting, cell-reactive, or targeted compstatin analog. In some aspects, the invention provides a long-acting compstatin analog comprising a compstatin analog moiety and an unstructured polypeptide, wherein the compstatin analog moiety is linked to the unstructured polypeptide. In some embodiments, multiple compstatin

analog moieties, which may be the same or different, are linked to an unstructured polypeptide. For example, between 2 and 10, e.g., 2, 3, or 4, 5, or more compstatin analog moieties, may be linked to an unstructured polypeptide in certain embodiments. In some aspects, an unstructured polypeptide serves as a clearance-reducing moiety. In some aspects, the invention provides a cell-reactive compstatin analog comprising a compstatin analog moiety, a cell-reactive moiety, and an unstructured polypeptide, wherein the compstatin analog moiety and the cell-reactive moiety are linked to the unstructured polypeptide. In some aspects, the invention provides a targeted compstatin analog comprising a compstatin analog moiety, a targeting moiety, and an unstructured polypeptide, wherein the compstatin analog moiety and the targeting moiety are linked to the unstructured polypeptide.

[00276] As used herein, the term "unstructured polypeptide" (UP) generally refers to polypeptides that comprise non-naturally occurring, substantially non-repetitive sequences that are composed mainly of small hydrophilic amino acids, with the sequence having a low degree of secondary or tertiary structure or, in some embodiments, no secondary or tertiary structure under physiological conditions. As used herein, the term "unstructured polypeptide" specifically excludes whole antibodies or antibody fragments (e.g., single-chain antibodies and Fc fragments). The term "unstructured polypeptide" includes, but is not limited to, unstructured recombinant polypeptides (URP) and extended recombinant polypeptides (XTEN), as those terms are used in Schellenberger et al., (2009) Nature Biotechnology, 27, 1186-90; USSN 11/715,276; USSN12/699,761; PCT/US2007/005952 (WO/2007/103515); PCT/US2010/023106 (WO/2010/091122); PCT/US2010/061590 (WO/2011/084808); PCT/US2011/031014 (WO2011123830); PCT/US2011/030992 (WO/2011/123813) and/or other patent applications or patents presently listed by the World Intellectual Property Organization and/or USPTO as assigned to Amunix Operating, Inc. or Amunix, Inc., collectively "Amunix". In any of the embodiments herein pertaining to a UP, the UP may in certain embodiments be or comprise an XTEN. Certain embodiments described herein refer to an XTEN by way of example, but analogous embodiments comprising other UPs are encompassed. UPs, e.g., XTENS, may be produced using standard recombinant DNA technology. Nucleic acids encoding a UP may be inserted into a suitable vector, which may be introduced into suitable host cells. A cell line may be generated. The cell line may be maintained, cultivated in appropriate culture medium, and used as a source of the UP. The UP may be expressed and purified using standard methods. In some embodiments a polypeptide comprises a UP and a compstatin analog as a fusion protein. In some embodiments the compstatin analog portion is separated from the UP by a polypeptide

spacer. In some such embodiments the compstatin analog comprises only amino acids that are genetically encodable, e.g., standard amino acids. In some embodiments the compstatin analog portion may be post-translationally modified on one or more side chains. In some embodiments a UP and a compstatin analog moiety are chemically conjugated, e.g., as described below. The compstatin analog may be produced using standard peptide synthesis and subsequently conjugated to the UP.

[00277] In certain embodiments a UP, e.g., an XTEN, exhibits one or more of the following properties: conformational flexibility, relatively high aqueous solubility, low or absent (undetectable) immunogenicity, and/or increased hydrodynamic radius as compared with typical mammalian polypeptides of the same length. In some embodiments a UP, e.g., an XTEN, has a relatively high degree of resistance to proteases found in blood, sometimes termed plasma proteases or serum proteases. In various embodiments such resistance is sufficient such that at least 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or more, of the UP polypeptide molecules remain intact (full length) upon incubation in blood, plasma, or serum at 37 degrees C for 48 hours, 72 hours, 96 hours, 120 hours, 144 hours, 168 hours, or more. For example, in some embodiments such resistance is sufficient such that at least 75% of the UP molecules remain intact upon incubation in blood, plasma, or serum at 37 degrees C for 48 hours, 72 hours, 96 hours, 120 hours, 144 hours, 168 hours, or more. Incubation may be performed using a UP at a concentration of between 1 microgram/ml to about 100 mg/ml in various embodiments. Samples may be analyzed at various time points. Size or intactness may be assessed using, e.g., chromatography (e.g., HPLC), mass spectrometry, Western blot, or any other suitable method. Such stability characteristics may be conferred on a moiety conjugated to the UP. For example, a long-acting compstatin analog comprising a UP may have stability characteristics similar to those of the UP to which it is conjugated. In some embodiments a UP sequence lacks a cleavage site for one or more proteases, e.g., one or more mammalian proteases that is normally present or may be present in blood or interstitial fluid. In some embodiments a UP sequence may comprise a cleavage site for one or more proteases, e.g., one or more mammalian proteases that is normally present or may be present in blood or interstitial fluid or intracellularly. In some embodiments the presence, absence, position, and/or identity of one or more protease cleavage sites may be selected to produce a desired half-life, degradation rate, release of an active moiety (e.g., a compstatin analog moiety), etc. In some embodiments a UP sequence is selected to lack a cleavage site for one or more proteases, e.g., one or more mammalian proteases that is normally present or may be present in blood or, in some embodiments,

interstitial fluid. Exemplary proteases include, e.g., coagulation/clotting factors, e.g., factor XIa, factor XIIa, kallikrein, factor VIIa, factor IXa, factor Xa, factor IIa (thrombin), elastase-2, matrix metalloproteases, e.g., MMP-12, MMP-13, MMP-17 and MMP-20. In some embodiments a UP sequence is free of such cleavage sites or contains no more than 1, 2, 3, 4, or 5 such cleavage sites. Cleavage sites are known to those of ordinary skill in the art. See, e.g., the MEROPs database (http://merops.sanger.ac.uk/; Rawlings, N.D., Barrett, A.J. & Bateman, A. (2012) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res 40, D343-D350) and references cited therein. In some embodiments a UP may be degraded by intracellular proteases. Such degradation may be apparent, for example, upon incubating a UP in vitro in a cell lysate.

[00278] In some embodiments a long-acting, cell-reactive, or targeted compstatin analog comprising a compstatin analog moiety and UP, e.g., an XTEN, exhibits (i) increased overall solubility, (ii) reduced rate of absorption when administered subcutaneously or intramuscularly, and/or (iii) one or more enhanced pharmacokinetic properties such as longer terminal half-life, reduced clearance, and/or increased area under the curve (AUC) as measured in molar units, when compared to a compstatin analog having the same sequence as the compstatin analog moiety but not linked to a UP, e.g., an XTEN, when administered by the same route, and, typically, in the same or an equivalent or substantially similar vehicle.

[00279] In some embodiments a UP is designed to assume a state similar in some respects to that assumed by denatured peptide sequences under physiological conditions. "Denatured" describes the state of a peptide in solution that is characterized by a large conformational freedom of the peptide backbone. As known in the art, most peptides and proteins adopt a denatured conformation in the presence of high concentrations of denaturants or at elevated temperature. Peptides in denatured conformation can have, for example, characteristic circular dichroism (CD) spectra and a lack of long-range interactions as determined, e.g., by nuclear magnetic resonance (NMR). In some embodiments a UP sequence, under physiological conditions, resembles denatured sequences that are largely or substantially devoid of secondary structure. In some embodiments a UP sequence is largely or substantially devoid of secondary structure under physiological conditions. Secondary structure refers to the general three-dimensional form of local segments of biopolymers such as proteins. In the case of proteins, secondary structure typically refers to presence of structural motifs such as alpha helices or beta sheets due to patterns of hydrogen bonds between backbone amino and carboxyl groups. Random coil refers the class of conformations that indicate an absence of regular secondary structure. "Largely devoid", as

used in this context, means that less than 50% of the amino acid residues of the UP sequence contribute to secondary structure as measured or determined by one or more methods described herein or methods considered equivalent by those of ordinary skill in the art. "Substantially devoid", as used in this context, means that at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, or at least about 96%, 97%, 98%, or 99% of the amino acid residues of the UP sequence do not contribute to secondary structure, as measured or determined by one or more methods described herein or equivalent method. In some aspects, an amino acid residue is considered not to contribute to secondary structure if it is not within a defined secondary structure element, such as an alpha helix or beta sheet. (n some aspects, an amino acid residue is considered not to contribute to secondary structure if it forms or is part of a turn or random coil.

[00280] A variety of methods are known in the art as being useful to determine the presence or absence of secondary and tertiary structures in polypeptides. For example, secondary structure can be measured spectrophotometrically, e.g., by circular dichroism spectroscopy in the far-ultraviolet (far-UV) spectral region (170-250 nm). Secondary structure elements, such as alpha-helix and beta-sheet, each give rise to a characteristic shape and magnitude of CD spectra. For example, a pronounced double minimum at 208 and 222 nm may indicate α-helical structure, whereas a single minimum at 204 nm or 217 nm may reflect random-coil or β-sheet structure, respectively. Another useful method is infrared spectroscopy, which detects differences in the bond oscillations of amide groups due to hydrogen-bonding. Secondary structure contents may be estimated using the chemical shifts of an NMR spectrum.

[00281] Secondary structure of a polypeptide can also or additionally be predicted using various computer programs or algorithms, such as the Chou-Fasman algorithm (Chou, P. Y., et al. (1974) Biochemistry, 13: 222-45), sometimes termed the "original Chou-Fasman algorithm" herein, or a modified version thereof (see, e.g., Chen H, Gu F, Huang Z (2006). "Improved Chou–Fasman method for protein secondary structure prediction". BMC Bioinformatics 7 (Suppl 4): S14), the Garnier-Osguthorpe-Robson ("GOR") algorithm (Garnier J, Gibrat J F, Robson B. (1996), Methods Enzymol 266:540-553), sometimes termed the "original GOR algorithm" or a modified version thereof (see, e.g., Kloczkowski, A., Ting, K.-L, Jernigan, R.L., Garnier, J., "Combining the GOR V algorithm with evolutionary information for protein secondary structure prediction from amino acid sequence", Proteins, 49, 154-166, 2002), an approach utilizing fragment database mining (FDM) (see, e.g., Cheng, H., Sen, T.Z., Kloczkowski, A., Margaritis, D., Jernigan, R.L., "Prediction of protein

secondary structure by mining structural fragment database", Polymer 46, 4314- 4321, 2005), a combination approach, such as combining GOR with FDM (see, e.g., Sen, T.Z., Cheng, H., Kloczkowski, A., Jernigan, R.L. "A Consensus Data Mining secondary structure prediction by combining GOR V and Fragment Database Mining", Protein Sci. 15(11):2499-2506, 2006) or others known in the art. Such algorithms can predict, for a given sequence, whether there exists some or no secondary structure, expressed as the total and/or percentage of residues of the sequence that form, for example, alpha-helices or beta-sheets or the percentage of residues of the sequence predicted to result in random coil formation (which lacks secondary structure).

[00282] Any one or more experimental methods, computer programs, or algorithms may be used to determine a level or location of secondary structure present in a protein, e.g., a UP. In some embodiments, e.g., if multiple experimental methods, structure prediction algorithms or programs are used, an average value for the amount (e.g., percentage) of amino acids that contribute to secondary structure may be used. In some embodiments, e.g., if multiple experimental methods, structure prediction algorithms or programs are used, an amino acid is designated as contributing to or not contributing to secondary structure according to the assignment made by the majority of such methods, algorithms or programs or made by the method, algorithm, or program that, in general, gives results most closely resembling those obtained using circular dichroism. If an amino acid is deemed equally likely to contribute or not contribute to secondary structure or cannot be reliably assigned as contributing or not contributing to secondary structure, embodiments are provided in which such amino acid is considered to contribute to secondary structure, and embodiments are provided in which such amino acid is considered not to contribute to secondary structure.

[00283] In some embodiments a UP sequence has an alpha helix percentage ranging from 0% to less than about 5% as determined by, e.g., Chou-Fasman algorithm or GOR algorithm. In some embodiments a UP sequence has a beta sheet percentage ranging from 0% to less than about 5% as determined by, e.g., Chou-Fasman algorithm or GOR algorithm. In some embodiments a UP sequence has an alpha helix percentage ranging from 0% to less than about 5% and a beta sheet percentage ranging from 0% to less than about 5% as determined by, e.g., Chou-Fasman algorithm or GOR algorithm. In some embodiments a UP sequence has an alpha helix percentage less than about 2%, a beta sheet percentage less than about 2%, or both an alpha helix percentage less than about 2% and a beta sheet percentage less than about 2% as determined by, e.g., Chou-Fasman algorithm or GOR algorithm. In some embodiments a UP sequence has a high degree of random coil percentage, as determined by

GOR algorithm. In some embodiments a UP sequence has at least about 80%, e.g., at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% random coil percentage, as determined by GOR algorithm. In some embodiments a Chou-Fasman algorithm is the original Chou-Fasman algorithm. In some embodiments a Chou-Fasman algorithm is a modified Chou-Fasman algorithm. In some embodiments a GOR algorithm is the original GOR algorithm. In some embodiments a GOR algorithm is GOR V.

[00284] In some embodiments, a UP sequence is substantially non-repetitive. In some embodiments a UP sequence is between about 100 to about 3000 amino acid residues long, wherein the sequence is substantially non-repetitive. In some embodiments a UP sequence has between about 100 to about 3000 amino acid residues in which no three contiguous amino acids in the sequence are identical except, in some embodiments, if the amino acid is serine, in which case no more than three contiguous amino acids are serine residues. In such embodiments, the UP sequence is termed "substantially non-repetitive". In some embodiments "substantially non-repetitive" means that there are few (e.g., less than 5) or no instances of four contiguous amino acids in the sequence that are identical amino acid types. In some embodiments "substantially non-repetitive" means that there are few (e.g., less than 5) or no instances of three contiguous amino acids in the sequence that are identical amino acid types with the proviso that, in some embodiments, at least one sequence of 3 serines may be present.

[00285] The degree of repetitiveness of a polypeptide or a gene can be measured by computer programs or algorithms or by other means known in the art. Repetitiveness in a polypeptide sequence can, for example, be assessed by determining the number of times shorter sequences of a given length occur within the polypeptide. For example, a polypeptide of 200 amino acid residues has 192 overlapping 9-amino acid sequences (or 9-mer "frames") and 198 overlapping 3-mer frames. The number of unique 9-mer or 3-mer sequences will depend on the amount of repetitiveness within the sequence. In some embodiments, a score is generated (hereinafter "subsequence score") that is reflective of the degree of repetitiveness of the subsequences in the overall polypeptide sequence. As used herein, a "subsequence score" means the sum of occurrences of each unique 3-mer frame across a 200 consecutive amino acid sequence of the polypeptide divided by the absolute number of unique 3-mer subsequences within the 200 amino acid sequence. Examples of such subsequence scores derived from the first 200 amino acids of repetitive and non-repetitive polypeptides are presented in Example 44 of WO2011028229. In some embodiments, a UP has a subsequence score less than 12, e.g., less than 10, e.g., less than 9, e.g., less than 8, e.g., less than 7, more

preferably less than 6, and most preferably less than 5. In the embodiments a UP that has a subsequence score less than about 10 (i.e., 9, 8, 7, etc.) is substantially non-repetitive". In some aspects, polypeptides having a non-repetitive sequence are, in general, more soluble and have less tendency to aggregate compared to polypeptides having repetitive sequences. Such properties in some embodiments facilitate the formulation of preparations containing high concentrations of polypeptide, in some cases exceeding 100 mg/ml. In some embodiments a UP sequence is designed to have a low degree of internal repetitiveness in order to reduce or substantially eliminate immunogenicity when administered to a mammal. In some embodiments a UP sequence may be characterized by an absence of or low number of short, repeated motifs largely limited to three amino acids, such as glycine, serine and glutamate. In some embodiments, a UP sequence may be characterized by no more than 0.1%, 0.5%, 1%, 2%, 3%, 4%, or 5% of such repeats in its sequence.

[00286] In some embodiments a UP comprises multiple units of shorter sequences, or motifs, in which the amino acid sequences of the motifs are non-repetitive. In some aspects, the non-repetitive criterion can be met despite the use of a "building block" approach using a library of sequence motifs that are multimerized to create the UP sequences. Thus, while a UP sequence may in some embodiments consist of multiple units of as few as four different types of sequence motifs, because the motifs themselves generally consist of non-repetitive amino acid sequences, the overall UP sequence is rendered substantially non-repetitive. In some embodiments at least some such sequence motifs are separated by one or more amino acid residues that are not part of the repeat motif building block. Such intervening amino acid residues may be selected so that the overall UP sequence meets various criteria, such as having specified proportions or ranges of particular amino acid types, excluding or having a low number of certain amino acid types, etc.

[00287] In some embodiments a UP has a non-repetitive sequence of greater than about 100 amino acid residues up to about 3000 amino acid residues wherein at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97%, at least about 98%, at least about 99%, or about 100% of the UP sequence consists of non-overlapping sequence motifs, wherein each of the motifs has about 9 to 36 amino acid residues. In some embodiments, at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97%, at least about 98%, at least about 99%, or about 100% of the UP sequence consists of non-overlapping sequence motifs wherein each of the motifs has 9 to 14 amino acid residues. In some embodiments, at least about 80%, or at least about 85%, or at least about 90%, or at least about 95%, or at least about 97%, at least

about 98%, at least about 99%, or about 100% of the UP sequence consists of non-overlapping sequence motifs wherein each of the motifs has 12 amino acid residues. In certain embodiments, the sequence motifs are often composed mainly of small hydrophilic amino acids, such that the overall sequence has an unstructured, flexible characteristic. Examples of amino acids that are included in UP sequences in certain embodiments are, e.g., arginine, lysine, threonine, alanine, asparagine, glutamine, aspartate, glutamate, serine, and glycine. In some embodiments a UP comprises mainly glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P) residues wherein the sequences are, in some embodiments, designed to be substantially non-repetitive. In some embodiments, UP sequences have predominately four to six types of amino acids selected from glycine (G), alanine (A), serine (S), threonine (T), glutamate (E), and proline (P) that are arranged in a substantially non-repetitive sequence that is greater than about 100 residues and up to about 3000 amino acid residues, e.g., greater than 400 residues and up to about 3000 residues in length. In some embodiments UP have sequences of greater than about 100 to about 3000 amino acid residues wherein at least about 80% of the sequence consists of non-overlapping sequence motifs wherein each of the motifs has 9 to 36 amino acid residues wherein each of the motifs consists of 4 to 6 types of amino acids selected from glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P), and wherein the content of any one amino acid type in the full-length UP does not exceed 30%. In some embodiments, at least about 90% of the UP sequence consists of non-overlapping sequence motifs wherein each of the motifs has 9 to 36 amino acid residues wherein the motifs consist of 4 to 6 types of amino acids selected from glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P), and wherein the content of any one amino acid type in the full-length UP does not exceed 30%. In other embodiments, at least about 90% of the UP sequence consists of non-overlapping sequence motifs wherein each of the motifs has 12 amino acid residues consisting of 4 to 6 types of amino acids selected from glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P), and wherein the content of any one amino acid type in the full-length UP does not exceed 30%. In yet other embodiments, at least about 90%, or about 91%, or about 92%, or about 93%, or about 94%, or about 95%, or about 96%, or about 97%, or about 98%, or about 99%, to about 100% of the UP sequence consists of non-overlapping sequence motifs wherein each of the motifs has 12 amino acid residues consisting of glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P), and wherein the content of any one amino acid type in the full-length UP does not exceed 30%.

[00288] In some embodiments, UPs comprise non-repetitive sequences of greater than about 100 to about 3000 amino acid residues wherein at least about 80%, or at least about 90%, or about 91%, or about 92%, or about 93%, or about 94%, or about 95%, or about 96%, or about 97%, or about 98%, or about 99% of the sequence consists of non-overlapping sequence motifs of 9 to 14 amino acid residues wherein the motifs consist of 4 to 6 types of amino acids selected from glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P), and wherein the sequence of any two contiguous amino acid residues in any one motif is not repeated more than twice in the sequence motif. In other embodiments, at least about 90%, or about 91%, or about 92%, or about 93%, or about 94%, or about 95%, or about 96%, or about 97%, or about 98%, or about 99% of a UP sequence consists of non-overlapping sequence motifs of 12 amino acid residues wherein the motifs consist of 4 to 6 types of amino acids selected from glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P), and wherein the sequence of any two contiguous amino acid residues in any one sequence motif is not repeated more than twice in the sequence motif. In some embodiments, at least about 90%, or about 91%, or about 92%, or about 93%, or about 94%, or about 95%, or about 96%, or about 97%, or about 98%, or about 99% of a UP sequence consists of non-overlapping sequence motifs of 12 amino acid residues wherein the motifs consist of glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P), and wherein the sequence of any two contiguous amino acid residues in any one sequence motif is not repeated more than twice in the sequence motif. In some embodiments, UPs consist of 12 amino acid sequence motifs wherein the amino acids are selected from glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P), and wherein the sequence of any two contiguous amino acid residues in any one sequence motif is not repeated more than twice in the sequence motif, and wherein the content of any one amino acid type in the full-length UP does not exceed 30%. In the foregoing embodiments hereinabove described in this paragraph, the UP sequence is substantially non-repetitive.

[00289] In some embodiments, a UP sequence is a non-repetitive UP sequence of greater than about 100 to about 3000 amino acid residues wherein at least about 80%, or at least about 90%, or about 91%, or about 92%, or about 93%, or about 94%, or about 95%, or about 96%, or about 97%, or about 98%, or about 99% to about 100% of the sequence consists of multiple units of two or more non-overlapping sequence motifs selected from the amino acid sequences of Table A. In some embodiments, the UP comprises non-overlapping sequence motifs in which about 80%, or at least about 90%, or about 91%, or about 92%, or about 93%, or about 94%, or about 95%, or about 96%, or about 97%, or about 98%, or about 99%

to about 100% of the sequence consists of two or more non-overlapping sequences selected from a single motif family of Table A, resulting in a "family" sequence in which the overall sequence remains substantially non-repetitive. Accordingly, in these embodiments, a UP sequence comprises multiple units of non-overlapping sequence motifs of the AD motif family, or the AE motif family, or the AF motif family, or the AG motif family, or the AM motif family, or the AQ motif family, or the BC family, or the BD family of sequences of Table A. In some embodiments, the UP comprises motif sequences from two or more of the motif families of Table A. In some embodiments a UP sequence may comprise multiple smaller sequence modules having the structural (sequence) features of a UP. For example, a UP may comprise between 2 and 50, 50 and 100, 100 and 250, or 250 and 500 copies of a smaller UP. In some embodiments a UP may comprise multiple modules having different sequences. In some embodiments a UP comprises at least 2 copies of each of at least 2 different modules. In some embodiments at least some of the modules may be separated by one or more amino acids that are not part of the module. In general, such amino acids may be any amino acids, provided that the complete resulting sequence satisfies the constraints of a UP. In some embodiments a UP sequence consists at least 80%, 85%, 95%, 95%, 96%, 97%, 98%, 99%, or more of sequences listed in Table A. In some embodiments different sequences may be concatenated in any order. In some embodiments at least some sequence motifs are separated by one or more, e.g., up to 5 amino acid residues selected from G, A, S, E, T, and P residues. In some embodiments, as discussed herein, one or more amino acid residues other than G, A, S, E, T, P, may be present, e.g., one or more lysine or cysteine residues.

[00290] Table A: UP Sequence Motifs of 12 Amino Acids in Length and Motif Families

Motif Family*	MOTIF SEQUENCE
AD	GESPGGSSGSES
AD	GSEGSSGPGESS
AD	GSSESGSSEGGP
AD	GSGGEPSESGSS
AE, AM	GSPAGSPTSTEE
AE, AM, AQ	GSEPATSGSETP
AE, AM, AQ	GTSESATPESGP
AE, AM, AQ	GTSTEPSEGSAP
AF, AM	GSTSESPSGTAP
AF, AM	GTSTPESGSASP
AF, AM	GTSPSGESSTAP
AF, AM	GSTSSTAESPGP
AG, AM	GTPGSGTASSSP
AG, AM	GSSTPSGATGSP
AG, AM	GSSPSASTGTGP
AG, AM	GASPGTSSTGSP
AQ	GEPAGSPTSTSE
AQ	GTGEPSSTPASE
AQ	GSGPSTESAPTE
AQ	GSETPSGPSETA
AQ	GPSETSTSEPGA
AQ	GSPSEPTEGTSA
BC	GSGASEPTSTEP
BC	GSEPATSGTEPS
BC	GTSEPSTSEPGA
BC	GTSTEPSEPGSA
BD	GSTAGSETSTEA
BD	GSETATSGSETA
BD	GTSESATSESGA
BD	GTSTEASEGSAS

*Denotes individual motif sequences that, when used together in various permutations, results in a "family sequence".

[00291] In some embodiments, a UP sequence is a non-repetitive UP sequence of greater than about 100 to about 3000 amino acid residues wherein at least about 80%, or at least In other embodiments a UP comprises a non-repetitive UP sequence of greater than about 100 to about 3000 amino acid residues, wherein at least about 80%, or at least about 90%, or about 91%, or about 92%, or about 93%, or about 94%, or about 95%, or about 96%, or about 97%, or about 98%, or about 99% to about 100% of the sequence consists of non-overlapping 36 amino acid sequence motifs selected from one or more of the polypeptide sequences of the Tables.

[00292] Certain of the Tables herein include polynucleotide sequences encoding various UP polypeptides. It will be understood that due to the degeneracy of the genetic code, many different nucleic acid sequences can encode the UP polypeptide. The provided sequences are exemplary. In some embodiments a sequence is codon optimized for expression in a cell type (e.g., bacterial, eukaryotic), genus, or species of interest.

[00293] In certain embodiments wherein a UP has less than 100% of its amino acids consisting of four to six amino acid selected from glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P), or less than 100% of the sequence consisting of the sequence motifs of Table A, or less than 100% sequence identity compared with an a sequence from Table A, the other amino acid residues may be selected from any other of the 14 standard L-amino acids. In some embodiments the other amino acid residues are selected from hydrophilic amino acids such that the UP sequence contains at least about 90%, or at least about 91%, or at least about 92%, or at least about 93%, or at least about 94%, or at least about 95%, or at least about 96%, or at least about 97%, or at least about 98%, or at least about 99% hydrophilic amino acids. The amino acids that are not glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) or proline (P) are interspersed throughout the UP sequence, are located within or between the sequence motifs, or are concentrated in one or more short stretches of the UP sequence in various embodiments. In cases where a UP comprises one or more amino acids other than glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P), in many embodiments the amino acids are not hydrophobic residues and do not substantially confer secondary structure on the UP. Hydrophobic residues that may be avoided or rarely used in a UP include tryptophan, phenylalanine, tyrosine, leucine, isoleucine, valine, and methionine. In some embodiments a UP sequence contains few (e.g., less than 5%) or none of the following amino acids: cysteine (e.g., so as to avoid disulfide formation and oxidation), methionine (e.g., so as to avoid oxidation), asparagine and glutamine (e.g., so as to avoid desamidation). In some embodiments, a UP comprising one or more other amino acids in addition to glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P) has a sequence with less than 5% of the residues contributing to alpha helices and beta sheets as measured by Chou-Fasman algorithm and have at least 90%, or at least about 95% or more random coil formation as measured by GOR algorithm.

[00294] In some embodiments a UP comprises the following sequence or a portion thereof:

GSTSESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSASPG
TSTPESGSASPGSTSESPSGTAPGSTSESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGT
SPSGESSTAPGTSPSGESSTAPGSTSSTAESPGPGTSPSGESSTAPGTSPSGESSTAPGST
SE.

[00295] Non-limiting examples of UP that may be used in various embodiments are presented in Table B1and/or Table B2, wherein the UPs are XTENs. In some embodiments

UP sequences may be composed at least in part of fragments (subsequences) or variants of one or more sequences in Table B1 or B2. In some embodiments a fragment is at least 10, 20, 30, 40, or 50 amino acids long, up to the complete length of a sequence. In some embodiments a fragment extends over at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or more of a sequence. In some embodiments UP sequences may be generated by concatenating sequence motfs from any one or more of Tables C, D, E, F, and G, optionally with spacers between at least some of the sequences. For example, multiple 12 amino acid sequences and/or 36 amino acid sequences may be concatenated, optionally with intervening spacers, to form a sequence of desired length and/or other properties. UP sequences composed of multiple UP sequence motifs, e.g., from any of Tables B1, B2, C, D, E, F, and/or G, may be referred to as "composite UP sequences". Sequences

[00296] In some embodiments the present invention provides long-acting compstatin analogs comprising a UP and at least one compstatin analog moiety, wherein the length of the UP sequence is greater than about 100 to about 3000 amino acid residues, e.g., greater than about 400 amino acid residues, wherein the UP confers enhanced pharmacokinetic properties on the compound in comparison to a compstatin analog comprising the same compstatin analog sequence not linked to a UP. In some embodiments, the UP sequence is about 100, or about 200, or about 300, or about 400, or about 500, or about 600, or about 700, or about 800, or about 900, or about 1000, or about 1500, or about 2000, or about 2500 or up to about 3000 amino acid residues in length. In some embodiments the length of a UP is a multiple of 12 or 36. Any UP length herein may be adjusted to the nearest multiple of 12 or 36 in various embodiments. In some embodiments a UP sequence is about 100 to about 150, about 150 to about 250, about 250 to about 350, about 350 to about 450, about 450 to about 550, about 550 to about 650, about 650 to about 750, about 750 to about 850, about 850 to about 950, about 950 to about 1050, about 1050 to about 1250, about 1250 to about 1500, about 1500 to about 2000, or about 2000 to about 3000 amino acid residues in length. In some embodiments a UP sequence has a molecular weight of between 20 kD and about 30 kD, about 30 kD and about 40 kD, about 40 kD and about 50 kD, about 50 kD and about 60 kD, about 60 kD and about 70 kD, about 70 kD and about 80 kD, about 80 kD and about 90 kD, about 90 kD and about 100 kD, about 100 kD to about 125 kD, about 125 kD to about 150 kD, about 150 kD to about 175 kD, about 175 kD to about 200 kD, e.g., up to about 300 kD.

[00297] In certain embodiments a UP sequence is between about 600 and about 1200 amino acids, e.g., between about 800 and 1000 amino acids. In certain embodiments a UP

sequence has a molecular weight of about 70 kD to about 100 kD, e.g., about 80 kD, about 85 kD, about 90 kD, about 95 kD.

[00298] In some embodiments a long-acting compstatin analog comprises a UP sequence that exhibits over at least a portion of its length (e.g., over at least 80%, 85%, 90%, 95%, 95%, 97%, 98%, 99%, or 100% of its length) at least about 80% sequence identity, or at least about 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity compared to a UP listed in Table B1 or B2 or to a fragment consisting of at least 50% of a sequence listed in Table B1 or B2. In some embodiments the In some embodiments multiple UP sequences fulfilling such criteria (and/or fulfilling other criteria set forth herein) are concatenated (in any order), optionally separated by spacer(s) to form a sequence having a selected number of amino acids and/or selected size or within a range of lengths or sizes set forth herein.

[00299] In some embodiments a spacer may be encodable by a nucleic acid comprising a restriction site. Exemplary spacers are listed in the following table. Inclusion of such spacers may be useful in various embodiments, e.g., where a UP is to be used as a component of a fusion protein, or wherein two or more UPs are to be combined to form a longer UP.

Spacer Sequence	Restriction Enzyme
GSPG	BsaI
ETET	BsaI
PGSSS	BbsI
GAP	AscI
GPA	FseI
GPSGP	SfiI
AAA	SacII
TG	AgeI
GT	KpnI

[00300] In some embodiments a UP is designed for optimized expression by inclusion of encoding nucleotides for an optimized N-terminal leader sequence (NTS) in the polynucleotide encoding the protein. In some embodiment, the N-terminal sequence has at least 90% sequence identity compared to the sequence of AE48 or AM48, AE624, or AE912 or AM923. In some embodiments the UP has the N-terminal residues described in Examples 14-17 of WO2011028229.

[00301] In some embodiments a UP comprises a first and a second UP sequence, wherein the cumulative total of the residues in the UP sequences is greater than about 400 up to about

3000 amino acid residues and the UP sequences may be identical or may be different in sequence. In certain embodiments of the foregoing, the UP polypeptide comprises a first and a second UP sequence wherein the sequences each exhibit at least about 80% sequence identity, or alternatively 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity compared to at least a first or a second UP selected from any of Tables B1, B2, C, D, E, F, or G. In various embodiments any of the UPs set forth in any of the afore-mentioned tables may be concatenated, optionally separated by spacers, to form a longer UP. For example, short sequences of 12 – 36 amino acids may be concatentated randomly, in a regular repeating pattern, or irregularly, to form a longer UP sequence. Examples where more than one UP is used include, but are not limited to embodiments in which a compstatin analog moiety is linked to a side chain of a UP sequence (e.g., via a reactive functional group, which may be a click chemistry group) and multiple such UP sequences are linked together, e.g., via N- and C-termini of the UPs. In some embodiments N- and/or C-termini of a UP sequence comprise compatible click chemistry groups, so that the UPs can be linked together in a modular fashion, e.g., as depicted schematically in Figure 16. In some embodiments it is envisioned to create long-acting compstatin analogs comprising a UP as shown schematically in Figure 16, where "drug" and "peptide" in Figure 16 would represent a compstatin analog moiety.

[00302] In some embodiments the length and/or sequence of a UP is selected to confer a target half-life on a long-acting compstatin analog administered to a subject. In some embodiments a long-acting compstatin analog comprises at least about cumulative 400, 500, 600, 700, 800, 900, or 1000 amino acid residues of UP sequence, e.g., about 550 to about 650, about 650 to about 750, about 750 to about 850, about 850 to about 950, about 950 to about 1050, about 1050 to about 1250, about 1250 to about 1500, about 1500 to about 2000, or about 2000 to about 3000 amino acid residues of UP sequence. The UP sequences may be contiguous or may be separated by one or more compstatin analog moieties, linkers, or non-UP sequences, in various embodiments. In some embodiments the total length of the UP sequences is selected to provide a molecular weight of at least about 10,000; 20,000; 30,000; 40,000; 50,000; 60,000; 70,000; 80,000; 90,000; and 100,000 daltons. In some embodiments the total length of UP sequences is selected to provide a molecular weight of between 20 kD and about 30 kD, about 30 kD and about 40 kD, about 40 kD and about 50 kD, about 50 kD and about 60 kD, about 60 kD and about 70 kD, about 70 kD and about 80 kD, about 80 kD and about 90 kD, about 90 kD and about 100 kD, about 100 kD to about 125 kD, about 125 kD to about 150 kD, about 150 kD to about 175 kD, about 175 kD to about 200 kD, e.g., up

to about 300 kD. In some embodiments the total length of the UP sequences is selected to provide a molecular weight of at least about 20 kD, up to about 100 kD, 120 kD, 140 kD, 160 kD, 180 kD, or 200 kD or up to 300 kD. In some embodiments UP sequence lengths are selected to provide a terminal half-life of at least 2, 3, 4, 5, 6, 7, 8, 9, 10 days, or more. In some embodiments a long-acting compstatin analog comprising a UP is designed to confer a slower rate of systemic absorption after subcutaneous or intramuscular administration to a subject as compared with a compstatin analog comprising UP sequences of different lengths. In some embodiments, the C_{max} is reduced in comparison to a comparable dose of a compstatin analog not linked to a UP, which may thereby contribute to keeping the plasma concentration within a desired window, e.g., the therapeutic window, for the compound.

[00303] In some embodiments a long-acting compstatin analog comprises UP sequences wherein the cumulative length of the UP component is greater than about 100 to about 3000 amino acid residues containing at least one polypeptide sequence segment selected from Tables B1 and B2 (or a composite UP sequence) and wherein at least about 90%, or at least about 91%, or at least about 92%, or at least about 93%, or at least about 94%, or at least about 95%, or at least about 96%, or at least about 97%, or at least about 98% or more of the remainder of the UP sequence contains or consists of hydrophilic amino acids and less than about 2% of the remainder of the UP consists of hydrophobic or aromatic amino acids or cysteine. In some embodiments, the UP contains multiple segments wherein the segments are identical or different. In some embodiments the cumulative length of the UP component is greater than about 100 up to about 3000 amino acid residues and comprises at least one sequence segment of at least about 100 to about 923, or at least about 100 to about 875, or at least about 100 to about 576, or at least about 100 to about 288, or at least about 100 to about 144 amino acid residues wherein the sequence segment(s) consists of at least three different types of amino acids and the sum of glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P) residues in the sequence segment(s) constitutes at least about 90%, or at least about 91%, or at least about 92%, or at least about 93%, or at least about 94%, or at least about 95%, or at least about 96%, or at least about 97%, or at least about 98%, or at least about 99% of the total amino acid sequence of the sequence segment and at least about 90%, or at least about 91%, or at least about 92%, or at least about 93%, or at least about 94%, or at least about 95%, or at least about 96%, or at least about 97%, or at least about 98% of the remainder of the UP sequence(s) consist of hydrophilic amino acids and less than about 2% of the remainder of the UP sequence(s) consists of hydrophobic or aromatic amino acids, or cysteine. In some embodiments the cumulative length of the UP

component is greater than about 100 to about 3000 amino acid residues and comprises at least one sequence segment of at least about 200 to about 923, or at least about 200 to about 875, or at least about 200 to about 576, or at least about 200 to about 288 amino acid residues wherein the sequence segment(s) the sum of glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P) residues in the sequence segment(s) constitutes at least about 90%, or at least about 91%, or at least about 92%, or at least about 93%, or at least about 94%, or at least about 95%, or at least about 96%, or at least about 97%, or at least about 98%, or at least about 99% of the total amino acid sequence of the sequence segment and wherein the subsequence score of the segment is less than 12, more preferably less than 10, more preferably less than 9, more preferably less than 8, more preferably less than 7, more preferably less than 6, and most preferably less than 5, and at least about 90%, or at least about 91%, or at least about 92%, or at least about 93%, or at least about 94%, or at least about 95%, or at least about 96%, or at least about 97%, or at least about 98% of the remainder of the UP sequence(s) consist of hydrophilic amino acids and less than about 2% or less than about 1% of the remainder of the UP sequence(s) consists of hydrophobic, aromatic or cysteine amino acids.

[00304] "Identity" or "percent identity" is a measure of the extent to which the sequence of two or more nucleic acids or polypeptides is the same. The percent identity between a sequence of interest A and a second sequence B may be computed by aligning the sequences, allowing the introduction of gaps to maximize identity, determining the number of residues (nucleotides or amino acids) that are opposite an identical residue, dividing by the length over which alignment is being calculated (including gaps), and multiplying by 100. Fractions may be rounded to the nearest whole number. Sequences can be aligned with the use of a variety of computer programs known in the art, e.g., computer programs such as BLAST2, BLASTN, BLASTP, Gapped BLAST, etc., may be used to generate alignments and/or to obtain a percent identity. The algorithm of Karlin and Altschul (Karlin and Altschul, Proc. Natl. Acad. Sci. USA 87:22264-2268, 1990) modified as in Karlin and Altschul, Proc. Natl. Acad Sci. USA 90:5873-5877,1993 is incorporated into the NBLAST and XBLAST programs of Altschul et al. (Altschul, et al., J. Mol. Biol. 215:403-410, 1990). In some embodiments, to obtain gapped alignments for comparison purposes, Gapped BLAST is utilized as described in Altschul et al. (Altschul, et al. Nucleic Acids Res. 25: 3389-3402, 1997). When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs may be used. See the Web site having URL www.ncbi.nlm.nih.gov and/or McGinnis, S. and Madden, TL, W20–W25 Nucleic Acids Research, 2004, Vol. 32,

Web server issue. Other suitable programs include CLUSTALW (Thompson JD, Higgins DG, Gibson TJ, Nuc Ac Res, 22:4673-4680, 1994) and GAP (GCG Version 9.1; which implements the Needleman & Wunsch, 1970 algorithm (Needleman SB, Wunsch CD, J Mol Biol, 48:443-453, 1970.) Percent identity may be evaluated over a window of evaluation. In some embodiments a window of evaluation may have a length of at least 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or more, e.g., 100%, of the length of the shortest of the sequences being compared. In some embodiments a window of evaluation is at least 100; 200; 300; 400; 500; 600; 700; 800; 900; 1,000; 1,200; 1,500; 2,000; 2,500; 3,000; 3,500; 4,000; 4,500; or 5,000 amino acids. In some embodiments no more than 20%, 10%, 5%, or 1% of positions in either sequence or in both sequences over a window of evaluation are occupied by a gap. In some embodiments no more than 20%, 10%, 5%, or 1% of positions in either sequence or in both sequences are occupied by a gap over a window of evaluation.

[00305] In some embodiments a UP comprises at least 40, 60, 80, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1100, 1200, 1300, 1400, 1500, or more contiguous amino acids in which the sum of G, D, A, S, T, E, and/or P residues in the UP constitutes at least 50%, 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% or more of the total amino acids of the URP. In some embodiments the remainder of the sequence, when present, consists of at least 90%, 95%, 96%, 97%, 98%, 99%, or 100% arginine and/or lysine. In some the remainder of the sequence, when present, does not contain methionine (M), cysteine (C), asparagine (N), and glutamine (Q) or contains no more than 1%, 2%, 3%, 4%, or 5% methionine, cysteine, asparagine, and glutamine (Q) in total. In some embodiments one of the amino acids of the amino acids selected from the group consisting of glycine (G), aspartate (D), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P) alone constitutes at least 20% of the total amino acids of the URP. In some embodiments one of the amino acids of the amino acids selected from the group consisting of one of the amino acids selected from the group consisting of glycine (G), aspartate (D), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P) alone constitutes at least 30% of the total amino acids of the URP. In some embodiments at least one of the amino acids of the amino acids selected from the group consisting of glycine (G), aspartate (D), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P) alone constitutes at least 20% of the total amino acids of the URP. In some embodiments at least one of the amino acids selected from the group consisting of glycine (G), aspartate (D), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P) alone constitutes at least

30% of the total amino acids of the UP. In some embodiments, the glycine residues contained in the URP may constitute at least about 50% of the total amino acids of the URP. In some embodiments a URP comprises at least three different types of amino acids selected from glycine (G), aspartate (D), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P). In some embodiments a UP is characterized in that the sum of glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P) residues constitutes more than about 80%, or more than about 90%, or more than about 95%, or more than about 96%, or more than about 97%, or more than about 98% of the total amino acid residues of the UP; the sum of asparagine and glutamine residues is less than 10%, or less than 5%, or less than 2% of the total amino acid sequence of the UP; the sum of methionine and tryptophan residues is less than 2% of the total amino acid sequence of the UP; the UP sequence is substantially non-repetitive such that (i) the UP sequence contains no three contiguous amino acids that are identical unless the amino acids are serine, (ii) at least about 80%, or at least about 90%, or at least about 95%, or at least about 96%, or at least about 97%, or at least about 98% of the UP sequence consists of non- overlapping sequence motifs, each of the sequence motifs comprising about 9 to about 14 amino acid residues, wherein any two contiguous amino acid residues do not occur more than twice in each of the sequence motifs; or (iii) the UP sequence has a subsequence score of less than 10, or less than 9, or less than 8, or less than 7, or less than 6, or less than 5; the UP sequence has greater than 90%, or greater than 95%, or greater than 96%, or greater than 97%, or greater than 98% random coil formation as determined by GOR algorithm; the UP sequence has less than 2% alpha helices and 2% beta-sheets as determined by Chou-Fasman algorithm; and the UP sequence lacks a predicted T-cell epitope when analyzed by TEPITOPE algorithm, wherein the TEPITOPE algorithm prediction for epitopes within the UP sequence is based on a score of -9. In some embodiments, no one type of amino acid selected from glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P) constitutes more than about 8%, or about 16%, or about 20%, or about 25%, or about 30% of the UP sequence. In some embodiments a UP sequence has at least at about 90%, 91%, 92%, 93%, or about 94%, or about 95%, or about 96%, or about 97%, or about 98%, or about 99% to about 100% sequence identity to a sequence found in Table 4, Table 8, Table 9, Table 10, Table 11, Table 12, or Table 13 of WO2011028229. In some embodiments a UP sequence has at least at about 90%, 91%, 92%, 93%, or about 94%, or about 95%, or about 96%, or about 97%, or about 98%, or about 99% to about 100% sequence identity to a sequence found in Table 4, Table 8, Table 9, Table 10,

Table 11, Table 12, Table 13, Table 14, Table 17 or Table 18 of WO2011123830. It will be understood that such examples are non-limiting.

[00306] In some embodiments a UP sequence comprises an N-terminal and/or a C-terminal sequence included to enhance expression, secretion, detection, or purification of a UP polypeptide. Such sequence(s) may be the same or different and may be included for the same or different purposes. Such sequence(s) may, in some embodiments, be removed in a final compound comprising the UP. In some embodiments any of the properties of a UP sequence may be evaluated excluding or including such an N-terminal and/or C-terminal sequence. In some embodiments a short UP sequence is incorporated as the N-terminal portion of a UP protein, e.g., as described in WO2011028229, which discloses that the expression of a fusion protein comprising an XTEN is enhanced in a host cell transformed with a suitable expression vector comprising an optimized N-terminal leader polynucleotide sequence. For example, a host cell transformed with such an expression vector comprising an optimized N-terminal leader sequence (NTS) in the gene may result in greatly enhanced expression compared to the expression of a corresponding protein from a polynucleotide not comprising the NTS, and may in at least some embodiments replace incorporation of a non-XTEN leader sequence that may otherwise be used to enhance expression. In some embodiment, the N-terminal portion of a UP polypeptide comprises a sequence that exhibits at least about 80%, e.g., at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or exhibits 100% sequence identity compared to the amino acid sequence of AE48 or AM48, the respective amino acid sequences of which are as follows:

AE48: MAEPAGSPTSTEEGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGS
AM48: MAEPAGSPTSTEEGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGS

[00307] In some embodiments a longer N-terminal sequence of a UP comprises at least about 91%, or at least about 92%, or at least about 93%, or at least about 94%, or at least about 95%, or at least about 96%, or at least about 97%, or at least about 98%, or at least 99%, or exhibits 100% sequence identity with an amino acid sequence selected from the group consisting of the sequences AE624, AE912, and AM923.

[00308] In any of the various embodiments an N-terminal sequence may in some embodiments have from about one to about six additional amino acid residues, e.g., selected from G, E, S, T, P, and A, which may be included to accommodate a restriction endonuclease restriction site.

[00309] In some embodiments, a UP polypeptide, e.g., an XTEN, has an unstructured characteristic imparted by incorporation of amino acid residues with a net charge and/or reducing the proportion of hydrophobic amino acids in the sequence. In some embodiments, overall net charge and/or net charge density may be controlled by modifying the content of charged amino acids in the sequences. In some embodiments, the net charge density of a UP, e.g., an XTEN, may be above +0.1 or below -0.1 charges/residue. By "net charge density" of a protein or peptide herein is meant the net charge divided by the total number of amino acids in the protein or peptide. In some embodiments, the net charge density may be about 0%, about 1%, about 2%, about 3%, about 4%, about 5%, about 6%, about 7%, about 8%, about 9%, about 10% about 11%, about 12%, about 13%, about 14%, about 15%, about 16%, about 17%, about 18%, about 19%, or about 20% or more.

[00310] In some embodiments a UP sequence is designed to have a net negative charge to minimize non-specific interactions between a UP-containing compound and various surfaces such as blood vessel walls, healthy tissues, or various receptors, which may have a net negative charge. In some embodiments a negative charge or reduced charge is conferred by incorporation of glutamic acid residues. In some embodiments a UP seuqence contains about 8, 10, 15, 20, 25, or up to 30% glutamic acid. In some embodiments the glutamic acid residues are spaced relatively uniformly across at least a portion of the sequence, e.g., across at least 30%, 50%, 70%, 90%, or about the entire sequence. In some embodiments a UP sequence contains about 10-80, or about 15-60, or about 20-50 glutamic residues per approximately 20 kDa of UP.

[00311] In most embodiments a UP sequence, e.g., an XTEN, generally has no or a low content of positively charged amino acids. In some embodiments a UP may have less than about 10% amino acid residues with a positive charge, or less than about 7%, or less than about 5%, or less than about 2%, or less than about 1% amino acid residues with a positive charge.

[00312] In some embodiments a UP sequence comprises one or more charged residues separated by other residues such as serine or glycine, which leads to better expression or purification behavior. Based on the net charge a UP may, in various non-limiting embodiments, have an isoelectric point (pI) of 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, or 6.5. In some embodiments a UP has an isoelectric point between 1.5 and 4.5. In some embodiments a UP carries a net negative charge under physiological conditions that may contribute to the unstructured conformation and/or reduced binding of the UP component to mammalian proteins and tissues.

[00313] In many embodiments, the content of hydrophobic amino acids in a UP is less than 5%, or less than 2%, or less than 1% hydrophobic amino acid content. In some embodiments the amino acid content of methionine and tryptophan in a UP is typically less than 5%, or less than 2%, or less than 1%. In some embodiments a UP has a sequence that has less than 10% amino acid residues with a positive charge, or less than about 7%, or less that about 5%, or less than about 2% amino acid residues with a positive charge. In some embodiments the sum of methionine and tryptophan residues is less than 2%, and the sum of asparagine and glutamine residues is less than 10% of the total UP sequence.

[00314] In some aspects, UP sequences, e.g., XTENs, have a low degree of immunogenicity or are substantially non-immunogenic. Without wishing to be bound by theory, various factors that may contribute to low immunogenicity may include, e.g., the non-repetitive sequence, unstructured conformation, high degree of solubility, low degree or lack of self-aggregation, low degree or lack of proteolytic sites within the sequence, and/or low degree or lack of epitopes (e.g., B cell and/or T cell epitopes) in the sequence. In some embodiments a UP sequence is characterized by a low degree or absence of conformational epitopes. Conformational epitopes may be formed by regions of the protein surface that are composed of multiple discontinuous amino acid sequences of the protein antigen. The precise folding of the protein brings these sequences into a well-defined, stable spatial configurations, or epitopes, that can be recognized as "foreign" by the host humoral immune system, resulting in the production of antibodies to the protein or the activation of a cell-mediated immune response. In some embodiment a lack of or low degree of immunogenicity is achieved by designing UP sequences that resist antigen processing in antigen presenting cells, and/or choosing sequences that do not bind well to MHC receptors (e.g., MHC Class II). In some embodiments substantially non-repetitive UP polypeptides have low or absent (undetectable) binding to MHC II receptors and avoid formation of epitopes for T-cell receptor or antibody binding, resulting in a low degree of immunogenicity. In some aspects, administration of compounds comprising UP sequences, using appropriate therapeutic practices and dosing, would generally not result in the formation of neutralizing antibodies to the UP sequence or a compound containing the UP sequence. In some embodiments UP sequences are substantially free of epitopes recognized by human T cells. The elimination of such epitopes for the purpose of generating less immunogenic proteins is known in the art; see for example WO 98/52976, WO 02/079232, WO 00/34317, and/or US20060228758 and/or references cited in any one or more of the foregoing. Assays for human T cell epitopes have been described (Stickler, M., et al. (2003) J Immunol Methods, 281: 95-108).

In some embodiments peptide sequences that can be oligomerized without generating T cell epitopes or non-human sequences that would be recognized as foreign are used. In some embodiments, this is achieved by testing direct repeats of these sequences for the presence of T-cell epitopes and/or for the occurrence of 6 to 15-mer and, in particular, 9-mer sequences that are not found in human proteins, and then altering the design of the sequence to eliminate or disrupt the epitope sequence. In some embodiments a UP sequence is designed to be substantially non-immunogenic by limiting the number of epitopes predicted to bind MHC receptors. The degree of predicted T-cell epitopes can be determined by epitope prediction algorithms such as, e.g., TEPITOPE (Sturniolo, T., et al. (1999) Nat Biotechnol, 17: 555-61) or others known in the art such as EpiMatrix, Multipred, MHC Thread, MHCPred, NetMHC, LpPep, BIMAS, and SYFPEITHI. (See, e.g., Kangueane, P., "Bioinformation Discovery: Data to Knowledge in Biology", Springer, 2009, for references.) The TEPITOPE score of a given peptide frame within a protein is the log of the K_d (dissociation constant, affinity, off-rate) of the binding of that peptide frame to multiple of the most common human MHC alleles, as disclosed in Sturniolo, T. et al. (1999) Nature Biotechnology 17:555). The score ranges over at least 20 logs, from about 10 to about -10 and can be reduced by avoiding hydrophobic amino acids that serve as anchor residues during peptide display on MHC, such as M, I, L, V, F. In some embodiments a UP sequence does not have a predicted T-cell epitope at a TEPITOPE score of about -5 or greater, or -6 or greater, or -7 or greater, or -8 or greater, or at a TEPITOPE score of -9 or greater. As used herein, a score of "-9 or greater" would encompass TEPITOPE scores of 10 to -9, inclusive, but would not encompass a score of -10, as -10 is less than -9.

[00315] In some embodiments a UP is rendered substantially non-immunogenic at least in part by limiting the number of known proteolytic sites from the sequence, thereby reducing the potential processing of the protein into small peptides that may bind to MHC II receptors. In some embodiments a UP is rendered substantially non-immunogenic by the use a sequence that is substantially devoid of secondary structure. In some embodiments the non-repetitive sequence and lack of epitopes limit the ability of B cells to bind to or be activated and/or reduce the strength of antibody binding, which may result in reduced likelihood of immune clearance.

[00316] In some aspects, a UP polypeptide has a high hydrodynamic radius that confers a corresponding increased apparent molecular weight to a compound comprising it. The "hydrodynamic radius" or "Stokes radius" is the effective radius (Rh in nm) of a molecule in a solution measured by assuming that it is a body moving through the solution and resisted by

the solution's viscosity. In some embodiments the hydrodynamic radius of the UP or compound is greater than the approximately 3-5 nm. The hydrodynamic radius of a protein is determined by its molecular weight as well as by its structure, including shape or compactness. In some embodiments a UP has a greater proportional hydrodynamic radius compared to polypeptides of a comparable sequence length and/or molecular weight that have secondary and/or tertiary structure, such as typical globular proteins. Methods for determining the hydrodynamic radius are well known in the art, such as by the use of size exclusion chromatography (SEC), as described in U.S. Pat. Nos. 6,406,632 and 7,294,513. In certain embodiments a UP or compound (e.g., long-acting compstatin analog) comprising a UP is designed such that the UP or compound can have a hydrodynamic radius of at least about 5 nm, or at least about 8 nm, or at least about 10 nm, or 12 nm, or at least about 15 nm. In some aspects, the large hydrodynamic radius conferred by the UP can lead to reduced renal clearance of the resulting fusion protein, leading to a corresponding increase in terminal half-life, an increase in mean residence time, and/or a decrease in renal clearance rate.

[00317] It will be understood that a UP may have any one or more UP properties described herein, in any combination, provided such properties are not mutually exclusive. For example, a UP may have an amino acid makeup selected from any of the exemplary amino acid makeups described herein, and a degree of secondary structure or lack thereof selected from any range or value set forth herein, a length selected from any of the lengths or length ranges set forth herein, and a molecular weight selected from any of the weights or weight ranges set forth herein, provided the resulting combination is possible without inconsistency. Each such combination is to be considered expressly set forth herein.

[00318] In some embodiment a UP of a chosen length and sequence (e.g., a sequence from Table B1 or B2 or a variant thereof or a composite sequence) may be used to create a long-acting compstatin analog that has, e.g., under physiological conditions, an apparent molecular weight of at least about 70 kDa, of at least about 80 kDa, at least about 90 kDa, at least about 100 kDa, at least about 120 kDa, at least about 150 kDa, at least about 200 kDa, at least about 300 kDa, at least about 500 kDa, at least about 800 kDa, at least about 1000 kDa, or at least about 1500 kDA, or at least about 1800 kDa, or at least about 2000 kDa, or more.

[00319] In various embodiments a long-acting compstatin analog comprising a UP may have any of the properties of a long-acting compstatin analog described herein in terms of increase in half-life, decrease in renal clearance, average activity per molecule, average activity per compstatin analog moiety, etc. In some embodiments a long-acting compstatin

analog comprising a UP may exhibit one or more properties equivalent or superior to that exhibited by a compstatin analog in which a PEG is used.

[00320] In some embodiments a UP comprises one or more amino acid residues comprising a side chain comprising a reactive functional group, such as an amine or thiol. For example, one or more lysine or cysteine residues may be incorporated into a UP. In some embodiments a limited number of amino acids with a positive charge, such as an amino acid having a side chain containing a primary or secondary amine group, e.g., lysine, are incorporated into a UP. In some aspects, presence of an amino acid with a reactive side chain permits conjugation between the reactive side chian and a reactive group on a peptide, a linker, or a reactive functional group on a drug or small molecule, or a non-polypeptide polymer, to be conjugated to the UP backbone. For example, in the case of lysine, such conjugation may occur between the epsilon amine of the lysine and a reactive group on a peptide, a linker, or a reactive functional group on a drug or small molecule, or a non-polypeptide polymer. In some embodiment a UP, e.g., an XTEN, has between about 1 to about 100 lysine residues, or about 1 to about 70 lysine residues, or about 1 to about 50 lysine residues, or about 1 to about 30 lysine residues, or about 1 to about 20 lysine residues, or about 1 to about 10 lysine residues. In some embodiments only a single lysine, or from about 1 to about 5 lysine residues, e.g., 1, 2, 3, 4, or 5 lysine residues, are incorporated. Lysine is sometimes used herein as an example, but other amino acids that comprise or may be modified to comprise a reactive side chain may be used additional or alternately. Any such residues may be located anywhere within the UP sequence. In some embodiments first and second lysines (or other amino acids comprising reactive side chains) are located near the N- and C-termini (e.g., up to about 1-5, 5-10, or 10-20, or 20-50 amino acids away from the termini). In some embodiments if three or more such amino acid residues are present, they are approximately equally spaced. In some embodiments a UP is chemically or enzymatically modified (e.g., after synthesis or after purification) to render one or more amino acid side chains reactive, e.g., to replace a naturally occurring side chain with a more reactive moiety to facilitate subsequent conjugation of a compstatin analog moiety or targeting moiety thereto.

[00321] In some embodiments a long-acting compstatin analog is generated using a lysine-containing UP or using a UP containing one or more other reactive side chains (e.g., cysteine-containing UP) and a compstatin analog moiety, wherein the maximum number of molecules of compstatin analog moiety attached to the UP component is determined by the numbers of lysines or other amino acids with reactive side chains (e.g., cysteine) in the UP sequence.

[00322] In some aspects, compounds are provided that comprise UP and compstatin analog moiety components linked in any of a variety of N- to C-terminus configurations. In some embodiments, one or more compstatin analog moieties are linked to one or more UPs either at the N-terminus or at the C-terminus, or at both termini, with or without a spacer. The resulting configuration may be considered a block copolymer, and the sequential arrangement of the compstatin analog moieties and the UP in the compound may assume a configuration akin to those known in the art of block copolymer chemistry. A compstatin analog moiety may be abbreviated as "M". When there is more than one compstatin analog moiety M, more than one UP, or more than one spacer or linker, each of the M, the UP, or spacer may have the same or different sequences, and the Ms and/or UPs may be linked either continuously or alternately (regular or irregular), optionally separated by spacer(s). Thus, in all of the formulae provided herein, when there is more than one M, UP, or spacer, each of the M, UP, and spacer may be the same or different in various embodiments. In some embodiments a compound has a single M moiety linked to one UP polypeptide. In some embodiments a compound comprises a M linked to two or more UP polypeptides, e.g., a UP is linked to the N- and to the C-terminus of the M moiety. In some such embodiments, one or more additional M moieties are linked to either or both of the UP sequences. In some embodiments two or more M moieties are linked to one UP polypeptide. In some embodiments two or more M moieties are linked to two or more UP polypeptides. Non-limiting examples of various configurations are shown below (separated by semicolons).

[00323] M-UP; UP-M; UP-M-UP; M-UP-UP; UP-UP-M; UP-M-UP-UP; UP-UP-M-UP; UP-UP-M-UP; M-UP-M; UP-M-M; M-M-UP; M-UP-M-M; M-UP-M-UP; UP-M-UP-M; UP-UP-M-UP-M; UP-UP-M-M; M-UP-UP-M; M-M-UP-UP; M-M-UP-UP-M; M-UP-M-UP-M.

[00324] Numerous other variations will be apparent to the ordinarily skilled artisan. In some embodiments any of the spacer and/or cleavage sequences disclosed herein or known in the art may be present between units. In some embodiments at least one M is replaced by a targeting moiety or comprises a cell-reactive moiety. In some embodiments, at least some M moieties are conjugated to a side chain of a UP. Conjugation to a side chain may be used instead of or in addition to conjugation to a N- or C-terminus. Thus, for example, any of the configurations above may further comprise, or may instead comprise, one or more compstatin analog moieties conjugated to a side chain of one or more UP component(s). Figure 14 schematically shows UPs with side chains comprising reactive functional groups. In some embodiments utilizing orthogonal conjugation sites, a compstatin analog moiety is

conjugated to a first reactive functional group type on a UP and a targeting moiety or second therapeutically active agent is conjugated to a second reactive functional group type on a UP. Figure 15 shows further details regarding particular strategies that may be employed to generate a conjugate.

[00325] In various embodiments any of the configurations may be produced using click chemistry or other chemical conjugation method. In certain embodiments at least one or more compstatin analog moieties is encoded in a polynucleotide sequence and expressed in the form of a fusion protein comprising a compstatin analog sequence and a UP sequence, optionally separated by a polypeptide spacer. Such compstatin analog sequences may be located anywhere with respect to the UP sequence(s). Any of the compstatin analog sequences disclosed herein may be conjugated to a UP. For example, embodiments are provided in which any of SEQ ID NO: 3-36, 37, 37A, 38A, 39A, 40A, or 41A are used In some embodiments any such sequence may comprise one or more additional amino acids, e.g., at the N- and/or C-terminus, wherein such amino acid comprises a reactive functional group, e.g., an amine or click chemistry group and optionally is attached to the remainder of the compstatin analog via a linker. For example, in some embodiments any such sequence comprises AEEAc-X at its C-terminus, wherein X is an amino acid comprising a side chain comprising a reactive functional group , e.g., AEEAc-Lys at the C-terminus. It will be understood that a blocking moiety, e.g., NH_2, may be present at the C-terminus. Suitable linkers and reactive functional groups that may be used for connecting a compstatin analog and a polypeptide such as a UP are extensively described above and in classes and subclasses herein. In some embodiments, a linker has multiple functional groups, wherein one functional group is connected to a compstatin analog and another is connected to a UP.

[00326] Any of the linking moieties and/or reactive functional groups disclosed herein may be used, in various embodiments, to produce a long-acting compstatin analog comprising a UP. In certain embodiments a long-acting compstatin analog comprising a UP comprises between 1 and 8 compstatin analog moieties, e.g., 1, 2, 3, 4, 5, 6, 7, or 8. In particular embodiments, a long-acting compstatin analog comprises a single compstatin analog moiety or two compstatin analog moieties or thee compstatin analog moieties. In some embodiments a long-acting compstatin analog comprises two UPs conjugated together, each of which comprises 1 or 2 compstatin analog moieties linked thereto.

[00327] In some embodiments a UP, e.g., an XTEN, comprising a thiol (a single thiol, or in some embodiments, multiple thiols, e.g., 2 or 3. In such embodiments the compstatin analog component to be reacted with the XTEN may comprise any thiol reactive group. For

example, in some embodiments a maleimide functionalized compstatin analog such as *Ac*-Ile1-Cys2*-Val3-(1-methyl-l-tryptophan)4-Gln5-Asp6-Trp7-Gly8-Ala9-His10-Arg11-Cys12*-Thr13-NH-$CH_2CH_2OCH_2CH_2OCH_2CO$-Lys(CO-(CH_2)5-maleimide)-NH_2 is used. In some embodiments a compstatin analog comprising a lysine, e.g., *Ac*-Ile1-Cys2*-Val3-(1-methyl-l-tryptophan)4-Gln5-Asp6-Trp7-Gly8-Ala9-His10-Arg11-Cys12*-Thr13-NH-$CH_2CH_2OCH_2CH_2OCH_2CO$-Lys-NH_2, is reacted with a heterobifunctional linker, wherein a first reactive functional group of the linker reacts with amine groups, and a second reactive functional group of the linker reacts with thiols. An example of such a linker is AMAS (N-(a-maleimidoacetoxy)-succinimide ester). One of ordinary skill in the art will appreciate that numerous reagents that react with amines on one side and a thiol on the other are available and may be used in various embodiments. Examples are found in, e.g., Hermanson, supra, or the Thermo Scientific Pierce Cross-Linking Technical Handbook 2009 (e.g., pp. 26-28). In some embodiments, the linker is first reacted with the peptide and then reacted (i.e., after attachment to the peptide) with the XTEN. In some embodiments a purification step may be performed in between the first and second reactions. In some embodiments the compstatin analog is used in excess, which may avoid the need for a purification step. In some embodiments, an amine-terminated XTEN is used. In some embodiments an orthogonal group is introduced to the compstatin analog. For example, instead of a lysine one may use a third cysteine. In some embodiments the third cysteine may be protected, e.g., to facilitate cyclizing the compstatin analog via disulfide bond without interference from the third cysteine. Then a bifunctional linker is employed, which reacts with a thiol on the compstatin analog and an amine on the XTEN. It will be understood that these represent only some of the various ways that a compstatin analog may be conjugated to a UP, e.g., an XTEN. As discussed above, use of click chemistry is contemplated. In some embodiments a compstatin analog moiety comprising a first click chemistry group is reacted with a UP, e.g., an XTEN, comprising a complementary second click chemistry group. For example, AAX6 (Figure 17) may be reacted with a compstatin analog moiety comprising an azide moiety.

[00328] Long-acting compstatin analogs comprising a UP may be used in any method or composition described herein pertaining to a long-acting compstatin analog. For example, long-acting compstatin analogs comprising a UP may be incorporated in a pharmaceutical composition and/or may be administered to treat a subject in need of treatment for a complement-mediated disorder. In some embodiments a long-acting compstatin analog comprising a UP is administered to a subject in need of treatment for a disease or condition causing complement-mediated damage to red blood cells. In some embodiments a long-

acting compstatin analog comprising a UP is used to treat a subject suffering from PNH or aHUS. In some embodiments a long-acting compstatin analog comprising a UP is used to treat a subject in need of treatment for a chronic respiratory disease (e.g., asthma, COPD). In some embodiments a long-acting compstatin analog is used to treat a subject in need of treatment for a Th17-associated disorder. In some embodiments a long-acting compstatin analog is used to treat a subject in need of treatment for an autoimmune disorder.

[00329] The invention further provides multimers, e.g., concatamers, comprising two or more (e.g., between 2 and 10) compstatin analogs comprising a CRM, wherein the average molecular weight of the resulting molecule (or the CRM components thereof) is between 20,000; 30,000; 40,000; 50,000; 60,000; 70,000; 80,000; 90,000; and 100,000 daltons. In some embodiments the average molecular weight of the resulting molecule (or the CRM components thereof) is at least 20,000 daltons, up to about 100,000; 120,000; 140,000; 160,000; 180,000; or 200,000 daltons. In some embodiments, the compstatin analogs comprising a CRM can be linked using any of the linking moieties described above. Compositions and methods for making long-acting compstatin analogs, and intermediates in the synthesis, are aspects of the invention.

[00330] A wide variety of methods and assays useful for detection of polymers, e.g., PEGs, POZs, and/or polypeptides and/or useful for measurement of physical and/or structural properties of of polymers, e.g., PEGs, POZs, and/or polypeptides are known in the art and may, if desired, be used to detect a compstatin analog, e.g., a cell-reactive, long-acting, targeted compstatin analog or a compstatin analog moiety. For example, methods and assays useful for determining properties such as aggregation, solubility, size, structure, melting properties, purity, presence of degradation products or contaminants, water content, hydrodynamic radius, etc., are available. Such methods include, e.g., analytical centrifugation, various types of chromatography such as liquid chromatography (e.g., HPLC-ion exchange, HPLC-size exclusion, HPLC-reverse phase), light scattering, capillary electrophoresis, circular dichroism, isothermal calorimetry, differential scanning calorimetry, fluorescence, infrared (IR), nuclear magnetic resonance (NMR), Raman spectroscopy, refractometry, UV/Visible spectroscopy, mass spectrometry, immunological methods, etc. It will be understood that methods may be combined. In some aspects, a cell-reactive, long-acting, or targeted comptatin analog (or composition comprising a cell-reactive, long-acting, or targeted compstatin analog) has one or more properties described herein, as assessed using any of the foregoing methods. In some aspects, methods useful to detect and/or quantify a long-acting compstatin analog are described herein.

[00331] VII. Targeted Compstatin Analogs

[00332] The invention provides targeted compstatin analogs that comprise a targeting moiety and a compstatin analog moiety, wherein the targeting moiety binds non-covalently to a target molecule. In some aspects, the invention provides targeted compstatin analogs analogous to the cell-reactive compstatin analogs described in Section VI, wherein the compounds comprise a targeting moiety in addition to, or instead of, a cell-reactive moiety. The targeting moiety can comprise, e.g., an antibody, polypeptide, peptide, nucleic acid (e.g., an aptamer), carbohydrate, small molecule, or supramolecular complex, that specifically binds to the target molecule. In some embodiments, the affinity (as measured by the equilibrium dissociation constant, Kd) of targeting moiety for the target molecule (as measured by the equilibrium dissociation constant, Kd) is 10^{-3} M or less, e.g., 10^{-4} M or less, e.g., 10^{-5} M or less, e.g., 10^{-6}M or less, 10^{-7}M or less, 10^{-8}M or less, or 10^{-9} M or less under the conditions tested, e.g., under physiological conditions.

[00333] In those embodiments of the invention in which the targeting moiety is an antibody, the antibody may be any immunoglobulin or a derivative thereof, which maintains binding ability, or any protein having a binding domain which is homologous or largely homologous to an immunoglobulin binding domain. Such proteins may be derived from natural sources, or partly or wholly synthetically produced (e.g., using recombinant DNA techniques, chemical synthesis, etc.). The antibody can be of any species, e.g., human, rodent, rabbit, goat, chicken, etc. The antibody may be a member of any immunoglobulin class, including any of the human classes: IgG, IgM, IgA, IgD, and IgE. In various embodiments of the invention the antibody may be a fragment of an antibody such as an Fab', F(ab').sub.2, scFv (single-chain variable) or other fragment that retains an antigen binding site, or a recombinantly produced scFv fragment, including recombinantly produced fragments. See, e.g., Allen, T., Nature Reviews Cancer, Vol. 2, 750-765, 2002, and references therein. Monovalent, bivalent or multivalent antibodies can be used. The antibody may be a chimeric antibody in which, for example, a variable domain of rodent origin is fused to a constant domain of human origin, thus retaining the specificity of the rodent antibody. In some embodiments, a human antibody or portion thereof is generated, for example, in rodents whose genome incorporates human immunoglobulin genes, using a display technology such as phage display, etc. In some embodiments, a humanized antibody is generated by grafting one or more complementarity determining region(s) from a non-human species (e.g., mouse) into a human antibody sequence. The antibody may be partially or completely humanized. See, e.g., Almagro JC, Fransson J.Humanization of antibodies. Front Biosci. 13:1619-33

(2008) for review of various methods of obtaining humanized antibodies that may be used to obtain a targeting moiety of use in the invention. An antibody may be polyclonal or monoclonal, though for purposes of the present invention monoclonal antibodies are generally preferred. In certain embodiments of the invention a F(ab')2 or F(ab') fragment is use while in other embodiments antibodies comprising an Fc domain are used. Methods for producing antibodies that specifically bind to virtually any molecule of interest are known in the art. For example, monoclonal or polyclonal antibodies can be purified from natural sources, e.g., from blood or ascites fluid of an animal that produces the antibody (e.g., following immunization with the molecule or an antigenic fragment thereof) or can be produced recombinantly, in cell culture. Methods of generating antibody fragments, e.g., by digestion, disulfide reduction, or synthesis are known in the art.

[00334] In various embodiments of the invention a targeting moiety can be any molecule that specifically binds to a target molecule through a mechanism other than an antigen-antibody interaction. Such a targeting moiety is referred to as a "ligand". For example, in various embodiments of the invention a ligand can be a polypeptide, peptide, nucleic acid (e.g., DNA or RNA), carbohydrate, lipid or phospholipid, or small molecule. In some embodiments a small molecule is an organic compound, whether naturally-occurring or artificially created, that has relatively low molecular weight and is not a protein, polypeptide, nucleic acid, or lipid, typically with a molecular weight of less than about 1500 g/mol and typically having multiple carbon-carbon bonds. In general, an aptamer is an oligonucleotide (e.g., RNA or DNA, optionally comprising one or more modified nucleosides (e.g., bases or sugars other than the 5 standard bases (A, G, C, T, U) or sugars (ribose and deoxribose) found most commonly in RNA and DNA), or modified internucleoside linkages (e.g., non-phosphodiester bonds) that, e.g., stabilize the molecule, e.g., by rendering it more resistant to degradation by nucleases) that binds to a particular protein. In some embodiments an oligonucleotide is up to about 100 nucleosides long, e.g., between 12 and 100 nucleosides long. Aptamers can be derived using an in vitro evolution process called SELEX, and methods for obtaining aptamers specific for a protein of interest are known in the art. See, e.g., Brody E N, Gold L. J Biotechnol. 2000 March; 74(1):5-13. In some embodiments, a peptide nucleic acid or locked nucleic acid is used.

[00335] In certain embodiments of the invention a targeting moiety comprises a peptide. In some embodiments a peptide that binds to a target molecule of interest is identified using a display technology such as phage display, ribosome display, yeast display, etc.

[00336] Small molecules can be used as ligands. Methods for identifying such ligands are known in the art. For example in vitro screening of small molecule libraries, including combinatorial libraries, and computer-based screening, e.g., to identify small organic compounds that bind to concave surfaces (pockets) of proteins, can identify small molecule ligands for numerous proteins of interest (Huang, Z., Pharm. & Ther. 86: 201-215, 2000).

[00337] In certain embodiments of the invention targeting moieties are not proteins or molecules that are typically used as carriers and conjugated to antigens for the purpose of raising antibodies. Examples are carrier proteins or molecules such as bovine serum albumin, keyhole limpet hemocyanin, bovine gamma globulin, and diphtheria toxin. In certain embodiments of the invention the targeting moiety is not an Fc portion of an immunoglobulin molecule. In some embodiments, a targeting moiety is part of a complex comprising one or more additional moieties to which it is covalently or noncovalently attached.

[00338] In various embodiments of the invention a target molecule can be any molecule produced by a cell (including any forms expressed on the cell surface or modified forms thereof resulting at least in part from extracellular modification). In some embodiments a target molecule is an extracellular substance present in or on a tissue. In some embodiments, a target molecule is characteristic of a particular diseased or physiological state or characteristic of one or more cell type(s) or tissue type(s). A target molecule is often a molecule at least partly present at the cell surface (e.g., a transmembrane or otherwise membrane-attached protein) so that at least a portion of the molecule is accessible to binding by an extracellular binding agent such as an antibody. A target molecule may, but need not be, cell type specific. For example, a cell type specific target molecule is often a protein, peptide, mRNA, lipid, or carbohydrate that is present at a higher level on or in a particular cell type or cell type(s) than on or in many other cell types. In some instances a cell type specific target molecule is present at detectable levels only on or in a particular cell type of interest. However, it will be appreciated that a useful cell type specific target molecule need not be absolutely specific for the cell type of interest in order to be considered cell type specific. In some embodiments, a cell type specific target molecule for a particular cell type is expressed at levels at least 3 fold greater in that cell type than in a reference population of cells which may consist, for example, of a mixture containing cells from a plurality (e.g., 5-10 or more) of different tissues or organs in approximately equal amounts. In some embodiments, the cell type specific target molecule is present at levels at least 4-5 fold, between 5-10 fold, or more than 10-fold greater than its average expression in a reference population. In some embodiments, detection or measurement of a cell type specific target

molecule allows one of ordinary skill in the art to distinguish a cell type or types of interest from cells of many, most, or all other types. In general, the presence and/or abundance of most target molecules may be determined using one or more standard techniques such as Northern blotting, in situ hybridization, RT-PCR, sequencing, immunological methods such as immunoblotting, immunodetection (e.g., by immunohistochemistry), or fluorescence detection following staining with fluorescently labeled antibodies (e.g., using FACS), oligonucleotide or cDNA microarray or membrane array, protein microarray analysis, mass spectrometry, etc.

[00339] In some embodiments, a target molecule is a channel, transporter, receptor, or other molecule at least in part exposed at the cell surface. In some embodiments a target molecule is an anion transporter or water channel (e.g., an aquaporin protein).

[00340] In some embodiments, the target molecule is a protein at least in part exposed at the surface of red blood cells, such as a glycophorin (e.g., glycophorin A, B, C, or D) or band 3.

[00341] In some embodiments, the target molecule is a protein at least in part exposed at the surface of endothelial cells. In some embodiments, the target molecule is present at the surface of normal, healthy vasculature. In some embodiments, the target molecule is present at the surface of activated endothelial cells. In some embodiments, the target molecule is present at the surface of activated endothelial cells but not at the surface of non-activated endothelial cells. In some embodiments a target molecule is a molecule whose expression or exposure is induced by a stimulus such as injury or inflammation. In some embodiments, a target molecule would be recognized as "non-self" by a recipient receiving a transplant containing cells that express the target molecule. In some embodiments, the target molecule is a carbohydrate xenoantigen to which antibodies are commonly found in human beings. In some embodiments the carbohydrate comprises a blood group antigen. In some embodiments the carbohydrate comprises a xenoantigen, e.g., an alpha-gal epitope (Galalpha1-3Galbeta1-(3)4GlcNAc-R) (see, e.g., Macher BA and Galili U. The Galalpha1, 3Galbeta1, 4GlcNAc-R (alpha-Gal) epitope: a carbohydrate of unique evolution and clinical relevance. Biochim Biophys Acta. 1780(2):75-88 (2008).

[00342] In some embodiments of the invention, a compstatin analog comprises both a targeting moiety and a CRM.

[00343] In some embodiments, a targeted compstatin analog comprises multiple targeting moieties, which can be the same or different. Different targeting moieties may bind to the same target molecule or to different target molecules. The invention provides a targeted

compstatin analog that is multivalent with respect to the targeting moiety, the compstatin analog, or both.

[00344] In general, the invention encompasses any method of producing a compound comprising a compstatin analog moiety and a targeting moiety, and the resulting compounds. In some embodiments, a targeted compstatin analog may be produced using methods generally similar to those described in Section VI, wherein a targeting moiety is used instead of, or in addition to, a cell-reactive moiety. In some embodiments, a targeted compstatin analog comprising a peptide as a targeting moiety is synthesized as a polypeptide chain comprising a compstatin analog moiety and a peptide targeting moiety. Optionally, the polypeptide chain comprises one or more spacer peptides between the compstatin analog moiety and the targeting moiety.

[00345] In some embodiments, a targeted compstatin analog has a molar activity of at least about 10%, 20%, or 30%, e.g., between 30% and 40%, between 30% and 50%, between 30% and 60%, between 30% and 70%, between 30% and 80%, between 30% and 90%, or more, of the activity of a corresponding compstatin analog having the same amino acid sequence (and, if applicable, one or more blocking moiet(ies)) but not comprising a targeting moiety. In some embodiments wherein a targeted compstatin analog comprises multiple compstatin analog moieties, the molar activity of the targeted compstatin analog is at least about 10%, 20%, or 30%, e.g., between 30% and 40%, between 30% and 50%, between 30% and 60%, between 30% and 70%, between 30% and 80%, between 30% and 90%, or more, of the sum of the activities of said compstatin analog moieties. Compositions and methods for making targeted compstatin analogs, and intermediates in the synthesis, are aspects of the invention.

[00346] **VIII. Uses**

[00347] Cell-reactive, long-acting, or targeted compstatin analogs have a wide variety of uses. Without limiting the invention in any way, certain uses of cell-reactive, long-acting, or targeted compstatin analogs, and related aspects of the invention, are described herein. In some embodiments, a cell-reactive, long-acting, or targeted compstatin analog is administered to a subject suffering from or at risk of complement-mediated damage to an organ, tissue, or cells. In some embodiments, a cell-reactive compstatin analog is contacted with an organ, tissue, or cells ex vivo and become covalently attached thereto. The organ, tissue, or cells are introduced into a subject and are protected from damage that would otherwise be caused by the recipient's complement system.

[00348] Compstatin analogs that do not bind covalently to cells can be used for purposes described herein. For example, a compstatin analog modified with a moiety that increases the

lifetime of the compound in the body and/or a compstatin analog comprising a moiety that targets the compstatin analog to a cell type or location susceptible to complement activation can be used, and the invention encompasses such uses. In some embodiments, a long-acting compstatin analog is used. In some embodiments a compstatin analog comprising a targeting moiety is used. In some embodiments, a compstatin analog comprising both a moiety that extends the lifetime of the compound in the body and a targeting moiety is used. Where the discussion below refers to a cell-reactive compstatin analog, the invention provides analogous compositions and methods relating to targeted compstatin analogs and (at least in those aspects pertaining to administration of a compstatin analog to a subject) embodiments in which a compstatin analog that does not comprise a targeting moiety or a cell-reactive moiety, optionally a long-acting compstatin analog, is used instead of, or in addition to, a cell-reactive compstatin analog.

[00349] Certain uses of interest include: (1) protecting red blood cells (RBCs) from compement-mediated damage in individuals with disorders such as paroxysmal nocturnal hemoglobinuria or atypical hemolytic uremic syndrome or other disorders characterized by complement-mediated RBC lysis; (2) protecting transplanted organs, tissues, and cells from complement-mediated damage; (3) reducing ischemia/reperfusion (I/R) injury (e.g., in individuals suffering from trauma, vascular obstruction, myocardial infarction, or other situations in which I/R injury may occur); and (4) protecting various body structures (e.g., the retina) or membranes (e.g., synovial membrane) that may be exposed to complement components from complement mediated damage in any of a variety of different complement-mediated disorders. The beneficial effects of inhibiting complement activation at the surface of cells or other body structures are not limited to those resulting directly from protection of the cells or structures themselves against direct complement-mediated damage (e.g., preventing cell lysis). For example, inhibiting complement activation using a cell-reactive compstatin analog may reduce the generation of anaphylotoxins and resulting influx/activation of neutrophils and other pro-inflammatory events and/or reduce potentially damaging release of intracellular contents, thereby potentially having beneficial effects on remote organ systems or throughout the body.

[00350] A. Blood cell protection

[00351] In some embodiments of the invention, a cell-reactive compstatin analog, cell-targeted compstatin analog, and/or non-targeted compstatin analog (e.g., a long-acting non-targeted compstatin analog) is used to protect blood cells against complement-mediated damage. The blood cells may be any cellular component of the blood, e.g., red blood cells

(RBCs), white blood cells (WBCs), and/or platelets. In some embodiments, a cell-targeted compstatin analog is targeted to a target molecule exposed at the cell surface of RBCs such as a glycophorin or band 3. A variety of disorders are associated with complement-mediated damage to blood cells. Such disorders can result, for example, from deficiencies or defects in one or more of an individual's cellular or soluble CRPs, e.g., due to (a) mutation(s) in the gene(s) encoding such proteins; (b) mutation(s) in genes required for production or proper function of one or more CRPs, and/or (c) presence of autoantibodies to one or more CRPs. Complement-mediated RBC lysis can result from the presence of autoantibodies against RBC antigens which may arise due to a diverse set of causes (often being idiopathic). Individuals having such mutation(s) in genes encoding CRPs and/or having antibodies against CRPs or against their own RBCs are at increased risk of disorders involving complement-mediated RBC damage. Individuals who have had one or more episodes characteristic of a disorder are at increased risk of a recurrence.

[00352] Paroxysmal nocturnal hemoglobinuria (PNH) is a relatively rare disorder comprising an acquired hemolytic anemia characterized by complement-mediated intravascular hemolysis, hemoglobinuria, bone marrow failure, and thrombophilia (propensity to develop blood clots). It affects an estimated 16 individuals per million worldwide, occurs in both sexes, and can arise at any age, frequently striking young adults (Bessler, M. & Hiken, J., Hematology Am Soc Hematol Educ Program, 104-110 (2008); Hillmen, P. Hematology Am Soc Hematol Educ Program, 116-123 (2008)). PNH is a chronic and debilitating disease punctuated by acute hemolytic episodes and results in significant morbidities and reduced life expectancy. In addition to anemia, many patients experience abdominal pain, dysphagia, erectile dysfunction, and pulmonary hypertension, and are at increased risk of renal failure and thromboembolic events.

[00353] PNH was first described as a distinct entity in the 1800s, but it was only in the 1950s, with discovery of the alternative pathway of complement activation, that the cause of hemolysis in PNH was firmly established (Parker CJ. Paroxysmal nocturnal hemoglobinuria: an historical overview. Hematology Am Soc Hematol Educ Program. 93-103 (2008)). CD55 and CD59 are normally attached to the cell membrane via glycosyl phosphatidylinositol (GPI) anchors (glycolipid structures that anchor certain proteins to the plasma membrane). PNH arises as a consequence of nonmalignant clonal expansion of hematopoietic stem cell(s) that have acquired a somatic mutation in the PIGA gene, which encodes a protein involved in synthesis of GPI anchors (Takeda J, et al. Deficiency of the GPI anchor caused by a somatic mutation of the PIG-A gene in paroxysmal nocturnal hemoglobinuria. Cell. 73:703-711

(1993)). Progeny of such stem cells are deficient in GPI-anchored proteins, including CD55 and CD59. This defect renders these cells susceptible to complement-mediated RBC lysis. Flow cytometric analysis using antibodies to GPI-anchored proteins is often used for diagnosis. It detects deficiency of GPI-anchored proteins at the cell surface and allows determination of the degree of deficiency and the proportion of affected cells (Brodsky RA. Advances in the diagnosis and therapy of paroxysmal nocturnal hemoglobinuria. Blood Rev. 22(2):65-74 (2008). PNH type III RBCs are completely deficient in GPI-linked proteins and are highly sensitive to complement whereas PNH type II RBCs have a partial deficiency and are less sensitive. FLAER is a fluorescently labeled inactive variant of proaerolysin (a bacterial toxin that binds GPI anchors) and is increasingly used together with flow cytometry for diagnosis of PNH. Lack of binding of FLAER to granulocytes is sufficient for diagnosis of PNH. In some embodiments, a cell-reactive compstatin analog protects PNH RBCs from deposition of C3b.

[00354] In some embodiments, a cell-reactive, long-acting, or targeted compstatin analog is administered to a subject suffering from atypical hemolytic syndrome (aHUS). aHUS is a chronic disorder characterized by microangiopathic hemolytic anemia, thrombocytopenia, and acute renal failureand is caused by inappropriate complement activation, often due to mutations in genes encoding complement regulatory proteins (Warwicker, P., et al.. Kidney Int 53, 836-844 (1998); Kavanagh, D. & Goodship, T. Pediatr Nephrol 25, 2431-2442 (2010). Mutations in the complement factor H (CFH) gene are the most common genetic abnormality in patients with aHUS, and 60–70% of these patients die or reach end stage renal failure within one year after disease onset (Kavanagh & Goodship, supra.) Mutations in factor I, factor B, C3, factor H-related proteins 1-5, and thrombomodulin have also been described. Other causes of aHUS include autoantibodies against complement regulatory proteins such as CFH. In some embodiments, a cell-reactive, long-acting, or targeted compstatin analog is administered to a subject that has been identified as having a mutation in factor I, factor B, C3, factor H-related proteins 1-5, or thrombomodulin or has been identified as having antibodies against a complement regulatory protein, e.g., CFH.

[00355] Complement-mediated hemolysis occurs in a diverse group of other conditions including autoimmune hemolytic anemias that involve antibodies that bind to RBCs and lead to complement-mediated hemolysis. For example, such hemolysis can occur in primary chronic cold agglutinin disease and certain reactions to drugs and other foreign substances (Berentsen, S., et al., Hematology 12, 361-370 (2007); Rosse, W.F., Hillmen, P. & Schreiber, A.D. Hematology Am Soc Hematol Educ Program, 48-62 (2004)). In some embodiments of

the invention a cell-reactive compstatin analog is administered to a subject suffering from or at risk of chronic cold agglutinin disease. In another embodiment, a cell-reactive compstatin analog is used to treat a subject suffering from or at risk of the HELLP syndrome, which is defined by the existence of hemolysis, elevated liver enzymes, and low platelet count and is associated with mutations in complement regulatory protein(s) in at least some subjects (Fakhouri, F., et al., 112: 4542-4545 (2008)).

[00356] In other embodiments, cell-reactive compstatin analogs are used to protect RBCs or other cellular components of blood to be transfused into a subject. Certain examples of such uses are discussed further in below. As noted above, targeted and/or long-acting compstatin analogs can be used in the above methods for inhibiting complement-mediated hemolysis and/or RBC damage. In some embodiments, a long-acting compstatin analog comprising a (CH2CH2O) moiety is used to treat PNH or aHUS.

[00357] B. Transplantation

[00358] Transplantation is a therapeutic approach of increasing importance, providing a means to replace organs and tissues that have been damaged through trauma, disease, or other conditions. Kidneys, liver, lungs, pancreas, and heart are among the organs that can be successfully transplanted. Tissues that are frequently transplanted include bones, cartilage, tendons, cornea, skin, heart valves, and blood vessels. Pancreatic islet or islet cell transplantation is a promising approach for treatment of diabetes, e.g., type I diabetes. For purposes of the invention, an organ, tissue, or cell (or population of cells) that is be transplanted, is being transplanted, or has been transplanted may be referred to as a "graft". For purposes hereof, a blood transfusion is considered a "graft".

[00359] Transplantation subjects the graft to a variety of damaging events and stimuli that can contribute to graft dysfunction and, potentially, failure. For example, ischemia-reperfusion (I/R) injury is a common and significant cause of morbidity and mortality in the case of many grafts (particularly solid organs) and can be a major determinant of likelihood of graft survival. Transplant rejection is one of the major risks associated with transplants between genetically different individuals and can lead to graft failure and a need to remove the graft from the recipient.

[00360] In some embodiments of the invention, a cell-reactive compstatin analog, cell-targeted compstatin analog, and/or a long-acting compstatin analog is used to protect a graft from complement-mediated damage. A cell-reactive compstatin analog reacts with cells of the graft, becomes covalently attached thereto, and inhibits complement activation. A cell-targeted compstatin analog binds to a target molecule in the graft (e.g., expressed by

endothelial cells or other cells in the graft) and inhibits complement activation. A target molecule may be, e.g., is a molecule whose expression is induced or stimulated by a stimulus such as injury or inflammation, molecule that would be recognized as "non-self" by the recipient, a carbohydrate xenoantigen to which antibodies are commonly found in human beings such as a blood group antigen or a xenoantigen, e.g., a molecule comprising an alpha-gal epitope. In some embodiments, a reduction in complement activation can be demonstrated by a reduction in average C4d deposition in blood vessels of grafts that have been contacted with a compstatin analog, e.g., a cell-reactive compstatin analog, as compared with the average level of C4d deposition in grafts that have not been contacted with a ccompstatin analog (e.g., in subjects who are matched with respect to the grafts and other therapy that they receive).

[00361] A graft can be contacted with a cell-reactive, long-acting, or targeted compstatin analog prior to, during, and/or after being transplanted, in various embodiments of the invention. For example, prior to transplantation a graft removed from a donor can be contacted with a liquid comprising a cell-reactive, long-acting, or targeted compstatin analog. For example, the graft can be bathed in and/or perfused with the solution. In another embodiment, a cell-reactive, long-acting, or targeted compstatin analog is administered to a donor prior to removal of the graft. In some embodiments, a cell-reactive, long-acting, or targeted compstatin analog is administered to a recipient during and/or after the introduction of the graft. In some embodiments, a cell-reactive compstatin, long-acting, or targeted analog is delivered locally to the transplanted graft. In some embodiments a cell-reactive compstatin analog is administered systemically, e.g., intravenously.

[00362] The invention provides a composition comprising: (a) an isolated graft; and (b) a cell-reactive, long-acting, or targeted compstatin analog. In some embodiments the composition further comprises a liquid solution suitable for contacting (e.g., suitable for rinsing, washing, bathing, perfusing, maintaining, or storing) a graft (e.g., an organ) such as an isolated graft that has been removed from a donor and is awaiting transplantation to a recipient. In some embodiments the invention provides a composition comprising: (a) a liquid solution suitable for contacting a graft (e.g., an organ); and (b) a cell-reactive, long-acting, or targeted compstatin analog. The liquid solution can be any liquid solution that is physiologically acceptable to the graft (e.g., appropriate osmotic composition, non-cytotoxic) and medically acceptable in view of the subsequent introduction of the graft into the recipient (e.g., preferably sterile or at least reasonably free from microorganisms or other contaminants) and compatible with the cell-reactive compstatin analog (i.e., will not destroy

the reactivity of the compstatin analog) or compatible with the long-acting or targeted compstatin analog. In some embodiments, a solution is any solution own in the art for any such purposes. In some embodiments, a liquid solution is Marshall's or Hyperosmolar Citrate (Soltran®, Baxter Healthcare), University of Wisconsin (UW) solution (ViaSpan™, Bristol Myers Squibb), Histidine Tryptophan Ketoglutarate (HTK) solution (Custodial®, Kohler Medical Limited), EuroCollins (Fresenius), and Celsior® (Sangstat Medical), Polysol, IGL-1, or AQIX® RS-1. Of course other solutions, e.g., containing equivalent or similar ingredients in the same or different concentrations could be used within the scope of physiologically acceptable compositions. In some embodiments a solution does not contain ingredient(s) with which the cell-reactive compstatin analog would be expected to significantly react, and any solution may be modified or designed to lack such ingredients. In some embodiments, the cell-reactive compstatin analog is present in the graft-compatible solution at a concentration of, e.g., between 0.01 mg/ml and 100 mg/ml or may be added to the solution to achieve such concentration.

[00363] In some embodiments, the invention provides a kit comprising: (a) a cell-reactive, long-acting, or targeted compstatin analog; and (b) a graft-compatible solution or solid (e.g., powder) components thereof. The cell-reactive, long-acting, or targeted compstatin analog may be provided in solid form (e.g., powder) or at least in part dissolved in a solution. In some embodiments the cell-reactive, long-acting, or targeted compstatin analog and/or graft-compatible solution are provided in predetermined amounts, so that when combined, a solution of appropriate concentration for contacting a graft with the cell-reactive, long-acting, or targeted compstatin analog is produced. In many embodiments the cell-reactive, long-acting, or targeted compstatin analog and graft-compatible solution or solid (e.g., powder) components thereof are in separate containers within the kit. In some embodiments the cell-reactive compstatin analog and components of a graft-compatible solution are both provided in solid (e.g., powder) form, either in separate containers or mixed. In some embodiments the kit comprises instructions for use, e.g., instructions for adding a cell-reactive, long-acting, or targeted compstatin analog to a graft-compatible solution and/or instructions for contacting a graft with a cell-reactive compstatin analog. Optionally the kit contains a label approved by a government agency responsible for regulating products used in transplantation, cell therapy, and/or blood transfusion.

[00364] The invention further provides a method of covalently attaching a compstatin analog to an isolated graft comprising contacting the isolated graft with a cell-reactive compstatin analog. The invention further provides an isolated graft having a compstatin

analog covalently attached thereto. Typically the isolated graft has many molecules of compstatin analog attached thereto. In some embodiments, a graft is or comprises a solid organ such as a kidney, liver, lung, pancreas, or heart. In some embodiments, a graft is or comprises bone, cartilage, fascia, tendon, ligament, cornea, sclera, pericardium, skin, heart valve, blood vessel, amniotic membrane, or dura mater. In some embodiments, a graft comprises multiple organs such as a heart-lung or pancreas-kidney graft. In some embodiments, a graft comprises less than a complete organ or tissue. For example, a graft may contain a portion of an organ or tissue, e.g., a liver lobe, section of blood vessel, skin flap, or heart valve. In some embodiments, a graft comprises a preparation comprising isolated cells or tissue fragments that have been isolated from their tissue of origin but retain at least some tissue architecture, e.g., pancreatic islets. In some embodiments, a preparation comprises isolated cells that are not attached to each other via connective tissue, e.g., hematopoietic stem cells or progenitor cells derived from peripheral and/or cord blood, or whole blood or any cell-containing blood product such as red blood cells (RBCs) or platelets. In some embodiments a graft is obtained from a deceased donor (e.g., a "donation after brain death" (DBD) donor or "donation after cardiac death" donor). In some embodiments, depending on the particular type of graft, a graft is obtained from a living donor. For example, kidneys, liver sections, blood cells, are among the types of grafts that can often be obtained from a living donor without undue risk to the donor and consistent with sound medical practice.

[00365] In some embodiments, a graft is a xenograft (i.e., the donor and recipient are of different species). In some embodiments a graft is an autograft (i.e., a graft from one part of the body to another part of the body in the same individual). In some embodiments, a graft is an isograft (i.e., the donor and recipient are genetically identical). In most embodiments, the graft is an allograft (i.e., the donor and receipient are genetically non-identical members of the same species). In the case of an allograft, the donor and recipient may or may not be genetically related (e.g., family members). Typically, the donor and recipient have compatible blood groups (at least ABO compatibility and optionally Rh, Kell and/or other blood cell antigen compatibility). The recipient's blood may have been screened for alloantibodies to the graft and/or the recipient and donor since the presence of such antibodies can lead to hyperacute rejection (i.e., rejection beginning almost immediately, e.g., within several minutes after the graft comes into contact with the recipient's blood). A complement-dependent cytoxicity (CDC) assay can be used to screen a subject's serum for anti-HLA antibodies. The serum is incubated with a panel of lymphcytes of known HLA phenotype. If

the serum contains antibodies against HLA molecules on the target cells, cell death due to complement-mediated lysis occurs. Using a selected panel of target cells allows one to assign specificity to the detected antibody. Other techniques useful for determining the presence or absence anti-HLA antibodies and, optionally, determining their HLA specificity, include ELISA assays, flow cytometry assays, microbead array technology (e.g., Luminex technology). The methodology for performing these assays is well known, and a variety of kits for performing them are commercially available.

[00366] In some embodiments a cell-reactive, long-acting, or targeted compstatin analog inhibits complement-mediated rejection. For example, in some embodiments a cell-reactive, long-acting, or targeted compstatin analog inhibits hyperacute rejection. Hyperacute rejection is caused at least in part by antibody-mediated activation of the recipient's complement system via the classical pathway and resulting MAC deposition on the graft. It typically results from the presence in the recipient of pre-existing antibodies that react with the graft. While it is desirable to attempt to avoid hyperacute rejection by appropriate matching prior to transplantation, it may not always possible to do so due, e.g., to time and/or resource constraints. Furthermore, some recipients (e.g., multiply transfused individuals, individuals who have previously received transplants, women who have had multiple pregnancies) may already have so many pre-formed antibodies, potentially including antibodies to antigens that are not typically tested for, that it can be difficult or perhaps almost impossible to obtain with confidence a compatible graft in a timely manner. Such individuals are at increased risk of hyperacute rejection.

[00367] In some embodiments, a cell-reactive, long-acting, or targeted compstatin analog inhibits acute rejection or graft failure. As used herein, "acute rejection" refers to rejection occurring between at least 24 hours, typically at least several days to a week, after a transplant, up to 6 months after the transplant. Acute antibody-mediated rejection (AMR) often involves an acute rise in donor-specific alloantibody (DSA) in the first few weeks after transplantation. Without wishing to be bound by any theory, it is possible that pre-existing plasma cells and/or the conversion of memory B cells to new plasma cells play a role in the increased DSA production. Such antibodies can result in complement-mediated damage to the graft, which can be inhibited by contacting the graft with a cell-reactive compstatin analog. Without wishing to be bound by any theory, inhibiting complement activation at the graft may reduce leukocyte (e.g., neutrophil) infiltration, another contributor to acute graft failure.

[00368] In some embodiments, a cell-reactive, long-acting, or targeted compstatin analog inhibits complement-mediated I/R injury to a graft. As discussed further below, I/R injury can occur upon reperfusion of tissue whose blood supply has been temporarily disrupted, as occurs in transplanted organs. Reducing I/R injury would reduce the likelihood of acute graft dysfunction or reduce its severity, and reduce the likelihood of acute graft failure.

[00369] In some embodiments, a cell-reactive, long-acting, or targeted compstatin analog inhibits chronic rejection and/or chronic graft failure. As used herein, "chronic rejection or graft failure" refers to rejection or failure occurring at least 6 months post-transplant, e.g., between 6 months and 1, 2, 3, 4, 5 years, or more post-transplant, often after months to years of good graft function. It is caused by a chronic inflammatory and immune response against the graft. For purposes hereof, chronic rejection can include chronic allograft vasculopathy, a term used to refer to fibrosis of the internal blood vessels of the transplanted tissue. As immunosuppressive regimens have reduced the incidence of acute rejection, chronic rejection is becoming more prominent as a cause of graft dysfunction and failure. There is increasing evidence that B-cell production of alloantibody is an important element in the genesis of chronic rejection and graft failure (Kwun J. and Knechtle SJ, Transplantation, 88(8):955-61 (2009). Earlier damage to the graft may be a contributing factor leading to chronic processes such as fibrosis that can ultimately lead to chronic rejection. Thus, inhibiting such earlier damage using a cell-reactive compstatin analog may delay and/or reduce the likelihood or severity of chronic graft rejection.

[00370] In some embodiments, a long-acting compstatin analog is administered to a graft recipient to inhibit graft rejection and/or graft failure.

[00371] C. Ischemia/Reperfusion Injury

[00372] Ischemia-reperfusion (I/R) injury is an important cause of tissue damage following trauma and in other conditions associated with temporary disruption of blood flow such as myocardial infarction, stroke, severe infection, vascular disease, aneurysm repair, cardiopulmonary bypass, and transplantation.

[00373] In the setting of trauma, systemic hypoxemia, hypotension, and local interruption of the blood supply resulting from contusions, compartment syndrome, and vascular injuries cause ischemia that damages metabolically active tissues. Restoration of the blood supply triggers an intense systemic inflammatory reaction that is often more harmful than the ischemia itself. Once the ischemic region is reperfused, factors that are produced and released locally enter the circulatory system and reach remote locations, sometimes causing significant damage to organs not affected by the original ischemic insult, such as the lungs

and intestine, leading to single and multiple organ dysfunction. Complement activation occurs soon after reperfusion and is a key mediator of post-ischemic damage, both directly and through its chemoattractive and stimulatory effects on neutrophils. All three major complement pathways are activated and, acting cooperatively or independently, are involved in I/R related adverse events affecting numerous organ systems. In some embodiments of the invention, a cell-reactive, long-acting, or targeted compstatin analog is administered to a subject who has recently (e.g., within the preceding 2, 4, 8, 12, 24, or 48 hours) experienced trauma, e.g., trauma that puts the subject at risk of I/R injury, e.g., due to systemic hypoxemia, hypotension, and/or local interruption of the blood supply. In some embodiments the cell-reactive compstatin analog may be administered intravascularly, optionally into a blood vessel that supplies an injured body part or directly to the body part. In some embodiments, the subject suffers from spinal cord injury, traumatic brain injury, burn, and/or hemorrhagic shock.

[00374] In some embodiments, a cell-reactive, long-acting, or targeted compstatin analog is administered to a subject prior to, during, or after a surgical procedure, e.g., a surgical procedure that is expected to temporarily disrupt blood flow to a tissue, organ, or portion of the body. Examples of such procedures include cardiopulmonary bypass, angioplasty, heart valve repair/replacement, aneurysm repair, or other vascular surgeries. The cell-reactive compstatin analog may be administered prior to, after, and/or during an overlapping time period with the surgical procedure.

[00375] In some embodiments, a cell-reactive, long-acting, or targeted compstatin analog is administered to a subject who has suffered an MI, thromboembolic stroke, deep vein thrombosis, or pulmonary embolism. The cell-reactive compstatin analog may be administered in combination with a thrombolytic agent such as tissue plasminogen activator (tPA) (e.g., alteplase (Activase), reteplase (Retavase), tenecteplase (TNKase)), anistreplase (Eminase), streptokinase (Kabikinase, Streptase), or urokinase (Abbokinase). The cell-reactive, long-acting, or targeted compstatin analog may be administered prior to, after, and/or during an overlapping time period with the thrombolytic agent.

[00376] In some embodiments, a cell-reactive, long-acting, or targeted compstatin analog is administered to a subject to treat I/R injury.

[00377] <u>D. Other Complement-Mediated Disorders</u>

[00378] In some embodiments, a cell-reactive, long-acting, or targeted compstatin analog is introduced into the eye for treatment of an eye disorder such as age-related macular degeneration (AMD), diabetic retinopathy, glaucoma, or uveitis. For example, a cell-reactive

compstatin analog may be introduced into the vitreous cavity (e.g., by intravitreal injection), for treatment of a subject at suffering from or at risk of AMD. In some embodiments a cell-reactive, long-acting, or targeted compstatin analog is introduced into the anterior chamber, e.g., to treat anterior uveitis.

[00379] In some embodiments a cell-reactive, long-acting, or targeted compstatin analog is used to treat a subject suffering from or at risk of an autoimmune disease, e.g., an autoimmune disease mediated at least in part by antibodies against one or more self antigens.

[00380] Cell-reactive, long-acting, or targeted compstatin analogs may be introduced into the synovial cavity, e.g., in a subject suffering from arthritis (e.g., rheumatoid arthritis). Of course they may be administered systemically in addition or

[00381] In some embodiments, a cell-reactive, long-acting, or targeted compstatin analog is used to treat a subject suffering from or at risk of an intracerebral hemorrhage.

[00382] In some embodiments a cell-reactive, long-acting, or targeted compstatin analog is used to treat a subject suffering from or at risk of myasthenia gravis.

[00383] In some embodiments a cell-reactive, long-acting, or targeted compstatin analog is used to treat a subject suffering from or at risk of neuromyelitis optica (NMO),

[00384] In some embodiments a cell-reactive, long-acting, or targeted compstatin analog is used to treat a subject suffering from or at risk of membranoproliferative glomerulitis (MPGN), e.g., MPGN type I, MPGN type II, or MPGH type III.

[00385] In some embodiments a cell-reactive, long-acting, or targeted compstatin analog is used to treat a subject suffering from or at risk of a neurodegenerative disease. In some embodiments a cell-reactive, long-acting, or targeted compstatin analog is used to treat a subject suffering from neuropathic pain or at risk of developing neuropathic pain. In some embodiments a cell-reactive, long-acting, or targeted compstatin analog is used to treat a subject suffering from or at risk of rhinosinusitis or nasal polyposis. In some embodiments a cell-reactive, long-acting, or targeted compstatin analog is used to treat a subject suffering from or at risk of cancer. In some embodiments a cell-reactive, long-acting, or targeted compstatin analog is used to treat a subject suffering from or at risk of sepsis. . In some embodiments a cell-reactive, long-acting, or targeted compstatin analog is used to treat a subject suffering from or at risk of adult respiratory distress syndrome.

[00386] In some embodiments a cell-reactive, long-acting, or targeted compstatin analog is used to treat a subject suffering from or at risk of anaphylaxis or infusion reaction. For example, in some embodiments a subject may be pretreated prior to, during, or after receiving a drug or a vehicle that may cause anaphylaxis or infusion reaction. In some embodiments a

subject at risk of or suffering from anaphylaxis from a food (e.g., peanut, shellfish, or other food allergens), insect sting (e.g., bee, wasp), is treated with a cell-reactive, long-acting, or targeted compstatin analog.

[00387] The cell-reactive long-acting, or targeted compstatin analog may be administered locally or systemically, in various embodiments of the invention.

[00388] In some embodiments, a cell-reactive, long-acting, or targeted compstatin analog is used to treat a respiratory disease, e.g., asthma or chronic obstructive pulmonary disease (COPD). The cell-reactive, long-acting, or targeted compstatin analog may, for example, be administered to the respiratory tract by inhalation, e.g., as a dry powder or via nebulization, or may be administered by injection, e.g., intravenously, intramuscularly, or subcutenously, in various embodiments. In some embodiments, a cell-reactive, long-acting, or targeted compstatin analog is used to treat severe asthma, e.g., asthma that is not sufficiently controlled by bronchodilators and/or inhaled corticosteroids.

[00389] In some aspects, methods of treating a complement-mediated disorder, e.g., a chronic complement-mediated disorder, are provided, the methods comprising administering a long-acting complement inhibitor to a subject in need of treatment for the disorder. The long-acting compstatin analog may be any long-acting compstatin analog described herein, in various embodiments. In some aspects, methods of treating a Th17-associated disorder are provided, the methods comprising administering a long-acting complement inhibitor to a subject in need of treatment for the disorder.

[00390] In some aspects, a "chronic disorder" is a disorder that persists for at least 3 months and/or is accepted in the art as being a chronic disorder. In many embodiments, a chronic disorder persists for at least 6 months, e.g., at least 1 year, or more, e.g., indefinitely. One of ordinary skill in the art will appreciate that at least some manifestations of various chronic disorders may be intermittent and/or may wax and wane in severity over time. A chronic disorder may be progressive, e.g., having a tendency to become more severe or affect larger areas over time. A number of chronic complement-mediated disorders are discussed herein. A chronic complement-mediated disorder may be any chronic disorder in which complement activation (e.g., excessive or inappropriate complement activation) is involved, e.g., as a contributing and/or at least partially causative factor. For convenience, disorders are sometimes grouped by reference to an organ or system that is often particularly affected in subjects suffering from the disorder. It will be appreciated that a number of disorders can affect multiple organs or systems, and such classification(s) are in no way limiting. Furthermore, a number of manifestations (e.g., symptoms) may occur in subjects suffering

from any of a number of different disorders. Non-limiting information regarding disorders of interest herein may be found, e.g., in standard textbooks of internal medicine such as Cecil Textbook of Medicine (e.g., 23rd edition), Harrison's Principles of Internal Medicine (e.g., 17th edition), and/or standard textbooks focusing on particular areas of medicine, particular body systems or organs, and/or particular disorders.

[00391] In some embodiments, a chronic complement-mediated disorder is a Th2-associated disorder. As used herein, a Th2-associated disorder is a disorder characterized by an excessive number and/or excessive or inappropriate activity of CD4+ helper T cells of the Th2 subtype ("Th2 cells") in the body or a portion thereof, e.g., in at least one tissue, organ, or structure. For example, there may be a predominance of Th2 cells relative to CD4+ helper T cells of the Th1 subtype ("Th1 cells") e.g., in at least one tissue, organ, or structure affected by a disorder. As known in the art, Th2 cells typically secrete characteristic cytokines such as interleukin-4 (IL-4), interleukin-5 (IL-5), and interleukin-13 (IL-13), while Th1 cells typically secrete interferon-γ (IFN- γ) and tumor necrosis factor β (TNF β). In some embodiments, a Th2-associated disorder is characterized by excessive production and/or amount of IL-4, IL-5, and/or IL-13, e.g., relative to IFN-γ and/or TNF β e.g., in at least some at least one tissue, organ, or structure

[00392] In some embodiments, a chronic complement-mediated disorder is a Th17-associated disorder. In some aspects, as described in further detail in PCT/US2012/043845, filed June 22, 2012, entitled "Methods of Treating Chronic Disorders with Complement Inhibitors", complement activation and Th17 cells participate in a cycle that involves dendritic cells and antibodies and that contributes to maintenance of a pathologic immunologic microenvironment underlying a range of disorders. Without wishing to be bound by any theory, the pathologic immunologic microenvironment, once established, is self-sustaining and contributes to cell and tissue injury. In some aspects, long-acting compstatin analogs are of use to treat Th17-associated disorders.

[00393] As used herein, a Th17-associated disorder is a disorder characterized by an excessive number and/or excessive or inappropriate activity of CD4+ helper T cells of the Th17 subtype ("Th17 cells") in the body or a portion thereof, e.g., in at least one tissue, organ, or structure. For example, there may be a predominance of Th17 cells relative to Th1 and/or Th2 cells, e.g., in at least one tissue, organ, or structure affected by a disorder. In some embodiments a predominance of Th17 cells is a relative predominance, e.g., the ratio of Th17 cells to Th1 cells and/or the ratio of Th17 cells to Th2 cells, is increased relative to

normal values. In some embodiments the ratio of Th17 cells to T regulatory cells ($CD4^+CD25^+$ regulatory T cells, also termed "Treg cells"), is increased relative to normal values. Formation of Th17 cells and/or activation of Th 17 cells is promoted by various cytokines, e.g., interleukin 6 (IL-6), interleukin 21 (IL-21), interleukin 23 (IL-23), and/or interleukin 1β (IL-1β). Formation of Th17 cells encompasses differentiation of precursor T cells, e.g., naïve CD4+ T cells, towards a Th17 phenotype and their maturation into functional Th17 cells. In some embodiments, formation of Th17 cells encompasses any aspect of development, proliferation (expansion), survival, and/or maturation of Th17 cells. In some embodiments, a Th17-associated disorder is characterized by excessive production and/or amount of IL-6, IL-21, IL-23, and/or IL-1β. Th17 cells typically secrete characteristic cytokines such as interleukin-17A (IL-17A), interleukin-17F (IL-17F), interleukin-21 (IL-21), and interleukin-22 (IL-22). In some embodiments, a Th17-associated disorder is characterized by excessive production and/or amount of a Th17 effector cytokine, e.g., IL-17A, IL-17F, IL-21, and/or IL-22. In some embodiments excessive production or amount of a cytokine is detectable in the blood. In some embodiments excessive production or amount of a cytokine is detectable locally, e.g., in at least one tissue, organ or structure. In some embodiments a Th17-associated disorder is associated with a decreased number of Tregs and/or decreased amount of a Treg-associated cytokine. In some embodiments a Th17 disorder is any chronic inflammatory disease, which term encompasses a range of ailments characterized by self-perpetuating immune insults to a variety of tissues and that seem to be dissociated from the initial insult that caused the ailment (which may be unknown). In some embodiments a Th17-associated disorder is any autoimmune disease. Many if not most "chronic inflammatory diseases" may in fact be auto-immune diseases. Examples of Th17-associated disorders include inflammatory skin diseases such as psoriasis and atopic dermatitis; systemic scleroderma and sclerosis; inflammatory bowel disease (IBD) (such as Crohn's disease and ulcerative colitis); Behcet's Disease; dermatomyositis; polymyositis; multiple sclerosis (MS); dermatitis; meningitis; encephalitis; uveitis; osteoarthritis; lupus nephritis; rheumatoid arthritis (RA), Sjorgen's syndrome, multiple sclerosis, vasculitis; central nervous system (CNS) inflammatory disorders, chronic hepatitis; chronic pancreatitis, glomerulonephritis; sarcoidosis; thyroiditis, pathologic immune responses to tissue/organ transplantation (e.g., transplant rejection); COPD, asthma, bronchiolitis, hypersensitivity pneumonitis, idiopathic pulmonary fibrosis (IPF), periodontitis, and gingivitis. In some embodiments a Th17 disease is a classically known auto-immmune disease such as Type I

diabetes or psoriasis. In some embodiments a Th17-associated disorder is age-related macular degeneration.

[00394] In some embodiments, a chronic complement-mediated disorder is an IgE-associated disorder. As used herein, an "IgE-associated disorder" is a disorder characterized by excessive and/or inappropriate production and/or amount of IgE, excessive or inappropriate activity of IgE producing cells (e.g., IgE producing B cells or plasma cells), and/or excessive and/or inappropriate activity of IgE responsive cells such as eosinophils or mast cells. In some embodiments, an IgE-associated disorder is characterized by elevated levels of total IgE and/or in some embodiments, allergen-specific IgE, in the plasma of a subject and/or locally.

[00395] In some embodiments, a chronic complement-mediated disorder is characterized by the presence of autoantibodies and/or immune complexes in the body, which may activate complement via, e.g., the classical pathway. Autoantibodies may, for example, bind to self antigens, e.g., on cells or tissues in the body. In some embodiments, autoantibodies bind to antigens in blood vessels, skin, nerves, muscle, connective tissue, heart, kidney, thyroid, etc. In some embodiments, a chronic complement-mediated disorder is not characterized by autoantibodies and/or immune complexes.

[00396] In some embodiments, a chronic complement-mediated disorder is a respiratory disorder. In some embodiments, a chronic respiratory disorder is asthma or chronic obstructive pulmonary disease (COPD). In some embodiments, a chronic respiratory disorder is pulmonary fibrosis (e.g., idiopathic pulmonary fibrosis), radiation-induced lung injury, allergic bronchopulmonary aspergillosis, hypersensitivity pneumonitis (also known as allergic alveolitis), eosinophilic pneumonia, interstitial pneumonia, sarcoid, Wegener's granulomatosis, or bronchiolitis obliterans. In some embodiments, the invention provides a method of treating a subject in need of treatment for a chronic respiratory disorder, e.g., asthma, COPD, pulmonary fibrosis, radiation-induced lung injury, allergic bronchopulmonary aspergillosis, hypersensitivity pneumonitis (also known as allergic alveolitis), eosinophilic pneumonia, interstitial pneumonia, sarcoid, Wegener's granulomatosis, or bronchiolitis obliterans, the method comprising administering a long-acting complement inhibitor to a subject in need of treatment for the disorder.

[00397] In some embodiments, a chronic complement-mediated disorder is allergic rhinitis, rhinosinusitis, or nasal polyposis. In some embodiments, the invention provides a method of treating a subject in need of treatment for allergic rhinitis, rhinosinusitis, or nasal

polyposis, the method comprising administering a long-acting complement inhibitor to a subject in need of treatment for the disorder.

[00398] In some embodiments, a chronic complement-mediated disorder is a disorder that affects the musculoskeletal system. Examples of such disorders include inflammatory joint conditions (e.g., arthritis such as rheumatoid arthritis or psoriatic arthritis, juvenile chronic arthritis, spondyloarthropathies Reiter's syndrome, gout). In some embodiments, a musculoskeletal system disorder results in symptoms such as pain, stiffness and/or limitation of motion of the affected body part(s). Inflammatory myopathies include dermatomyositis, polymyositis, and various others are disorders of chronic muscle inflammation of unknown etiology that result in muscle weakness. In some embodiments, a chronic complement-mediated disorder is myasthenia gravis. In some embodiments, the invention provides a method of treating any of the foregoing disorders affecting the musculoskeletal system, the method comprising administering a long-acting complement inhibitor to a subject in need of treatment for the disorder.

[00399] In some embodiments, a chronic complement-mediated disorder is a disorder that affects the integumentary system. Examples of such disorders include, e.g., atopic dermatitis, psoriasis, pemphigus, systemic lupus erythematosus, dermatomyositis, scleroderma, sclerodermatomyositis, Sjögren syndrome, and chronic urticaria. In some aspects, the invention provides a method of treating any of the foregoing disorders affecting the integumentary system, the method comprising administering a long-acting complement inhibitor to a subject in need of treatment for the disorder.

[00400] In some embodiments, a chronic complement-mediated disorder affects the nervous system, e.g., the central nervous system (CNS) and/or peripheral nervous system (PNS). Examples of such disorders include, e.g., multiple sclerosis, other chronic demyelinating diseases, amyotrophic lateral sclerosis, chronic pain, stroke, allergic neuritis, Huntington's disease, Alzheimer's disease, and Parkinson's disease. In some embodiments, the invention provides a method of treating any of the foregoing disorders affecting the nervous system, the method comprising administering a complement inhibitor according to a dosing schedule described herein to a subject in need of treatment for the disorder.

[00401] In some embodiments, a chronic complement-mediated disorder affects the circulatory system. For example, in some embodiments the disorder is a vasculitis or other disorder associated with vessel inflammation, e.g., blood vessel and/or lymph vessel inflammation. In some embodiments, a vasculitis is polyarteritis nodosa, Wegener's granulomatosis, giant cell arteritis, Churg-Strauss syndrome, microscopic polyangiitis,

Henoch-Schonlein purpura, Takayasu's arteritis, Kawasaki disease, or Behcet's disease. In some embodiments, a subject, e.g., a subject in need of treatment for vasculitis, is positive for antineutrophil cytoplasmic antibody (ANCA).

[00402] In some embodiments, a chronic complement-mediated disorder affects the gastrointestinal system. For example, the disorder may be inflammatory bowel disease, e.g., Crohn's disease or ulcerative colitis. In some embodiments, the invention provides a method of treating a chronic complement-mediated disorder that affects the gastrointestinal system, the method comprising administering a long-acting complement inhibitor to a subject in need of treatment for the disorder.

[00403] In some embodiments, a chronic complement-mediated disorder is a thyroiditis (e.g., Hashimoto's thryoiditis, Graves' disease, post-partum thryoiditis), myocarditis, hepatitis (e.g., hepatitis C), pancreatitis, glomerulonephritis (e.g., membranoproliferative glomerulonephritis or membranous glomerulonephritis), or panniculitis.

[00404] In some embodiments, the invention provides methods of treating a subject suffering from chronic pain, the methods comprising administering a long-acting complement inhibitor to a subject in need thereof. In some embodiments, a subject suffers from neuropathic pain. Neuropathic pain has been defined as pain initiated or caused by a primary lesion or dysfunction in the nervous system, in particular, pain arising as a direct consequence of a lesion or disease affecting the somatosensory system. For example, neuropathic pain may arise from lesions that involve the somatosensory pathways with damage to small fibres in peripheral nerves and/or to the spino-thalamocortical system in the CNS. In some embodiments, neuropathic pain arises from autoimmune disease (e.g., multiple sclerosis), metabolic disease (e.g., diabetes), infection (e.g., viral disease such as shingles or HIV), vascular disease (e.g., stroke), trauma (e.g., injury, surgery), or cancer. For example, neuropathic pain can be pain that persists after healing of an injury or after cessation of a stimulus of peripheral nerve endings or pain that arises due to damage to nerves. Exemplary conditions of or associated with neuropathic pain include painful diabetic neuropathy, post-herpetic neuralgia (e.g., pain persisting or recurring at the site of acute herpes zoster 3 or more months after the acute episode), trigeminal neuralgia, cancer related neuropathic pain, chemotherapy-associated neuropathic pain, HIV-related neuropathic pain (e.g., from HIV neuropathy), central/post-stroke neuropathic pain, neuropathy associated with back pain, e.g., low back pain (e.g., from radiculopathy such as spinal root compression, e.g., lumbar root compression, which compression may arise due to disc herniation), spinal stenosis, peripheral nerve injury pain, phantom limb pain, polyneuropathy, spinal cord injury related pain,

myelopathy, and multiple sclerosis. In certain embodiments of the invention a complement inhibitor is administered according to an inventive dosing schedule to treat neuropathic pain in a subject with one or more of the afore-mentioned conditions.

[00405] In some embodiments, a chronic complement-mediated disorder is a chronic eye disorder. In some embodiments, the chronic eye disorder is characterized by macular degeneration, choroidal neovascularization (CNV), retinal neovascularization (RNV), ocular inflammation, or any combination of the foregoing. Macular degeneration, CNV, RNV, and/or ocular inflammation may be a defining and/or diagnostic feature of the disorder. Exemplary disorders that are characterized by one or more of these features include, but are not limited to, macular degeneration related conditions, diabetic retinopathy, retinopathy of prematurity, proliferative vitreoretinopathy, uveitis, keratitis, conjunctivitis, and scleritis. Macular degeneration related conditions include, e.g., age-related macular degeneration (AMD). In some embodiments, a subject is in need of treatment for wet AMD. In some embodiments, a subject is in need of treatment for dry AMD. In some embodiments, a subject is in need of treatment for geographic atrophy (GA). In some embodiments, a subject is in need of treatment for ocular inflammation. Ocular inflammation can affect a large number of eye structures such as the conjunctiva (conjunctivitis), cornea (keratitis), episclera, sclera (scleritis), uveal tract, retina, vasculature, and/or optic nerve. Evidence of ocular inflammation can include the presence of inflammation-associated cells such as white blood cells (e.g., neutrophils, macrophages) in the eye, the presence of endogenous inflammatory mediator(s), one or more symptoms such as eye pain, redness, light sensitivity, blurred vision and floaters, etc. Uveitis is a general term that refers to inflammation in the uvea of the eye, e.g., in any of the structures of the uvea, including the iris, ciliary body or choroid. Specific types of uveitis include iritis, iridocyclitis, cyclitis, pars planitis and choroiditis. In some embodiments, a subject is in need of treatment for geographic atrophy (GA). In some embodiments, the chronic eye disorder is an eye disorder characterized by optic nerve damage (e.g., optic nerve degeneration), such as glaucoma.

[00406] As noted above, in some embodiments the chronic respiratory disease is asthma. Information regarding risk factors, epidemiology, pathogenesis, diagnosis, current management of asthma, etc., may be found, e.g., in "Expert Panel Report 3: Guidelines for the Diagnosis and Management of Asthma". National Heart Lung and Blood Institute. 2007. http://www.nhlbi.nih.gov/guidelines/asthma/asthgdln.pdf. ("NHLBI Guidelines"; www.nhlbi.nih.gov/guidelines/asthma/asthgdln.htm), Global Initiative for Asthma, Global Strategy for Asthma Management and Prevention 2010 "GINA Report") and/or standard

textbooks of internal medicine such as Cecil Textbook of Medicine (20th edition), Harrison's Principles of Internal Medicine (17th edition), and/or standard textbooks focusing on pulmonary medicine. Asthma is a chronic inflammatory disorder of the airways in which many cells and cellular elements play a role, such as, mast cells, eosinophils, T lymphocytes, macrophages, neutrophils, and epithelial cells Asthmatic individuals experience recurrent episodes associated with symptoms such as wheezing, breathlessness (also termed dyspnea or shortness of breath), chest tightness, and coughing. These episodes are usually associated with widespread but variable airflow obstruction that is often reversible, either spontaneously or with treatment. The inflammation also causes an associated increase in the existing bronchial hyperresponsiveness to a variety of stimuli. Airway hyperresponsiveness (an exaggerated bronchoconstrictor response to stimuli) is a typical feature of asthma. In general, airflow limitation results from bronchoconstriction and airway edema. Reversibility of airflow limitation may be incomplete in some patients with asthma. For example, airway remodeling can lead to fixed airway narrowing. Structural changes can include thickening of the sub-basement membrane, subepithelial fibrosis, airway smooth muscle hypertrophy and hyperplasia, blood vessel proliferation and dilation, and mucous gland hyperplasia, and hypersecretion.

[00407] Individuals with asthma may experience exacerbations, which are identified as events characterized by a change from the individual's previous status. Severe asthma exacerbations can be defined as events that require urgent action on the part of the individual and his/her physician to prevent a serious outcome, such as hospitalization or death from asthma. For example, a severe asthma exacerbation may require use of systemic corticosteroids (e.g., oral corticosteroids) in a subject whose asthma is usually well controlled without OCS or may require an increase in a stable maintenance dose. Moderate asthma exacerbations can be defined as events that are troublesome to the subject, and that prompt a need for a change in treatment, but that are not severe. These events are clinically identified by being outside the subject's usual range of day-to-day asthma variation.

[00408] Current medications for asthma are typically categorized into two general classes: long-term control medications ("controller medications") such as inhaled corticosteroids (ICS), oral corticosteroids (OCS), long-acting bronchodilators (LABAs), leukotriene modifiers (e.g., leukotriene receptor antagonists or leukotriene synthesis inhibitors, anti-IgE antibodies (omalizumab (Xolair®)), cromolyn and nedocromil, which are used to achieve and maintain control of persistent asthma and quick-relief medications such as short-acting bronchodilators (SABAs), which are used to treat acute symptoms and exacerbations. For

purposes of the present invention, these treatments may be referred to as "conventional therapy". Treatment of exacerbations may also include increasing the dose and/or intensity of controller medication therapy. For example, a course of OCS can be used to regain asthma control. Current guidelines mandate daily administration of controller medication or, in many cases, administration of multiple doses of controller medication each day for subjects with persistent asthma (with the exception of Xolair, which is administered every 2 or 4 weeks).

[00409] A subject is generally considered to have persistent asthma if the subject suffers from symptoms on average more than twice a week and/or typically uses a quick relief medication (e.g., SABA) more than twice a week for symptom control. "Asthma severity" can be classified based on the intensity of treatment required to control the subject's asthma once relevant comorbidities have been treated and inhaler technique and adherence have been optimized (see, e.g., GINA Report; Taylor, DR, Eur Respir J 2008; 32:545–554). The description of treatment intensity can be based on the medications and doses recommended in the stepwise treatment algorithm found in guidelines such as NHLBI Guidelines 2007, GINA Report, and their predecessors and/or in standard medical textbooks. For example, asthma can be classified as intermittent, mild, moderate, or severe as indicated in Table X, where "treatment" refers to treatment sufficient to achieve subject's best level of asthma control. (It will be understood that the categories of mild, moderate, and severe asthma in general imply persistent rather than intermittent asthma). One of ordinary skill in the art will appreciate that Table X is exemplary, and that not all of these medications will be available in all healthcare systems, which may affect the assessment of asthma severity in some environments. It will also be appreciated that other emerging or new approaches may affect the classification of mild/moderate asthma. However, the same principle, of mild asthma being defined by the ability to achieve good control using very low-intensity treatment and severe asthma being defined by the requirement for high-intensity treatment, can still be applied. Asthma severity can also or alternately be classified based on intrinsic intensity of the disease in the absence of treatment (see, e.g., NHBLI Guidelines 2007). Assessment can be made on the basis of current spirometry and the patient's recall of symptomsover the previous 2–4 weeks. Parameters of current impairment and future risk may be assessed and included in a determination of the level of asthma severity. In some embodiments, asthma severity is defined as shown in Figure 3.4(a), 3.4(b), 3.4(c) of the NHBLI Guidelines, for individuals 0-4, 5-11, or $\geq$ 12 years of age, respectively.

Table X: Treatment-based Asthma Classification

Asthma Classification	Treatment
Intermittent	SABA as needed (typically no more than twice a week)
Mild	Low-dose ICS or other low-intensity treatment (e.g., LTRA, cromolyn, nedocromil, theophylline)
Moderate	Low to moderate dose ICS and LABA or other extra treatment
Severe	High-intensity treatment (high-dose ICS and LABA ± oral corticosteroids and/or other extra treatment)

[00410] ''Asthma control'' refers to the extent to which the manifestations of asthma have been reduced or removed by treatment (whether pharmacological or non-pharmacological). Asthma control can be assessed based on factors such as symptom frequency, nighttime symptoms, objective measures of lung function such as spirometry parameters (e.g., $\%FEV_1$ of predicted, FEV_1 variability, requirement for use of SABA for symptom control. Parameters of current impairment and future risk may be assessed and included in a determination of the level of asthma control. In some embodiments, asthma control is defined as shown in Figure 4.3(a), 4.3(b), or 4.3(c) of NHBLI Guidelines, for individuals 0-4, 5-11, or $\geq$ 12 years of age, respectively.

[00411] In general, one of ordinary skill in the art can select an appropriate means of determining asthma severity level and/or degree of control, and any classification scheme considered reasonable by those of ordinary skill in the art can be used.

[00412] In some embodiments of the invention, a subject suffering from persistent asthma is treated with a complement inhibitor using an inventive dosing regimen. In some embodiments, the subject suffers from mild or moderate asthma. In some embodiments, the subject suffers from severe asthma. In some embodiments, a subject has asthma that is not well controlled using conventional therapy. In some embodiments, a subject has asthma that, when treated using conventional therapy, requires use of ICS in order to be well controlled. In some embodiments, a subject has asthma that fails to be well controlled despite use of ICS. In some embodiments, a subject has asthma that, if treated using conventional therapy, would require use of OCS in order to be well controlled. In some embodiments, a subject has asthma that fails to be well controlled despite use of high intensity conventional therapy that includes OCS. In some embodiments of the invention a long-acting complement inhibitor is administered as a controller medication or allow the subject to avoid using or reduce their dose of a conventional controller medication.

[00413] In some embodiments, the subject suffers from allergic asthma, which is the case for most asthmatic individuals. In some embodiments, an asthmatic subject is considered to have allergic asthma if a non-allergic trigger for the asthma (e.g., cold, exercise) is not known

and/or is not identified in a standard diagnostic evaluation. In some embodiments, an asthmatic subject is considered to have allergic asthma if the subject (i) reproducibly develops asthma symptoms (or worsening of asthma symptoms) following exposure to an allergen or allergen(s) to which the subject is sensitive; (ii) exhibits IgE specific for an allergen or allergen(s) to which the subject is sensitive; (iii) exhibits a positive skin-prick test to an allergen or allergen(s) to which the subject is sensitive; and/or (iv) exhibits other symptom(s) of characteristic(s) consistent with atopy such as allergic rhinitis, eczema, or elevated total serum IgE. It will be appreciated that a specific allergic trigger may not be identified but may be suspected or inferred if the subject experiences worsening symptoms in particular environments, for example.

[00414] Allergen challenge by inhalation is a technique that is widely used in evaluating allergic airway disease. Inhalation of allergen leads to cross-linking of allergen-specific IgE bound to IgE receptors on, e.g., mast cells and basophils. Activation of secretory pathways ensues, resulting in release of mediators of bronchoconstriction and vascular permeability. Individuals with allergic asthma may develop various manifestations following allergen challenge, e.g., early asthmatic response (EAR), late asthmatic response (LAR), airway hyperreactivity (AHR), and airway eosinophilia, each of which can be detected and quantified as known in the art. For example, airway eosiphophilia may be detected as an increase in eosinophils in sputum and/or BAL fluid. The EAR, sometimes referred to as the immediate asthmatic response (IAR), is a response to allergen challenge by inhalation that becomes detectable shortly after the inhalation, typically within 10 minutes (min) of the inhalation, e.g., as a decrease in FEV_1. The EAR typically reaches a maximum within 30 min and resolves within 2 -3 hours (h) post-challenge. For example, a subject may be considered to exhibit a "positive" EAR if his/her FEV_1 decreases by at least 15%, e.g., at least 20%, within this time window relative to baseline FEV_1 (where "baseline" in this context refers to conditions before the challenge, e.g., conditions equivalent to the subject's usual condition when not experiencing an asthma exacerbation and not exposed to allergic stimuli to which the subjectis sensitive). The late asthmatic response (LAR) typically starts between 3 h and 8 h post-challenge and is characterized by cellular inflammation of the airway, increased bronchiovascular permeability, and mucus secretion. It is typically detected as a decrease in FEV_1, which may be greater in magnitude than that associated with the EAR and potentially more clinically important. For example, a subject may be considered to exhibit a "positive" LAR if his/her FEV_1 decreases by at least 15%, e.g., at least 20%, relative to baseline FEV_1 within the relevant time period as compared with

baseline FEV_1. A delayed airway response (DAR) may occur beginning between about 26 and 32 h, reaching a maximum between about 32 and 48 h and resolving within about 56 h after the challenge (Pelikan, Z. Ann Allergy Asthma Immunol. 2010, 104(5):394-404).

[00415] In some embodiments, the chronic respiratory disorder is chronic obstructive pulmonary disease (COPD). COPD encompasses a spectrum of conditions characterized by airflow limitation that is not fully reversible even with therapy and is usually progressive. Symptoms of COPD include dyspnea (breathlessness), decreased exercise tolerance, cough, sputum production, wheezing, and chest tightness. Persons with COPD can experience episodes of acute (e.g., developing over course of less than a week and often over the course of 24 hours or less) worsening of symptoms (termed COPD exacerbations) that can vary in frequency and duration and are associated with significant morbidity. They may be triggered by events such as respiratory infection, exposure to noxious particles, or may have an unknown etiology. Smoking is the most commonly encountered risk factor for COPD, and other inhalational exposures can also contribute to development and progression of the disease. The role of genetic factors in COPD is an area of active research. A small percentage of COPD patients have a hereditary deficiency of alpha-1 antitrypsin, a major circulating inhibitor of serine proteases, and this deficiency can lead to a rapidly progressive form of the disease.

[00416] Characteristic pathophysiologic features of COPD include narrowing of and structural changes in the small airways and destruction of lung parenchyma (in particular around alveoli), most commonly due to chronic inflammation. The chronic airflow limitation observed in COPD typically involves a mixture of these factors, and their relative importance in contributing to airflow limitation and symptoms varies from person to person. The term "emphysema" refers to enlargement of the air spaces (alveoli) distal to the terminal bronchioles, with destruction of their walls. It should be noted that the term "emphysema" is often used clinically to refer to the medical condition associated with such pathological changes. Some individuals with COPD have chronic bronchitis, which is defined in clinical terms as a cough with sputum production on most days for 3 months of a year, for 2 consecutive years. Further information regarding risk factors, epidemiology, pathogenesis, diagnosis, and current management of COPD may be found, e.g., in "Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease" (updated 2009) available on the Global Initiative on Chronic Obstructive Pulmonary Disease, Inc. (GOLD) website (www.goldcopd.org), also referred to herein as the "GOLD Report", the American Thoracic Society/European Respiratory Society Guidelines (2004) available on

the ATS website at www.thoracic.org/clinical/copd-guidelines/resources/copddoc.pdf, referred to herein as "ATC/ERS COPD Guidelines" and standard textbooks of internal medicine such as Cecil Textbook of Medicine (20th edition), Harrison's Principles of Internal Medicine (17th edition), and/or standard textbooks focusing on pulmonary medicine.

[00417] In some embodiments methods disclosed herein inhibit (interfere with, disrupt) the DC-Th17-B-Ab-C-DC cycle discussed above. For example, administration of a complement inhibitor may break the cycle by which complement stimulates DC cells to promote the Th17 phenotype. As a result, the number and/or activity of Th17 cells diminishes, which in turn reduces the amount of Th17-mediated stimulation of B cells and polyclonal antibody production. In some embodiments, these effects result in "resetting" the immunological microenvironment to a more normal, less pathological state. As described in Example 1, evidence supporting the capacity of complement inhibition to have a prolonged inhibitory effect on Th17-associated cytokine production was obtained in an animal model of asthma.

[00418] In some embodiments, inhibiting the DC-Th17-B-Ab-C-DC cycle has a disease-modifying effect. Without wishing to be bound by any theory, rather than merely treating symptoms of a disorder, inhibiting the DC-Th17-B-Ab-C-DC cycle may interfere with fundamental pathologic mechanisms that may contribute to ongoing tissue damage even when symptoms are well controlled and/or that may contribute to exacerbations of the disease. In some embodiments, inhibiting the DC-Th17-B-Ab-C-DC cycle causes a chronic disorder to go into remission. In some embodiments, remission refers to a state of absence or substantial absence of disease activity in a subject with a chronic disorder, with the possibility of return of disease. In some embodiments remission may be sustained for a prolonged period of time (e.g., at least 6 months, e,g., 6-12 months, 12-24 months, or more) in the absence of continued therapy or with a reduced dose or increased dosing interval. In some aspects, inhibition of complement may change the immunological micro-environment of a tissue that is rich in Th17 cells and modify it into a micro-environment that is rich in regulatory T cells (Tregs). Doing so could allow the immune system to "reset" itself and go into a state of remission. In some embodiments, for example, remission may be sustained until occurrence of a triggering event. A triggering event may be, for example, an infection (which may result in production of polyclonal antibodies that react both with an infectious agent and a self protein), exposure to particular environmental conditions (e.g., high levels of air pollutants such as ozone or particulate matter or components of smoke such as cigarette smoke, allergens), etc. Genetic factors may play a role. For example, individuals having

particular alleles of genes encoding complement components may have a higher baseline level of complement activity, a more reactive complement system and/or a lower baseline level of endogenous complement regulatory protein activity. In some embodiments an individual has a genotype associated with increased risk of AMD. For example, the subject may have a polymorphism in a gene encoding a complement protein or complement regulatory protein, e.g., CFH, C3, factor B, wherein the polymorphism is associated with an increased risk of AMD.

[00419] In some embodiments an immunologic microenvironment may become progressively more polarized towards a pathological state over time, e.g., in a subject who has not yet developed symptoms of a chronic disorder or in a subject who has developed the disorder and has been treated as described herein. Such a transition may occur stochastically (e.g., due at least in part to apparently random fluctuations in antibody levels and/or affinity) and/or as a result of accumulated "sub-threshold" trigger events that are not of sufficient intensity to trigger a symptomatic outbreak of a disorder.

[00420] In some embodiments it is contemplated that a relatively short course of a long-acting compstatin analog, e.g., between 1 week and 6 weeks, e.g., about 2 – 4 week, may provide a long-lasting benefit. In some embodiments a remission is achieved for a prolonged period of time, e.g., 1-3 months, 3-6 months, 6-12 months, 12-24 months, or more. In some embodiments a subject may be monitored and/or treated prophylactically before recurrence of symptoms. For example, a subject may be treated prior to or upon exposure to a triggering event. In some embodiments a subject may be monitored,e g., for an increase in a biomarker, e.g., a biomarker comprising an indicator of Th17 cells or Th17 cell activity, or complement activation, and may be treated upon increase in the level of such biomarker. See, e.g., PCT/US2012/043845 for further discussion.

[00421] IX. Compositions and Administration

[00422] The invention provides a variety of compositions comprising a cell-reactive, long-acting, or targeted compstatin analog. In various embodiments, a composition can have any feature or combination of features discussed herein so long as they are not mutually exclusive. The invention provides embodiments of such compositions, and methods of use thereof, in which the compstatin analog is any compstatin analog.

[00423] In some embodiments, a composition comprises a purified cell-reactive, long-acting, or targeted compstatin analog. Purification can be achieved using a variety of approaches that can be selected by one of ordinary skill in the art based to achieve a desired degree of purity with respect to various components present in the composition prior to

purification. For example, filtration, high performance liquid chromatography, affinity chromatography, and/or other approaches and combinations thereof can be used. In some embodiments, the composition comprises at least 80%, 85%, 90%, 95%, 98%, 99%, or more cell-reactive, long-acting, or targeted compstatin analog as a percentage of the total compstatin analog by weight. In some embodiments, the composition comprises at least 80%, 85%, 90%, 95%, 98%, 99%, or more cell-reactive, long-acting, or targeted compstatin analog as a percentage of the total compstatin analog on a molar basis. In some embodiments, a composition consists or consists essentially of a cell-reactive, long-acting, or targeted compstatin analog.

[00424] In some embodiments, a composition comprising a cell-reactive compstatin analog and a compound comprising a cell-reactive functional group is characterized in that the ratio of the cell-reactive compstatin analog to the compound comprising the cell-reactive functional group on a molar basis is at least 10:1, 20:1, 50:1, 100:1, 500:1, 1,000:1, or more. In some embodiments the composition comprises at least 80%, 85%, 90%, 95%, 98%, 99%, or more cell-reactive compstatin analog as a percentage of the total compstatin analog by weight. In some embodiments the composition comprises at least 80%, 85%, 90%, 95%, 98%, 99%, or more cell-reactive compstatin analog as a percentage of the total compstatin analog on a molar basis. In some embodiments a composition comprises at least 80%, 85%, 90%, 95%, 98%, 99%, or more cell-reactive compstatin analog by weight. In some embodiments a composition comprises at least 80%, 85%, 90%, 95%, 98%, 99%, or more long-acting compstatin analog by weight. In some embodiments a composition comprises at least 80%, 85%, 90%, 95%, 98%, 99%, or more targeted compstatin analog by weight. In some embodiments a composition comprises at least 80%, 85%, 90%, 95%, 98%, 99%, or more targeted compstatin analog by weight. In some embodiments weight is dry weight.

[00425] In some aspects, the invention provides a pharmaceutical grade composition comprising a cell-reactive, long-acting, or targeted compstatin analog. The pharmaceutical grade composition can have any of the above-mentioned characteristics in terms of purity in various embodiments. The pharmaceutical grade composition is sufficiently free of endotoxin, heavy metals, and unidentified and/or uncharacterized substances so as to be acceptable, without further purification, as a pharmaceutical composition suitable for administration to a human subject or for the manufacture of a pharmaceutical composition to be administered to a human subject. In some embodiments, the pharmaceutical grade composition is sterile.

[00426] Suitable preparations, e.g., substantially pure preparations of a cell-reactive, long-acting, or targeted compstatin analog or other active agent, may be combined with pharmaceutically acceptable carriers or vehicles, etc., to produce an appropriate pharmaceutical composition. The term "pharmaceutically acceptable carrier or vehicle" refers to a non-toxic carrier or vehicle that does not destroy the pharmacological activity of the compound with which it is formulated. One of skill in the art will understand that a carrier or vehicle is "non-toxic" if it is compatible with administration to a subject in an amount appropriate to deliver the compound without causing undue toxicity. Pharmaceutically acceptable carriers or vehicles that may be used in the compositions of this invention include, but are not limited to, water, physiological saline, Ringer's solution, sodium acetate or potassium acetate solution, 5% dextrose, and the like. The composition may include other components as appropriate for the formulation desired, e.g., as discussed herein. Supplementary active compounds, e.g., compounds independently useful for treating a subject suffering from a complement-mediated disorder, can also be incorporated into the compositions. The invention provides such pharmaceutical compositions comprising a cell-reactive, long-acting, or targeted compstatin analog and, optionally, a second active agent useful for treating a subject suffering from a complement-mediated disorder.

[00427] In some embodiments, the invention provides a pharmaceutically acceptable composition suitable for administration to humans, packaged together with a label approved by a government agency responsible for regulating pharmaceutical agents, e.g., the U.S. Food & Drug Administration. In some embodiments, the invention provides a pharmaceutical kit or pack comprising: (a) a pharmaceutically acceptable cell-reactive, long-acting, or targeted compstatin analog in solid form; (b) a pharmaceutically acceptable carrier or vehicle. Optionally the kit or pack contains instructions for dissolving the cell-reactive, long-acting, or targeted compstatin analog in the carrier. In some embodiments a pharmaceutical kit or pack is provided. The pack or kit comprises sufficient amount of pharmaceutical composition for at least 1 dose, e.g., between 1 and 200 doses or any intervening number or subrange. In some embodiments a pharmaceutical pack or kit comprises one or more needles and, optionally, one or more syringes. In some embodiments at least one prefilled syringe is provided. In some embodiments one or more unit dosage forms or premeasured aliquots are provided. In some embodiments instructions for administration, which in some embodiments comprise instructions for self-administration, e.g., via subcutaneous injection, are provided.

[00428] A pharmaceutical composition can be administered to a subject by any suitable route of administration including, but not limited to, intravenous, intramuscular, subcutaneously, by inhalation, by nasal delivery, intrathecally, intracranially, intraarterially, orally, rectally, transdermally, intradermally, subdermally, etc. In some embodiments, a composition comprising a cell-reactive, long-acting, or targeted compstatin analog is administered intravenously. In some embodiments, a composition comprising a cell-reactive, long-acting, or targeted compstatin analog is administered intra-arterially. The composition can be administered locally, either into the vascular system supplying an organ or tissue, or extra-vascularly in the vicinity of an organ or tissue. It will be understood that "administration" encompasses directly administering a compound or composition to a subject, instructing a third party to administer a compound or composition to a subject, prescribing or suggesting a compound or composition to a subject (e.g., for self-administration), self-administration, and, as appropriate, other means of making a compound or composition available to a subject.

[00429] Pharmaceutical compositions suitable for injectable use (e.g., intravenous administration) or by pump or catheter typically include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. Sterile solutions can be prepared by incorporating the compound in the required amount in an appropriate solvent, optionally with one or a combination of ingredients such as buffers such as acetates, citrates, lactates or phosphates; agents for the adjustment of tonicity such as sodium chloride or dextrose; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid, glutathione, or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; and other suitable ingredients etc., as desired, followed by filter-based sterilization. One of skill in the art will be aware of numerous physiologically acceptable compounds that may be included in a pharmaceutical composition. Other useful compounds include, for example, carbohydrates, such as glucose, sucrose, lactose; dextrans; amino acids such as glycine; polyols such as mannitol. These compounds may, for example, serve as bulking agents and/or stabilizers, e.g., in a powder and/or when part of the manufacture or storage process involves lyophilization. Surfactant(s) such as Tween-80, Pluronic-F108/F68, deoxycholic acid, phosphatidylcholine, etc., may be included in a composition, e.g., to increase solubility or to provide microemulsion to deliver hydrophobic drugs. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide, if desired. The parenteral preparation can be enclosed in ampoules, disposable syringes or infusion bags or multiple dose vials

made of glass or plastic. Preferably solutions for injection are sterile and acceptably free of endotoxin.

[00430] Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and appropriate other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation can include vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient, e.g., from a previously sterile-filtered solution thereof.

[00431] Oral administration may be used in certain embodiments. Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules, e.g., gelatin capsules. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring. A liquid composition can also be administered orally. Formulations for oral delivery may incorporate agents to improve stability within the gastrointestinal tract and/or to enhance absorption.

[00432] For administration by inhalation, a compstatin analog may be delivered in the form of an aerosol spray from a pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide. A metered dose inhaler or nebulizer may be used. The aerosol may comprise liquid particles or dry aerosol (e.g., dry powders, large porous particles, etc.).

[00433] For topical application, a compstatin analog may be formulated in a suitable ointment containing the active component suspended or dissolved in one or more carriers. Carriers for topical administration include, but are not limited to, mineral oil, liquid petrolatum, white petrolatum, propylene glycol, polyoxyethylene, polyoxypropylene compound, emulsifying wax and water. Alternatively, the pharmaceutically acceptable compositions can be formulated as a suitable lotion or cream containing a compstatin analog suspended or dissolved in one or more pharmaceutically acceptable carriers. Suitable carriers

include, but are not limited to, mineral oil, sorbitan monostearate, polysorbate 60, cetyl esters wax, cetearyl alcohol, 2-octyldodecanol, benzyl alcohol, and water.

[00434] Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated may be used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished, e.g., through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are typically formulated into ointments, salves, gels, or creams as generally known in the art.

[00435] The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

[00436] In certain embodiments of the invention, a compstatin analog or other active compound is prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. For example, a compstatin analog may be incorporated into or encapsulated in a microparticle or nanoparticle formulation. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, polyethers, polylactic acid, PLGA, etc. Liposomes or other lipid-based particles can be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811 and/or other references listed herein. Depot formulations containing a compstatin analog may be used. The compstatin analog is released from the depot over time, e.g., so as to provide a therapeutic concentration for longer than if the compound was administered intravenously. In some aspects, a CRM confers depot properties on a compstatin analog. One of ordinary skill in the art will appreciate that the materials and methods selected for preparation of a controlled release formulation, implant, etc., should be such as to retain activity of the compound.

[00437] It will be appreciated that the compstatin analog and/or additional active agent(s) can be provided as a pharmaceutically acceptable salt. Pharmaceutically acceptable salts include those derived from pharmaceutically acceptable inorganic and organic acids and bases. Examples of suitable acid salts include acetate, adipate, alginate, aspartate, benzoate, benzenesulfonate, bisulfate, butyrate, citrate, camphorate, camphorsulfonate, cyclopentanepropionate, digluconate, dodecylsulfate, ethanesulfonate, formate, fumarate,

glucoheptanoate, glycerophosphate, glycolate, hemisulfate, heptanoate, hexanoate, hydrochloride, hydrobromide, hydroiodide, 2-hydroxyethanesulfonate, lactate, maleate, malonate, methanesulfonate, 2-naphthalenesulfonate, nicotinate, nitrate, oxalate, palmoate, pectinate, persulfate, 3-phenylpropionate, phosphate, picrate, pivalate, propionate, salicylate, succinate, sulfate, tartrate, thiocyanate, tosylate and undecanoate. Also, pharmaceutically-acceptable salts can be prepared as alkaline metal or alkaline earth salts, such as sodium, potassium or calcium salts, if appropriate depending on the identity of the active agent.

[00438] It will be understood that the pharmaceutically acceptable carriers, compounds, and preparation methods mentioned herein are exemplary and non-limiting. See, e.g., Remington: The Science and Practice of Pharmacy. 21st Edition. Philadelphia, PA. Lippincott Williams & Wilkins, 2005, for additional discussion of pharmaceutically acceptable compounds and methods of preparing pharmaceutical compositions of various types.

[00439] A pharmaceutical composition can be administered in an amount effective to achieve a desired beneficial effect. In some embodiments, an effective amount is sufficient to provide one or more of the following benefits: (i) reduction in at least one symptom or sign of a complement-mediated disorder; (ii) increased quality of life; (iii) reduced hospitalization; (iv) reduced mortality. One of ordinary skill in the art will appreciate that the particular beneficial effect will depend at least in part on various factors, such as the particular disorder being treated. One of ordinary skill in the art will be aware of the symptoms and signs that may occur in subjects with complement-mediated disorders. Examples of symptoms and signs of various complement disorders are provided herein. For example, in some embodiments, e.g., wherein a subject suffers from PNH or aHUS, a beneficial effect is a reduction in complement-mediated red blood cell lysis. In some aspects, a beneficial effect is statistically significant and/or therapeutically meaningful within the judgement of one or ordinary skill in the art.

[00440] In certain embodiments of the invention a pharmaceutical composition comprising a cell-reactive, long-acting, or targeted compstatin analog is administered parenterally. In some embodiments, the composition is administered intravenously. In some embodiments, the composition is administered by intravenous injection. In some embodiments the composition is administered as an IV bolus or an IV infusion. In some embodiments the composition is administered as an IV drip. In some embodiments the composition is administered as an IV bolus followed by an IV infusion or IV drip. In some embodiments an IV infusion is administered over about 1, 2, 3, 4, 5, 15, 20, 30, 60, or 120 minutes. In some

embodiments an IV drip is administered over more than about 60 minutes, e.g., over about 1, 2, 3, or more hours. In some embodiments, a total amount of between about 0.1 mg/kg/day and about 2,000 mg/kg/day of compstatin analog is administered, e.g., between about 1 mg/kg/day and about 1,000 mg/kg/day, e.g., between about 5 mg/kg/day and about 500 mg/kg/day. In some embodiments, a total amount of between about 10 mg/kg/day and about 100 mg/kg/day of compstatin analog is administered, e.g., between about 10 mg/kg/day and about 50 mg/kg/day e.g., between about 10 mg/kg/day and about 20 mg/kg/day. . In some embodiments, between about 0.5 mg/kg/day to about 10 mg/kg/day of compstatin analog is administered. In some embodiments, between about 1 mg/kg/day to about 5 mg/kg/day of compstatin analog is administered. In some embodiments, between about 1 mg/kg/day to about 3 mg/kg/day of compstatin analog is administered. In some embodiments, between about 3 mg/kg/day to about 5 mg/kg/day of compstatin analog is administered. In some embodiments, between about 5 mg/kg/day to about 7.5 mg/kg/day of compstatin analog is administered. In some embodiments, between about 7.5 mg/kg/day to about 10 mg/kg/day of compstatin analog is administered. It will be appreciated that a variety of different dosing regimens could be used to administer a desired total daily amount. For example, a desired amount of compstatin analog could be administered in a single administration or in multiple administrations, e.g., during a 24 hour period. For example, a subject could receive two or more doses within a 24 hour period, which doses could be administered over the same length of time or over different lengths of time. In some embodiments, a cell-reactive, long-acting, or targeted compstatin analog is administered at time intervals greater than 24 hours. For example, doses could be administered on average every other day, every 3-4 days, weekly, every other week, etc., in various embodiments. In some embodiments, covalently attached, long-acting, or targeted compstatin analogs protect cells, tissues, organs, for a period of weeks or months without need for retreatment. For example, subjects may be maintained with retreatment at intervals of between 1-2 weeks, 2-4 weeks, 4-6 weeks, 6-8 weeks, or even longer. In some embodiments subcutaneous administration is used to administer at least some doses. For example, administration of approximately 0.1- 5 mg/kg/day, e.g., about 0.5-2 mg/kg/day is contemplated in some embodiments, e.g., in a volume of about 0.25 ml - 2 mL, e.g., a volume of about 1 ml. In some embodiments the concentration is about 50 mg/ml to about 300 mg/ml, e.g., about 50 mg/ml – about 100 mg/ml or about 100 mg/ml – about 200 mg/ml. In some embodiments administration is daily. In some embodiments administration is 1 or 2 times per day. As described further in the Examples, daily subcutaneous administration of an exemplary long-acting compstatin analog readily

achieived blood levels well above 5 micromolar. In some embodiments, intramuscular administration is used to deliver similar amounts of compound. In some embodiments a long-acting compstatin analog is administered using a therapeutically effective amount to a subject, wherein such administration results in blood concentrations of the compound that achieve a level above at least 1 μM, at least 2 μM, at least 2.5 μM, , at least 3 μM, at least 4 μM, at least 5 μM, at least 6 μM, at least 7 μM, at least 8 μM, at least 9 μM , at least 10 μM , at least 11 μM, at least 12 μM, or at least 13 μM, att least 14 μM, at least 15 μM, at least 16 μM, at least 18 μM, or at least about 20 μM, or at least about 25 μM or within any range between4 μM and about 15 μM or about 20 μM or about 25 μM. In some embodiments such level is maintained for at least about 24 hours, or at least about 48 hours, or at least about 72 hours, or at least about 96 hours, or at least about 120 hours, or at least about 144 hours following a single IV injection or following daily subcutaneous injections for about 5-7 days. Sustained levels may be achieved for longer, e.g., up to about 10 days, 12 days, 14 days, or more. In some embodiments a subject is treated so as to maintain a steady state level of about 1.0 μM, about 2.0 μM, about 2.5 μM, about 3.0 μM, about 3.5 μM, about 4.0 μM, about 4.5 μM, about 5.0 μM, about 5.5 μM, about 6.0 μM, about 6.5 μM, about 7.0 μM, about 7.5 μM, about 8.0 μM, about 8.5 μM, about 9.0 μM, about 9.5 μM, or about 10 μM. In some embodiments a steady state level has a value between about 1.0 μM and about 10.0 μM, e.g., between about 2.0 μM and about 5.0 μM, between about 2.5 μM and about 5.0 μM, between about 5.0 μM and about 7.5 μM, or between about 7.5 μM, and about 10 μM, or any intervening value within any of the afore-mentioned ranges. In some embodiments a concentration is sufficient to substantially inhibit lysis of red blood cells of PNH patients exposed to human serum in vitro, e.g., using a modified Ham's assay using human serum (see, e.g., Example 8). In some embodiments a concentration is sufficient to reduce by at least 50%, 60%, 70%, 80%, 90%, or more, lysis of red blood cells of PNH patients exposed to human serum in vitro, e.g., using a modified Ham's assay using human serum (see, e.g., Example 8). In some embodiments a Ham's assay may be performed using human serum adjusted to a magnesium level at about 0.005 mol/L and a pH lowered to about 6.2 to activate complement. Examples 18 and 19 present data confirming the ability of compstatin analogs described herein to inhibit lysis of RBCs from PNH patients.

[00441] In some aspects compstatin analogs, e.g., long-acting compstatin analogs, may protect red blood cells of PNH patients from accumulating significant amounts of C3 and/or products of C3 activation on their surface. For example, PNH red blood cells that are protected from complement-mediated lysis by compstatin analogs, e.g., long-acting

compstatin analogs, may also be protected from accumulating significant amounts of C3 and/or products of C3 activation on their surface. As known in the art, eculizumab (Soliris®, Alexion Pharmaceuticals. Inc.), is a humanized anti-C5 monoclonal antibody that is approved for treatment of PNH and aHUS in a number of countries (see, e.g., Dmytrijuk A, FDA report: eculizumab (Soliris) for the treatment of patients with paroxysmal nocturnal hemoglobinuria. Oncologist. 2008 Sep;13(9):993-1000. doi: 10.1634/theoncologist.2008-0086. Epub 2008 Sep 10; Westra D., A new era in the diagnosis and treatment of atypical haemolytic uraemic syndrome. Neth J Med. 2012 Apr;70(3):121-9). It has been reported that when PNH RBCs are exposed to eculizumab they may exhibit accumulation of significant amounts of C3 and/or products of C3 activation on their surface, which may contribute to the clearance and/or extravascular hemolysis of these cells (e.g., in the spleen) and may thus at least in part account for persistent hematologic abnormalities, e.g., persistent anemia, observed in some PNH patients despite treatment with eculizumab. Without wishing to be bound by any theory, this may occur due to inhibition of formation of the MAC by eculizumab, which protects cells from MAC-mediated lysis but does not inhibit C3 activation or deposition of C3 and/or products of C3 activation and leaves PNH cells vulnerable to surface C3 activation and deposition of C3 and/or products of C3 activation due to their lack of GPI-anchored complement inhibiting proteins. Without wishing to be bound by any theory, the ability of compstatin analogs described herein to inhibit C3 activation and thereby inhibit the production of C3 activation products may afford a significant advantage. In some embodiments a subject who has been or is being treated with eculizumab and continues to exhibit evidence of hemolysis, e.g., clinically significant hemolysis, such as causing anemia and/or requiring transfusion is treated with a comptatin analog described herein. In some embodiments a compstatin analog is used at a concentration sufficient such that the level of C3 and/or C3 activation products on PNH RBCs exposed to the compstatin analog (in vitro (e.g., in a Ham's assay) or in vivo) is within the range exhibited by normal RBCs from healthy subjects. In some embodiments the level of C3 and/or C3 activation products on PNH RBCs exposed to the compstatin analog (in vitro (e.g., in a Ham's assay) or in vivo) is within about 1.2, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, or 5.0 of the average level or upper limit of normal. In some embodiments the level of C3 and/or C3 activation products on PNH RBCs exposed to a compstatin analog (in vitro (e.g., in a Ham's assay) or in vivo) is less than the level of of C3 and/or C3 activation products on PNH RBCs exposed to Soliris at a concentration that provides equivalent protection against complement-mediated lysis. In some embodiments the level of C3 and/or C3 activation products on PNH RBCs exposed to a

compstatin analog (in vitro or in vivo) is no more than about 5%, 10%, 15%, 20%, 25%, 30%, 40%, or 50% of the level of of C3 and/or C3 activation products on PNH RBCs exposed to eculizumab at a concentration that provides equivalent protection against complement-mediated lysis. In some embodiments the level of C3 and/or C3 activation products on PNH RBCs exposed to the compstatin analog (in vitro or in vivo) is within about 1.2, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, or 5.0 of the average level or upper limit of a normal range. In some embodiments the PNH cells comprise or consist ofType II PNH cells, Type III PNH cells, or a mixture thereof. In some embodiments the RBCs are at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, or more Type III and/or Type II RBCs. In some embodiments the cells may comprise some Type I cells. In some embodiments RBCs may be classified as Type I, II, or III based on the level of a GPI-anchored protein such as CD59 on their surface, which may be measured using flow cytometry, immunofluorescence, or ELISA, e.g., using an antibody (e.g., a monoclonal antibody) or other binding agent that binds to said GPI-anchored protein. In some aspects inhibition of deposition ofC3 and/or products of C3 activation on cells or surfaces may be used as an indicator of efficacy of a compstatin analog in other complement-mediated diseases such as aHUS, other complement-mediated hemolytic diseases, or other complement-mediated diseases. For example, in some embodiments a compstatin analog inhibits such deposition on endothelial cells in a subject with aHUS. In some embodiments the level of C3 and/or C3 activation product(s) may be measured using flow cytometry, immunofluorescence, or ELISA, e.g., using an antibody (e.g., a monoclonal antibody) or other binding agent that binds to C3 and/or to one or more C3 activation product(s). In some embodiments a C3 activation product is C3b, C3c, or C3d. In some embodiments a binding agent binds to C3d. In some embodiments a binding agent binds to C3d and at least one other C3 activation product. In some embodiments PNH patient RBCs contacted with a compstatin analog in vitro (e.g., in a Ham's assay) are protected from activated complement such that the relative proportion (percentages) of Type I, Type II, and Type III cells or the relative proportion or percentages of Type III and Type I, Type II and Type I, or Type III and Type II, are approximately the same as in a control assay in which inactivated complement (e.g., heat inactivated complement) is used. In some embodiments PNH patient RBCs contacted with a compstatin analog in vitro (e.g., in a Ham's assay) are protected from activated complement such that the relative proportion or percentages of Type I, Type II, and Type III cells or the relative proportion or percentages of Type III and Type I, Type II and Type I, or Type III and Type II, are within 5% of the proportions or percentages obtained in a control assay in which inactivated complement (e.g., heat inactivated

complement) is used. In some embodiments complement may be inactivated by heat inactivation, which may be performed by heating complement components or serum or plasma containing complement components to 56 degrees C or higher.

[00442] In some embodiments measurements of LDH (an enzyme that isabundant in red blood cells and can function as a marker for hemolysis), one or more hematologic parameters such as hematocrit, hemoglobin, and/or reticulocyte measurements may additionally or alternately be used in determining the amount of lysis. In some embodiments one or more such methods may be used to determine the amount of lysis of RBCs, e.g., RBCs that are susceptible to complement-mediated lysis, e.g., PNH patient cells, aHUS patient cells, cells from subjects with other complement-mediated hematologic disorders, cells exposed to abnormally high levels of complement activation. In some aspects, the disclosure provides a method comprising contacting one or more cells in vitro or in vivo with a compstatin analog described herein and measuring the effect of the compstatin analog on one or more indicators of complement-mediated cell damage and/or cell surface complement activation or deposition. In some embodiments contacting the one or more cells for a sufficient time at a sufficient concentration results in a reduction of an abnormally high value or an increase in an abnormally low value to within a normal range or to within 5%, 10%, 15%, 20%, or 25% of the lower or upper limit of a normal range.

[00443] In some aspects, the disclosure provides a method of selecting or modifying a dosing regimen or one or more components of a dosing regimen for a patient with a complement-mediated hemolytic disease such as PNH. The one or more components of a dosing regimen may comprise a dose, dosing interval, route of administration (e.g., IV or subcutaneous), or combination thereof. A dose may be a loading dose, maintenance dose, or both. In some embodiments, one or more blood samples may be obtained from a patient and a dosing regimen or component thereof for a compstatin analog, e.g., a long-acting compstatin analog, may be selected or modified to achieve a desired level of protection of the patient's RBCs from lysis and/or from accumulation of C3 and/or C3 activation product(s) in vitro. In some embodiments one or more doses of a compstatin analog, e.g., a long-acting compstatin analog, may be administered to a patient, and one or more blood samples may be subsequently obtained from a patient and assessed for level of C3 and/or C3 activation product(s) on their surface. In some embodiments a dosing regimen or component thereof, e.g., a dose, dosing interval, or route of administration, may be selected or modified to achieve a desired level of protection of the patient's RBCs from lysis and/or from accumulation of C3 and/or C3 activation products in vitro or in vivo. In some embodiments a

dosing regimen or component thereof, e.g., dose, dosing interval, or route of administration, may be selected or modified to achieve a desired level of protection of the patient's RBCs from extravascular clearance and/or extravascular lysis in vivo. A desired level may be, e.g., a level that is accepted in the art as providing a clinically meaningful benefit, a level that provides a clinically meaningful benefit to a particular patient, a level that is within the normal range, a level selected by a medical practitioner, or any other selected level. A normal range for a parameter may be known in the art and/or may be a reference range established by a laboratory, e.g., a clinical laboratory, wherein the value of the relevant parameter as measured in at least 95%, 96%, 97%, 98%, or 99% of the general population or at least 95%, 96%, 97%, 98%, or 99% of healthy individuals (which may optionally be matched for one or more demographic variables such as gender, age, etc.) or biological specimens obtained therefrom (such as blood samples) would fall within the reference range. A reference range may be established using a sample population representative of the general population or representative of healthy individuals.

[00444] In some embodiments a long-acting compstatin analog comprising a CRM is designed to confer a slower rate of systemic absorption after subcutaneous or intramuscular administration to a subject as compared with a compstatin analog not comprising a CRM. In some embodiments particular CRM properties, e.g., length, are selected to confer a desired rate of systemic absorption after subcutaneous or intramuscular administration as compared with at least some other CRMs. In some embodiments, the Cmax is reduced in comparison to a comparable dose of a compstatin analog not linked to a CRM, which may thereby contribute to keeping the plasma concentration within a desired window, e.g., the therapeutic window, for the compound. In some embodiments a long-acting compstatin analog composition is characterized in that a dose, when administered subcutaneously, appears fully absorbed within about 1, 2, 3, 4, 6, 8, 12, 15, 30, 45, 60, 90, or 120 hours following administration based on visual observation at the injection site.

[00445] It will be understood that there may be an initial treatment phase during which treatment is more frequent and/or in which higher doses are administered. For example, in a subject with PNH or aHUS, it may require several doses to achieve protection of a substantial fraction of the subject's RBCs. After that, lower doses and/or less frequent dosing could be used, e.g., to protect newly formed RBCs and/or to replenish protection of existing RBCs. Of course similar approaches may be followed for treatment of any disease where appropriate. In some embodiments treatment is started using IV administration and then switched to subcutaneous, intramuscular, or intradermal for maintenance therapy. Depending on the

disease, treatment may continue at intervals for, e.g.., months, years, or indefinitely. Appropriate doses and dosing regimen depend at least in part upon the potency and half-life of the compstatin analog (or other active agent), and may optionally be tailored to the particular recipient, for example, through administration of increasing doses until a preselected desired response is achieved, such as a desired degree of complement inhibition and/or cell protection. If desired, the specific dose level for any particular subject may be selected based at least in part upon a variety of factors including the activity of the specific compound employed, the particular condition being treated, the age, body weight, general health, route of administration, the rate of excretion, any drug combination, and/or the degree of complement protein expression or activity measured in one or more samples obtained from the subject.

[00446] The invention encompasses administration of a compstatin analog in combination with additional therapy. Such additional therapy may include administration of any agent(s) used in the art or potentially useful for treating a subject suffering from the disease.

[00447] When two or more therapies (e.g., compounds or compositions) are used or administered "in combination" with each other, they may be given at the same time, within overlapping time periods, or sequentially (e.g., separated by up to 2 weeks in time), in various embodiments of the invention. They may be administered via the same route or different routes. In some embodiments, the compounds or compositions are administered within 48 hours of each other. In some embodiments, a compstatin analog can be given prior to or after administration of the additional compound(s), e.g., sufficiently close in time that the compstatin analog and additional compound(s) are present at useful levels within the body at least once. In some embodiments, the compounds or compositions are administered sufficiently close together in time such that no more than 90% of the earlier administered composition has been metabolized to inactive metabolites or eliminated, e.g., excreted, from the body, at the time the second compound or composition is administered.

[00448] In some embodiments, a composition that includes both the cell-reactive compstatin analog and additional compound(s) is administered.

[00449] Example 1: Development of PEGylated Compstatin Analogs that Retain Substantial Complement Inhibiting Activity

[00450] A compstatin analog having the amino acid sequence of the compstatin analog of SEQ ID NO: 28, but incorporating an AEEAc-Lys moiety located C-terminal to the Thr residue of SEQ ID NO: 28 for purposes of subsequent conjugation of an NHS ester activated PEG to the amino group of the Lys side chain was synthesized. The compound was

synthesized using standard methods. Briefly, amino acids (including AEEAc) were obtained as Fmoc-protected amino acids, in which the α-amino group of each amino acid was protected with Fmoc. Side chain functional groups were also blocked with various appropriate protective groups. Synthesis was accomplished following the solid phase methodology described by Merrifield (J. Amer. Chem. Soc. 85, 2149 (1963)). Chain assembly was performed on solid phase, at the conclusion of which the N-terminus was acetylated; the peptide was then cleaved from the solid phase and simultaneously deprotected via acidolysis using TFA and amidated. The linear peptide was then oxidized and purified. The resulting compstatin analog is represented as follows *Ac*-Ile-Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-AEEAc-Lys-*NH2* (SEQ ID NO: 51), abbreviated as CA28-AEEAc-Lys. Note that for purposes of brevity, the N-terminal acetyl group and C-terminal amino groups are omitted in this abbreviation. Monofunctional, linear NHS-ester activated PEGs with molecular weights of 30 kD and 40 kD (NOF America Corp. White Plains, NY, Cat. No. SUNBRIGHT® ME-300GS and Cat. No. SUNBRIGHT® ME-400GS), respectively, were coupled to the lysine side chain of CA28-AEEAc-Lys, resulting in long-acting compstatin analogs represented as follows: CA28-AEEAc-Lys-(PEG30k) and CA28-AEEAc-Lys-(PEG40k), and purified. Note that the number after the term "PEG" and preceding the letter "k" represents the molecular weight of the PEG moiety in kilodaltons, and the "k" is an abbreviation for kD). CA28-AEEAc-Lys-(PEG30k) is also referred to as CA28-1. CA28-AEEAc-Lys-(PEG40k) is also referred to as CA28-2.

[00451] Inhibitory activity of the synthesized compounds was assessed by measuring the effect of the compounds on complement activation via the classical pathway using a standard complement inhibition assay. The protocol measures C3b deposition in an ELISA format. C3b deposition monitored using this method is generated through complement activated by the classical pathway. Briefly, 96-well plates are coated with BSA. Human plasma, chicken ovalbumin (OVA), polyclonal anti-OVA antibodies and compound being tested (referred to as "drug") are added and incubated, followed by addition of Anti-human C3 HRP-conjugated antibody. After an additional incubation, substrate is added and signal detected. Details of the protocol are as follows:

[00452] Protocol for Classical Complement Inhibition Assay

Materials:

- Ninety-six well plate (polystyrene plate, Thermo Scientific, 9205)
- Chicken OVA (Sigma A5503-5G)

- Rabbit anti-chicken OVA (Abcam ab1221)
- Blocking buffer (Startingblock buffer, Thermo Scientific 37538)
- Veronal Buffer (5X concentration, Lonza 12-624E)
- Human plasma (collected with Lepirudin at 50 ug/ml final concentration)
- Goat anti-human C3 HRP-conjugated Ab (MP Biomedicals, 55237)
- Tween-20 Wash Buffer (0.05% Tween 20-PBS buffer)
- TMB (Peroxidase substrate, BD 555214) – 1:1 mixture of BD 51-2607KC and 51-2606KC.
- 1M H_2SO_4

Protocol:

1. Add 100 ul/well of 1% chicken OVA (in PBS)
2. Incubate overnight @ 4°C or room temperature for 1-2 hr.
3. Remove by shaking and tapping the plate.
4. Block by adding 200ul of blocking buffer
5. Incubate for 1 h at room temp
6. Remove by shaking and tapping the plate
7. Add 100 ul of 1:1000 dilution of Polyclonal anti-chicken OVA in blocking buffer
8. Incubate for 1h at room temp
9. Wash twice with wash buffer
10. Add 50 ul VB^{++} to wells #2 to 12
11. Add 100ul of starting drug dilution (2X in VB^{++}) to well 1.
12. Serially dilute (1:2) the drug from wells 1 to 10 as follow
    a. Take 50ul of solution from the originating well
    b. Add this to the next well
    c. Mix by pipetting several times
    d. Repeat up to well #10

    Note: from well #10 remove 50ul and discard.
13. Add 50ul of 2X plasma (1:37.5 dilution of original plasma) dilution to wells 1 to 11
14. Incubate for 1h
15. Wash with wash buffer
16. Add 100ul of 1/1000 dilution of anti-C3-HRP Ab in blocking buffer
17. Incubate for 1h
18. Wash with wash buffer
19. Add 100ul of TMB to all wells
20. Incubate for 5-10 min in dark
21. Add 50 ul 1M H_2SO_4
22. Read the plate at 450nm

VB^{++}

Formula:

Barbital 5 mM
NaCl 72.5 mM
$MgCl_2$ 0.5 mM
$CaCl_2$ 0.15 mM
PH 7.4

Stock solutions:

Veronal Buffer (5X)

	Prod #	MW	For 500ml
9 mM Sodium Barbitone	Sigma B0500	206.17	927 mg
15.5 mM diethylbarbituric acid	Sigma B0375	184.19	1.42 grams

Mg-Cl2 (200X)

	Prod #	MW	For 50 ml
100 mM $MgCl_2$-$6H_2O$	Sigma M0250	203.30	1.00 gram

$CaCl_2$ (500x)

	Prod #	MW	For 50 ml
75mM $CaCl_2$	Sigma C7902	147.01	551.28 mg

To prepare 50 ml of working buffer:

- Weight 210 mg NaCl
- Add 10ml of 5X VB
- Add 100ul of $CaCl_2$ (500X)
- Add 250ul MgCl (200X)
- Adjust volume to 50 ml with H_2O
- Adjust pH to 7.4

[00453] Data was analyzed using GraphPad Prism5 software. Data sets from each experiment were normalized to percent activation compared to the 100% activation control corresponding to the well to which no compound is added. Drug concentration values (X values) were transformed to their logarithms, and percent activation (Pa) (Y values) was transformed to percent inhibition (Pi) using the following formula Pi=100-Pa (Yi=100-Ya). The percent inhibition was plotted against the drug concentration and the resulting data set was fit to a sigmoidal-dose response function [Y=Bottom+(Top-Bottom)/(1+10 ((Log EC-X)))]. IC_{50} values were obtained from the fit parameters.

[00454] Results are presented in Figure 1, and the IC_{50} values are shown in Table 2 (in Example 2). As indicated, CA28-1 and CA28-2 displayed about 30% of the activity of CA28 on a molar basis.

[00455] Example 2: Development of Long-Acting Compstatin Analogs that Demonstrate Increased Molar Activity

[00456] Eight-arm NHS-ester activated PEG with molecular weight of 40 kD (NOF America Corp. White Plains, NY, Cat. No. SUNBRIGHT® HGEO-400GS; chemical formula: hexaglycerol octa(succinimidyloxyglutaryl) polyoxyethylene) was coupled to the lysine side chain of CA28-AEEAc-Lys, resulting in long-acting compstatin analogs represented as follows: $(CA28\text{-}AEEAc)_8$-PEG40k, also referred to as CA28-3. Complement inhibiting activity of CA28-3 was tested using the assay described in Example 1. Results are plotted in Figure 1 and IC50 value is listed in Table 2, both as a function of CA28 concentration. The concentration of CA28 was calculated using the extinction coefficient of CA28 at 283 nm (10208.14 L·mol-1·cm-1). Based on other analysis (UV absorption vs. mass of material, and elemental CHN% analysis) it was concluded that there are 7.5 CA28 moieties per molecule of CA28-3. Thus, the activity of CA28-3 on a molar basis is 7.5-fold higher than shown in Figure 1 and Table 2. Thus, the IC50 value in Table 2 is 7.5-fold higher than the actual IC50 of CA28-3 on a molar basis. The IC50 of CA28-3 on a molar basis is calculated as about 0.26 (lower than that of the parent compound CA28). Figure 2 shows percent complement activation inhibiting activity of CA28 and long-acting compstatin analogs CA28-2 and CA28-3, as a function of CA28-3 concentration (μM), i.e., the activity of CA28-3 has been corrected to account for the fact that the compound contains 7.5 CA28 moieties. On a molar basis, the complement inhibiting activity of CA28-3 exceeds that of CA28.

Table 2

	CA28	CA28-1	CA28-2	CA28-3
IC50	0.3909	1.264	1.288	1.927

[00457] The solubility of CA28-1, CD28-2, and CA28-3 in water with or without a variety of buffer substances and/or excipients was observed to exceed that of the parent compound CA28.

[00458] Example 3: Long-Acting Compstatin Analogs that Demonstrate Dramatically Increased Plasma Half-Life and Cmax

[00459] This Example describes determination of pharmacokinetic parameters of long-acting compstatin analogs CA28-2 and CA28-3 following administration to Cynomolgus monkeys.

[00460] Dosing and Sample Collection

[00461] CA28-2 and CA28-3 were administered at time 0 via intravenous injection into female Cynomolgus monkeys (three per group, 2-5 yrs old, 2.9-3.5 kg). Compounds were administered at 50 mg/kg in 5% dextrose in water at a concentration of 25 mg/ml. Blood specimens (~1 mL each) were collected from the femoral vein at the following timepoints: Pre-dose, 5 min, 15 min, 30 min, 1 hour (h), 4 h, 8 h, 24 h, 48 h, 96 h (4days), and 192 h (8 days) post dose. Specimens were collected via direct venipuncture and placed into a red top serum tube containing no anticoagulant, and kept at room temperature for at least 30 minutes. Blood samples were centrifuged at a temperature of 4°C at 3000 x g for 5 minutes. Samples were maintained chilled throughout processing. Serum samples were collected after centrifugation and placed into sample tubes. Samples were stored in a freezer set to maintain -60°C to -80°C. All animals showed normal activity throughout the study. No compound-related abnormalities were noted in the animals throughout the study.

[00462] Sample Analysis. Plasma samples obtained as described above were analyzed by LC/MS/MS using the following methods to determine the concentration of compound: 50 µL of sample was mixed with internal standard (CA28-AEEAc-Arg) and then 100 µL of 1 M NH_4OAc, pH 3.5 with HOAc was added and mixed. Then 250 µL of acetonitrile was added and mixed. The sample was centrifuged and supernatant poured into another tube and dried. The sample was reconstituted and injected onto the LC/MS/MS system. Mobile phase A was 5 mM NH_4OAc with 0.1% FA and Mobile Phase B was 90:10 (ACN:50 mM NH_4OAc) with 0.1% FA. The LC column was the Intrada WP-RP 2x150 mm, 3 µ. Quantitation was on an Applied Biosystems API-4000 triple quadrupole mass spectrometer operated in positive ion mode. In-source collision induced dissociation (CID) was used to fragment the compound in the mass spectrometer source and the *m/z* 144 ion was mass selected in Q1, fragmented, and the *m/z* 77 ion mass selected in Q3 and detected. Data was processed using Analyst 1.4.2 software.

[00463] Results. The serum concentrations in micrograms/ml of CA28-2 and CA28-3 at each time point are presented in Table 3 below. Data for each of 3 monkeys that received the indicated compound are shown. Average values and standard deviations are readily

calculated. There was notable consistency between animals. CA28 are historical data obtained in a previous study in which CA28 was administered intravenously to Cynomolgus monkeys. In that study, CA28 was detected in samples using HPLC.

Table 3

	Serum Concentration in ug/mL						
Time (days)	CA28-3(8-arm 40k PEG) (50 mg/kg)			CA28 (200 mg/kg)	CA28-2 (linear 40k PEG) (50 mg/kg)		
0.0035	1600	1330	1300		1460	1660	1610
0.01	1600	1220	1480		1360	1430	1530
0.02	1510	1170	1270	34	1310	1510	1500
0.042	1270	1030	1220	17	1290	1340	1540
0.167	926	893	934	9	1200	1210	1390
0.333	797	714	792	5		1190	1180
1	621	479	558	1.5	927	853	881
2	384	355	360		612	733	760
4	280	252	262		461	458	424
8	151	136	136		268	282	293

[00464] Results for each compound were averaged and are plotted in Figure 3. A remarkable increase in half-life and Cmax was observed for both CA28-2 and CA28-3 compared to CA28. The terminal half-lives of both CA28-2 and CA28-3 were around 4 – 4.5 days. Based on these data, it is expected that intravenous administration at approximately 1-2 week dosing intervals will provide sustained levels of compound and effectively inhibit complement activation in human subjects, though shorter or longer dosing intervals may be used.

[00465] Example 4: Long-Acting Compstatin Analog Comprising HSA as a Clearance Reducing Moiety

[00466] Side chain lysines of human serum albumin (HSA) were converted to thiols using 2-iminothiolane and reacted with a compstatin analog comprising a maleimide as a reactive functional group: *Ac*-Ile-Cys*-Val-Trp(1-Me)-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-AEEAc-Lys-(C(=O)-$(CH_2)_5$-Mal)-NH_2 (SEQ ID NO: 68). The resulting long-acting compstatin analog (CA28-4) was tested in vitro for complement inhibiting activity (Figure 4) as described in Example 1 and in vivo for pharmacokinetic properties as described in Example 3. Pharmacokinetic parameters of CA28-4 following administration to Cynomolgus monkeys were determined as described in preceding example. Results are shown in Figure 5

(along with results for CA28, CA28-1, CA28-2, and CA28-3). PK data for CA28-4 are presented in Table 4.

Table 4

Serum Concentration in ug/mL

Time (hr)			
0.0035	1790.0	1445.00	1395.00
0.0100	1195.0	915.50	885.00
0.0200	900.0	504.50	553.50
0.0420	449.0	267.50	295.00
0.1670	194.0	164.00	158.50
0.3330	150.0	163.00	119.50
1.0000	97.2	86.00	78.05
2.0000	73.3	51.55	57.40
4.0000	43.1	29.20	34.15
8.0000	24.1	16.25	20.00

[00467] Example 5: Synthesis and Activity of PEG-Based Compstatin Analogs Using Different NHS-Activated PEGs

[00468] A compstatin analog having the amino acid sequence of the compstatin analog of SEQ ID NO: 28, but incorporating an AEEAc-Lys moiety located C-terminal to the Thr residue of SEQ ID NO: 28 for purposes of subsequent conjugation of an NHS ester activated PEG to the amino group of the Lys side chain was synthesized as described in Example 1. The resulting compstatin analog is represented as follows *Ac*-Ile-Cys*-Val-(1Me)Trp-Gln-Asp-Trp-Gly-Ala-His-Arg-Cys*-Thr-AEEAc-Lys-NH_2 (SEQ ID NO: 51), abbreviated as CA28-AEEAc-Lys. Monofunctional, linear monomethoxy-NHS-activated ester/carbonate PEGs with molecular weights of 40 kD and differing in terms of the NHS carboxylate attachment chemistry (NOF America Corp. White Plains, NY, Cat. Nos. SUNBRIGHT® ME-400CS, SUNBRIGHT® ME-400GS, SUNBRIGHT® ME-400HS, SUNBRIGHT® ME-400TS) were coupled to the lysine side chain of CA28-AEEAc-Lys via an amide bond. (The Lys reside is Lys15 since the AEEAc linker contains an amino acid residue.) All compounds were acetylated on the N-terminus, amidated on the C-terminus, and cyclized via a disulfide bond between Cys2 and Cys12. (The acetylation, amidation, and cyclization were performed prior to coupling to the PEG.) The compounds were prepared as trifluoroacetate

salts and were purified. The compounds are represented as shown in the following table (Table 5). The letters CS, GS, HS, and TS represent the different linker moieties between the PEG moiety and the NHS moiety as indicated in further detail in Table 5. It will be understood that various names and abbreviations for each compound may be used interchangeably. Note that CA28-2 (see Example 1) is the same as CA28-2GS.

Table 5: PEG-Based Compstatin Analogs Containing One Compstatin Analog Moiety

Compound Abbreviation and ID	Compound Name*,†	Activated PEG
CA28-2CS	CA28-AEEAc-LysCS = CA28-AEEAc-Lys(mPEG40K-succinyl)	PEG40K: Methoxy-PEG-CO(CH_2)$_2$COO-NHS (NOF Sunbright 400CS) Chemical Name: α-Succinimidyloxysuccinyl-ω-methoxy, polyoxyethylene CAS#:78274-32-5
CA28-2GS (also referred to as CA28-2)	CA28-AEEAc-LysGS CA28-AEEAc-Lys(mPEG40K-pentanedioyl)	PEG40K: Methoxy-PEG-CO(CH_2)$_3$COO-NHS (NOF Sunbright 400GS) Chemical Name: α-Succinimidyloxyglutaryl-ω-methoxy, polyoxyethylene CAS#: 111575-54-3
CA28-2HS	CA28-AEEAc-LysHS CA28-AEEAc-Lys(mPEG40K-hexanoyl)	PEG40K: Methoxy-PEG-(CH_2)$_5$COO-NHS (NOF Sunbright 400HS) Chemical Name: Poly(oxy-1,2-ethanediyl), α-methyl-ω-{2-[(2,5-dioxo-1-pyrrolidinyl)oxy]-6-oxohexyloxy}-
CA28-2TS	CA28-AEEAc-LysTS CA28-AEEAc-Lys(mPEG40K-carbonyl)	PEG40K: Methoxy-PEG-COO-NHS (NOF Sunbright 400TS) Chemical Name: α-Succinimidyl carbonyl-ω-methoxy, polyoxyethylene CAS# 135649-01-3

*AEEAc=8=Amino-3,6-dioxa-octanoyl

†Compounds were prepared as trifluoroacetate salts but other counterions could be used

[00469] Compounds were analyzed by reverse phase HPLC. Figure 6 shows a representative chromatogram for one of the compounds. A VariTide RPC column was used. Eluent A was 0.1% TFA in water; Eluent B was 0.1% TFA in 50% CAN/40% water. Flow rate was 1.000 ml/min with a gradient of 0%B to 100%B over 40 minutes.. The peak with a retention time of 33.68 minutes represents the PEGylated compound and has a relative area of 96.50%.

[00470] Inhibitory activity of the compounds was assessed by measuring the effect of the compounds on complement activation via the classical pathway using a standard complement inhibition assay as described in Example 1. Results are plotted in Figure 7.

These results represent a combination of two separate experiments. The compounds showed notably similar complement inhibiting activity.

[00471] Example 6: Synthesis and Activity of Bifunctionalized PEG-Based Compstatin Analogs

[00472] Bifunctional, linear monomethoxy-NHS-activated ester/carbonate PEGs with molecular weights of 40 kD and differing in terms of the NHS carboxylate attachment chemistry were obtained from NOF America Corp. (White Plains, NY). The activated PEGs were coupled to the lysine side chain of CA28-AEEAc-Lys via an amide bond such that two CA28-AEEAc-Lys moieties were coupled to each PEG chain. All compounds were acetylated on the N-terminus and amidated on the C-terminus of the CA28-AEEAc-Lys moieties, and cyclized via a disulfide bond between Cys2 and Cys12. (The acetylation, amidation, and cyclization were performed prior to coupling to the PEG.) The compounds were prepared as acetate salts and were purified. The compounds are represented as shown in the following table (Table 6).

[00473] Table 6: Bifunctionalized PEG-Based Compstatin Analogs

Compound Abbreviation and ID	Compound Name[*,†]	Activated PEG
CA28-2CS-BF	CA28-AEEAc-LysCS = CA28-AEEAc-Lys(mPEG40K-succinyl)	PEG40K: NHS-OCO(CH_2)$_2$COO-PEG-CO(CH_2)$_2$COO-NHS Chemical Name: α-Succinimidyloxysuccinyl-ω-succinimidyloxysuccinyloxy, polyoxyethylene CAS#: 85419-94-9
CA28-2GS-BF	CA28-AEEAc-LysGS CA28-AEEAc-Lys(mPEG40K-pentanedioyl)	PEG40K: NHS-OCO(CH_2)$_3$COO-PEG-CO(CH_2)$_3$COO-NHS Chemical Name: α-Succinimidyloxyglutaryl -ω-succinimidyloxyglutaryloxy-, polyoxyethylene CAS#: 154467-38-6
CA28-2HS-BF	CA28-AEEAc-LysHS CA28-AEEAc-Lys(mPEG40K-hexanoyl)	PEG40K: NHS-OCO(CH_2)$_5$O-PEG-(CH_2)$_5$COO-NHS Chemical Name: α-[6-[(2,5-dioxo-1-pyrrolidinyl)oxy]-6-oxohexyl]-ω-[6-[(2,5-dioxo-1-pyrrolidinyl)oxy]-6-oxohexyloxy]-, polyoxyethylene
CA28-2TS-BF	CA28-AEEAc-LysTS CA28-AEEAc-Lys(mPEG40K-carbonyl)	PEG40K: NHS-OCO-PEG-COO-NHS Chemical Name: α-Succinimidyl carbonyl-ω-Succinimidyl carbonyl, polyoxyethylene

*AEEAc=8=Amino-3,6-dioxa-octanoyl
†Compounds were prepared as acetate salts but other counterions could be used

[00474] Inhibitory activity of CA28-2GS-BF was assessed by measuring the effect of the compound on complement activation via the classical pathway using a standard complement inhibition assay as described in Example 1 and analyzed as described in Example 1. Results are plotted in Figure 8. As described above, CA28-2GS-BF contains two compstatin analog moieties per molecule. Although the activity per compstatin analog moiety of CA28-2GS-BF is less than the activity of an individual CA28 molecule, the activity of the two compounds on a molar basis is virtually identical over a broad range of concentrations.

[00475] Example 7: Subcutaneous Administration of Bifunctionalized PEG-Based Compstatin Analog

[00476] This Example describes determination of pharmacokinetic parameters of long-acting compstatin analog CA28-2GS-BF following administration to Cynomolgus monkeys either via a single intravenous (IV) injection or with repeated (once daily) subcutaneous administration for seven days.

[00477] Dosing and Sample Collection CA28-2GS was administered at time 0 via intravenous injection or via repeat subcutaneous injection (daily, for seven days) into male Cynomolgus monkeys. Six non-naïve male Cynomolgus monkeys, age 1-5 years, ranging in weight from 4.6 to 5.3 kilograms, were used in thes study (three per group). The animals were healthy at the start of the trial.seven days. The study was not blinded. Animals were supplied with water ad libitum and a commercial diet twice daily prior to initiation of the study. Food was supplied to the animals per facility SOP prior to the study. Animals were not fasted. Animals were dosed via intravenous and subcutaneous administration at time 0 on the appropriate day. A size 22 gauge needle was used for the subcutaneous administration. The compound was administered at 50 mg/kg in 5% dextrose in water at a concentration of 25 mg/ml. Blood specimens (~1 mL each) were collected from the femoral vein at the following timepoints: Day 1: Pre-dose, 5 min, 15 min, 30 min, 1 hour (h), 4 h, 8 h. Days 2 – 9: 0 min. Day 16: Final sample based on Day 1 dosing. Each blood sample (~1.0 mL) was collected from the monkey's femoral or saphenous vein via direct venipuncture and placed into a red top serum tube containing no anticoagulant, and kept at room temperature for at least 30 minutes. Blood samples were centrifuged at a temperature of 4°C at 3000 xg for 5 minutes. Samples were maintained chilled throughout processing. Serum samples

were collected after centrifugation and placed into sample tubes. Samples were stored in a freezer set to maintain -60°C to -80°C. Serum samples and leftover dosing solutions were shipped frozen on dry ice for analysis.

[00478] The site of each subcutaneous administration was observed to see how fast the injection volume was absorbed and also to see if the formulation left behind a lump or fully went away. The dose sites were observed at each collection timepoint and in the afternoon of days 2-7. All doses were absorbed during the duration of the study. Based on the observations it is estimated that doses were absorbed within fifteen minutes after administration. All animals showed normal activity throughout the study. No compound-related abnormalities were noted in the animals throughout the study.

[00479] Sample Analysis. Plasma samples obtained as described above were analyzed by LC/MS/MS leveraging CID (collision induced degradation) similarly to the method described in Example 3.

[00480] Results. Serum concentration vs time for CA28-2GS-BF when administered IV or subcutaneously as described above are plotted in Figure 9, The data points represent all PEGylated CA28 compound detected. CA28 data shown on Figure 9 are historical data obtained in a previous study in which CA28 was administered intravenously to Cynomolgus monkeys. In that study, CA28 was detected in samples using HPLC.

[00481] A peak serum concentration of 500 μg/ml (11 μM) was achieved by subcutaneous administration of CA28-2GS-BF. The terminal half-life of CA28-2GS-BF was approximately 5 days when administered either IV or subcutaneously. Results are summarized in the tables below:

Table 7
Summary of Study Sample Concentrations for CA28-2GS-BF in Monkey Serum (IV @ 50 mg/kg on Day 0)

	CA28-2GS-BF Conc. (µg/mL)		
Timepoint	Animal 1	Animal 2	Animal 3
5 min	1850	1550	2030
15 min	1760	1440	2000
30 min	1560	1380	1810
1 hr	1650	1330	1710
4 hr	1270	1000	1510
8 hr	1050	913	1360
Day 2	684	661	711
Day 3	541	471	538
Day 4	463	417	492
Day 5	366	384	389
Day 6	346	331	358
Day 7	303	306	311
Day 8	257	252	259
Day 9	217	252	233
Day 15	92.8	107	95.6

Table 8
Summary of Study Sample Concentrations for CA28-2GS-BF in Monkey Serum (SQ @ 7 mg/kg/day x 7 days)

	CA28-2GS-BF Conc. (µg/mL)		
Timepoint	Animal 1	Animal 2	Animal 3
5 min	0	5.18	3.55
15 min	0	5.57	3.47
30 min	0	4.66	3.93
1 hr	0	5.41	3.56
4 hr	4.41	12.6	11.1
8 hr	32.0	15.3	21.5
Day 2	54.2	56.6	53.5
Day 3	135	117	122
Day 4	248	260	234
Day 5	398	316	311
Day 6	447	391	419
Day 7	564	412	448
Day 8	591	432	468
Day 9	596	423	455
Day 15	152	241	199

[00482] Example 8: Inhibition of Complement-Mediated Lysis of Red Blood Cells from Patients with PNH

[00483] A modified Ham's test is performed to measure the ability of compstatin analogs to inhibit complement-mediated lysis of red blood cells from patients with PNH in vitro. Complement is activated by acidified serum with added magnesium to lyse the PNH red cells. The incubation is performed for 90 minutes. The read out is flow cytometry for PNH red cells using standard markers. Heat inactivated serum is used as a control (produces no hemolysis). Acidified serum in the absence of added complement inhibitor produces maximum lysis. The experiment is performed with serial two-fold dilutions of compstatin analogs CA28, CA28-2, CA28-2CS, CA28-2CS-BD, CA28-2GS, CA28-2GS-BF, CA28-2HS, CA28-2HS-BF, CA28-2TS, CA28-2GS-BF, and CA28-3. The concentration of each compound required to fully block hemolysis in vitro is determined. Red blood cells are also stained for C3 fragment deposition using anti-C3 polyclonal antibodies that do not contain any bridge leading to agglutination (e.g., either Ab4214 or Ab14396, both commercially available FITC-conjugated Abcam, Cambridge, United Kingdom) in order to measure the

ability of the compounds to inhibit deposition of C3 fragments on PNH red blood cells. Results are compared with those obtained with eculizumab using the same assays.

[00484] Example 9: Long-Acting Compstatin Analogs in Patients with PNH

[00485] A cohort of subjects diagnosed with PNH is divided into 4 groups. Subjects in Groups 1 and 2 are treated with intravenous administration of CA28-2 or CA28-3, respectively, at a dose of between 5 mg/kg and 20 mg/kg, at time intervals between 1 and 2 weeks. Optionally, treatment is started at more frequent time intervals and then reduced in frequency for maintenance therapy. Subjects in Group 3 are treated with eculizumab according to the recommended dosing regimen. Group 4 serves as a control (no complement inhibitor therapy). Intravascular hemolysis (based on LDH measurement and/or (51)Cr labeling of RBCs), reticulocytosis (an indicator of anemia), hematocrit, hemoglobin concentration in the blood, opsonization of red blood cells (deposition of products of C3 activation, such as C3b, on red blood cells, which may be detected using flow cytometry), PNH symptoms, transfusion requirements, thromboembolic events, haemolysis-associated nitric oxide depletion, measures of pulmonary hypertension, quality of life, and survival are monitored over time. Results are compared between groups and with historical data from control PNH patients obtained in clinical trials of eculizumab. An improvement in persistent anemia (e.g., as evidenced by reduced reticulocytosis, reduced evidence of hemolysis, increased hematocrit, increased hemoglobin), improved quality of life, reduced PNH symptoms, reduced transfusion requirements, reduced thromboembolic events, reduced haemolysis-associated nitric oxide depletion, reduced measures of pulmonary hypertension increased quality of life, and/or increased survival, in subjects receiving CA28-2 (Group 1) or CA28-3 (Group 2), as compared with subjects in Group 4 is indicative of efficacy.

[00486] Example 10: Long-Acting Compstatin Analogs in Patients with PNH

[00487] Example 9 is repeated with the modification that subjects are individuals with PNH who remain transfusion-dependent and/or continue to have a hemoglobin below a cutoff (such as 9.0 g/dL) despite treatment with eculizumab. Results are compared among groups.

[00488] Example 11: Long-Acting Compstatin Analogs in Patients with aHUS

[00489] A cohort of subjects diagnosed with aHUS is divided into 4 groups. Subjects in groups 1 and 2 are treated with intravenous administration of CA28-2 or CA28-3, respectively, at a dose of between 5 mg/kg and 20 mg/kg, at time intervals between 1 and 2

weeks. Optionally, treatment is started at more frequent time intervals and then reduced in frequency for maintenance therapy. Subjects in Group 3 are treated with eculizumab according to the recommended dosing regimen. Intravascular hemolysis (based on LDH measurement), opsonization of red blood cells (deposition of products of C3 activation, such as C3b, on red blood cells), aHUS symptoms, renal function, need for plasma exchange or dialysis, quality of life, and survival are monitored over time. Results are compared between groups and with historical data from control aHUS patients obtained in clinical trials of eculizumab. Reduced evidence of hemolysis, improved quality of life, reduced aHUS symptoms, reduced need for plasma exchange or dialysis, increased quality of life, and/or increased survival, in subjects receiving CA28-2 or CA28-3, as compared with subjects in group 4 are indicative of efficacy.

[00490] **Example 12**: Examples 8 – 11 are repeated using CA28-2GS-BF, CA28-2HS, CA28-2HS-BF, CA28-2TS, and CA28-2GS-TS-BF.

[00491] **Example 14**: Examples 9 – 12 are repeated using CA28-2GS-BF, CA28-2HS, CA28-2HS-BF, CA28-2TS, and CA28-2GS-TS-BF administered daily by subcutaneous injection.

[00492] **Example 14**: Examples 8-11 are repeated using additional long-acting compstatin analogs.

[00493] **Example 15**: Examples 8-11 are repeated using cell-reactive compstatin analogs.

[00494] **Example 16**: **Complement Activation Inhibiting Activity of a Long-Acting Compstatin Analog**

[00495] CA28 and CA28-AEEAc-Lys were synthesized as described above. CA28-2TS-BF was synthesized using a reactive bifunctional PEG of the TS type in terms of the NHS carboxylate attachment chemistry, which was linked to two molecules of CA28-AEEAc-Lys via the primary amine of the lysine side chain. The complement activation inhibitory activity of CA28 and CA28-2TS-BF was assessed by measuring the effect of the compounds on complement activation via the classical pathway and via the alternative pathways using standard complement inhibition assays. The protocol for the classical pathway activation assay is described in Example 1. The protocol for alternative pathway activation also measures C3b deposition in an ELISA format and is described below. C3b deposition monitored using this method is generated through complement activated by the alternative pathway by lipopolysaccharide (LPS). Briefly, 96-well plates are coated with LPS. Compound being tested (referred to as "drug") is added, followed by addition of plasma or serum as a source of complement, and incubated. This is followed by addition of anti-

human C3 HRP-conjugated antibody. After an additional incubation, substrate is added and signal detected. Details of the protocol are as follows:

[00496] ELISA-based Assay for Alternative Complement Pathway Activation

Materials:

- Ninety six-well ELISA plate (Corning 3590)
- LPS from *Salmonella typhosa* – Sigma L7136 (40ug/ml in PBS)
- BSA 1% in PBS – Calbiochem #126626 1/30 dilution
- Veronal Buffer + 10 mM $MgCl_2$ + 10 mM EGTA (VB-Mg EGTA)
- Human plasma (collected with Lepirudin at 5ug/ml final concentration)
- Anti-human C3 HRP-conjugated Ab (Poli to C3-HRP Ab, Cappel 55237)
- Tween-20 Wash Buffer (0.05% in PBS)
- TMB (Peroxidase substrate) – 1:1 mixture of BD 51-2607KC and 51-2606KC.
- 3M H_2SO_4
- Micro-plate Reader

Protocol:

1. Add 50ul/well of LPS at 40ug/ml (in PBS)
2. Incubate for 2 hours at room temp
3. Remove by shaking and tapping the plate.
4. Block by adding 200ul of 1% BSA/PBS
5. Incubate for 1 h at room temp
6. Remove by shaking and tapping the plate
7. Add 50 ul VB-Mg EGTA to wells #2 to 12
8. Add 100ul of starting drug dilution (2x in VB-Mg EGTA) to well 1.
9. Serially dilute (1:2) the drug from wells 1 to 10 as follow
    a. Take 50ul of solution from the originating well
    b. Add this to the next well
    c. Mix by pipetting several times
    d. Repeat up to well #10

Note: from well #10 remove 50ul and discard.

10. Add 50ul of 2x plasma dilution to wells 1 to 11
11. Incubate for 1h
12. Wash twice with wash buffer
13. Add 50ul of 1/1000 dilution of C3-HRP Ab in 1% BSA/PBS
14. Incubate for 1h
15. Add 100ul of TMB to all wells
16. Incubate for 30min
17. Add 50 ul 3M H_2SO_4
18. Read the plate at 450nm

Formula for VB Mg EGTA

Barbital	5 mM
NaC1	72.5 mM
MgC12	10 mM
EGTA	10 mM
pH	7.3-7.4

Stock solutions:

Veronal Buffer (5X)

	Prod #	MW	For 500ml
9 mM Sodium Barbitone	Sigma B0500	206.17	927 mg
15.5 mM diethylbarbituric acid	Sigma B0375	184.19	1.42 grams

Mg-Cl2 (10X)

	Prod #	MW	For 50 ml
100 mM $MgCl_2$-$6H_2O$	Sigma M0250	203.30	1.00 gram

EGTA (10x)

	Prod #	MW	For 25 ml
100mM EGTA	Sigma E8145	468.3	1.17 grams

To prepare 20 ml of working buffer:

- Weight 84 mg NaCl
- Add 4ml of 5X VB
- Add 2ml of EDTA 10X
- Add 2ml MgCl 10X
- Adjust volume to 20 ml with H_2O
- Adjust pH to 7.4

[00497] Results

[00498] Figure 10(A) shows percent inhibition of classical complement activation inhibiting activity by CA28 and CA28-2TS-BF as a function of molar concentration of the compounds. Figure 10(B) shows percent inhibition of alternative complement activation inhibiting activity by CA28 and CA28-2TS-BF as a function of molar concentration of the compounds. Raw data are tabulated in Table 9 below (4 replicates of each condition). Based on the inhibition curves shown in the figures and underlying data, the complement inhibiting activity of CA28-2TS-BF is at least as great as that of CA28 on a molar basis within the experimental error of the assay. These results further confirm the suitability of long-acting compstatin analogs described herein, e.g., for therapeutic purposes.

[00499] Table 9

AP Inhibition

% Inhibition

Conc uM	CA28					CA28-2TS-BF		
25	90.07908	89.9353	90.65421	89.21639	90.25768	91.10484	90.68126	90.96365
12.5	88.92883	89.21639	89.21639	90.94177	90.11649	89.12814	89.26933	88.56336
6.25	87.05967	89.07261	88.20992	88.64127	88.28098	87.29263	87.43382	87.43382
3.125	85.62186	87.49101	87.49101	86.77211	84.04518	84.46877	85.17473	85.03353
1.5625	78.81236	81.30841	82.02732	70.38102	77.40911	69.5023	79.80939	81.08012
0.78125		58.01582	55.57153	57.15313	61.73668	66.39605	72.74974	61.3131
0.390625	18.04458	12.4371	23.65205	24.94608	56.79492	44.9347	56.65372	40.59891
0.1953	15.02516		15.31273	15.31273	22.34381	23.89693	26.01483	20.6485
0.09765	2.803741	14.30625	5.685844	5.623158	13.87222	6.953766	10.62479	0.3176956
0.0488	-1.365921	5.679367	0.790802	-2.803734	0.5000748	2.012009	0.3176956	-2.929749
0	-2.803734	-1.509697	1.078362	3.235085	0.6000748	7.094963	2.153198	-0.5294724

CP Inhibition

% Inhibition

Conc uM	CA28					CA28-2TS-BF		
25	83.69539	79.90365		81.00482	82.09877		81.32716	77.31482
12.5	78.65483	67.24019	81.69305	83.75774	83.48766	80.70988	78.54939	76.08025
6.25	84.58362	84.03304	80.45423	81.28011	81.94444	78.39507	79.93827	74.84568
3.125	83.62009	81.59305	79.90365	81.8307	79.93827	77.9321	77.46913	76.69753
1.5625	76.04955	76.80014	79.76601	78.94012	71.60493	70.52469	74.53703	75.61728
0.78125	71.50723	69.85547	73.98486	72.33911	72.37654	68.20988	72.0679	71.14198
0.390625	58.84377	72.05782	68.89195	63.11081	79.16666	70.37037	71.2963	62.19136
0.1953	42.60152	44.39091			56.51234		47.58519	50.92593
0.09765	24.7075	24.15692	23.05575	35.03098	41.66667	35.03086	48.91975	42.12963
0.0488	15.7605	12.59464	15.20992	27.04749	12.50001	26.23457	26.23457	23.61111
0	-20.99105	7.088783	12.04405	1.858231	5.09259	-0.1543198	-0.9259262	-4.012352

[00500] Example 17: Pharmacokinetic Properties of Long-Acting Compstatin Analog Administered by the Intravenous or Subcutaneous Route

[00501] This Example describes determination of pharmacokinetic parameters of long-acting compstatin analog CA28-2TS-BF following administration to Cynomolgus monkeys with a single intravenous (IV) injection, single subcutaneous administration, or with once daily subcutaneous administration for seven days. CA28-2TS-BF was synthesized using a reactive bifunctional PEG of the TS type in terms of the NHS carboxylate attachment chemistry, which was linked to two molecules of CA28-AEEAc-Lys via the primary amine of the lysine side chain.

[00502] Dosing and Sample Collection

[00503] CA28-2TS-BF was administered to Cynomolgous monkeys at time 0 via intravenous injection into the saphenous vein or via single subcutaneous injection or repeat subcutaneous injection (once daily, for seven days). Six

non-naïve female Cynomolgus monkeys, age 2-5 years, ranging in weight from 2.6 to 3.9 kilograms, were used in this study (three per group). The animals were healthy at the start of the trial. The study was not blinded. Animals were supplied with water ad libitum and a commercial diet twice daily prior to initiation of the study. Food was supplied to the animals per facility SOP prior to the study. Animals were not fasted. Animals were dosed at 7 mg/kg via intravenous or subcutaneous administration at time 0 on the appropriate day(s). Dosing solution concentration was 3.5 mg/mL for IV administration and 25 mg/mL for subcutaneous administration. Dosing volume was 2 mL/kg for IV administration and 0.28 mL/kg for subcutaneous administration. A size 23G3/4 gauge needle was used for subcutaneous administration. The compound was administered in 5% dextrose in water.

[00504] Blood specimens (~0.5 - 1 mL) were collected from the femoral vein at the following timepoints: Day 1: Pre-dose, 5 min, 15 min, 30 min, 1 hour (h), 4 h, 8 h. Days 2 – 9: 0 min. Day 15: Final sample based on Day 1 dosing. Each blood sample was collected from the monkey's femoral vein via direct venipuncture and placed into a red top serum tube containing no anticoagulant, and kept at room temperature for at least 30 minutes. Blood samples were centrifuged at a temperature of 4°C at 3000 xg for 5 minutes. Samples were maintained chilled throughout processing. Serum samples were collected after centrifugation and placed into sample tubes. Samples were stored in a freezer set to maintain -60°C to -80°C. Serum samples and leftover dosing solutions were shipped frozen on dry ice for analysis.

[00505] The site of each subcutaneous administration was observed to see how fast the injection volume was absorbed and also to see if the formulation left behind a lump or fully went away. The dose sites for the animals receiving subcutaneous injections were observed in the evening of each dosing day. The dose site did not appear to have a lump and was fully absorbed by that time based on visual inspection. All animals were observed twice daily and showed normal activity throughout the study. No compound-related abnormalities were noted in the animals throughout the study.

[00506] Sample Analysis. Plasma samples obtained as described above were analyzed by LC/MS/MS leveraging CID (collision induced degradation) similarly to the method described in Example 3.

[00507] Results.

[00508] Serum concentrations vs time for CA28-2TS-BF when administered IV or subcutaneously as described above are plotted in Figure 11. The data points represent all PEGylated CA28 compound detected. CA28 data shown on Figure 11 are historical data

obtained in a previous study in which CA28 was administered intravenously to Cynomolgus monkeys. In that study, CA28 was detected in samples using HPLC/MS.

[00509]

[00510] A peak serum concentration of about 500 micrograms/mLwas achieved by subcutaneous administration of CA28-2TS-BF once daily for 7 days. The terminal half-life of CA28-2TS-BF was approximately 8 days when administered either IV or by single subcutaneous injection. Raw data are provided in Tables 10(A) (IV administration) and 10(B) (subcutaneous administration) below. (In Figure 11 and Tables 10(A) and 10(B), the day of dosing is considered day 0).

Table 10(A)

Time (days)	CA28 (IV) 200 mg/kg	CA28-2TS-BF (single dose IV) 7 mg/kg		
0.0035		232	190	214
0.01		216	190	209
0.02	34	221	177	199
0.042	17	211	175	183
0.167	9	190	152	185
0.333	5	212	191	154
1	1.5	180	130	150
2		141	116	126
3		128	98.7	113
4		114	89.1	95.8
5		105	75	87.1
6		95.1	67	74.4
7		83.3	61.4	69.9
8		86	52.8	68.4
14		51	30.8	39.7

Table 10(B)

Time (days)	CA28-2TS-BF (single dose SC) 7 mg/kg			CA28-2TS-BF (7 x daily SC) 7 mg/kg/day		
0.0035	BQL	BQL	BQL	BQL	BQL	BQL
0.01	1.42	BQL	1.7	BQL	BQL	BQL
0.02	3.55	1.64	3.8	BQL	BQL	BQL
0.042	6.1	3.46	7.05	2.32	1.14	2.44
0.167	15	12.2	20.6	14.7	5.76	12.1
0.333	25	25.8	32.9	32.9	18	25.8
1	70	76.5	76	80.5	80	66.5
2	107	101	96.6	196	185	169
3	111	103	99.8	391	286	292
4	108	98.9	99.5	455	377	405
5	99.4	97.6	101	427	404	486
6	86.8	87	81.6	490	483	568
7	75.2	83.2	78.6	607	502	564
8	67.5	73.4	72.2	495	481	570
14	38.3	44.5	40.7	322	298	397

BQL = below quantification limit

[00511]

[00512] As noted above, CA28-2TS-BF was synthesized using a reactive bifunctional PEG of the TS type, resulting in formation of a carbamate after reaction with the primary amine of lysine. CA28-2GS-BF was synthesized using a reactive bifunctional PEG of the GS type in terms of the NHS carboxylate attachment chemistry, resulting in formation of an amide after reaction with the primary amine of lysine. The compound also contains an ester linkage, which is absent in CA28-2TS-BF. It is notable that the terminal half-life of about 8 days achieved with CA28-2TS-BF in this experiment was considerably greater than that of CA28-2GS-BF, which was found to have a half-life of about 5 days in a similar experiment (see Example 8).

[00513] Example 18: Compstatin Analogs Inhibit C3 Deposition on Red Blood Cells of PNH Patientsand Protects against Complement-mediated Lysis

[00514] A modified Ham's test was performed to assess the ability of compstatin analogs to protect PNH RBCs from complement-mediated lysis. RBCs from a patient with PNH were exposed to acidified human serum (as a source of complement components) and magnesium (Mg^{2+}, needed for alternative pathway activation) in the absence of complement inhibitors or in the presence of varying amounts of compstatin analogs CA28 or CA28-2GS-BF. Exposure to heat inactivated human serum was used as a control representing no significant complement-mediated lysis as complement is inactivated by heat. Exposure to acidified human serum and magnesium (Mg^{2+}) in the absence of complement inhibitors (pane labeled Mg^{2+}) was used as a control representing maximum lysis.

[00515] Following incubation, cells were stained with antibodies to CD59 and C3d. Cd59 level permitted the classification of the PNH RBCs as Type I, Type II, or Type III. Staining for C3d, a product of C3 activation and cleavage was used as a marker of C3 and C3 activation product deposition (loading). Flow cytometric analysis was performed to assess CD59 and C3d on RBC surfaces and to quantify the percentages of Type I, Type II, and Type III cells present in various samples.

[00516] Results of a dilution experiment demonstrating the effect of different concentrations of CA28 on C3 deposition and cell percentages are shown in Fig. 12(A). Results of a dilution experiment demonstrating the effect of different concentrations CA28-2GS-BF on C3 deposition and cell percentages are shown in Fig. 12(B). The results are presented quantitatively in Table 11 below. Type I cells (shown in orange in Fig. 12) have normal levels of CD59. Type III cells (shown in blue in Fig. 12) have essentially no detectable CD59. These cells are very susceptible to complement-mediated lysis. Type II

cells (shown in purple in Fig. 12) have reduced levels of CD59 as compared with normal or Type I cells and have an intermediate sensitivity to complement-mediated lysis. In the presence of complement activation Type III cells rapidly lyse. Reduction in or absence of lysis may be evidenced by an increased presence of Type III cells, as is evident as a higher percentage of Type III cells in the no lysis panel compared with the panel in(presence of Mg^{2+} (max. lysis)) in both Figures 12(A) and 12(B). In other words, there are relatively fewer Type III cells in the positive control than in the negative control. Type II cells may eventually lyse in the presence of activated complement but can accumulate a considerable amount of C3 activation products such as as C3d before they do so. Reduction in or absence of lysis may be evidenced by increased levels of C3 or C3 activation products on Type II cells, as is evident by comparing the level of C3d on Type II cells in the no lysis panel with the level of C3d on Type II cells in the max. lysis panel in both Figures 12(A) and 12(B). In other words, there is more C3d on cells in the max. lysis panel than in the no lysis panel. Type I cells have functional CD59, so they deactivate convertase and therefore do not accumulate as much C3d as Type II cells. However, the amount of C3d they accumulate can be used as a surrogate indicator for the amount of lysis of the more vulnerable cells (Type II and III). Thus, reduced C3d on Type I cells is indicative of protection against lysis. A shift in the relative percentages of Type I, II, and III cells from the percentages present in the max. lysis control panels (Mg^{2+}) towards the percentages present in the no lysis control panels (heat inactivated serum) is indicative of protection against complement-mediated lysis. These percentages are shown in the table below. The column labeled %C'3 in Table 11 refers to the percentage of cells deemed "positive" for presence of C3 and C3 activation products ("C3 loading"). As can be seen in Figures 12(A) and (B) and Table 11, CA28 and CA28-2GS-BF demonstrated similar protection of PNH red cell lysis over the concentrations tested, with virtually no C3 loading on PNH red cells at 100 micrograms/ml compound or higher concentrations. Note that the percentages of Type III, II, and I cells in the presence of 100 ug/ml or more compstatin analog were essentially the same as in the no lysis control, indicating complete protection from complement-mediated lysis as determined by this assay. Concentrations below 100 ug/ml but above 60 ug/ml, e.g., at least 70 ug/ml, at least 80 mg/ml, or at least 90 ug/ml, but below 100 ug/ml were not tested in this experiment but may also provide significant protection. 100 micrograms/ml CA28-2GS-BF represents a concentration of about 2.5 micromolar, which is readily achievable in vivo as described herein.

[00517] Table 11: Percentages and C3 loading of Type I, II, and III PNH RBCs in the absence or presence of compstatin analogs (concentrations in micrograms/ml are shown)

	Type III %	% C3	Type II %	% C3	Type I %	%C3
Heat inactivated (no lysis)	37.09	0.29	51.79	1.22	11.12	0.08
Magnesium added (maximum lysis)	15.75	0.66	64.7	13.48	19.55	3.2
CA28 1	19.42	1.2	64.81	10.08	15.77	1.7
CA28 4	18.43	0.83	64.8	10.42	16.68	1.78
CA28 8	17.83	0.7	66	11.92	16.17	1.88
CA28 15	22.11	0.86	62.9	8.83	15	1.83
CA28 25	20.53	0.75	64.51	8.3	14.96	1.84
CA28 100	37.96	0.1	51.5	0.92	10.64	0.08
CA28 500	37.29	0.09	52.06	0.83	10.64	0.14
CA28-2GS-BF 20	19.19	0.69	65.35	12.28	15.46	2.19
CA28-2GS-BF 40	15.15	1.21	68.96	15.12	15.89	2.94
CA28-2GS-BF 50	13.87	1.29	69.46	15.94	16.67	2.28
CA28-2GS-BF 60	17.94	1.44	66.97	13.04	15.09	2.05
CA28-2GS-BF 100	35.32	0.11	53.93	0.86	10.75	0.07
CA28-2GS-BF 200	37.43	0.02	51.99	0.28	10.58	0.02
CA28-2GS-BF 500	37.87	0.05	51.87	0.3	10.26	0.04

[00518] Example 19: Effect of Compstatin Analog and Soliris on C3 Deposition on Red Blood Cells from PNH Patient

[00519] A similar experiment to that described in Example 18 was performed to further demonstrate the protective effect of compstatin analog CA28-2GS-BF and compare it with that of anti-C5 antibody Soliris. A modified Ham's assay as in Example 18 was performed using PNH RBCs incubated in the presence of activated complement either in the absence of complement inhibitor (left panel) or in the presence of Soliris (middle panel) or CA28-2GS-BF (50 ug/ml) (right panel). Flow cytometry was performed after antibody staining using antibodies against CD59 and C3d. Results are shown in Figure 13. In this figure, quadrant 1 (Q1) and quadrant 3 (Q3) represent Type III cells. Quadrant 2 (Q2) and quadrant 4 (Q4) represent Types I and II cells. Q1 and Q2 represent cells with a significant and abnormally high amount of C3 activation product (e.g., C3d) deposition. Q3 and Q4 represent cells without significant C3d deposition or somewhat elevated level (right portion of of Q4) but less so than Q2 cells. The percentages of cells in the different quadrants is presented below each panel in Figure 13 and in Table 12 below.

[00520] **Table 12**

	No Inhibitor			Eculizumab			CA28-2GS-BF	
Population	# Events	%Parent		# Events	%Parent		# Events	%Parent
Q1	233	0.09		90,146	36.79		23	0.01
Q2	7,992	3.22		9,609	3.92		18	0.01
Q3	9,853	3.97		4,591	1.87		153,187	61.54
Q4	230,241	92.72		140,689	57.42		95,713	38.45

[00521] As can be seen, in the absence of inhibitors the great majority of cells lie in Q4 (Type I or Type II with low levels of C3 activation product deposition). Type III cells would have mainly been lysed, so their percentages (Q1 and Q3) are low. Q2 cells that accumulate C3 deposition products eventually lyse, so their number stays relatively low. In the presence of eculizumab, Type III cells are protected from lysis at least initially, but accumulate C3 activation products (e.g., C3d) as shown by the high percentage of Q1 cells as compared with the no inhibitor panel (36.79% vs 0.09%). The relative proportion of Q2+Q4 cells (Type I and II) is lower as a result of the increased survival of Type III cells. However, it is evident that significant deposition of C3 activation products (e.g., C3d) occurs on Type III cells, which may lead eventually to lysis or to clearance (in vivo). PNH RBCs treated with CA28-2GS-BF (right panel) exhibit essentially no deposition of C3d regardless of whether they are Type I, II, or III, in contrast to the results with eculizumab. The percentage of cells in Q1 and Q2 is negligible. There is a dramatic increase in the percentage of Type III cells (61.55%) as compared with results with no inhibitor or with eculizumab, indicating (together with the lack of C3d deposition) enhanced protection from lysis by CA28-2GS-BF.

* * * * *

[00522] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. The scope of the present invention is not intended to be limited to the above Description, but rather is as set forth in the appended claims. It will be appreciated that the invention is in no way dependent upon particular results achieved in any specific example or with any specific embodiment. Articles such as “a”, “an” and “the” may mean one or more than one unless indicated to the contrary or otherwise evident from the context. Claims or descriptions that include “or” between one or more members of a group are considered satisfied if one, more than one, or all of the group members are present in,

employed in, or otherwise relevant to a given product or process unless indicated to the contrary or otherwise evident from the context. The invention includes embodiments in which exactly one member of the group is present in, employed in, or otherwise relevant to a given product or process. For example, and without limitation, it is understood that where claims or description indicate that a residue at a particular position may be selected from a particular group of amino acids or amino acid analogs, the invention includes individual embodiments in which the residue at that position is any of the listed amino acids or amino acid analogs. The invention also includes embodiments in which more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process. Furthermore, it is to be understood that the invention encompasses all variations, combinations, and permutations in which one or more limitations, elements, clauses, descriptive terms, etc., from one or more of the listed claims or from the description above is introduced into another claim. For example, any claim that is dependent on another claim can be modified to include one or more elements, limitations, clauses, or descriptive terms, found in any other claim that is dependent on the same base claim. Furthermore, where the claims recite a composition, it is to be understood that methods of administering the composition according to any of the methods disclosed herein, and methods of using the composition for any of the purposes disclosed herein are included within the scope of the invention, and methods of making the composition according to any of the methods of making disclosed herein are included within the scope of the invention, unless otherwise indicated or unless it would be evident to one of ordinary skill in the art that a contradiction or inconsistency would arise. Methods of treating a subject can include a step of providing a subject in need of such treatment (e.g., a subject who has had, or is at increased risk of having, a disease), a step of diagnosing a subject as having a disease and/or a step of selecting a subject for treatment with a cell-reactive compstatin analog.Where elements are presented as lists, it is to be understood that each subgroup of the elements is also disclosed, and any element(s) can be removed from the group. For purposes of conciseness only some of these embodiments have been specifically recited herein, but the invention includes all such embodiments. It should also be understood that, in general, where the invention, or aspects of the invention, is/are referred to as comprising particular elements, features, etc., certain embodiments of the invention or aspects of the invention consist, or consist essentially of, such elements, features, etc. Discussion of various diseases, disorders, and conditions under various headings herein is for convenience and is not intended to limit the invention.

[00523] Where ranges are given, endpoints are included. Furthermore, it is to be understood that unless otherwise indicated or otherwise evident from the context and understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value or subrange within the stated ranges in different embodiments of the invention, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise. Any particular embodiment, aspect, element, feature, etc., of the present invention may be explicitly excluded from the claims even if such exclusion is not set forth explicitly herein. For example, any compstatin analog, functional group, linking portion, clearance-reducing moiety, disease, or indication can be explicitly excluded.

[0001] Table B1 follows:

XTEN Name	Amino Acid Sequence
AE48	MAEPAGSPTSTEEGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGS
AM48	MAEPAGSPTSTEEGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGS
AE144	GSEPATSGSETPGTSESATPESGPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGSEP ATSGSETPGSEPATSGSETPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGSEPATSG SETPGTSTEPSEGSAP
AF144	GTSTPESGSASPGTSPSGESSTAPGTSPSGESSTAPGSTSSTAESPGPGSTSESPSGTAPGSTSS TAESPGPGTSPSGESSTAPGTSTPESGSASPGSTSSTAESPGPGTSPSGESSTAPGTSPSGESST APGTSPSGESSTAP
AE288	GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTST EPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPT STEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGP GSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEP ATSGSETPGTSESATPESGPGTSTEPSEGSAP
AF504	GASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSXP SASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASPGTS STGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSS PGSSTPSGATGSPGSXPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSPGA SPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGT SSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTG SPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGS STPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSP
AF540	GSTSSTAESPGPGSTSSTAESPGPGSTSESPSGTAPGSTSSTAESPGPGSTSSTAESPGPGTSTP ESGSASPGSTSESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGSTSESPSGTAPGTSPSGESST APGSTSESPSGTAPGSTSESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGSTSESPSGTAPGS TSESPSGTAPGTSTPESGSASPGSTSESPSGTAPGTSTPESGSASPGSTSSTAESPGPGSTSSTA ESPGPGTSTPESGSASPGTSTPESGSASPGSTSESPSGTAPGTSTPESGSASPGTSTPESGSASP GSTSESPSGTAPGSTSESPSGTAPGSTSESPSGTAPGSTSSTAESPGPGTSTPESGSASPGTSTP ESGSASPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGSTSESPSGTAPGSTSESPSGT APGTSTPESGSASPGTSPSGESSTAPGSTSSTAESPGPGTSPSGESSTAPGSTSSTAESPGPGT STPESGSASPGSTSESPSGTAP
AD576	GSSESGSSEGGPGSGGEPSESGSSGSSESGSSEGGPGSSESGSSEGGPGSSESGSSEGGPGSSE SGSSEGGPGSSESGSSEGGPGESPGGSSGSESGSEGSSGPGESSGSSESGSSEGGPGSSESGSS EGGPGSSESGSSEGGPGSGGEPSESGSSGESPGGSSGSESGESPGGSSGSESGSGGEPSESGS SGSSESGSSEGGPGSGGEPSESGSSGSGGEPSESGSSGSEGSSGPOESSGESPGGSSGSESGSG GEPSESGSSGSGGEPSESGSSGSGGEPSESGSSGSSESGSSEGGPGESPGGSSGSESGESPGGS SGSESGESPGGSSGSESGESPGGSSGSESGESPGGSSGSESGSSESGSSEGGPGSGGEPSESGS SGSEGSSGPGESSGSSESGSSEGGPGSGGEPSESGSSGSSESGSSEGGPGSGGEPSESGSSGES PGGSSGSESGESPGGSSGSESGSSESGSSEGGPGSGGEPSESGSSGSSESGSSEGGPGSGGEPS ESGSSGSGGEPSESGSSGESPGGSSGSESGSEGSSGPGESSGSSESGSSEGGPGSEGSSGPOES S
AE576	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTST EPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATP ESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAP GTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTST EPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSG SETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAP GTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEP ATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATP ESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP
AE576	GSTSSTAESPGPGSTSSTAESPGPGSTSESPSGTAPGSTSSTAESPGPGSTSSTAESPGPGTSTP ESGSASPGSTSESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGSTSESPSGTAPGTSPSGESST APGSTSESPSGTAPGSTSESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGSTSESPSGTAPGS TSESPSGTAPGTSTPESGSASPGSTSESPSGTAPGTSTPESGSASPGSTSSTAESPGPGSTSSTA ESPGPGTSTPESGSASPGTSTPESGSASPGSTSESPSGTAPGTSTPESGSASPGTSTPESGSASP GSTSESPSGTAPGSTSESPSGTAPGSTSESPSGTAPGSTSSTAESPGPGTSTPESGSASPGTSTP ESGSASPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGSTSESPSGTAPGSTSESPSGT APGTSTPESGSASPGTSPSGESSTAPGSTSSTAESPGPGTSPSGESSTAPGSTSSTAESPGPGT STPESGSASPGSTSESPSGTAPGSTSSTAESPGPGTSTPESGSASPGTSTPESGSASP

[0002]

XTEN Name	Amino Acid Sequence
AE624	MAEPAGSPTSTEEGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGSPAGSPTSTEEGT SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESA TPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS APGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGT STEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESA TPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGT STEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTST EEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP
AD836	GSSESGSSEGGPGSSESGSSEGGPGESPGGSSGSESGSGGEPSESGSSGESPGGSSGSESGESP GGSSGSESGSSESGSSEGGPGSSESGSSEGGPGSSESGSSEGGPGESPGGSSGSESGESPGGSS GSESGESPGGSSGSESGSSESGSSEGGPGSSESGSSEGGPGSSESGSSEGGPGSSESGSSEGGP GSSESGSSEGGPGSSESGSSEGGPGSGGEPSESGSSGESPGGSSGSESGESGSG GEPSESGSSGSEGSSGPGESSGSSESGSSEGGPGSGGEPSESGSSGSGESSGPGESSGSSESGS SEGGPGSGGEPSESGSSGESPGGSSGSESGSGGEPSESGSSGSGGEPSESGSSGSSESGSSEG GPGSGGEPSESGSSGSGGEPSESGSSGSEGSSGPGESSGESPGGSSGSESGSEGSSGPGESSGS EGSSGPGESSGSGGEPSESGSSGSSESGSSEGGPGSSESGSSEGGPGESPGGSSGSESGSGGE PSESGSSGSEGSSGPGESSGESPGGSSGSESGSEGSSGPGSSESGSSEGGPGESPGGSSGSES SEGSSGPGESSGSEGSSGPGESSGSEGSSGPGESSGSGGEPSESGSSGSGGEPSESGSSGESPG GSSGSESGESPGGSSGSESGSGGEPSESGSSGSEGSSGPGESSGESPGGSSGSESGSSESGSSE GGPGSSESGSSEGGPGSSESGSSEGGPGSGGEPSESGSSGSSESGSSEGGPGESPGGSSGSES GSGGEPSESGSSGSSESGSSEGGPGESPGGSSGSESGSGGEPSESGSSGESPGGSSGSESGSQ GEPSESGSS
AE864	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTST EPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATP ESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAP GTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTST EPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSG SETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAP GTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEP ATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATP ESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTST EPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPT STEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGP GSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEP ATSGSETPGTSESATPESGPGTSTEPSEGSAP
AF864	GSTSESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSTP ESGSASPGSTSESPSGTAPGSTSESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGTSPSGESST APGTSPSGESSTAPGSTSSTAESPGPGTSPSGESSTAPGTSPSGESSTAPGSTSSTAESPGPGT STPESGSASPGTSTPESGSASPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGSTSSTA ESPGPGTSTPESGSASPGSTSESPSGTAPGTSPSGESSTAPGSTSSTAESPGPGTSPSGESSTAP GTSTPESGSASPGSTSSTAESPGPGSTSSTAESPGPGSTSSTAESPGPGSTSSTAESPGPGTSPS GESSTAPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGPXXXGASASGAPSTXXXXSESPSG TAPGSTSESPSGTAPGSTSESPSGTAPGSTSESPSGTAPGSTSESPSGTAPGSTSESPSGTAPG TSTPESGSASPGTSPSGESSTAPGTSPSGESSTAPGSTSSTAESPGPGTSPSGESSTAPGTSTPE SGSASPGSTSESPSGTAPGSTSESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGTSTPESGSAS PGTSTPESGSASPGSTSESPSGTAPGTSTPESGSASPGSTSSTAESPGPGSTSESPSGTAPGSTS ESPSGTAPGTSPSGESSTAPGSTSSTAESPGPGTSPSGESSTAPGTSTPESGSASPGTSPSGESS TAPGTSPSGESSTAPGTSPSGESSTAPGSTSSTAESPGPGSTSSTAESPGPGTSPSGESSTAPG SSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSP
AG864	GASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSP SASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASPGTS STGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSS PGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSPGAS PGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTS STGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGS PGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSS

XTEN Name	Amino Acid Sequence
	TPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGT SSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTG SPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGS STPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGS GTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSST GSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPG SSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSP
AM875	GTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGSTSSTAESPGPGTSTPESGSASPGSTS ESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSTPESGSASPGSEPATSGSETPGTSESATP ESGPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAP GSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTST EPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGSPAGSPT STEEGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTSTEPSEGSAPGTSTEPSEGSAP GSEPATSGSETPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGASASGAPSTGGTSES ATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSSTAESPGPGSTSESPSGTAPGTSPSGESS TAPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSEPATSGSETPGTSESATPESGPG SEPATSGSETPGSTSSTAESPGPGSTSSTAESPGPGTSPSGESSTAPGSEPATSGSETPGSEPA TSGSETPGTSTEPSEGSAPGSTSSTAESPGPGTSTPESGSASPGSTSESPSGTAPGTSTEPSEGS APGTSTEPSEGSAPGTSTEPSEGSAPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGS EPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSSTPSGATGSPGSSPSASTGTGPGASPGT SSTGSPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAP
AE912	MAEPAGSPTSTEEGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGSPAGSPTSTEEGT SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESA TPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS APGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGT STEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESA TPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGT STEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTST EEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGS EPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGS PTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTST EEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSSEPATSGSETPGS EPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESA TPESGPGTSTEPSEGSAP
AM923	MAEPAGSPTSTEEGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGTSTEPSEGSAPGS EPATSGSETPGSPAGSPTSTEEGSTSSTAESPGPGTSTPESGSASPGSTSESPSGTAPGSTSESP SGTAPGTSTPESGSASPGTSTPESGSASPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTE EGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTS TEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSTEPSEGSAPGTSTEPS EGSAPGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGSSTPSGATGS PGTPGSGTASSSPGSSTPSGATGSPGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGSP AGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGASASGAPSTGGTSESATPESGPGSPAGSP TSTEEGSPAGSPTSTEEGSTSSTAESPGPGSTSESPSGTAPGTSPSGESSTAPGTPGSGTASSS PGSSTPSGATGSPGSSPSASTGTGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGST SSTAESPGPGSTSSTAESPGPGTSPSGESSTAPGSEPATSGSETPGSEPATSGSETPGTSTEPSE GSAPGSTSSTAESPGPGTSTPESGSASPGSTSESPSGTAPGTSTEPSEGSAPGTSTEPSEGSAP GTSTEPSEGSAPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSEPATSGSETPGTSE SATPESGPGSPAGSPTSTEEGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTSESATP ESGPGTSTEPSEGSAPGTSTEPSEGSAP
AM1318	GTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGSTSSTAESPGPGTSTPESGSASPGSTS ESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSTPESGSASPGSEPATSGSETPGTSESATP ESGPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAP GSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTST EPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGSPAGSPT STEEGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTSTEPSEGSAPGTSTEPSEGSAP GSEPATSGSETPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGPEPTGPAPSGGSEPA TSGSETPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSPAGSPTSTEEGSPAGSPTS

XTEN Name	Amino Acid Sequence
	TEEGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSTSSTAESPGPGSTSESPSGTAPG TSPSGESSTAPGSTSESPSGTAPGSTSESPSGTAPGTSPSGESSTAPGTSTEPSEGSAPGTSESA TPESGPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSESATPESGPGTSTEPSEGS APGTSESATPESGPGTSTEPSEGSAPGTSPSGESSTAPGTSPSGESSTAPGTSPSGESSTAPGT STEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGSSPSASTGTGPGSSTPSGATGSPGSSTPS GATGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASASGAPSTGGTSPSGESST APGSTSSTAESPGPGTSPSGESSTAPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGS SPSASTGTGPGSSTPSGATGSPGASPGTSSTGSPGTSTPESGSASPGTSPSGESSTAPGTSPSG ESSTAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGSTSESPSGTAPGSTSESPSGT APGTSTPESGSASPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT SESATPESGPGSEPATSGSETPGSSTPSGATGSPGASPGTSSTGSPGSSTPSGATGSPGSTSES PSGTAPGTSPSGESSTAPGSTSSTAESPGPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASS SPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAP
BC 864	GTSTEPSEPGSAGTSTEPSEPGSAGSEPATSGTEPSGSGASEPTSTEPGSEPATSGTEPSGSEP ATSGTEPSGSEPATSGTEPSGSGASEPTSTEPGTSTEPSEPGSAGSEPATSGTEPSGTSTEPSE PGSAGSEPATSGTEPSGSEPATSGTEPSGTSTEPSEPGSAGTSTEPSEPGSAGSEPATSGTEPS GSEPATSGTEPSGTSEPSTSEPGAGSGASEPTSTEPGTSEPSTSEPGAGSEPATSGTEPSGSEP ATSGTEPSGTSTEPSEPGSAGTSTEPSEPGSAGSGASEPTSTEPGSEPATSGTEPSGSEPATSG TEPSGSEPATSGTEPSGSEPATSGTEPSGTSTEPSEPGSAGSEPATSGTEPSGSGASEPTSTEP GTSTEPSEPGSAGSEPATSGTEPSGSGASEPTSTEPGTSTEPSEPGSAGSGASEPTSTEPGSEP ATSGTEPSGSGASEPTSTEPGSEPATSGTEPSGSGASEPTSTEPGTSTEPSEPGSAGSEPATSG TEPSGSGASEPTSTEPGTSTEPSEPGSAGSEPATSGTEPSGTSTEPSEPGSAGSEPATSGTEPS GTSTEPSEPGSAGTSTEPSEPGSAGTSTEPSEPGSAGTSTEPSEPGSAGTSTEPSEPGSAGTST EPSEPGSAGTSEPSTSEPGAGSGASEPTSTEPGTSTEPSEPGSAGTSTEPSEPGSAGTSTEPSE PGSAGSEPATSGTEPSGSGASEPTSTEPGSEPATSGTEPSGSEPATSGTEPSGSEPATSGTEPS GSEPATSGTEPSGTSEPSTSEPGAGSEPATSGTEPSGSGASEPTSTEPGTSTEPSEPGSAGSEP ATSGTEPSGSGASEPTSTEPGTSTEPSEPGSA
BD864	GSETATSGSETAGTSESATSESGAGSTAGSETSTEAGTSESATSESGAGSETATSGSETAGS ETATSGSETAGTSTEASEGSASGTSTEASEGSASGTSESATSESGAGSETATSGSETAGTSTE ASEGSASGSTAGSETSTEAGTSESATSESGAGTSESATSESGAGSETATSGSETAGTSESATS ESGAGTSTEASEGSASGSETATSGSETAGSETATSGSETAGTSTEASEGSASGSTAGSETST EAGTSESATSESGAGTSTEASEGSASGSETATSGSETAGSTAGSETSTEAGSTAGSETSTEA GSETATSGSETAGTSESATSESGAGTSESATSESGAGSETATSGSETAGTSESATSESGAGTS ESATSESGAGSETATSGSETAGSETATSGSETAGTSTEASEGSASGSTAGSETSTEAGSETA TSGSETAGTSESATSESGAGSTAGSETSTEAGSTAGSETSTEAGSTAGSETSTEAGTSTEASE GSASGSTAGSETSTEAGSTAGSETSTEAGTSTEASEGSASGSTAGSETSTEAGSETATSGSE TAGTSTEASEGSASGTSESATSESGAGSETATSGSETAGTSESATSESGAGTSESATSESGA GSETATSGSETAGTSESATSESGAGSETATSGSETAGTSTEASEGSASGTSTEASEGSASGST AGSETSTEAGSTAGSETSTEAGSETATSGSETAGTSESATSESGAGTSESATSESGAGSETA TSGSETAGSETATSGSETAGSETATSGSETAGTSTEASEGSASGTSESATSESGAGSETATS GSETAGSETATSGSETAGTSESATSESGAGTSESATSESGAGSETATSGSETA
Y288	GEGSGEGSEGEGSEGSGEGEGSEGSGEGEGGSEGSEGEGGSEGSEGEGGSEGSEGEGSGEG SEGEGGSEGSEGEGSGEGSEGEGSEGGSEGEGGSEGSEGEGSEGSEGEGSGEGSEGEGSE GSGEGEGSGEGSEGEGSEGSGEGEGSGEGSSEGEGSEGSGEGEGSEGSGEGEGGSEGSEGEG SEGSGEGEGGEGSGEGEGSGEGSEGEGGEGSEGEGSGEGGEGEGSEGGSEGEGGSEGGE GEGSEGSGEGEGSEGGSEGEGSEGGSEGEGSEGSGEGEGSEGSGE
Y576	GEGSGEGSEGEGSEGSGEGEGSEGSGEGEGGSEGSEGEGSEGSGEGEGGEGSGEGEGSGE GSEGEGGEGSEGEGSGEGGEGEGSEGGSEGEGGSEGGSEGEGSEGGSEGEGGSEGSEG SEGGSEGEGSEGSGEGEGSEGSGEGEGSEGSGEGEGSEGSGEGEGSEGGSEGEGGSEGSEG EGSGEGSEGEGGSEGSEGEGGEGSEGEGSGEGSEGEGGSEGSEGEGGSEGGSEGEGGEGS GEGEGSEGSGEGEGSGEGSEGEGSEGSGEGEGSEGSGEGEGGSEGSEGEGSGEGSEGEGSE GSGEGEGSEGSGEGEGGSEGSEGEGGSEGSEGEGGSEGSEGEGGSEGSEGEGSGEGSEGE GSGEGSEGEGSEGSGEGEGSEGSGEGEGGSEGSEGEGSEGSGEGEGGEGSGEGEGSGEGSE GEGGGEGSEGEGSEGSGEGEGSEGSGEGEGSEGGSEGEGGSEGSEGEGSEGGSEGEGSEG GSEGEGSEGSGEGEGSEGSGEGEGSEGSGEGEGGSEGGSEGEGGSEGSEGEGSEGGSEGEG GEGSGEGEGGGEGSEGEGSEGSGEGEGSGEGSE

[0003] Table B2

XTEN Name	Amino Acid Sequence
AE42_1	TEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGS
AE42_2	PAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSG
AE42_3	SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSP
AG42_1	GAPSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGPSGP
AG42_2	GPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASP
AG42_3	SPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGA
AG42_4	SASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATG
AE48	MAEPAGSPTSTEEGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGS
AM48	MAEPAGSPTSTEEGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGS
AE144	GSEPATSGSETPGTSESATPESGPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGSEPATSG SETPGSEPATSGSETPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTST EPSEGSAP
AF144	GTSTPESGSASPGTSPSGESSTAPGTSPSGESSTAPGSTSSTAESPGPGTSESPSGTAPGSTSSTAE SPGPGTSPSGESSTAPGTSTPESGSASPGSTSSTAESPGPGTSPSGESSTAPGTSPSGESSTAPGTSP SGESSTAP
AG144_1	PGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGT SSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPG TPGSGTASSS
AG144_2	SGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTG PGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSAS TGTGPGASP
AG144_3	GTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTS STGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGA SPGTSSTGSP
AG144_4	GTPGSGTASSSPGSSTPSGATGSPGASPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTS STGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGS STPSGATGSP
AE288	GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSE GSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSP AGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSET PGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESAT PESGPGTSTEPSEGSAP
AG288_1	ASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSS TPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTG PGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGT SSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPG ASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGS
AG288_2	PGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPS GATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPG ASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSS TGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTP GSGTASSSPGSSTPSGATGS
AG288_3	GSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTS STGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGT PGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTG TGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASP GTSSTGSPGTPGSGTASSSP
AF504	GASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSXPSAS TGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGA SPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGAT GSPGSXPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASP GTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGS PGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPS GATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPG SSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSP
AF540	GSTSSTAESPGPGSTSSTAESPGPGSTSESPSGTAPGSTSSTAESPGPGSTSSTAESPGPGTSTPESG SASPGSTSESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGSTSESPSGTAPGTSPSGESSTAPGSTS ESPSGTAPGSTSESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGSTSESPSGTAPGSTSESPSGTAP GTSTPESGSASPGSTSESPSGTAPGTSTPESGSASPGSTSSTAESPGPGSTSSTAESPGPGTSTPESG

XTEN Name	Amino Acid Sequence
	SASPGTSTPESGSASPGSTSESPSGTAPGTSTPESGSASPGTSTPESGSASPGSTSESPSGTAPGSTS ESPSGTAPGSTSESPSGTAPGTSSTAESPGPGTSTPESGSASPGTSTPESGSASPGSTSESPSGTAP GSTSESPSGTAPGTSTPESGSASPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSPSGES STAPGSTSSTAESPGPGTSPSGESSTAPGSTSSTAESPGPGTSTPESGSASPGSTSESPSGTAP
AD576	GSSESGSSEGGPGSGGEPSESGSSGSSESGSSEGGPGSSESGSSEGGPGSSESGSSEGGPGSSESGS SEGGPGSSESGSSEGGPGESPGGSSGSESGSEGSSGPGESSGSSESGSSEGGPGSSESGSSEGGPGS SESGSSEGGPGSGGEPSESGSSGESPGGSSGSESGESPGGSSGSESGSGGEPSESGSSGSSESGSSE GGPGSGGEPSESGSSGSGGEPSESGSSGSEGSSGPGESSGESPGGSSGSESGSGGEPSESGSSGSGG EPSESGSSGSGGEPSESGSSGSSESGSSEGGPGESPGGSSGSESGESPGGSSGSESGESPGGSSGSES GESPGGSSGSESGESPGGSSGSESGSSESGSSEGGPGSGGEPSESGSSGSEGSSGPGESSGSSESGS SEGGPGSGGEPSESGSSGSSESGSSEGGPGSGGEPSESGSSGESPGGSSGSESGESPGGSSGSESGS SESGSSEGGPGSGGEPSESGSSGSSESGSSEGGPGSGGEPSESGSSGSGGEPSESGSSGESPGGSSG SESGSEGSSGPGESSGSSESGSSEGGPGSEGSSGPGESS
AE576	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE GSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTST EPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGP GTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATP ESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTST EPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAP GSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSG SETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPA GSPTSTEEGTSESATPESGPGTSTEPSEGSAP
AF576	GSTSSTAESPGPGSTSSTAESPGPGSTSESPSGTAPGSTSSTAESPGPGSTSSTAESPGPGTSTPESG SASPGSTSESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGSTSESPSGTAPGTSPSGESSTAPGSTS ESPSGTAPGSTSESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGSTSESPSGTAPGSTSESPSGTAP GTSTPESGSASPGSTSESPSGTAPGTSTPESGSASPGSTSSTAESPGPGSTSSTAESPGPGTSTPESG SASPGTSTPESGSASPGSTSESPSGTAPGTSTPESGSASPGTSTPESGSASPGSTSESPSGTAPGSTS ESPSGTAPGSTSESPSGTAPGSTSSTAESPGPGTSTPESGSASPGTSTPESGSASPGSTSESPSGTAP GSTSESPSGTAPGTSTPESGSASPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSPSGES STAPGSTSSTAESPGPGTSPSGESSTAPGSTSSTAESPGPGTSTPESGSASPGSTSESPSGTAPGSTS STAESPGPGTSTPESGSASPGTSTPESGSASP
AG576	PGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPS GATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPG ASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSS TGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSS TPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTG SPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGS GTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSP GSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSG ATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGS
AE624	MAEPAGSPTSTEEGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGSPAGSPTSTEEGTSESA TPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGS EPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESA TPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGS PAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGS APGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGT STEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPES GPGTSTEPSEGSAP
AD836	GSSESGSSEGGPGSSESGSSEGGPGESPGGSSGSESGSGGEPSESGSSGESPGGSSGSESGESPGGS SGSESGSSESGSSEGGPGSSESGSSEGGPGSSESGSSEGGPGSSESGSSEGGPGESPGGSSGSESGE SPGGSSGSESGSSESGSSEGGPGSSESGSSEGGPGSSESGSSEGGPGSSESGSSEGGPGSSESGSSE GGPGSSESGSSEGGPGSGGEPSESGSSGESPGGSSGSESGESPGGSSGSESGSGGEPSESGSSGSEG SSGPGESSGSSESGSSEGGPGSGGEPSESGSSGSEGSSGPGESSGSSESGSSEGGPGSGGEPSESGS SGESPGGSSGSESGSGGEPSESGSSGSGGEPSESGSSGSSESGSSEGGPGSGGEPSESGSSGSGGEP SESGSSGSEGSSGPGESSGESPGGSSGSESGSEGSSGPGESSGSEGSSGPGESSGSGGEPSESGSSG SSESGSSEGGPGSSESGSSEGGPGESPGGSSGSESGSGGEPSESGSSGSEGSSGPGESSGESPGGSS GSESGSEGSSGPGSSESGSSEGGPGSGGEPSESGSSGSEGSSGPGESSGSEGSSGPGESSGSEGSSG

XTEN Name	Amino Acid Sequence
	PGESSGSGGEPSESGSSGSGGGEPSESGSGSGESPGGSGSESGESPGGSGSESGESPGGSGSESGESPGGSGSES EGSSGPGESGSGESPGGSGSESGSESGSSEGGPGESPGGSGSESGSGGEPSESGSSGSSESGSSEGGPGESPGGSSGSESGSGGSSGSESGSSEGGPGESPGGSSGSESGSGGEPSESGSSGSESPGGSSGSESGSGGEPSESGSS
AE864	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE GSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTST EPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGP GTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATP ESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTST EPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAP GSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSG SETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPA GSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGP GSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSG SETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSE SATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAP GTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP
AF864	GSTSESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSTPESG SASPGSTSESPSGTAPGSTSESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGTSPSGESSTAPGTSP SGESSTAPGSTSSTAESPGPGTSPSGESSTAPGTSPSGESSTAPGSTSSTAESPGPGTSTPESGSASP GTSTPESGSASPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGSTSSTAESPGPGTSTPESG SASPGSTSESPSGTAPGTSPSGESSTAPGSTSSTAESPGPGTSPSGESSTAPGTSTPESGSASPGTS STAESPGPGSTSSTAESPGPGSTSSTAESPGPGSTSSTAESPGPGTSPSGESSTAPGSTSESPSGTAP GSTSESPSGTAPGTSTPESGPXXXGASASGAPSTXXXXSESPSGTAPGSTSESPSGTAPGSTSESPS GTAPGSTSESPSGTAPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSPSGESSTAPGTSP SGESSTAPGSTSSTAESPGPGTSPSGESSTAPGTSTPESGSASPGSTSESPSGTAPGSTSESPSGTAP GTSPSGESSTAPGSTSESPSGTAPGTSTPESGSASPGTSTPESGSASPGSTSESPSGTAPGTSTPESG SASPGSTSSTAESPGPGSTSESPSGTAPGSTSESPSGTAPGTSPSGESSTAPGSTSSTAESPGPGTSP SGESSTAPGTSTPESGSASPGTSPSGESSTAPGTSPSGESSTAPGTSPSGESSTAPGSTSSTAESPGP GSTSSTAESPGPGTSPSGESSTAPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSP
AG864	GASPGTSSTGSPGSSPSASTGTGPGSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSPSAS TGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGA SPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGAT GSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASP GTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTG SPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPS GATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPG SSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSS TGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTP GSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGT GPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSPS ASTGTGPGASPGTSSTGSPGASPGTSSTGSPGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSTP GSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSP
AM875	GTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGTSSTAESPGPGTSTPESGSASPGTSESPS GTAPGSTSESPSGTAPGTSTPESGSASPGTSTPESGSASPGSEPATSGSETPGTSESATPESGPGSPA GSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSTEPSEGSAPGTSTEPSE GSAPGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGSTPSGATGSPGTP GSGTASSSPGSTPSGATGSPGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTE EGSPAGSPTSTEEGTSTEPSEGSAPGASASGAPSTGGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGTSSTAESPGPGSTSESPSGTAPGTSPSGESSTAPGTPGSGTASSSPGSSTPSGATGSPGSS PSASTGTGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSSTAESPGPGTSSTAESPG PGTSPSGESSTAPGSEPATSGSETPGSEPATSGSETPGTSTEPSEGSAPGTSSTAESPGPGTSTPES GSASPGSTSESPSGTAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSSTPSGATGSPGSS PSASTGTGPGASPGTSSTGSPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSSTPSGATG SPGSSPSASTGTGPGASPGTSSTGSPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAP
AE912	MAEPAGSPTSTEEGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGSPAGSPTSTEEGTSESA TPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGS EPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGS

XTEN Name	Amino Acid Sequence
	AP/SPAG/SPTSTEEG/TSTEPSEG/SAP/GTSTEPSEG/SAP/GTSESATPESG/PG/TSTEPSEG/SAP/GTSESA TPESG/PG/SEPATSG/SETPG/TSTEPSEG/SAP/GTSTEPSEG/SAP/GTSESATPESG/PG/TSESATPESG/PG/S PAG/SPTSTEEG/TSESATPESG/PG/SEPATSG/SETPG/TSESATPESG/PG/TSTEPSEG/SAP/GTSTEPSEG/S AP/GTSTEPSEG/SAP/GTSTEPSEG/SAP/GTSTEPSEG/SAP/GTSTEPSEG/SAP/GSPAG/SPTSTEEG/TSTEP SEG/SAP/GTSESATPESG/PG/SEPATSG/SETPG/TSESATPESG/PG/SEPATSG/SETPG/TSESATPESG/PG/T STEPSEG/SAP/GTSESATPESG/PG/SPAG/SPTSTEEG/SPAG/SPTSTEEG/SPAG/SPTSTEEG/TSESATPES G/PG/TSTEPSEG/SAP/GTSESATPESG/PG/SEPATSG/SETPG/TSESATPESG/PG/SEPATSG/SETPG/TSESA TPESG/PG/TSTEPSEG/SAP/GSPAG/SPTSTEEG/TSESATPESG/PG/SEPATSG/SETPG/TSESATPESG/PG/S PAG/SPTSTEEG/SPAG/SPTSTEEG/TSTEPSEG/SAP/GTSESATPESG/PG/TSESATPESG/PG/TSESATPES G/PG/SEPATSG/SETPG/SEPATSG/SETPG/SPAG/SPTSTEEG/TSTEPSEG/SAP/GTSTEPSEG/SAP/GSEPAT SG/SETPG/TSESATPESG/PG/TSTEPSEG/SAP
AM923	MAEPAG/SPTSTEEG/ASPG/TSSTG/SPG/SSTPSGATG/SPG/SSTPSGATG/SPG/TSTEPSEG/SAP/GSEPAT SG/SETPG/SPAG/SPTSTEEG/TSSTAESPG/PG/TSTPESG/SASPG/TSESPSG/TAPG/TSESPSG/TAPG/T STPESG/SASPG/TSTPESG/SASPG/SEPATSG/SETPG/TSESATPESG/PG/SPAG/SPTSTEEG/TSTEPSEG/S APG/TSESATPESG/PG/TSTEPSEG/SAP/GTSTEPSEG/SAP/GSPAG/SPTSTEEG/TSTEPSEG/SAP/GTSTEP SEG/SAP/GTSESATPESG/PG/TSESATPESG/PG/TSTEPSEG/SAP/GTSTEPSEG/SAP/GTSESATPESG/PG/T STEPSEG/SAP/GSEPATSG/SETPG/SPAG/SPTSTEEG/SSTPSGATG/SPG/TPG/SGTASSSPG/SSTPSGAT GSPG/TSTEPSEG/SAP/GTSTEPSEG/SAP/GSEPATSG/SETPG/SPAG/SPTSTEEG/SPAG/SPTSTEEG/TSTE PSEG/SAP/GASASGAPSTGG/TSESATPESG/PG/SPAG/SPTSTEEG/SPAG/SPTSTEEG/TSSTAESPG/PG/T STSESPSG/TAPG/TSPSG/ESSTAPG/TPG/SG/TASSSPG/SSTPSGATG/SPG/SSPSASTG/TG/PG/SEPATSG/S ETPG/TSESATPESG/PG/SEPATSG/SETPG/STSSTAESPG/PG/STSSTAESPG/PG/TSPSG/ESSTAPG/SEPA TSG/SETPG/SEPATSG/SETPG/TSTEPSEG/SAP/GTSSTAESPG/PG/TSTPESG/SASPG/TSTSESPSG/TAPG/ TSTEPSEG/SAP/GTSTEPSEG/SAP/GTSTEPSEG/SAP/GSSTPSGATG/SPG/SSPSASTG/TG/PG/ASPG/TSST G/SPG/SEPATSG/SETPG/TSESATPESG/PG/SPAG/SPTSTEEG/SSTPSGATG/SPG/SSPSASTG/TG/PG/ASP G/TSSTG/SPG/TSESATPESG/PG/TSTEPSEG/SAP/GTSTEPSEG/SAP
AM1318	GTSTEPSEG/SAP/GSEPATSG/SETPG/SPAG/SPTSTEEG/STSSTAESPG/PG/TSTPESG/SASPG/STSESPS GTAPG/TSESPSG/TAPG/TSTPESG/SASPG/TSTPESG/SASPG/SEPATSG/SETPG/TSESATPESG/PG/SPA G/SPTSTEEG/TSTEPSEG/SAP/GTSESATPESG/PG/TSTEPSEG/SAP/GTSTEPSEG/SAP/G/SPAG/SPTSTEE GTSTEPSEG/SAP/GTSTEPSEG/SAP/GTSESATPESG/PG/TSESATPESG/PG/TSTEPSEG/SAP/GTSTEPSE GSAP/GTSESATPESG/PG/TSTEPSEG/SAP/GSEPATSG/SETPG/SPAG/SPTSTEEG/SSTPSGATG/SPG/TP G/SGTASSSPG/SSTPSGATG/SPG/TSTEPSEG/SAP/GTSTEPSEG/SAP/G/SEPATSG/SETPG/SPAG/SPTSTE EG/SPAG/SPTSTEEG/TSTEPSEG/SAP/GPEPTG/PAPSG/GSEPATSG/SETPG/TSESATPESG/PG/SPAG/SP TSTEEG/TSESATPESG/PG/SPAG/SPTSTEEG/SPAG/SPTSTEEG/TSESATPESG/PG/SPAG/SPTSTEEG/SP AG/SPTSTEEG/TSSTAESPG/PG/TSTESPSG/TAPG/TSPSG/ESSTAPG/STSESPSG/TAPG/TSESPSG/TA PG/TSPSG/ESSTAPG/TSTEPSEG/SAP/GTSESATPESG/PG/TSESATPESG/PG/SEPATSG/SETPG/TSESAT PESG/PG/TSESATPESG/PG/TSTEPSEG/SAP/GTSESATPESG/PG/TSTEPSEG/SAP/G/TSPSG/ESSTAP/GTS PSG/ESSTAPG/TSPSG/ESSTAP/GTSTEPSEG/SAP/G/SPAG/SPTSTEEG/TSTEPSEG/SAP/GSSPSASTG/TG PG/SSTPSGATG/SPG/SSTPSGATG/SPG/SSTPSGATG/SPG/SSTPSGATG/SPG/ASPG/TSSTG/SPG/ASASG APSTGG/TSPSG/ESSTAP/GTSSTAESPG/PG/TSPSG/ESSTAP/GTSESATPESG/PG/TSTEPSEG/SAP/GTS TEPSEG/SAP/G/SPSASTG/TG/PG/SSTPSGATG/SPG/ASPG/TSSTG/SPG/TSTPESG/SASPG/TSPSG/ESSTA PG/TSPSG/ESSTAP/GTSESATPESG/PG/SEPATSG/SETPG/TSTEPSEG/SAP/GTSESPSG/TAPG/TSESP SG/TAPG/TSTPESG/SASPG/SPAG/SPTSTEEG/TSESATPESG/PG/TSTEPSEG/SAP/G/SPAG/SPTSTEEG/TS ESATPESG/PG/SEPATSG/SETPG/SSTPSGATG/SPG/ASPG/TSSTG/SPG/SSTPSGATG/SPG/STSESPSG/TA PG/TSPSG/ESSTAP/GTSSTAESPG/PG/SSTPSGATG/SPG/ASPG/TSSTG/SPG/TPG/SGTASSSPG/SPAG/SP TSTEEG/SPAG/SPTSTEEG/TSTEPSEG/SAP
BC 864	GTSTEPSEP/GSAG/TSTEPSEP/GSAG/SEPATSG/TEPSG/SGASEPTSTEP/GSEPATSG/TEPSG/SEPATSG TEPSG/SEPATSG/TEPSG/SGASEPTSTEP/GTSTEPSEP/GSAG/SEPATSG/TEPSG/TSTEPSEP/GSAG/SEP ATSG/TEPSG/SEPATSG/TEPSG/TSTEPSEP/GSAG/TSTEPSEP/GSAG/SEPATSG/TEPSG/SEPATSG/TEPS GTSEPSTSEP/AG/SGASEPTSTEP/GTSEPSTSEP/AG/SEPATSG/TEPSG/SEPATSG/TEPSG/TSTEPSE P/GSAG/TSTEPSEP/GSAG/SGASEPTSTEP/GSEPATSG/TEPSG/SEPATSG/TEPSG/SEPATSG/TEPSG/SEP ATSG/TEPSG/TSTEPSEP/GSAG/SEPATSG/TEPSG/SGASEPTSTEP/GTSTEPSEP/GSAG/SEPATSG/TEPS G/SGASEPTSTEP/GTSTEPSEP/GSAG/SGASEPTSTEP/GSEPATSG/TEPSG/SGASEPTSTEP/GSEPATSG TEPSG/SGASEPTSTEP/GTSTEPSEP/GSAG/SEPATSG/TEPSG/SGASEPTSTEP/GTSTEPSEP/GSAG/SEP ATSG/TEPSG/TSTEPSEP/GSAG/SEPATSG/TEPSG/TSTEPSEP/GSAG/TSTEPSEP/GSAG/TSTEPSEP/GSA GTSTEPSEP/GSAG/TSTEPSEP/GSAG/TSTEPSEP/GSAG/TSEPSTSEP/AG/SGASEPTSTEP/GTSTEPSE P/GSAG/TSTEPSEP/GSAG/TSTEPSEP/GSAG/SEPATSG/TEPSG/SGASEPTSTEP/GSEPATSG/TEPSG/SEP ATSG/TEPSG/SEPATSG/TEPSG/SEPATSG/TEPSG/TSEPSTSEP/AG/SEPATSG/TEPSG/SGASEPTSTEP GTSTEPSEP/GSAG/SEPATSG/TEPSG/SGASEPTSTEP/GTSTEPSEP/GSA
BD868	GSETATSG/SETAG/TSESATSESG/AGSTAG/SETSTEAG/TSESATSESG/AGSETATSG/SETAG/SETAT SG/SETAG/TSTEASEG/SASG/TSTEASEG/SASG/TSESATSESG/AGSETATSG/SETAG/TSTEASEG/SAS

[0004]

XTEN Name	Amino Acid Sequence
	GSTAGSETSTEAGTSESATSESGAGTSESATSESGAGSETATSGSETAGTSESATSESGAGTSTEA SEGSASGSETATSGSETAGSETATSGSETAGTSTEASEGSASGSTAGSETSTEAGTSESATSESGA GTSTEASEGSASGSETATSGSETAGSTAGSETSTEAGSTAGSETSTEAGTSESATSESGAGSETATSGSETA TSESGAGTSESATSESGAGSETATSGSETAGTSESATSESGAGTSESATSESGAGSETATSGSETA GSETATSGSETAGTSTEASEGSASGSTAGSETSTEAGSETATSGSETAGTSESATSESGAGSTAGS ETSTEAGSTAGSETSTEAGSTAGSETSTEAGTSTEASEGSASGSTAGSETSTEAGSTAGSETSTEA GTSTEASEGSASGSTAGSETSTEAGSETATSGSETAGTSTEASEGSASGTSESATSESGAGSETAT SGSETAGTSESATSESGAGTSESATSESGAGSETATSGSETAGTSESATSESGAGSETATSGSETA GTSTEASEGSASGTSTEASEGSASGSTAGSETSTEAGSTAGSETSTEAGSETATSGSETAGTSESA TSESGAGTSESATSESGAGSETATSGSETAGSETATSGSETAGSETATSGSETAGTSTEASEGSAS GTSESATSESGAGSETATSGSETAGSETATSGSETAGTSESATSESGAGTSESATSESGAGSETAT SGSETA
AE948	GTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPT STEEGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGTST EPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSTEPSEGSAP GSEPATSGSETPGSEPATSGSETPGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSETPGTSTEPSE GSAPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGTSE SATPESGPGSEPATSGSETPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGP GSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPT STEEGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSE SATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETP GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSETPGSEPATSG SETPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGSPA GSPTSTEEGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAP GSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATP ESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGP
AE1044	GSEPATSGSETPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGTSTEPSE GSAPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGTSESATPESGPGTSE SATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGTSTEPSEGSAPGSPAGSPTSTEE GSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATP ESGPGTSESATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGTSTEPSEGSAPGTSE SATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAP GTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGSEPATSG SETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSEP ATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAP GTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSETPGTSESATPESGPGTSESATP ESGPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGSPAGSPTSTEEGTSE SATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAP GSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSESATPESGPGSEPATSGSETPGTSTEPSE GSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGTSESATPESG PGTST
AE1140	GSEPATSGSETPGSEPATSGSETPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGSEPATSG SETPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGTSE SATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAP GSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGTSTEPSE GSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP GTSTEPSEGSAPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGSEPATSG SETPGTSESATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSPA GSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSTEPSEGSAP GSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGSEPATSG SETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSTEPSEGSAPGTST EPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEE GTSESATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSETPGSEPATSG SETPGTSESATPESGPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTST EPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGSPAGSPTSTEE GSPA
AE1236	GSPAGSPTSTEEGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSTEPSE GSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGSP

XTEN Name	Amino Acid Sequence
	AG$PTSTEEGTSESATPESGPGSPAG$PTSTEEGTSTEPSEGSAPGSPAG$PTSTEEGTSTEPSEGSA PGSPAG$PTSTEEGTSTEPSEGSAPGPAGPTSTEEGPAGPTSTEEGTSTEPSEGSAPGSEPATS GSETPGSPAG$PTSTEEGTSESATPESGPGSEPATSG$ETPGSEPATSG$ETPGTSESATPESGPGSP AG$PTSTEEGTSESATPESGPGSEPATSG$ETPG$EPATSG$ETPGTSESATPESGPGTSESATPESG PGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSPAG$PTSTEEGTSTEPSEGSAPGSEPATS GSETPGSPAG$PTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGPAGPTSTEEGPAGPTSTEEGSP AGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSESATPESGPGTSTEPSEGSA PGTSESATPESGPGSPAG$PTSTEEGTSESATPESGPGTSESATPESGPGTSTEPSEGSAPGSEPATS GSETPGTSTEPSEGSAPGSPAG$PTSTEEG$ESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSE PATSGSETPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGSPAG$PTSTEEGTSESATPESG PGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSTEPSEGSAPGSEPATS GSETPGTSTEPSEGSAPGSPAG$PTSTEEG$PAG$PTSTEEGTSESATPESGPGSEPATSGSETPGSP AG$PTSTEEG$PAG$PTSTEEGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSA PGSEP
AE1332	GSPAG$PTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGTSTEPSE GSAPGTSESATPESGPGSPAG$PTSTEEGTSTEPSEGSAPGTSESATPESGPGSPAG$PTSTEEGTST EPSEGSAPGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSETPGSEPATSGSETP GTSESATPESGPGSPAG$PTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSESATP ESGPGSPAG$PTSTEEGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGSPAG$PT STEEG$PAG$PTSTEEGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSETPGTSTEPSEGSAPGSEP ATSGSETPGSPAG$PTSTEEGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGTSESATPESGP GTSESATPESGPGTSTEPSEGSAPGSPAG$PTSTEEGTSESATPESGPGTSTEPSEGSAPGSEPATSG SETPGSPAG$PTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTST EPSEGSAPGSPAG$PTSTEEGTSESATPESGPGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETP GSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGSPAG$PTSTEEGTSTEPSEGSAPGTSTEPSE GSAPGSPAG$PTSTEEG$PAG$PTSTEEGTSTEPSEGSAPGSEPATSGSETPGSPAG$PTSTEEGTSE SATPESGPGSEPATSGSETPGSEPATSGSETPGTSESATPESGPGTSESATPESGPGTSTEPSEGSAP GTST
AE1428	GSEPATSGSETPGSPAG$PTSTEEGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSE GSAPGTSESATPESGPGSPAG$PTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSPAG$PTSTEEGTST EPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAP GTSTEPSEGSAPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGSPAG$PTSTEEGTSTEPSE GSAPGTSESATPESGPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAG$PTSTEEG$P AG$PTSTEEGTSTEPSEGSAPGSEPATSGSETPGSPAG$PTSTEEGTSTEPSEGSAPGSPAG$PTSTE EGPAGPTSTEEGTSESATPESGPGSEPATSG$ETPGSEPATSG$ETPGTSTEPSEGSAPGSPAG$P TSTEEG$PAG$PTSTEEGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGTSESATPESGPGSE PATSGSETPGSEPATSGSETPGTSESATPESGPGTSESATPESGPGTSESATPESGPGTSESATPESG PGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSESAT PESGPGTSTEPSEGSAPGTSESATPESGPGSPAG$PTSTEEGTSTEPSEGSAPGSEPATSGSETPGTS TEPSEGSAPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGSPAG$PTSTEEGTSESATPESG PGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSESATPESGPGTSESATPESGPGTSTEPS EGSAPGSEPATSGSETPGSPAG$PTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSP AG$PTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAG$PTSTEEGTSESATPESG PGSPA
AE1524	GTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAG$PTSTEEG$PAG$PT STEEGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSE SATPESGPGTSESATPESGPGSPAG$PTSTEEGTSTEPSEGSAPGSEPATSGSETPGTSTEPSEGSAP GSPAG$PTSTEEGTSTEPSEGSAPGSPAG$PTSTEEGPAGPTSTEEGPAGPTSTEEGTSESATP ESGPGSEPATSGSETPGTSESATPESGPGTSESATPESGPGSPAG$PTSTEEGTSTEPSEGSAPGSEP ATSGSETPGTSTEPSEGSAPGSPAG$PTSTEEG$PAG$PTSTEEGTSTEPSEGSAPGSEPATSGSETP GSPAG$PTSTEEGTSESATPESGPGSEPATSGSETPGSPAG$PTSTEEGTSESATPESGPGTSESATP ESGPGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTST EPSEGSAPGSPAG$PTSTEEGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSETPGSPAG$PTSTEE GTSESATPESGPGTSTEPSEGSAPGSPAG$PTSTEEG$PAG$PTSTEEGTSESATPESGPGSPAG$PT STEEGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSETPGSEPATSGSETPGTSESATPESGPGTSE SATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGP GSPAG$PTSTEEG$PAG$PTSTEEGTSTEPSEGSAPGTSESATPESGPGSPAG$PTSTEEGTSTEPSE GSAPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGSEP

XTEN Name	Amino Acid Sequence
	ATSGSETPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGP GSPA
AE1620	GSEPATSGSETPGTSTEPSEGSAPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGTSESATP ESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTST EPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSESATPESGPGTSTEPSEGSAPGTSTEPSE GSAPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGTSE SATPESGPGTSESATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGSPAGSPTSTEEGTSESATP ESGPGSEPATSGSETPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPA GSPTSTEEGTSTEPSEGSAPGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETP GTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGSPAGSPT STEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSESATPESGPGSPA GSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETP GSEPATSGSETPGTSESATPESGPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSG SETPGTSESATPESGPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTST EPSEGSAPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEE GTST
AE1716	GTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGSPAGSPT STEEGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGTSESATPESGPGTSESATPESGPGTST EPSEGSAPGSEPATSGSETPGSEPATSGSETPGTSESATPESGPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGSPAGSPT STEEGTSESATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGTSE SATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEE GTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSTEPSE GSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSTEPSEGSAPGTST EPSEGSAPGSEPATSGSETPGTSESATPESGPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAP GSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSETPGTSESATP ESGPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGSPA GSPTSTEEGTSESATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEE GSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSTEPSE GSAPGSPAGSPTSTEEGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEP ATSGSETPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGP GTSE
AE1812	GTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSTEPSE GSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSE SATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETP GTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGTSESATP ESGPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEP ATSGSETPGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSETPGTSESATPESGPGTSESATPESGP GTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGTSESATP ESGPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPA GSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSESATPESGPGSEPATSGSETP GTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATP ESGPGSEPATSGSETPGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSE SATPESGPGTSESATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAP GTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSESATP ESGPGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTST EPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGSPAGSPTSTEEGTSTEPSEGSAP GSEP
AE1908	GSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGSPAGSPT STEEGTSESATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGTSTEPSEGSAPGTSE SATPESGPGSEPATSGSETPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGP GTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSESATPESGPGTSTEPSE GSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSEP ATSGSETPGSEPATSGSETPGTSESATPESGPGTSESATPESGPGTSESATPESGPGTSESATPESGP GTSESATPESGPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSG SETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSE SATPESGPGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSESATPESGP GTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSTEPSE

XTEN Name	Amino Acid Sequence
	GSAPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGTSE SATPESGPGTSESATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGSPAGSPTSTEEGSPAGSPTSTEE GSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATP ESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGSPA GSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGSPAGSPTSTEEGTSESATPESGP GSEP
AE3004A	GTSTEPSEGSAPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGSPAGSPT STEEGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGP GTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSE GSAPGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGSEP ATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETP GTSESATPESGPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSESATPESGPGSEPATSG SETPGTSESATPESGPGTSESATPESGPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEP ATSGSETPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEE GTSESATPESGPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATP ESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGSPAGSPTSTEEGTSE SATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEE GTSTEPSEGSAPGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGSEPATSG SETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGSPAGSPTSTEEGTSESATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGP GTSE
AG3048	GSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGTPGSGT ASSSPGTPGSGTASSSPGSSPSASTGTGPGTPGSGTASSSPGSSPSASTGTGPGSSTPSGATGSPGSS TPSGATGSPGSPSASTGTGPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTG PGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGASPG TSSTGSPGTPGSGTASSSPGTPGSGTASSSPGSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSP GASPGTSSTGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGSSPSAS TGTGPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGS STPSGATGSPGSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTAS SSPGSSPSASTGTGPGSSTPSGATGSPGASPGTSSTGSPGSSTPSGATGSPGTPGSGTASSSPGSSPS ASTGTGPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSP GSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGSTPSG ATGSPGSSPSASTGTGPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGS STPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTAS SSPGASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSP GSSTPSGATGSPGTPGSGTASSSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPS GATGSP
AG1044	GTPGSGTASSSPGTPGSGTASSSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGTPGSGT ASSSPGSSPSASTGTGPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGSS TPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGTPGSGTASSSPGASPGTSSTG SPGSSTPSGATGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGTPGSGTASSSPGSSPSA STGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPG TPGSGTASSSPGSSPSASTGTGPGSSTPSGATGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGA TGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGSSPSASTGTGPGTP GSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGSTPSGATGSPGTPGSGTASSSPGSSTPSGATG SPGASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGTPGS GTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGP GASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSG ATGSPGSSTPSGATGSPGASPGTSSTGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGS SPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSST GSPGTPGSGTASSSPGTPGSGTASSSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPS ASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSPSASTGTGPGASPGTSSTGSPGSSTPSGATGSP GTPGSGTASSSPGSST
AG1140	GASPGTSSTGSPGSSPSASTGTGPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSG ATGSPGTPGSGTASSSPGASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGSSTPSGATGSPGS STPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTG TGPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGTPGSGTASSSPGSSP SASTGTGPGSSTPSGATGSPGSSPSASTGTGPGSSTPSGATGSPGASPGTSSTGSPGSSPSASTGTG PGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSG TASSSPGTPGSGTASSSPGASPGTSSTGSPGSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGS

XTEN Name	Amino Acid Sequence
	ASPGTSSTGSPGSSTPSGATGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSPSAST GTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGAS PGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGTPGSGTASS SPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGS GTASSSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSP GSSTPSGATGSPGASPGTSSTGSPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSPGSSPSAS TGTGPGSSPSASTGTGPGSSTPSGATGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGA SPGTSSTGSPGSSPSASTGTGPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGT GPGASPGTSSTGSPGSST
AG1236	GSSPSASTGTGPGTPGSGTASSSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGASPGTS STGSPGSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGS SPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTAS SSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGTPG SGTASSSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTG PGASPGTSSTGSPGSSPSASTGTGPGSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPS GATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPG ASPGTSSTGSPGSSTPSGATGSPGTPGSGTASSSPGSSPSASTGTGPGSSTPSGATGSPGTPGSGTA SSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGASP GTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGTPGSGTASSSPGASPGTSSTGSPGTPGSGTASSS PGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTS STGSPGSSTPSGATGSPGTPGSGTASSSPGTPGSGTASSSPGSSPSASTGTGPGTPGSGTASSSPGS STPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGSSTPSGAT GSPGTPGSGTASSSPGSSPSASTGTGPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSST PSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSPSASTGTGPGSSPSASTGTG PGASPGTSSTGSPGASP
AG1332	GSSTPSGATGSPGSSPSASTGTGPGTPGSGTASSSPGSSPSASTGTGPGASPGTSSTGSPGSSPSAS TGTGPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGSSTPSGATGSPGA SPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGT GPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGTPGSGTASSSPGASPG TSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSP GSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGSSPSASTGTGPGSSTPSG ATGSPGSSPSASTGTGPGSSTPSGATGSPGTPGSGTASSSPGSSPSASTGTGPGSSTPSGATGSPGA SPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGASPGTSSTGSPGSSTPSGATGSPGASPGTSST GSPGTPGSGTASSSPGTPGSGTASSSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSPS ASTGTGPGSSTPSGATGSPGTPGSGTASSSPGTPGSGTASSSPGSSPSASTGTGPGSSTPSGATGSP GSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSG ATGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGT PGSGTASSSPGTPGSGTASSSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGAT GSPGTPGSGTASSSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGTPGSGTASSSPGSSPS ASTGTGPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSP GASPGTSSTGSPGTPG
AG1428	GTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGSSTPSGATGSPGTPGSGTASSSPGTPGSGT ASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGSSPSASTGTGPGSSTPSGATGSPGA SPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTAS SSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGSSPS ASTGTGPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSP GASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGSSPSASTGTGPGASPGTSSTGSPGSSTPSG ATGSPGASPGTSSTGSPGSSPSASTGTGPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGA SPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGT GPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGTPGSGTASSSPGTPGSGTASSSPGSSPS ASTGTGPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGASPGTSSTGSP GASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGSSPSASTGTGPGASPGTSSTGSPGSSTPSG ATGSPGASPGTSSTGSPGSSPSASTGTGPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGA SPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGT GPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGTPGSGTASSSPGTPGSGTASSSPGSSPS ASTGTGPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGASPGTSSTGSP GASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGSSPSASTGTGPGASPGTSSTGSPGSSTPSG ATGSPGASPGTSSTGSPGSSPSASTGTGPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGA SPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGT GPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGTPGSGTASSSPGTPGSGTASSSPGSSPS ASTGTGPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGASPGTSSTGSP GASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGSSPSASTGTGPGASPGTSSTGSPGSSTPSG ATGSPGASPGTSSTGSPGSSPSASTGTGPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGA SPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGT GPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGTPGSGTASSSPGTPGSGTASSSPGSSPS ASTGTGPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGASPGTSSTGSP GASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGSSPSASTGTGPGASPGTSSTGSPGSSTPSG ATGSPGASPGTSSTGSPGSSPSASTGTGPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGA SPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGT GPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGTPGSGTASSSPGTPGSGTASSSPGSSPS ASTGTGPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGASPGTSSTGSP GASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGSSPSASTGTGPGASPGTSSTGSPGSSTPSG ATGSPGASPGTSSTGSPGSSPSASTGTGPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGA SPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGT GPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGTPGSGTASSSPGTPGSGTASSSPGSSPS ASTGTGPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGASPGTSSTGSP GASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGSSPSASTGTGPGASPGTSSTGSPGSSTPSG ATGSPGASPGTSSTGSPGSSPSASTGTGPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGA SPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGT GPGTPGSGTASSSPGASP
AG1528	GSSTPSGATGSPGTPGSGTASSSPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGTPGSGT ASSSPGSSTPSGATGSPGSSPSASTGTGPGSSTPSGATGSPGTPGSGTASSSPGTPGSGTASSSPGSS PSASTGTGPGSSTPSGATGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATG

XTEN Name	Amino Acid Sequence
	SPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSPSASTGTGPGASPG TSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGP GSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGSSPSASTGTGPGASPGTS STGSPGASPGTSSTGSPGTPGSGTASSSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGS STPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGTPGSGTASSSPGASPGTSST GSPGSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGSST PSGATGSPGSSTPSGATGSPGTPGSGTASSSPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGS PGSSPSASTGTGPGTPGSGTASSSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSG TASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGSSTPSGATGSPG TPGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGTPGSGTA SSSPGSSPSASTGTGPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSP SASTGTGPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGS PGSSTPSGATGSPGTPG
AG1620	GSSTPSGATGSPGSSTPSGATGSPGTPGSGTASSSPGSSPSASTGTGPGTPGSGTASSSPGASPGTS STGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGSSPSASTGTGPGT PGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGSSTPSGATGSPGASPGTSST GSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSST PSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSPSASTGTG PGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSPSASTGTGPGSSPSAS TGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSPSASTGTGPGSSTPSGATGSPGA SPGTSSTGSPGSSTPSGATGSPGTPGSGTASSSPGSSPSASTGTGPGASPGTSSTGSPGSSTPSGAT GSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGSST PSGATGSPGTPGSGTASSSPGSSPSASTGTGPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGS PGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGASPGTSSTGSPGSSPSAS TGTGPGTPGSGTASSSPGASPGTSSTGSPGSSPSASTGTGPGSSTPSGATGSPGSSPSASTGTGPGS STPSGATGSPGSSPSASTGTGPGTPGSGTASSSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGAT GSPGTPGSGTASSSPGSSPSASTGTGPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSST PSGATGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGTPGSGTASSS PGSSTPSGATGSPGSST
AG1716	GASPGTSSTGSPGSSPSASTGTGPGSSTPSGATGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSG ATGSPGSSTPSGATGSPGSSPSASTGTGPGSSTPSGATGSPGTPGSGTASSSPGSSPSASTGTGPGS STPSGATGSPGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGASPGTSST GSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSTPSGATGSPGTPGSGTASSSPGSSPS ASTGTGPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGTPGSGTASSSPGSSTPSGATGSP GTPGSGTASSSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSPSAS TGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGT PGSGTASSSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGASPGTSST GSPGASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGSSPSASTGTGPGSSPSASTGTGPGSSPS ASTGTGPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGASPGTSSTGSPGSSPSASTGTGP GASPGTSSTGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGT ASSSPGASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGSS TPSGATGSPGSSPSASTGTGPGSSTPSGATGSPGTPGSGTASSSPGSSPSASTGTGPGSSTPSGATG SPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSPSA STGTGPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGASPGTSSTGSPG ASPGTSSTGSPGTPG
AG1812	GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSPSAS TGTGPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGSSPSASTGTGPGT PGSGTASSSPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTG TGPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGTPGSGTASSSPGSSPSASTGTGPGSST PSGATGSPGTPGSGTASSSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATGS PGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGT SSTGSPGASPGTSSTGSPGSSTPSGATGSPGTPGSGTASSSPGSSPSASTGTGPGTPGSGTASSSPG ASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSAST GTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGSS TPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASS SPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPG TSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSTPSGATGSPGTPGSGTASSSP GSSPSASTGTGPGASPGTSSTGSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSG ATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSPSASTGTGPGS SPSASTGTGPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSST

XTEN Name	Amino Acid Sequence
	GSPGSSTPSGATGSPGASP
AG1908	GSSPSASTGTGPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSPSASTGTGPGSSPSAS TGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGTPGSGTASSSPGASPGTSSTGSPGT PGSGTASSSPGTPGSGTASSSPGSSPSASTGTGPGSSTPSGATGSPGSSPSASTGTGPGASPGTSST GSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGASP GTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGSSPSASTGTGPGSSTPSGATGSPGASPGTSSTGS PGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGASPGTSSTGSPGSSPSAS TGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGS STPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGTPGSGTASSSPGSSPSASTG TGPGSSPSASTGTGPGSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGTPG SGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGS PGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGTPGSG TASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPG TPGSGTASSSPGSSPSASTGTGPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGSSTPSGA TGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSPSASTGTGPGSSTPSGATGSPGAS PGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTG SPGSSPSASTGTGPGSSP
AG2898A	GSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSG ATGSPGSSPSASTGTGPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGS STPSGATGSPGTPGSGTASSSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTG TGPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGTPG SGTASSSPGSSPSASTGTGPGSSTPSGATGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGS PGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGTPGSGTASSSPGTPGSG TASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGTPGSGTASSSPG SSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGSSPSAST GTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGAS PGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTG SPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSTP SGATGSPGSSTPSGATGSPGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSP GSSPSASTGTGPGSPSASTGTGPGASPGTSSTGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGT ASSSPGTPGSGTASSSPGSSPSASTGTGPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGS PSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSTPSGATG PGASP
AE72B	SPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSE ETPG
AE72C	TSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEG SAPG
AE108A	TEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSES ATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTS
AE108B	GSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSG SETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAP
AE144A	STEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPES GPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAG SPTSTEEGS
AE144B	SEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTS TEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTE PSEGSAPG
AE180A	TSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTS TEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSET PGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATS
AE216A	PESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSP AGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESG PGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATS GSETPGTSESAT
AE252A	ESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEP ATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGP GTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSE
AE288A	TPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGT

XTEN Name	Amino Acid Sequence
	SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPES GPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESA TPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGS EPATSGSETPGTSESA
AE324A	PESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTS ESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSA PGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSE PATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATS
AE360A	PESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTS ESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSET PGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTS ESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSA PGSEPATSGSETPGTSESAT
AE396A	PESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSP AGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSET PGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTS ESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPS
AE432A	EGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSE PATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTE EGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSE PATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESG PGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPS EGSAPGTSTEPSEGSAPGSEPATS
AE468A	EGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSE PATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESG PGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSP AGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTE EGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATS GSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESAT
AE504A	EGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTS TEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESG PGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESAT PESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTS ESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGTSESATPESGPGTSESAT PESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSE PATSGSETPGTSESATPESGPGTSTEPS
AE540A	TPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGT STEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTST EEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGT SESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT SESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEP
AE576A	TPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGT STEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPAT SGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGS PAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPES GPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPAT SGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGT SESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGS

XTEN Name	Amino Acid Sequence
	APGTSTEPSEGSAPGSEPATSGSETPGTSESA
AE612A	GSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTS ESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSA PGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTS ESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSA PGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPS EGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSP AGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSET PGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESAT
AE648A	PESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTS ESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESG PGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPS EGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSE PATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTE EGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSE PATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESG PGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPS EGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESAT
AE684A	EGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTS ESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESG PGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPS EGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTS TEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESG PGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESAT PESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTS ESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESG PGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESAT PESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSE PATS
AE720A	TSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPG SPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPE SGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPA GSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAP GTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTST EPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGP GTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATP ESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPA GSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGP GSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTE
AE756A	TSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPG SPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPE SGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPA GSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAP GTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTST EPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGP GTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATP ESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPA GSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGP GSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSG SETPGTSES
AE792A	EGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTS TEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESG PGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESAT PESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSA

XTEN Name	Amino Acid Sequence
	PGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESAT
AE828A	PESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSESAT
AG72A	GPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASS
AG72B	GSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSP
AG72C	SPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGA
AG108A	SASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASP
AG108B	PGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSS
AG144A	PGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSS
AG144B	PSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASP
AG180A	TSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSG
AG252A	TSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPG
AG288A	TSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGTPGS
AG324A	TSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTP
AG360A	TSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGT

XTEN Name	Amino Acid Sequence
	ASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGT PGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSST GSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSST PSGATGSPGSSTPSGATGSPGASPG
AG396A	GATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPG ASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGA TGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSS PSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGT GPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPG TSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGPGSSTPSGATGSPGASPGT
AG432A	GATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPG SSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSS TGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTP GSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGT GPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSPS ASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGP GSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGASPG
AG468A	TSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPG GSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSG ATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGA SPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGAT GSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSPS ASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGP GSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGASPG
AG504A	TSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSP GSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSG ATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGA SPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSTPSGAT GSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSPS ASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGP GSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGASPGTS STGSPGSSPSASTGTGPGTPGSGTASSSPGSSTP
AG580A	TSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSP GASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSG ATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTAS SSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSTP SGATGSPGSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSP GSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSAS TGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSTPSGATGSPGA SPG
AG576A	TSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGP GTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGATGSPGSSTPSG ATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGS SPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSST GSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSTPSGATGSPGTPGSGTASSSPGSST PSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSPSASTGTGPGSSPSASTGTG PGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGT SSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPG TPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPG
AG612A	STGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGS SPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSST GSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPS ASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSP GSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSPS ASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSP GSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSPS ASTGTGPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSP GASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTAS SSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSPSASTGTGPGSSPS ASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGP

XTEN Name	Amino Acid Sequence
	GASPGTS
AG648A	GTASSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSP GASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTS STGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGA SPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGAT GSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPG SGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSS PGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPS GATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPG SSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSAST GTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTP
AG684A	TSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSP GSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTS STGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGA SPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTAS SSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASP GTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGS PGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPS GATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPG SSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGA TGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSS PSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPG
AG720A	TSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSTPSGATGSP GSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTS STGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSS PSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGAT GSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSST PSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGS PGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSG TASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPG SSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSAST GTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGAS PGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTAS SSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPS ASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGP GASPGTSSTGSPGASPG
AG756A	TSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGP GASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTS STGSPGTPGSGTASSSPGSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGS SPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSST GSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPS ASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSP GSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSG ATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGA SPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTAS SSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPS ASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGP GASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPG
AG792A	TSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGP GASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTS STGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGA SPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGAT GSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGASP GTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGS PGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPS GATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPG SSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGA TGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSS PSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPG
AG828A	TSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGP GASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTS

XTEN Name	Amino Acid Sequence
	STGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGS SPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSST GSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPS ASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSP GSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSG ATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGA SPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTAS SSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPS ASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGP GASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSAS TGTGPGTPGSGTASSSPGSSTP

[0005] Table C (starts on next page)

Amino acid sequence	Nucleotide sequence
GSGGEPSESGSSGESPGG SSGSESGESPGGSSGSES	GGTTCTGGTGGCGAACCGTCCGAGTCTGGTAGC TCAGGTGAATCTCCGGGTGGCTCTAGCGGTTCC GAGTCAGGTGAATCTCCTGGTGGTTCCAGCGGT TCCGAGTCA
GSEGSSGPGESSGESPGG SSGSESGSSESGSSEGGP	GGTAGCGAAGGTTCTTCTGGTCCTGGCGAGTCT TCAGGTGAATCTCCTGGTGGTTCCAGCGGTTCT GAATCAGGTTCCTCCGAAAGCGGTTCTTCCGAG GGCGGTCCA
GSSESGSSEGGPGSSESG SSEGGPGESPGGSSGSES	GGTTCCTCTGAAAGCGGTTCTTCCGAAGGTGGT CCAGGTTCCTCTGAAAGCGGTTCTTCTGAGGGT GGTCCAGGTGAATCTCCGGGTGGCTCCAGCGGT TCCGAGTCA
GSGGEPSESGSSGSSESG SSEGGPGSGGEPSESGSS	GGTTCCGGTGGCGAACCGTCTGAATCTGGTAGC TCAGGTTCTTCTGAAAGCGGTTCTTCCGAGGGT GGTCCAGGTTCTGGTGGTGAACCTTCCGAGTCT GGTAGCTCA
GSSESGSSEGGPGSEGSS GPGESSGSEGSSGPGESS	GGTTCTTCCGAAAGCGGTTCTTCTGAGGGTGGT CCAGGTAGCGAAGGTTCTTCCGGTCCAGGTGAG TCTTCAGGTAGCGAAGGTTCTTCTGGTCCTGGT GAATCTTCA
GSSESGSSEGGPGESPGG SSGSESGSEGSSGPGESS	GGTTCCTCTGAAAGCGGTTCTTCCGAGGGTGGT CCAGGTGAATCTCCAGGTGGTTCCAGCGGTTCT GAGTCAGGTAGCGAAGGTTCTTCTGGTCCAGGT GAATCCTCA
GSGGEPSESGSSGSGGEP SESGSSGSEGSSGPGESS	GGTTCTGGTGGTGAACCGTCTGAGTCTGGTAGC TCAGGTTCCGGTGGCGAACCATCCGAATCTGGT AGCTCAGGTAGCGAAGGTTCTTCCGGTCCAGGT GAGTCTTCA
GSSESGSSEGGPGSEGSS GPGESSGESPGGSSGSES	GGTTCTTCCGAAAGCGGTTCTTCCGAAGGCGGT CCAGGTAGCGAAGGTTCTTCTGGTCCAGGCGAA TCTTCAGGTGAATCTCCTGGTGGCTCCAGCGGT TCTGAGTCA
GSSESGSSEGGPGSSESG SSEGGPGSSESGSSEGGP	GGTTCCTCCGAAAGCGGTTCTTCTGAGGGCGGT CCAGGTTCCTCCGAAAGCGGTTCTTCCGAGGGC GGTCCAGGTTCTTCTGAAAGCGGTTCTTCCGAG GGCGGTCCA
GSGGEPSESGSSGSEGSS GPGESSGSSESGSSEGGP	GGTTCCGGTGGCGAACCGTCCGAATCTGGTAGC TCAGGTAGCGAAGGTTCTTCTGGTCCAGGCGAA TCTTCAGGTTCCTCTGAAAGCGGTTCTTCTGAG GGCGGTCCA
GSGGEPSESGSSGSSESG SSEGGPGSGGEPSESGSS	GGTTCTGGTGGTGAACCGTCCGAATCTGGTAGC TCAGGTTCTTCCGAAAGCGGTTCTTCTGAAGGT GGTCCAGGTTCCGGTGGCGAACCTTCTGAATCT GGTAGCTCA
GSGGEPSESGSSGSSESG SSEGGPGESPGGSSGSES	GGTTCTGGTGGCGAACCGTCCGAATCTGGTAGC TCAGGTTCCTCCGAAAGCGGTTCTTCTGAAGGT GGTCCAGGTGAATCTCCAGGTGGTTCTAGCGGT TCTGAATCA
GSGGEPSESGSSGESPGG SSGSESGSEGSSGPGESS	GGTTCTGGTGGCGAACCGTCTGAGTCTGGTAGC TCAGGTGAATCTCCTGGTGGCTCCAGCGGTTCT

Amino acid sequence	Nucleotide sequence
	GAATCAGGTAGCGAAGGTTCTTCTGGTCCTGGT GAATCTTCA
GSGGEPSESGSSGESPGG SSGSESGSGGEPSESGSS	GGTTCCGGTGGCGAACCTTCCGAATCTGGTAGC TCAGGTGAATCTCCGGGTGGTTCTAGCGGTTCT GAGTCAGGTTCTGGTGGTGAACCTTCCGAGTCT GGTAGCTCA
GSSESGSSEGGPGSSESG SSEGGPGSSESGSSEGGP	GGTTCCTCTGAAAGCGGTTCTTCTGAGGGCGGT CCAGGTTCTTCCGAAAGCGGTTCTTCCGAGGGC GGTCCAGGTTCTTCCGAAAGCGGTTCTTCTGAA GGCGGTCCA
GESPGGSSGSESGSEGSS GPGESSGSEGSSGPGESS	GGTGAATCTCCGGGTGGCTCCAGCGGTTCTGAG TCAGGTAGCGAAGGTTCTTCCGGTCCGGGTGAG TCCTCAGGTAGCGAAGGTTCTTCCGGTCCTGGT GAGTCTTCA
GSGGEPSESGSSGSGGEP SESGSSGSSESGSSEGGP	GGTTCTGGTGGCGAACCTTCCGAATCTGGTAGC TCAGGTTCCGGTGGTGAACCTTCTGAATCTGGT AGCTCAGGTTCTTCTGAAAGCGGTTCTTCCGAG GGCGGTCCA
GSGGEPSESGSSGSGGEP SESGSSGSGGEPSESGSS	GGTTCCGGTGGTGAACCTTCTGAATCTGGTAGC TCAGGTTCCGGTGGCGAACCATCCGAGTCTGGT AGCTCAGGTTCCGGTGGTGAACCATCCGAGTCT GGTAGCTCA
GSGGEPSESGSSGSSESG SSEGGPGSEGSSGPGESS	GGTTCCGGTGGCGAACCTTCTGAATCTGGTAGC TCAGGTTCCTCCGAAAGCGGTTCTTCTGAGGGC GGTCCAGGTAGCGAAGGTTCTTCTGGTCCGGGC GAGTCTTCA
GSGGEPSESGSSGSEGSS GPGESSGSGGEPSESGSS	GGTTCCGGTGGTGAACCGTCCGAGTCTGGTAGC TCAGGTAGCGAAGGTTCTTCTGGTCCGGGTGAG TCTTCAGGTTCTGGTGGCGAACCGTCCGAATCT GGTAGCTCA
GSGGEPSESGSSGESPGG SSGSESGSGGEPSESGSS	GGTTCTGGTGGCGAACCGTCCGAATCTGGTAGC TCAGGTGAATCTCCTGGTGGTTCCAGCGGTTCC GAGTCAGGTTCTGGTGGCGAACCTTCCGAATCT GGTAGCTCA
GSSESGSSEGGPGSGGEP SESGSSGSSESGSSEGGP	GGTTCTTCCGAAAGCGGTTCTTCCGAGGGCGGT CCAGGTTCCGGTGGTGAACCATCTGAATCTGGT AGCTCAGGTTCTTCTGAAAGCGGTTCTTCTGAA GGTGGTCCA
GSEGSSGPGESSGESPGG SSGSESGSEGSSGPGESS	GGTAGCGAAGGTTCTTCCGGTCCTGGTGAGTCT TCAGGTGAATCTCCAGGTGGCTCTAGCGGTTCC GAGTCAGGTAGCGAAGGTTCTTCTGGTCCTGGC GAGTCCTCA
GSSESGSSEGGPGSSESG SSEGGPGSSESGSSEGGP	GGTTCCTCTGAAAGCGGTTCTTCCGAAGGCGGT CCAGGTTCTTCCGAAAGCGGTTCTTCTGAGGGC GGTCCAGGTTCCTCCGAAAGCGGTTCTTCTGAG GGTGGTCCA
GSGGEPSESGSSGESPGG SSGSESGESPGGSSGSES	GGTTCTGGTGGCGAACCTTCCGAGTCTGGTAGC TCAGGTGAATCTCCGGGTGGTTCTAGCGGTTCC GAGTCAGGTGAATCTCCGGGTGGTTCCAGCGGT TCTGAGTCA
GSGGEPSESGSSGSEGSS GPGESSGESPGGSSGSES	GGTTCTGGTGGCGAACCATCTGAGTCTGGTAGC TCAGGTAGCGAAGGTTCTTCCGGTCCAGGCGAG TCTTCAGGTGAATCTCCTGGTGGCTCCAGCGGT TCTGAGTCA
GESPGGSSGSESGESPGG SSGSESGESPGGSSGSES	GGTGAATCTCCTGGTGGTTCCAGCGGTTCCGAG TCAGGTGAATCTCCAGGTGGCTCTAGCGGTTCC GAGTCAGGTGAATCTCCTGGTGGTTCTAGCGGT TCTGAATCA
GSEGSSGPGESSGSEGSS	GGTAGCGAAGGTTCTTCCGGTCCAGGTGAATCT

Amino acid sequence	Nucleotide sequence
GPGESSGSGGEPSESGSS	TCAGGTAGCGAAGGTTCTTCTGGTCCTGGTGAA TCCTCAGGTTCCGGTGGCGAACCATCTGAATCT GGTAGCTCA
GSGGEPSESGSSGSEGSS GPGESSGESPGGSSGSES	GGTTCTGGTGGCGAACCATCCGAATCTGGTAGC TCAGGTAGCGAAGGTTCTTCTGGTCCTGGCGAA TCTTCAGGTGAATCTCCAGGTGGCTCTAGCGGT TCCGAATCA
GSGGEPSESGSSGSSESG SSEGGPGSGGEPSESGSS	GGTTCCGGTGGTGAACCGTCCGAATCTGGTAGC TCAGGTTCCTCTGAAAGCGGTTCTTCCGAGGGT GGTCCAGGTTCCGGTGGTGAACCTTCTGAGTCT GGTAGCTCA
GSSESGSSEGGPGSGGEP SESGSSGSEGSSGPGESS	GGTTCCTCTGAAAGCGGTTCTTCTGAGGGCGGT CCAGGTTCTGGTGGCGAACCATCTGAATCTGGT AGCTCAGGTAGCGAAGGTTCTTCCGGTCCGGGT GAATCTTCA
GSGGEPSESGSSGSEGSS GPGESSGSEGSSGPGESS	GGTTCTGGTGGTGAACCGTCCGAATCTGGTAGC TCAGGTAGCGAAGGTTCTTCTGGTCCTGGCGAG TCTTCAGGTAGCGAAGGTTCTTCTGGTCCTGGT GAATCTTCA
GSGGEPSESGSSGSSESG SSEGGPGSGGEPSESGSS	GGTTCTGGTGGCGAACCATCCGAGTCTGGTAGC TCAGGTTCTTCCGAAAGCGGTTCTTCCGAAGGC GGTCCAGGTTCTGGTGGTGAACCGTCCGAATCT GGTAGCTCA
GSGGEPSESGSSGESPGG SSGSESGESPGGSSGSES	GGTTCCGGTGGCGAACCTTCCGAATCTGGTAGC TCAGGTGAATCTCCGGGTGGTTCTAGCGGTTCC GAATCAGGTGAATCTCCAGGTGGTTCTAGCGGT TCCGAATCA
GSGGEPSESGSSGSGGEP SESGSSGESPGGSSGSES	GGTTCCGGTGGTGAACCATCTGAGTCTGGTAGC TCAGGTTCCGGTGGCGAACCGTCCGAGTCTGGT AGCTCAGGTGAATCTCCGGGTGGTTCCAGCGGT TCCGAATCA
GSEGSSGPGESSGSSESG SSEGGPGSEGSSGPGESS	GGTAGCGAAGGTTCTTCTGGTCCGGGCGAATCC TCAGGTTCCTCCGAAAGCGGTTCTTCCGAAGGT GGTCCAGGTAGCGAAGGTTCTTCCGGTCCTGGT GAATCTTCA
GSSESGSSEGGPGESPGG SSGSESGESPGGSSGSES	GGTTCCTCTGAAAGCGGTTCTTCTGAAGGCGGT CCAGGTGAATCTCCGGGTGGCTCCAGCGGTTCT GAATCAGGTGAATCTCCTGGTGGCTCCAGCGGT TCCGAGTCA
GSEGSSGPGESSGSEGSS GPGESSGSGGEPSESGSS	GGTAGCGAAGGTTCTTCTGGTCCAGGCGAGTCT TCAGGTAGCGAAGGTTCTTCCGGTCCTGGCGAG TCTTCAGGTTCCGGTGGCGAACCGTCCGAATCT GGTAGCTCA

[0007] Table D

Amino acid sequence	Nucleotide sequence
GSPAGSPTSTEEGTSE SATPESGPGTSTEPSE GSAP	GGTAGCCCGGCAGGCTCTCCGACCTCTACTGAGGAA GGTACTTCTGAAAGCGCAACCCCGGAGTCCGGCCCA GGTACCTCTACCGAACCGTCTGAGGGCAGCGCACCA
GTSTEPSEGSAPGTST EPSEGSAPGTSTEPSE GSAP	GGTACTTCTACCGAACCGTCCGAAGGCAGCGCTCCA GGTACCTCTACTGAACCTTCCGAGGGCAGCGCTCCA GGTACCTCTACCGAACCTTCTGAAGGTAGCGCACCA
GTSTEPSEGSAPGTSE SATPESGPGTSESATP ESGP	GGTACCTCTACCGAACCGTCTGAAGGTAGCGCACCA GGTACCTCTGAAAGCGCAACTCCTGAGTCCGGTCCA GGTACTTCTGAAAGCGCAACCCCGGAGTCTGGCCCA
GTSTEPSEGSAPGTSE SATPESGPGTSESATP ESGP	GGTACTTCTACTGAACCGTCTGAAGGTAGCGCACCA GGTACTTCTGAAAGCGCAACCCCGGAATCCGGCCCA GGTACCTCTGAAAGCGCAACCCCGGAGTCCGGCCCA
GSEPATSGSETPGTSE SATPESGPGSPAGSPT STEE	GGTAGCGAACCGGCAACCTCCGGCTCTGAAACCCCA GGTACCTCTGAAAGCGCTACTCCTGAATCCGGCCCA GGTAGCCCCGGCAGGTTCTCCGACTTCCACTGAGGAA
GTSESATPESGPGSEP ATSGSETPGTSTEPSE GSAP	GGTACTTCTGAAAGCGCAACCCCTGAATCCGGTCCA GGTAGCGAACCGGCTACTTCTGGCTCTGAGACTCCA GGTACTTCTACCGAACCGTCCGAAGGTAGCGCACCA
GSPAGSPTSTEEGSPA GSPTSTEEGSEPATSG SETP	GGTAGCCCGGCTGGCTCTCCAACCTCCACTGAGGAA GGTAGCCCGGCTGGCTCTCCAACCTCCACTGAAGAA GGTAGCGAACCGGCTACCTCCGGCTCTGAAACTCCA
GSPAGSPTSTEEGTSE SATPESGPGTSTEPSE GSAP	GGTAGCCCGGCTGGCTCTCCTACCTCTACTGAGGAA GGTACTTCTGAAAGCGCTACTCCTGAGTCTGGTCCA GGTACCTCTACTGAACCGTCCGAAGGTAGCGCTCCA
GSPAGSPTSTEEGSPA GSPTSTEEGTSTEPSE GSAP	GGTAGCCCTGCTGGCTCTCCGACTTCTACTGAGGAA GGTAGCCCGGCTGGTTCTCCGACTTCTACTGAGGAA GGTACTTCTACCGAACCTTCCGAAGGTAGCGCTCCA

Amino acid sequence	Nucleotide sequence
GTSESATPESGPGTST	GGTACTTCTGAAAGCGCTACTCCGGAGTCCGGTCCA
EPSEGSAPGTSTEPSE	GGTACCTCTACCGAACCGTCCGAAGGCAGCGCTCCA
GSAP	GGTACTTCTACTGAACCTTCTGAGGGTAGCGCTCCA
GTSTEPSEGSAPGSPA	GGTACCTCTACCGAACCTTCCGAAGGTAGCGCTCCA
GSPTSTEEGTSTEPSE	GGTAGCCCGGCAGGTTCTCCTACTTCCACTGAGGAA
GSAP	GGTACTTCTACCGAACCTTCTGAGGGTAGCGCACCA
GSEPATSGSETPGSPA	GGTAGCGAACCGGCTACTTCCGGCTCTGAGACTCCA
GSPTSTEEGTSESATP	GGTAGCCCTGCTGGCTCTCCGACCTCTACCGAAGAA
ESGP	GGTACCTCTGAAAGCGCTACCCCTGAGTCTGGCCCA
GTSTEPSEGSAPGTSE	GGTACTTCTACCGAACCTTCCGAGGGCAGCGCACCA
SATPESGPGTSESATP	GGTACTTCTGAAAGCGCTACCCCTGAGTCCGGCCCA
ESGP	GGTACTTCTGAAAGCGCTACTCCTGAATCCGGTCCA
GTSTEPSEGSAPGSEP	GGTACTTCTACTGAACCGTCTGAAGGCAGCGCACCA
ATSGSETPGSPAGSPT	GGTAGCGAACCGGCTACTTCCGGTTCTGAAACCCCA
STEE	GGTAGCCCAGCAGGTTCTCCAACTTCTACTGAAGAA
GSPAGSPTSTEEGTSE	GGTAGCCCTGCTGGCTCTCCAACCTCCACCGAAGAA
SATPESGPGSEPATSG	GGTACCTCTGAAAGCGCAACCCCTGAATCCGGCCCA
SETP	GGTAGCGAACCGGCAACCTCCGGTTCTGAAACCCCA
GTSESATPESGPGSPA	GGTACTTCTGAAAGCGCTACTCCTGAGTCCGGCCCA
GSPTSTEEGSPAGSPT	GGTAGCCCGGCTGGCTCTCCGACTTCCACCGAGGAA
STEE	GGTAGCCCGGCTGGCTCTCCAACTTCTACTGAAGAA
GTSTEPSEGSAPGTSE	GGTACCTCTACTGAACCTTCTGAGGGCAGCGCTCCA
SATPESGPGTSTEPSE	GGTACTTCTGAAAGCGCTACCCCCGGAGTCCGGTCCA
GSAP	GGTACTTCTACTGAACCGTCCGAAGGTAGCGCACCA
GSPAGSPTSTEEGTST	GGTAGCCCCGGCAGGCTCTCCGACTTCCACCGAGGAA
EPSEGSAPGSEPATSG	GGTACCTCTACTGAACCTTCTGAGGGTAGCGCTCCA
SETP	GGTAGCGAACCGGCAACCTCTGGCTCTGAAACCCCA
GSPAGSPTSTEEGTST	GGTAGCCCAGCAGGCTCTCCGACTTCCACTGAGGAA
EPSEGSAPGTSTEPSE	GGTACTTCTACTGAACCTTCCGAAGGCAGCGCACCA
GSAP	GGTACCTCTACTGAACCTTCTGAGGGCAGCGCTCCA
GSEPATSGSETPGTSE	GGTAGCGAACCTGCTACCTCCGGTTCTGAAACCCCA
SATPESGPGSPAGSPT	GGTACCTCTGAAAGCGCAACTCCGGAGTCTGGTCCA
STEE	GGTAGCCCTGCAGGTTCTCCTACCTCCACTGAGGAA
GTSESATPESGPGTST	GGTACCTCTGAAAGCGCTACTCCGGAGTCTGGCCCA
EPSEGSAPGTSTEPSE	GGTACCTCTACTGAACCGTCTGAGGGTAGCGCTCCA
GSAP	GGTACTTCTACTGAACCGTCCGAAGGTAGCGCACCA
GSPAGSPTSTEEGTST	GGTAGCCCGGCTGGTTCTCCGACTTCCACCGAGGAA
EPSEGSAPGTSTEPSE	GGTACCTCTACTGAACCTTCTGAGGGTAGCGCTCCA
GSAP	GGTACCTCTACTGAACCTTCCGAAGGCAGCGCTCCA
GTSTEPSEGSAPGTST	GGTACTTCTACCGAACCGTCCGAGGGCAGCGCTCCA
EPSEGSAPGTSTEPSE	GGTACTTCTACTGAACCTTCTGAAGGCAGCGCTCCA
GSAP	GGTACTTCTACTGAACCTTCCGAAGGTAGCGCACCA
GSEPATSGSETPGTSE	GGTAGCGAACCTGCAACCTCTGGCTCTGAAACCCCA
SATPESGPGTSTEPSE	GGTACCTCTGAAAGCGCTACTCCTGAATCTGGCCCA
GSAP	GGTACTTCTACTGAACCGTCCGAGGGCAGCGCACCA
GTSTEPSEGSAPGSPA	GGTACTTCTACCGAACCGTCCGAGGGTAGCGCACCA
GSPTSTEEGTSTEPSE	GGTAGCCCAGCAGGTTCTCCTACCTCCACCGAGGAA
GSAP	GGTACTTCTACCGAACCGTCCGAGGGTAGCGCACCA
GSEPATSGSETPGTSE	GGTAGCGAACCGGCAACCTCCGGCTCTGAAAACTCCA
SATPESGPGSEPATSG	GGTACTTCTGAAAGCGCTACTCCGGAATCCGGCCCA
SETP	GGTAGCGAACCGGCTACTTCCGGCTCTGAAACCCCA
GSEPATSGSETPGTSE	GGTAGCGAACCGGCAACTTCCGGCTCTGAAACCCCA
SATPESGPGSEPATSG	GGTACTTCTGAAAGCGCTACTCCTGAGTCTGGCCCA
SETP	GGTAGCGAACCTGCTACCTCTGGCTCTGAAACCCCA
GSEPATSGSETPGSEP	GGTAGCGAACCGGCAACCTCTGGCTCTGAAACTCCA
ATSGSETPGTSTEPSE	GGTAGCGAACCTGCAACCTCCGGCTCTGAAACCCCA
GSAP	GGTACTTCTACTGAACCTTCTGAGGGCAGCGCACCA
GTSESATPESGPGSEP	GGTACTTCTGAAAGCGCTACCCCGGAATCTGGCCCA
ATSGSETPGSEPATSG	GGTAGCGAACCGGCTACTTCTGGTTCTGAAACCCCA

Amino acid sequence	Nucleotide sequence
SETP	GGTAGCGAACCGGCTACCTCCGGTTCTGAAACTCCA
GTSTEPSEGSAPGTST EPSEGSAPGTSESATP ESGP	GGTACCTCTACTGAACCTTCCGAAGGCAGCGCTCCA GGTACCTCTACCGAACCGTCCGAGGGCAGCGCACCA GGTACTTCTGAAAGCGCAACCCCTGAATCCGGTCCA
GSEPATSGSETPGTSE SATPESGPGTSESATP ESGP	GGTAGCGAACCGGCAACCTCTGGCTCTGAAACCCCA GGTACCTCTGAAAGCGCTACTCCGGAATCTGGTCCA GGTACTTCTGAAAGCGCTACTCCGGAATCCGGTCCA
GSEPATSGSETPGSEP ATSGSETPGTSTEPSE GSAP	GGTAGCGAACCTGCTACCTCCGGCTCTGAAACTCCA GGTAGCGAACCGGCTACTTCCGGTTCTGAAACTCCA GGTACCTCTACCGAACCTTCCGAAGGCAGCGCACCA
GSEPATSGSETPGTST EPSEGSAPGSEPATSG SETP	GGTAGCGAACCTGCTACTTCTGGTTCTGAAACTCCA GGTACTTCTACCGAACCGTCCGAGGGTAGCGCTCCA GGTAGCGAACCTGCTACTTCTGGTTCTGAAACTCCA
GTSTEPSEGSAPGTST EPSEGSAPGSEPATSG SETP	GGTACCTCTACCGAACCGTCCGAGGGTAGCGCACCA GGTACCTCTACTGAACCGTCTGAGGGTAGCGCTCCA GGTAGCGAACCGGCAACCTCCGGTTCTGAAACTCCA
GTSTEPSEGSAPGSEP ATSGSETPGSPAGSPT STEE	GGTACTTCTACTGAACCTTCCGAAGGTAGCGCTCCA GGTAGCGAACCTGCTACTTCTGGTTCTGAAACCCCA GGTAGCCCGGCTGGCTCTCCGACCTCCACCGAGGAA
GSEPATSGSETPGSPA GSPTSTEEGTSESATP ESGP	GGTAGCGAACCGGCTACTTCCGGCTCTGAGACTCCA GGTAGCCCAGCTGGTTCTCCAACCTCTACTGAGGAA GGTACTTCTGAAAGCGCTACCCCTGAATCTGGTCCA
GTSESATPESGPGSEP ATSGSETPGTSESATP ESGP	GGTACCTCTGAAAGCGCAACTCCTGAGTCTGGCCCA GGTAGCGAACCTGCTACCTCCGGCTCTGAGACTCCA GGTACCTCTGAAAGCGCAACCCCGGAATCTGGTCCA

[0008] Table E

Amino acid sequence	Nucleotide sequence
GTSTPESGSASPGTSP	GGTACTTCTACTCCGGAAAGCGGTTCCGCATCTCCA
SGESSTAPGTSPSGES	GGTACTTCTCCTAGCGGTGAATCTTCTACTGCTCCAG
STAP	GTACCTCTCCTAGCGGCGAATCTTCTACTGCTCCA
GTSPSGESSTAPGSTS	GGTACTTCTCCGAGCGGTGAATCTTCTACCGCACCA
STAESPGPGTSPSGES	GGTTCTACTAGCTCTACCGCTGAATCTCCGGGCCCAG
STAP	GTACTTCTCCGAGCGGTGAATCTTCTACTGCTCCA
GSTSSTAESPGPGTSP	GGTTCCACCAGCTCTACTGCTGAATCTCCTGGTCCAG
SGESSTAPGTSTPESG	GTACCTCTCCTAGCGGTGAATCTTCTACTGCTCCAGG
SASP	TACTTCTACTCCTGAAAGCGGCTCTGCTTCTCCA
GSTSSTAESPGPGSTS	GGTTCTACCAGCTCTACTGCAGAATCTCCTGGCCCAG
STAESPGPGTSPSGES	GTTCCACCAGCTCTACCGCAGAATCTCCGGGTCCAG
STAP	GTACTTCCCCTAGCGGTGAATCTTCTACCGCACCA
GSTSSTAESPGPGTSP	GGTTCTACTAGCTCTACTGCTGAATCTCCTGGCCCAG
SGESSTAPGTSTPESG	GTACTTCTCCTAGCGGTGAATCTTCTACCGCTCCAGG
SASP	TACCTCTACTCCGGAAAGCGGTTCTGCATCTCCA
GSTSSTAESPGPGTST	GGTTCTACCAGCTCTACCGCAGAATCTCCTGGTCCAG
PESGSASPGSTSESPS	GTACCTCTACTCCGGAAAGCGGCTCTGCATCTCCAG
GTAP	GTTCTACTAGCGAATCTCCTTCTGGCACTGCACCA
GSTSSTAESPGPGTST	GGTTCTACTAGCTCTACTGCAGAATCTCCTGGCCCAG
PESGSASPGTSTPESG	GTACCTCTACTCCGGAAAGCGGCTCTGCATCTCCAG
SASP	GTACTTCTACCCCTGAAAGCGGTTCTGCATCTCCA
GSTSESPSGTAPGTSP	GGTTCTACCAGCGAATCTCCTTCTGGCACCGCTCCAG
SGESSTAPGSTSESPS	GTACCTCTCCTAGCGGCGAATCTTCTACCGCTCCAGG
GTAP	TTCTACTAGCGAATCTCCTTCTGGCACTGCACCA
GSTSSTAESPGPGSTS	GGTTCCACCAGCTCTACTGCAGAATCTCCGGGCCCA
STAESPGPGTSPSGES	GGTTCTACTAGCTCTACTGCAGAATCTCCGGGTCCAG
STAP	GTACTTCTCCTAGCGGCGAATCTTCTACCGCTCCA
GSTSSTAESPGPGTST	GGTTCCACTAGCTCTACTGCAGAATCTCCTGGCCCAG
PESGSASPGSTSESPS	GTACCTCTACCCCTGAAAGCGGCTCTGCATCTCCAG
GTAP	GTTCTACCAGCGAATCCCCGTCTGGCACCGCACCA
GSTSSTAESPGPGSTS	GGTTCTACTAGCTCTACTGCTGAATCTCCGGGTCCAG
STAESPGPGTSPSGES	GTTCTACCAGCTCTACTGCTGAATCTCCTGGTCCAGG
STAP	TACCTCCCCGAGCGGTGAATCTTCTACTGCACCA
GSTSSTAESPGPGSTS	GGTTCTACCAGCTCTACCGCTGAATCTCCTGGCCCAG
ESPSGTAPGSTSSTAE	GTTCTACCAGCGAATCCCCGTCTGGCACCGCACCAG
SPGP	GTTCTACTAGCTCTACCGCTGAATCTCCGGGTCCA
GSTSSTAESPGPGSTS	GGTTCTACCAGCTCTACCGCAGAATCTCCTGGCCCA
STAESPGPGSTSSTAE	GGTTCCACTAGCTCTACCGCTGAATCTCCTGGTCCAG
SPGP	GTTCTACTAGCTCTACCGCTGAATCTCCTGGTCCA
GSTSESPSGTAPGSTS	GGTTCTACTAGCGAATCCCCTTCTGGTACTGCTCCAG
STAESPGPGSTSSTAE	GTTCCACTAGCTCTACCGCTGAATCTCCTGGCCCAGG
SPGP	TTCCACTAGCTCTACTGCAGAATCTCCTGGTCCA

Amino acid sequence	Nucleotide sequence
GSTSESPSGTAPGSTS ESPSGTAPGSTSESPS GTAP	GGTTCTACTAGCGAATCTCCTTCTGGTACCGCTCCAG GTTCTACCAGCGAATCCCCGTCTGGTACTGCTCCAGG TTCTACCAGCGAATCTCCTTCTGGTACTGCACCA
GSTSSTAESPGPGSTS STAESPGPGSTSSTAE SPGP	GGTTCCACCAGCTCTACTGCTGAATCTCCTGGCCCAG GTTCTACCAGCTCTACTGCTGAATCTCCGGGCCCAGG TTCCACCAGCTCTACCGCTGAATCTCCGGGTCCA
GSTSSTAESPGPGSTS STAESPGPGTSPSGES STAP	GGTTCCACTAGCTCTACCGCAGAATCTCCTGGTCCAG GTTCTACTAGCTCTACTGCTGAATCTCCGGGTCCAGG TACCTCCCCTAGCGGCGAATCTTCTACCGCTCCA
GSSPSASTGTGPGSST PSGATGSPGSSTPSGA TGSP	GGTTCTAGCCCTTCTGCTTCCACCGGTACCGGCCCAG GTAGCTCTACTCCGTCTGGTGCAACTGGCTCTCCAGG TAGCTCTACTCCGTCTGGTGCAACCGGCTCCCCA
GTSPSGESSTAPGTST PESGSASPGSTSSTAE SPGP	GGTACTTCCCCTAGCGGTGAATCTTCTACTGCTCCAG GTACCTCTACTCCGGAAAGCGGCTCCGCATCTCCAG GTTCTACTAGCTCTACTGCTGAATCTCCTGGTCCA
GSTSSTAESPGPGSTS STAESPGPGTSTPESG SASP	GGTTCTACTAGCTCTACCGCTGAATCTCCGGGTCCAG GTTCTACCAGCTCTACTGCAGAATCTCCTGGCCCAGG TACTTCTACTCCGGAAAGCGGTTCCGCTTCTCCA
GTSPSGESSTAPGSTS STAESPGPGTSTPESG SASP	GGTACTTCTCCTAGCGGTGAATCTTCTACCGCTCCAG GTTCTACCAGCTCTACTGCTGAATCTCCTGGCCCAGG TACTTCTACCCCGGAAAGCGGCTCCGCTTCTCCA
GSTSESPSGTAPGSTS STAESPGPGSTSSTAE SPGP	GGTTCTACTAGCGAATCCCCTTCTGGTACTGCACCAG GTTCTACCAGCTCTACTGCTGAATCTCCGGGCCCAGG TTCCACCAGCTCTACCGCAGAATCTCCTGGTCCA
GSTSSTAESPGPGSTS ESPSGTAPGSTSSTAE SPGP	GGTTCCACCAGCTCTACCGCTGAATCTCCGGGCCCA GGTTCTACCAGCGAATCCCCTTCTGGCACTGCACCA GGTTCTACTAGCTCTACCGCAGAATCTCCGGGCCCA
GSTSSTAESPGPGTSP SGESSTAPGTSTPESG SASP	GGTTCTACCAGCTCTACTGCTGAATCTCCGGGTCCAG GTACTTCCCCGAGCGGTGAATCTTCTACTGCACCAG GTACTTCTACTCCGGAAAGCGGTTCCGCTTCTCCA
GSTSESPSGTAPGSTS ESPSGTAPGTSPSGES STAP	GGTTCTACCAGCGAATCTCCTTCTGGCACCGCTCCAG GTTCTACTAGCGAATCCCCGTCTGGTACCGCACCAG GTACTTCTCCTAGCGGCGAATCTTCTACCGCACCA
GSTSESPSGTAPGSTS ESPSGTAPGTSTPESG SASP	GGTTCTACCAGCGAATCCCCTTCTGGTACTGCTCCAG GTTCTACCAGCGAATCCCCTTCTGGCACCGCACCAG GTACTTCTACCCCTGAAAGCGGCTCCGCTTCTCCA
GTSTPESGSASPGSTS STAESPGPGSTSSTAE SPGP	GGTACCTCTACTCCTGAAAGCGGTTCTGCATCTCCAG GTTCCACTAGCTCTACCGCAGAATCTCCGGGCCCAG GTTCTACTAGCTCTACTGCTGAATCTCCTGGCCCA
GSTSESPSGTAPGSTS ESPSGTAPGTSPSGES STAP	GGTTCTACCAGCGAATCCCCTTCTGGCACTGCACCA GGTTCTACTAGCGAATCCCCTTCTGGTACCGCACCAG GTACTTCTCCGAGCGGCGAATCTTCTACTGCTCCA
GSTSSTAESPGPGSTS STAESPGPGSTSESPS GTAP	GGTTCTACTAGCTCTACCGCTGAATCTCCTGGCCCAG GTTCCACTAGCTCTACCGCAGAATCTCCGGGCCCAG GTTCTACTAGCGAATCCCCTTCTGGTACCGCTCCA
GSTSSTAESPGPGSTS STAESPGPGTSTPESG SASP	GGTTCCACCAGCTCTACTGCAGAATCTCCTGGCCCA GGTTCTACTAGCTCTACCGCAGAATCTCCTGGTCCAG GTACCTCTACTCCTGAAAGCGGTTCCGCATCTCCA
GSTSSTAESPGPGSTS STAESPGPGSTSESPS GTAP	GGTTCTACTAGCTCTACTGCTGAATCTCCGGGCCCAG GTTCTACTAGCTCTACCGCTGAATCTCCGGGTCCAGG TTCTACTAGCGAATCTCCTTCTGGTACCGCTCCA
GTSPSGESSTAPGSTS ESPSGTAPGSTSSTAE SPGP	GGTACCTCCCCCGAGCGGTGAATCTTCTACTGCACCA GGTTCTACTAGCGAATCCCCTTCTGGTACTGCTCCAG GTTCCACCAGCTCTACTGCAGAATCTCCGGGTCCA
GSTSESPSGTAPGTSP SGESSTAPGSTSSTAE SPGP	GGTTCTACTAGCGAATCCCCGTCTGGTACTGCTCCAG GTACTTCCCCTAGCGGTGAATCTTCTACTGCTCCAGG TTCTACCAGCTCTACCGCAGAATCTCCGGGTCCA
GSTSSTAESPGPGSTS ESPSGTAPGTSPSGES	GGTTCTACCAGCTCTACCGCTGAATCTCCTGGCCCAG GTTCTACTAGCGAATCTCCGTCTGGCACCGCACCAG

Amino acid sequence	Nucleotide sequence
STAP	GTACTTCCCCTAGCGGTGAATCTTCTACTGCACCA
GSTSESPSGTAPGSTS ESPSGTAPGTSTPESG SASP	GGTTCTACTAGCGAATCTCCTTCTGGCACTGCACCAG GTTCTACCAGCGAATCTCCGTCTGGCACTGCACCAG GTACCTCTACCCCTGAAAGCGGTTCCGCTTCTCCA
GTSTPESGSASPGSTS ESPSGTAPGSTSSTAE SPGP	GGTACCTCTACTCCGGAAAGCGGTTCCGCATCTCCA GGTTCTACCAGCGAATCCCCGTCTGGCACCGCACCA GGTTCTACTAGCTCTACTGCTGAATCTCCGGGCCCA
GSTSSTAESPGPGTSP SGESSTAPGTSPSGES STAP	GGTTCTACTAGCTCTACTGCAGAATCTCCGGGCCCA GGTACCTCTCCTAGCGGTGAATCTTCTACCGCTCCAG GTACTTCTCCGAGCGGTGAATCTTCTACCGCTCCA
GTSPSGESSTAPGTSP SGESSTAPGTSPSGES STAP	GGTACCTCCCCTAGCGGCGAATCTTCTACTGCTCCAG GTACCTCTCCTAGCGGCGAATCTTCTACCGCTCCAGG TACCTCCCCTAGCGGTGAATCTTCTACCGCACCA
GSTSSTAESPGPGTST PESGSASPGSTSESPS GTAP	GGTTCCACTAGCTCTACTGCTGAATCTCCTGGCCCAG GTACTTCTACTCCGGAAAGCGGTTCCGCTTCTCCAGG TTCTACTAGCGAATCTCCGTCTGGCACCGCACCA
GSTSESPSGTAPGTSP SGESSTAPGTSPSGES STAP	GGTTCTACTAGCGAATCTCCGTCTGGCACTGCTCCAG GTACTTCTCCTAGCGGTGAATCTTCTACCGCTCCAGG TACTTCCCCTAGCGGCGAATCTTCTACCGCTCCA
GSTSESPSGTAPGTST PESGSASPGSTSSTAE SPGP	GGTTCTACTAGCGAATCTCCTTCTGGTACCGCTCCAG GTACTTCTACCCCTGAAAGCGGCTCCGCTTCTCCAGG TTCCACTAGCTCTACCGCTGAATCTCCGGGTCCA
GSTSSTAESPGPGSTS STAESPGPGSTSESPS GTAP	GGTTCCACTAGCTCTACTGCTGAATCTCCTGGCCCAG GTTCTACCAGCTCTACCGCTGAATCTCCTGGCCCAGG TTCTACCAGCGAATCTCCGTCTGGCACCGCACCA
GSTSESPSGTAPGTST PESGSASPGTSTPESG SASP	GGTTCTACTAGCGAATCCCCGTCTGGTACCGCACCA GGTACTTCTACCCCGGAAAGCGGCTCTGCTTCTCCAG GTACTTCTACCCCGGAAAGCGGCTCCGCATCTCCA
GSTSESPSGTAPGTST PESGSASPGTSTPESG SASP	GGTTCTACTAGCGAATCCCCGTCTGGTACTGCTCCAG GTACTTCTACTCCTGAAAGCGGTTCCGCTTCTCCAGG TACCTCTACTCCGGAAAGCGGTTCTGCATCTCCA

[0009] Table F

Amino acid sequence	Nucleotide sequence
GASPGTSSTGSPGTPG SGTASSSPGSSTPSGA TGSP	GGTGCATCCCCGGGCACTAGCTCTACCGGTTCTCCA GGTACTCCTGGTAGCGGTACTGCTTCTTCTTCTCCAG GTAGCTCTACTCCTTCTGGTGCTACTGGTTCTCCA
GSSTPSGATGSPGSSP SASTGTGPGSSTPSGA TGSP	GGTAGCTCTACCCCTTCTGGTGCTACCGGCTCTCCAG GTTCTAGCCCGTCTGCTTCTACCGGTACCGGTCCAGG TAGCTCTACCCCTTCTGGTGCTACTGGTTCTCCA
GASPGTSSTGSPGSSP SASTGTGPGSSTPSGA TGSP	GGTGCATCTCCGGGTACTAGCTCTACCGGTTCTCCAG GTTCTAGCCCTTCTGCTTCCACTGGTACCGGCCCAGG TAGCTCTACCCCGTCTGGTGCTACTGGTTCCCCA
GTPGSGTASSSPGSST PSGATGSPGASPGTSS TGSP	GGTACTCCGGGCAGCGGTACTGCTTCTTCCTCTCCAG GTAGCTCTACCCCTTCTGGTGCAACTGGTTCCCCAGG TGCATCCCCTGGTACTAGCTCTACCGGTTCTCCA
GTPGSGTASSSPGASP GTSSTGSPGSRPSAST GTGP	GGTACCCCTGGCAGCGGTACTGCTTCTTCTTCTCCAG GTGCTTCCCCTGGTACCAGCTCTACCGGTTCTCCAGG TTCTAGACCTTCTGCATCCACCGGTACTGGTCCA
GASPGTSSTGSPGSST PSGATGSPGASPGTSS TGSP	GGTGCATCTCCTGGTACCAGCTCTACCGGTTCTCCAG GTAGCTCTACTCCTTCTGGTGCTACTGGCTCTCCAGG TGCTTCCCCGGGTACCAGCTCTACCGGTTCTCCA
GTPGSGTASSSPGSST PSGATGSPGSSTPSGA TGSP	GGTACCCCGGGCAGCGGTACCGCATCTTCCTCTCCA GGTAGCTCTACCCCGTCTGGTGCTACCGGTTCCCCAG GTAGCTCTACCCCGTCTGGTGCAACCGGCTCCCCA
GASPGTSSTGSPGASP GTSSTGSPGASPGTSS TGSP	GGTGCATCTCCGGGCACTAGCTCTACTGGTTCTCCAG GTGCATCCCCTGGCACTAGCTCTACTGGTTCTCCAGG TGCTTCTCCTGGTACCAGCTCTACTGGTTCTCCA
GSSTPSGATGSPGSSP SASTGTGPGASPGTSS TGSP	GGTAGCTCTACTCCGTCTGGTGCAACCGGCTCCCCA GGTTCTAGCCCGTCTGCTTCCACTGGTACTGGCCCAG GTGCTTCCCCGGGCACCAGCTCTACTGGTTCTCCA
GSSTPSGATGSPGSST PSGATGSPGTPGSGT ASSSP	GGTAGCTCTACTCCTTCTGGTGCTACCGGTTCCCCAG GTAGCTCTACTCCTTCTGGTGCTACTGGTTCCCCAGG TACTCCGGGCAGCGGTACTGCTTCTTCCTCTCCA
GSSTPSGATGSPGSST PSGATGSPGASPGTSS TGSP	GGTAGCTCTACTCCGTCTGGTGCAACCGGTTCCCCAG GTAGCTCTACTCCTTCTGGTGCTACTGGCTCCCCAGG TGCATCCCCTGGCACCAGCTCTACCGGTTCTCCA
GTPGSGTASSSPGSSP SASTGTGPGSSTPSGA TGSP	GGTACTCCTGGTAGCGGTACCGCATCTTCCTCTCCAG GTTCTAGCCCTTCTGCATCTACCGGTACCGGTCCAGG TAGCTCTACTCCTTCTGGTGCTACTGGCTCTCCA
GASPGTSSTGSPGSSP	GGTGCTTCCCCGGGCACTAGCTCTACCGGTTCTCCAG

Amino acid sequence	Nucleotide sequence
SASTGTGPGTPGSGT ASSSP	GTTCTAGCCCTTCTGCATCTACTGGTACTGGCCCAGG TACTCCGGGCAGCGGTACTGCTTCTTCCTCTCCA
GSSTPSGATGSPGSST PSGATGSPGASPGTSS TGSP	GGTAGCTCTACTCCGTCTGGTGCTACCGGCTCTCCAG GTAGCTCTACCCCTTCTGGTGCAACCGGCTCCCCAGG TGCTTCTCCGGGTACCAGCTCTACTGGTTCTCCA
GTPGSGTASSSPGSST PSGATGSPGSSPSAST GTGP	GGTACCCCTGGCAGCGGTACCGCTTCTTCCTCTCCAG GTAGCTCTACCCCGTCTGGTGCTACTGGCTCTCCAGG TTCTAGCCCGTCTGCATCTACCGGTACCGGCCCA
GSSTPSGATGSPGTPG SGTASSSPGTPGSGTA SSSP	GGTAGCTCTACTCCTTCTGGTGCAACCGGCTCCCCAG GTACCCCGGGCAGCGGTACCGCATCTTCCTCTCCAG GTACTCCGGGTAGCGGTACTGCTTCTTCTTCTCCA
GTPGSGTASSSPGSST PSGATGSPGASPGTSS TGSP	GGTACCCCGGGTAGCGGTACTGCTTCTTCCTCTCCAG GTAGCTCTACCCCTTCTGGTGCAACCGGCTCTCCAGG TGCTTCTCCGGGCACCAGCTCTACCGGTTCTCCA
GSSTPSGATGSPGSST PSGATGSPGASPGTSS TGSP	GGTAGCTCTACCCCGTCTGGTGCTACCGGCTCTCCAG GTAGCTCTACCCCGTCTGGTGCAACCGGCTCCCCAG GTGCATCCCCGGGTACTAGCTCTACCGGTTCTCCA
GASPGTSSTGSPGTPG SGTASSSPGSSTPSGA TGSP	GGTGCTTCTCCGGGCACCAGCTCTACTGGTTCTCCAG GTACCCCGGGCAGCGGTACCGCATCTTCTTCTCCAG GTAGCTCTACTCCTTCTGGTGCAACTGGTTCTCCA
GSSPSASTGTGPGSST PSGATGSPGTPGSGT ASSSP	GGTTCTAGCCCGTCTGCTTCCACCGGTACTGGCCCAG GTAGCTCTACCCCGTCTGGTGCAACTGGTTCCCCAGG TACCCCTGGTAGCGGTACCGCTTCTTCTTCTCCA
GASPGTSSTGSPGSSP SASTGTGPGSSTPSGA TGSP	GGTGCTTCTCCGGGCACCAGCTCTACTGGTTCTCCAG GTTCTAGCCCTTCTGCATCCACCGGTACCGGTCCAGG TAGCTCTACCCCTTCTGGTGCAACCGGCTCTCCA
GASPGTSSTGSPGSST PSGATGSPGSSTPSGA TGSP	GGTGCATCCCCGGGCACCAGCTCTACCGGTTCTCCA GGTAGCTCTACCCCGTCTGGTGCTACCGGCTCTCCAG GTAGCTCTACCCCGTCTGGTGCTACTGGCTCTCCA
GTPGSGTASSSPGSST PSGATGSPGTPGSGT ASSSP	GGTACCCCTGGTAGCGGTACTGCTTCTTCCTCTCCAG GTAGCTCTACTCCGTCTGGTGCTACCGGTTCTCCAGG TACCCCGGGTAGCGGTACCGCATCTTCTTCTCCA
GSSPSASTGTGPGSST PSGATGSPGSSTPSGA TGSP	GGTTCTAGCCCTTCTGCTTCCACCGGTACTGGCCCAG GTAGCTCTACCCCTTCTGGTGCTACCGGCTCCCCAGG TAGCTCTACTCCTTCTGGTGCAACTGGCTCTCCA
GASPGTSSTGSPGSSP SASTGTGPGSSTPSGA TGSP	GGTGCTTCTCCTGGCACCAGCTCTACTGGTTCTCCAG GTTCTAGCCCTTCTGCTTCTACCGGTACTGGTCCAGG TTCTAGCCCTTCTGCATCCACTGGTACTGGTCCA
GTPGSGTASSSPGASP GTSSTGSPGASPGTSS TGSP	GGTACTCCTGGCAGCGGTACCGCTTCTTCTTCTCCAG GTGCTTCTCCTGGTACTAGCTCTACTGGTTCTCCAGG TGCTTCTCCGGGCACTAGCTCTACTGGTTCTCCA
GSSTPSGATGSPGASP GTSSTGSPGSSTPSGA TGSP	GGTAGCTCTACCCCGTCTGGTGCTACCGGTTCCCCAG GTGCTTCTCCTGGTACTAGCTCTACCGGTTCTCCAGG TAGCTCTACCCCGTCTGGTGCTACTGGCTCTCCA
GSSTPSGATGSPGTPG SGTASSSPGSSTPSGA TGSP	GGTAGCTCTACCCCGTCTGGTGCTACTGGTTCTCCAG GTACTCCGGGCAGCGGTACTGCTTCTTCCTCTCCAGG TAGCTCTACCCCTTCTGGTGCTACTGGCTCTCCA
GASPGTSSTGSPGSSP SASTGTGPGSSTPSGA TGSP	GGTGCATCTCCTGGTACCAGCTCTACTGGTTCTCCAG GTTCTAGCCCTTCTGCTTCTACCGGTACCGGTCCAGG TAGCTCTACTCCTTCTGGTGCTACCGGTTCTCCA
GSSTPSGATGSPGSST PSGATGSPGSSTPSGA TGSP	GGTAGCTCTACCCCGTCTGGTGCTACTGGCTCTCCAG GTAGCTCTACTCCTTCTGGTGCTACTGGTTCCCCAGG TAGCTCTACCCCGTCTGGTGCAACTGGCTCTCCA
GASPGTSSTGSPGTPG SGTASSSPGASPGTSS TGSP	GGTGCATCCCCGGGTACCAGCTCTACCGGTTCTCCA GGTACTCCTGGCAGCGGTACTGCATCTTCCTCTCCAG GTGCTTCTCCGGGCACCAGCTCTACTGGTTCTCCA
GSSTPSGATGSPGSSP SASTGTGPGASPGTSS TGSP	GGTAGCTCTACTCCTTCTGGTGCAACTGGTTCTCCAG GTTCTAGCCCGTCTGCATCCACTGGTACCGGTCCAGG TGCTTCCCCTGGCACCAGCTCTACCGGTTCTCCA

[0010]

Amino acid sequence	Nucleotide sequence
GASPGTSSTGSPGSST PSGATGSPGSSPSAST GTGP	GGTGCATCTCCTGGTACTAGCTCTACTGGTTCTCCAG GTAGCTCTACTCCGTCTGGTGCAACCGGCTCTCCAGG TTCTAGCCCTTCTGCATCTACCGGTACTGGTCCA
GTPGSGTASSSPGSST PSGATGSPGASPGTSS TGSP	GGTACTCCTGGCAGCGGTACCGCATCTTCCTCTCCAG GTAGCTCTACTCCGTCTGGTGCAACTGGTTCCCCAGG TGCTTCTCCGGGTACCAGCTCTACCGGTTCTCCA
GSSTPSGATGSPGTPG SGTASSSPGSSTPSGA TGSP	GGTAGCTCTACCCCGTCTGGTGCAACCGGCTCCCCA GGTACTCCTGGTAGCGGTACCGCTTCTTCTTCTCCAG GTAGCTCTACTCCGTCTGGTGCTACCGGCTCCCCA
GSSPSASTGTGPGSSP SASTGTGPGASPGTSS TGSP	GGTTCTAGCCCTTCTGCATCCACCGGTACCGGCCCAG GTTCTAGCCCGTCTGCTTCTACCGGTACTGGTCCAGG TGCTTCTCCGGGTACTAGCTCTACTGGTTCTCCA
GTTPGSGTASSSPGSST PSGATGSPGSNPSAST GTGP	GGTACCCCCGGGTAGCGGTACCGCTTCTTCTTCTCCAG GTAGCTCTACTCCGTCTGGTGCTACCGGCTCTCCAGG TTCTAACCCTTCTGCATCCACCGGTACCGGCCCA
GSSPSASTGTGPGSST PSGATGSPGASPGTSS TGSP	GGTTCTAGCCCTTCTGCATCTACTGGTACTGGCCCAG GTAGCTCTACTCCTTCTGGTGCTACCGGCTCTCCAGG TGCTTCTCCGGGTACTAGCTCTACCGGTTCTCCA
GSSTPSGATGSPGASP GTSSTGSPGTPGSGTA SSSP	GGTAGCTCTACCCCTTCTGGTGCAACCGGCTCTCCAG GTGCATCCCCGGGTACCAGCTCTACCGGTTCTCCAG GTACTCCGGGTAGCGGTACCGCTTCTTCCTCTCCA
GSSTPSGATGSPGSST PSGATGSPGSSTPSGA TGSP	GGTAGCTCTACTCCGTCTGGTGCAACCGGTTCCCCAG GTAGCTCTACCCCTTCTGGTGCAACCGGCTCCCCAGG TAGCTCTACCCCTTCTGGTGCAACTGGCTCTCCA
GASPGTSSTGSPGTPG SGTASSSPGSSTPSGA TGSP	GGTGCTTCTCCTGGCACTAGCTCTACCGGTTCTCCAG GTACCCCTGGTAGCGGTACCGCATCTTCCTCTCCAGG TAGCTCTACTCCTTCTGGTGCTACTGGTTCCCCA
GSSTPSGATGSPGSSP SASTGTGPGSSPSAST GTGP	GGTAGCTCTACCCCGTCTGGTGCTACTGGCTCCCCAG GTTCTAGCCCTTCTGCATCCACCGGTACCGGTCCAGG TTCTAGCCCGTCTGCATCTACTGGTACTGGTCCA
GASPGTSSTGSPGSSP SASTGTGPGSSPSAST GTGP	GGTGCTTCCCCGGGCACCAGCTCTACTGGTTCTCCAG GTTCTAGCCCGTCTGCTTCTACTGGTACTGGTCCAGG TTCTAGCCCTTCTGCTTCCACTGGTACTGGTCCA
GASPGTSSTGSPGSSP SASTGTGPGTPGSGT ASSSP	GGTGCTTCCCCGGGTACCAGCTCTACCGGTTCTCCAG GTTCTAGCCCTTCTGCTTCTACCGGTACCGGTCCAGG TACCCCTGGCAGCGGTACCGCATCTTCCTCTCCA

[0011] Table G

Sequence Trimmed	Protein Sequence
GGTACCCCGGGCAGCGGTACCGCATCTTCCTCTCCAG	GTPGSGTASSSPGSST
GTAGCTCTACCCCGTCTGGTGCTACCGGTTCCCCAGG	PSGATGSPGSSTPSGA
TAGCTCTACCCCGTCTGGTGCAACCGGCTCCCCAGGT	TGSPGSPAGSPTSTEE
AGCCCGGCTGGCTCTCCTACCTCTACTGAGGAAGGTA	GTSESATPESGPGTST
CTTCTGAAAGCGCTACTCCTGAGTCTGGTCCAGGTAC	EPSEGSAPGSSPSAST
CTCTACTGAACCGTCCGAAGGTAGCGCTCCAGGTTCT	GTGPGSSPSASTGTGP
AGCCCTTCTGCATCCACCGGTACCGGCCCAGGTTCTA	GASPGTSSTGSPGTST
GCCCGTCTGCTTCTACCGGTACTGGTCCAGGTGCTTCT	EPSEGSAPGTSTEPSE
CCGGGTACTAGCTCTACTGGTTCTCCAGGTACCTCTA	GSAPGSEPATSGSETP
CCGAACCGTCCGAGGGTAGCGCACCAGGTACCTCTAC	
TGAACCGTCTGAGGGTAGCGCTCCAGGTAGCGAACC	
GGCAACCTCCGGTTCTGAAACTCCA	

Continues on next page

Sequence Trimmed	Protein Sequence
GGTTCTACCAGCGAATCCCCTTCTGGCACTGCACCAG	GSTSESPSGTAPGSTS
GTTCTACTAGCGAATCCCCTTCTGGTACCGCACCAGG	ESPSGTAPGTSPSGES
TACTTCTCCGAGCGGCGAATCTTCTACTGCTCCAGGT	STAPGTSTEPSEGSAP
ACCTCTACTGAACCTTCCGAAGGCAGCGCTCCAGGTA	GTSTEPSEGSAPGTSE
CCTCTACCGAACCGTCCGAGGGCAGCGCACCAGGTAC	SATPESGPGASPGTSS
TTCTGAAAGCGCAACCCCTGAATCCGGTCCAGGTGCA	TGSPGSSTPSGATGSP
TCTCCTGGTACCAGCTCTACCGGTTCTCCAGGTAGCTC	GASPGTSSTGSPGSTS
TACTCCTTCTGGTGCTACTGGCTCTCCAGGTGCTTCCC	ESPSGTAPGSTSESPS
CGGGTACCAGCTCTACCGGTTCTCCAGGTTCTACTAG	GTAPGTSTPESGSASP
CGAATCTCCTTCTGGCACTGCACCAGGTTCTACCAGC	
GAATCTCCGTCTGGCACTGCACCAGGTACCTCTACCC	
CTGAAAGCGGTTCCGCTTCTCCA	
GGTACTTCTACCGAACCTTCCGAGGGCAGCGCACCAG	GTSTEPSEGSAPGTSE
GTACTTCTGAAAGCGCTACCCCTGAGTCCGGCCCAGG	SATPESGPGTSESATP
TACTTCTGAAAGCGCTACTCCTGAATCCGGTCCAGGT	ESGPGTSTEPSEGSAP
ACCTCTACTGAACCTTCTGAGGGCAGCGCTCCAGGTA	GTSESATPESGPGTST
CTTCTGAAAGCGCTACCCCCGGAGTCCGGTCCAGGTAC	EPSEGSAPGTSTEPSE
TTCTACTGAACCGTCCGAAGGTAGCGCACCAGGTACT	GSAPGSEPATSGSETP
TCTACTGAACCTTCCGAAGGTAGCGCTCCAGGTAGCG	GSPAGSPTSTEEGASP
AACCTGCTACTTCTGGTTCTGAAACCCCAGGTAGCCC	GTSSTGSPGSSPSAST
GGCTGGCTCTCCGACCTCCACCGAGGAAGGTGCTTCT	GTGPGSSPSASTGTGP
CCTGGCACCAGCTCTACTGGTTCTCCAGGTTCTAGCC	
CTTCTGCTTCTACCGGTACTGGTCCAGGTTCTAGCCCT	
TCTGCATCCACTGGTACTGGTCCA	
GGTAGCGAACCGGCAACCTCTGGCTCTGAAACCCCAG	GSEPATSGSETPGTSE
GTACCTCTGAAAGCGCTACTCCGGAATCTGGTCCAGG	SATPESGPGTSESATP
TACTTCTGAAAGCGCTACTCCGGAATCCGGTCCAGGT	ESGPGSTSESPSGTAP
TCTACCAGCGAATCTCCTTCTGGCACCGCTCCAGGTT	GSTSESPSGTAPGTSP
CTACTAGCGAATCCCCGTCTGGTACCGCACCAGGTAC	SGESSTAPGASPGTSS
TTCTCCTAGCGGCGAATCTTCTACCGCACCAGGTGCA	TGSPGSSPSASTGTGP
TCTCCGGGTACTAGCTCTACCGGTTCTCCAGGTTCTAG	GSSTPSGATGSPGSST
CCCTTCTGCTTCCACTGGTACCGGCCCAGGTAGCTCT	PSGATGSPGSSTPSGA
ACCCCGTCTGGTGCTACTGGTTCCCCAGGTAGCTCTA	TGSPGASPGTSSTGSP
CTCCGTCTGGTGCAACCGGTTCCCCAGGTAGCTCTAC	
TCCTTCTGGTGCTACTGGCTCCCCAGGTGCATCCCCTG	
GCACCAGCTCTACCGGTTCTCCA	
GGTGCTTCTCCGGGCACCAGCTCTACTGGTTCTCCAG	GASPGTSSTGSPGSSP
GTTCTAGCCCTTCTGCATCCACCGGTACCGGTCCAGG	SASTGTGPGSSTPSGA
TAGCTCTACCCCTTCTGGTGCAACCGGCTCTCCAGGT	TGSPGTSESATPESGP
ACTTCTGAAAGCGCTACCCCCGGAATCTGGCCCAGGTA	GSEPATSGSETPGSEP
GCGAACCGGCTACTTCTGGTTCTGAAACCCCAGGTAG	ATSGSETPGTSESATP
CGAACCGGCTACCTCCGGTTCTGAAACTCCAGGTACT	ESGPGTSTEPSEGSAP
TCTGAAAGCGCTACTCCGGAGTCCGGTCCAGGTACCT	GTSTEPSEGSAPGTST
CTACCGAACCGTCCGAAGCCAGCGCTCCAGGTACTTC	EPSEGSAPGTSTEPSE
TACTGAACCTTCTGAGGGTAGCGCTCCAGGTACCTCT	GSAPGSEPATSGSETP
ACCGAACCGTCCGAGGGTAGCGCACCAGGTACCTCTA	
CTGAACCGTCTGAGGGTAGCGCTCCAGGTAGCGAACC	
GGCAACCTCCGGTTCTGAAACTCCA	
GGTACCTCTACCGAACCTTCCGAAGGTAGCGCTCCAG	GTSTEPSEGSAPGSPA
GTAGCCCGGCAGGTTCTCCTACTTCCACTGAGGAAGG	GSPTSTEEGTSTEPSE
TACTTCTACCGAACCTTCTGAGGGTAGCGCACCAGGT	GSAPGTSESATPESGP
ACCTCTGAAAGCGCAACTCCTGAGTCTGGCCCAGGTA	GSEPATSGSETPGTSE
GCGAACCTGCTACCTCCGGCTCTGAGACTCCAGGTAC	SATPESGPGSPAGSPT
CTCTGAAAGCGCAACCCCGGAATCTGGTCCAGGTAGC	STEEGTSESATPESGP
CCGGCTGGCTCTCCTACCTCTACTGAGGAAGGTACTT	GTSTEPSEGSAPGSEP
CTGAAAGCGCTACTCCTGAGTCTGGTCCAGGTACCTC	ATSGSETPGTSTEPSE
TACTGAACCGTCCGAAGGTAGCGCTCCAGGTAGCGA	GSAPGSEPATSGSETP
ACCTGCTACTTCTGGTTCTGAAACTCCAGGTACTTCTA	
CCGAACCGTCCGAGGGTAGCGCTCCAGGTAGCGAAC	
CTGCTACTTCTGGTTCTGAAACTCCA	

Sequence Trimmed	Protein Sequence
GGTACTTCTACCGAACCGTCCGAAGGCAGCGCTCCAG	GTSTEPSEGSAPGTST
GTACCTCTACTGAACCTTCCGAGGGCAGCGCTCCAGG	EPSEGSAPGTSTEPSE
TACCTCTACCGAACCTTCTGAAGGTAGCGCACCAGGT	GSAPGTSTEPSEGSAP
ACTTCTACCGAACCGTCCGAAGGCAGCGCTCCAGGTA	GTSTEPSEGSAPGTST
CCTCTACTGAACCTTCCGAGGGCAGCGCTCCAGGTAC	EPSEGSAPGTSTEPSE
CTCTACCGAACCTTCTGAAGGTAGCGCACCAGGTACT	GSAPGTSESATPESGP
TCTACCGAACCTTCCGAGGGCAGCGCACCAGGTACTT	GTSESATPESGPGTST
CTGAAAGCGCTACCCCTGAGTCCGGCCCAGGTACTTC	EPSEGSAPGSEPATSG
TGAAAGCGCTACTCCTGAATCCGGTCCAGGTACTTCT	SETPGSPAGSPTSTEE
ACTGAACCTTCCGAAGGTAGCGCTCCAGGTAGCGAAC	
CTGCTACTTCTGGTTCTGAAACCCCAGGTAGCCCGGC	
TGGCTCTCCGACCTCCACCGAGGAA	
GGTACTTCTACCGAACCGTCCCAGGGCAGCGCTCCAG	GTSTEPSEGSAPGTST
GTACTTCTACTGAACCTTCTGAAGGCAGCGCTCCAGG	EPSEGSAPGTSTEPSE
TACTTCTACTGAACCTTCCGAAGGTAGCGCACCAGGT	GSAPGSTSESPSGTAP
TCTACCAGCGAATCCCCTTCTGGTACTGCTCCAGGTTC	GSTSESPSGTAPGTST
TACCAGCGAATCCCCTTCTGGCACCGCACCAGGTACT	PESGSASPGSEPATSG
TCTACCCCTGAAAGCGGCTCCGCTTCTCCAGGTAGCG	SETPGTSESATPESGP
AACCTGCAACCTCTGGCTCTGAAACCCCAGGTACCTC	GTSTEPSEGSAPGTST
TGAAAGCGCTACTCCTGAATCTGGCCCAGGTACTTCT	EPSEGSAPGTSESATP
ACTGAACCGTCCGAGGGCAGCGCACCAGGTACTTCTA	ESGPGTSESATPESGP
CTGAACCGTCTGAAGGTAGCGCACCAGGTACTTCTGA	
AAGCGCAACCCCGGAATCCGGCCCAGGTACCTCTGA	
AAGCGCAACCCCGGAGTCCGGCCCA	
GGTAGCTCTACCCCTTCTGGTGCTACCGGCTCTCCAG	GSSTPSGATGSPGSSP
GTTCTAGCCCGTCTGCTTCTACCGGTACCGGTCCAGG	SASTGTGPGSSTPSGA
TAGCTCTACCCCTTCTGGTGCTACTGGTTCTCCAGGTA	TGSPGSPAGSPTSTEE
GCCCTGCTGGCTCTCCGACTTCTACTGAGGAAGGTAG	GSPAGSPTSTEEGTST
CCCGGCTGGTTCTCCGACTTCTACTGAGGAAGGTACT	EPSEGSAPGASPGTSS
TCTACCGAACCTTCCGAAGGTAGCGCTCCAGGTGCTT	TGSPGSSPSASTGTGP
CCCCGGGCACTAGCTCTACCGGTTCTCCAGGTTCTAG	GTPGSGTASSSPGSTS
CCCTTCTGCATCTACTGGTACTGGCCCAGGTACTCCG	STAESPGPGTSPSGES
GGCAGCGGTACTGCTTCTTCCTCTCCAGGTTCTACTAG	STAPGTSTPESGSASP
CTCTACTGCTGAATCTCCTGGCCCAGGTACTTCTCCTA	
GCGGTGAATCTTCTACCGCTCCAGGTACCTCTACTCC	
GGAAAGCGGTTCTGCATCTCCA	
GGTACCTCTACTGAACCTTCTGAGGGCAGCGCTCCAG	GTSTEPSEGSAPGTSE
GTACTTCTGAAAGCGCTACCCCGGAGTCCGGTCCAGG	SATPESGPGTSTEPSE
TACTTCTACTGAACCGTCCGAAGGTAGCGCACCAGGT	GSAPGTSTEPSEGSAP
ACTTCTACTGAACCGTCTGAAGGTAGCGCACCAGGTA	GTSESATPESGPGTSE
CTTCTGAAAGCGCAACCCCCGGAATCCGGCCCAGGTAC	SATPESGPGTPGSGTA
CTCTGAAAGCGCAACCCCCGGAGTCCGGCCCAGGTACT	SSSPGASPGTSSTGSP
CCTGGCAGCGGTACCGCTTCTTCTTCTCCAGGTGCTTC	GASPGTSSTGSPGSPA
TCCTGGTACTAGCTCTACTGGTTCTCCAGGTGCTTCTC	GSPTSTEEGSPAGSPT
CGGGCACTAGCTCTACTGGTTCTCCAGGTAGCCCTGC	STEEGTSTEPSEGSAP
TGGCTCTCCGACTTCTACTGAGGAAGGTAGCCCGGCT	
GGTTCTCCGACTTCTACTGAGGAAGGTACTTCTACCG	
AACCTTCCGAAGGTAGCGCTCCA	
GGTAGCCCAGCAGGCTCTCCGACTTCCACTGAGGAAG	GSPAGSPTSTEEGTST
GTACTTCTACTGAACCTTCCGAAGGCAGCGCACCAGG	EPSEGSAPGTSTEPSE
TACCTCTACTGAACCTTCTGAGGGCAGCGCTCCAGGT	GSAPGTSTEPSEGSAP
ACCTCTACCGAACCGTCTGAAGGTAGCGCACCAGGTA	GTSESATPESGPGTSE
CCTCTGAAAGCGCAACTCCTGAGTCCGGTCCAGGTAC	SATPESGPGTPGSGTA
TTCTGAAAGCGCAACCCCCGGAGTCTGGCCCAGGTACC	SSSPGSSTPSGATGSP
CCGGGTAGCGGTACTGCTTCTTCCTCTCCAGGTAGCT	GASPGTSSTGSPGTST
CTACCCCTTCTGGTGCAACCGGCTCTCCAGGTGCTTCT	EPSEGSAPGTSESATP
CCGGGCACCAGCTCTACCGGTTCTCCAGGTACCTCTA	ESGPGTSTEPSEGSAP
CTGAACCTTCTGAGGGCAGCGCTCCAGGTACTTCTGA	
AAGCGCTACCCCCGGAGTCCGGTCCAGGTACTTCTACT	
GAACCGTCCGAAGGTAGCGCACCA	

Sequence Trimmed	Protein Sequence
GGTAGCGAACCGGCAACCTCCGGCTCTGAAACTCCAG	GSEPATSGSETPGTSE
GTACTTCTGAAAGCGCTACTCCGGAATCCGGCCCAGG	SATPESGPGSEPATSG
TAGCGAACCGGCTACTTCCGGCTCTGAAACCCCAGGT	SETPGSSTPSGATGSP
AGCTCTACCCCGTCTGGTGCAACCGGCTCCCCAGGTA	GTPGSGTASSSPGSST
CTCCTGGTAGCGGTACCGCTTCTTCTTCTCCAGGTAGC	PSGATGSPGASPGTSS
TCTACTCCGTCTGGTGCTACCGGCTCCCCAGGTGCAT	TGSPGSSTPSGATGSP
CTCCTGGTACCAGCTCTACCGGTTCTCCAGGTAGCTCT	GASPGTSSTGSPGSEP
ACTCCTTCTGGTGCTACTGGCTCTCCAGGTGCTTCCCC	ATSGSETPGTSTEPSE
GGGTACCAGCTCTACCGGTTCTCCAGGTAGCGAACCT	GSAPGSEPATSGSETP
GCTACTTCTGGTTCTGAAACTCCAGGTACTTCTACCG	
AACCGTCCGAGGGTAGCGCTCCAGGTAGCGAACCTG	
CTACTTCTGGTTCTGAAACTCCA	
GGTACCTCTACTGAACCTTCCGAAGGCAGCGCTCCAG	GTSTEPSEGSAPGTST
GTACCTCTACCGAACCGTCCGAGGGCAGCGCACCAG	EPSEGSAPGTSESATP
GTACTTCTGAAAGCGCAACCCCTGAATCCGGTCCAGG	ESGPGSPAGSPTSTEE
TAGCCCTGCTGGCTCTCCGACTTCTACTGAGGAAGGT	GSPAGSPTSTEEGTST
AGCCCGGCTGGTTCTCCGACTTCTACTGAGGAAGGTA	EPSEGSAPGSPAGSPT
CTTCTACCGAACCTTCCGAAGGTAGCGCTCCAGGTAG	STEEGTSTEPSEGSAP
CCCGGCTGGTTCTCCGACTTCCACCGAGGAAGGTACC	GTSTEPSEGSAPGASP
TCTACTGAACCTTCTGAGGGTAGCGCTCCAGGTACCT	GTSSTGSPGSSPSAST
CTACTGAACCTTCCGAAGGCAGCGCTCCAGGTGCTTC	GTGPGSSPSASTGTGP
CCCGGGCACCAGCTCTACTGGTTCTCCAGGTTCTAGC	
CCGTCTGCTTCTACTGGTACTGGTCCAGGTTCTAGCCC	
TTCTGCTTCCACTGGTACTGGTCCA	
GGTAGCTCTACCCCGTCTGGTGCTACCGGTTCCCCAG	GSSTPSGATGSPGASP
GTGCTTCTCCTGGTACTAGCTCTACCGGTTCTCCAGGT	GTSSTGSPGSSTPSGA
AGCTCTACCCCGTCTGGTGCTACTGGCTCTCCAGGTA	TGSPGSPAGSPTSTEE
GCCCTGCTGGCTCTCCAACCTCCACCGAAGAAGGTAC	GTSESATPESGPGSEP
CTCTGAAAGCGCAACCCCTGAATCCGGCCCAGGTAGC	ATSGSETPGASPGTSS
GAACCGGCAACCTCCGGTTCTGAAACCCCAGGTGCAT	TGSPGSSTPSGATGSP
CTCCTGGTACTAGCTCTACTGGTTCTCCAGGTAGCTCT	GSSPSASTGTGPGSTS
ACTCCGTCTGGTGCAACCGGCTCTCCAGGTTCTAGCC	ESPSGTAPGSTSESPS
CTTCTGCATCTACCGGTACTGGTCCAGGTTCTACCAG	GTAPGTSTPESGSASP
CGAATCCCCTTCTGGTACTGCTCCAGGTTCTACCAGC	
GAATCCCCTTCTGGCACCGCACCAGGTACTTCTACCC	
CTGAAAGCGGCTCCGCTTCTCCA	
GGTTCTACCAGCGAATCTCCTTCTGGCACCGCTCCAG	GSTSESPSGTAPGSTS
GTTCTACTAGCGAATCCCCGTCTGGTACCGCACCAGG	ESPSGTAPGTSPSGES
TACTTCTCCTAGCGGCGAATCTTCTACCGCACCAGGT	STAPGTSESATPESGP
ACCTCTGAAAGCGCTACTCCGGAGTCTGGCCCAGGTA	GTSTEPSEGSAPGTST
CCTCTACTGAACCGTCTGAGGGTAGCGCTCCAGGTAC	EPSEGSAPGTSTEPSE
TTCTACTGAACCGTCCGAAGGTAGCGCACCAGGTACC	GSAPGTSESATPESGP
TCTACTGAACCTTCTGAGGGCAGCGCTCCAGGTACTT	GTSTEPSEGSAPGSST
CTGAAAGCGCTACCCCGGAGTCCGGTCCAGGTACTTC	PSGATGSPGASPGTSS
TACTGAACCGTCCGAAGGTAGCGCACCAGGTAGCTCT	TGSPGSSTPSGATGSP
ACCCCGTCTGGTGCTACCGGTTCCCCAGGTGCTTCTCC	
TGGTACTAGCTCTACCGGTTCTCCAGGTAGCTCTACC	
CCGTCTGGTGCTACTGGCTCTCCA	
GGTTCTACTAGCTCTACTGCAGAATCTCCGGGCCCAG	GSTSSTAESPGPGTSP
GTACCTCTCCTAGCGGTGAATCTTCTACCGCTCCAGG	SGESSTAPGTSPSGES
TACTTCTCCGAGCGGTGAATCTTCTACCGCTCCAGGTT	STAPGSTSSTAESPGP
CTACTAGCTCTACCGCTGAATCTCCGGGTCCAGGTTC	GSTSSTAESPGPGTST
TACCAGCTCTACTGCAGAATCTCCTGGCCCAGGTACT	PESGSASPGTSPSGES
TCTACTCCGGAAAGCGGTTCCGCTTCTCCAGGTACTT	STAPGSTSSTAESPGP
CTCCTAGCGGTGAATCTTCTACCGCTCCAGGTTCTACC	GTSTPESGSASPGSTS
AGCTCTACTGCTGAATCTCCTGGCCCAGGTACTTCTA	STAESPGPGSTSESPS
CCCCGGAAAGCGGCTCCGCTTCTCCAGGTTCTACCAG	GTAPGTSPSGESSTAP
CTCTACCGCTGAATCTCCTGGCCCAGGTTCTACTAGC	
GAATCTCCGTCTGGCACCGCACCAGGTACTTCCCCTA	
GCGGTGAATCTTCTACTGCACCA	

Sequence Trimmed	Protein Sequence
GGTACCTCTACTCCGGAAAGCGGTTCCGCATCTCCAG GTTCTACCAGCGAATCCCCGTCTGGCACCGCACCAGG TTCTACTAGCTCTACTGCTGAATCTCCGGGCCCAGGT ACCTCTACTGAACCTTCCGAAGGCAGCGCTCCAGGTA CCTCTACCGAACCGTCCGAGGGCAGCGCACCAGGTAC TTCTGAAAGCGCAACCCCCTGAATCCGGTCCAGGTACC TCTGAAAGCGCTACTCCGGAGTCTGGCCCAGGTACCT CTACTGAACCGTCTGAGGGTAGCGCTCCAGGTACTTC TACTGAACCGTCCGAAGGTAGCGCACCAGGTACTTCT GAAAGCGCTACTCCGGAGTCCGGTCCAGGTACCTCTA CCGAACCGTCCGAAGGCAGCGCTCCAGGTACTTCTAC TGAACCTTCTGAGGGTAGCGCTCCC	GTSTPESGSASPGSTS ESPSGTAPGSTSSTAE SPGPGTSTEPSEGSAP GTSTEPSEGSAPGTSE SATPESGPGTSESATP ESGPGTSTEPSEGSAP GTSTEPSEGSAPGTSE SATPESGPGTSTEPSE GSAPGTSTEPSEGSAP
GGTACCTCTACCGAACCGTCCGAGGGTAGCGCACCAG GTACCTCTACTGAACCGTCTGAGGGTAGCGCTCCAGG TAGCGAACCGGCAACCTCCGGTTCTGAAACTCCAGGT ACTTCTACTGAACCGTCTGAAGGTAGCGCACCAGGTA CTTCTGAAAGCGCAACCCCCGGAATCCGGCCCAGGTAC CTCTGAAAGCGCAACCCCCGGAGTCCGGCCCAGGTGC ATCTCCGGGTACTAGCTCTACCGGTTCTCCAGGTTCTA GCCCTTCTGCTTCCACTGGTACCGGCCCAGGTAGCTC TACCCCGTCTGGTGCTACTGGTTCCCCAGGTAGCTCT ACTCCGTCTGGTGCAACCGGTTCCCCAGGTAGCTCTA CTCCTTCTGGTGCTACTGGCTCCCCAGGTGCATCCCCT GGCACCAGCTCTACCGGTTCTCCA	GTSTEPSEGSAPGTST EPSEGSAPGSEPATSG SETPGTSTEPSEGSAP GTSESATPESGPGTSE SATPESGPGASPGTSS TGSPGSSPSASTGTGP GSSTPSGATGSPGSST PSGATGSPGSSTPSGA TGSPGASPGTSSTGSP
GGTAGCGAACCGGCAACCTCTGGCTCTGAAACTCCAG GTAGCGAACCTGCAACCTCCGGCTCTGAAACCCCAGG TACTTCTACTGAACCTTCTGAGGGCAGCGCACCAGGT AGCGAACCTGCAACCTCTGGCTCTGAAACCCCAGGTA CCTCTGAAAGCGCTACTCCTGAATCTGGCCCAGGTAC TTCTACTGAACCGTCCGAGGGCAGCGCACCAGGTAGC TCTACTCCGTCTGGTGCTACCGGCTCTCCAGGTAGCTC TACCCCTTCTGGTGCAACCGGCTCCCCAGGTGCTTCTC CGGGTACCAGCTCTACTGGTTCTCCAGGTAGCTCTAC CCCGTCTGGTGCTACCGGTTCCCCAGGTGCTTCTCCTG GTACTAGCTCTACCGGTTCTCCAGGTAGCTCTACCCC GTCTGGTGCTACTGGCTCTCCA	GSEPATSGSETPGSEP ATSGSETPGTSTEPSE GSAPGSEPATSGSETP GTSESATPESGPGTST EPSEGSAPGSSTPSGA TGSPGSSTPSGATGSP GASPGTSSTGSPGSST PSGATGSPGASPGTSS TGSPGSSTPSGATGSP
GGTACTTCTACCGAACCGTCCGAGGGCAGCGCTCCAG GTACTTCTACTGAACCTTCTGAAGGCAGCGCTCCAGG TACTTCTACTGAACCTTCCGAAGGTAGCGCACCAGGT ACTTCTGAAAGCGCTACTCCGGAGTCCGGTCCAGGTA CCTCTACCGAACCGTCCGAAGGCAGCGCTCCAGGTAC TTCTACTGAACCTTCTGAGGGTAGCGCTCCAGGTTCT ACTAGCGAATCTCCGTCTGGCACTGCTCCAGGTACTT CTCCTAGCGGTGAATCTTCTACCGCTCCAGGTACTTCC CCTAGCGGCGAATCTTCTACCGCTCCAGGTAGCCCGG CTGGCTCTCCTACCTCTACTGAGGAAGGTACTTCTGA AAGCGCTACTCCTGAGTCTGGTCCAGGTACCTCTACT GAACCGTCCGAAGGTAGCGCTCCA	GTSTEPSEGSAPGTST EPSEGSAPGTSTEPSE GSAPGTSESATPESGP GTSTEPSEGSAPGTST EPSEGSAPGSTSESPS GTAPGTSPSGESSTAP GTSPSGESSTAPGSPA GSPTSTEEGTSESATP ESGPGTSTEPSEGSAP
GGTACTTCTACTGAACCTTCCGAAGGTAGCGCTCCAG GTAGCGAACCTGCTACTTCTGGTTCTGAAACCCCAGG TAGCCCGGCTGGCTCTCCGACCTCCACCGAGGAAGGT AGCCCCGGCAGGCTCTCCGACCTCTACTGAGGAAGGTA CTTCTGAAAGCGCAACCCCCGGAGTCCGGCCCAGGTAC CTCTACCGAACCGTCTGAGGGCAGCGCACCAGGTACC TCTACTGAACCTTCCGAAGGCAGCGCTCCAGGTACCT CTACCGAACCGTCCGAGGGCAGCGCACCAGGTACTTC TGAAAGCGCAACCCCTGAATCCGGTCCAGGTAGCTCT ACTCCGTCTGGTGCAACCGGCTCCCCAGGTTCTAGCC CGTCTGCTTCCACTGGTACTGGCCCAGGTGCTTCCCC GGGCACCAGCTCTACTGGTTCTCCA	GTSTEPSEGSAPGSEP ATSGSETPGSPAGSPT STEEGSPAGSPTSTEE GTSESATPESGPGTST EPSEGSAPGTSTEPSE GSAPGTSTEPSEGSAP GTSESATPESGPGSST PSGATGSPGSSPSAST GTGPGASPGTSSTGSP

Sequence Trimmed	Protein Sequence
GGTAGCGAACCGGCTACTTCCGGCTCTGAGACTCCAG GTAGCCCTGCTGGCTCTCCGACCTCTACCGAAGAAGG TACCTCTGAAAGCGCTACCCCTGAGTCTGGCCCAGGT ACCTCTACTGAACCTTCCGAAGGCAGCGCTCCAGGTA CCTCTACCGAACCGTCCGAGGGCAGCGCACCAGGTAC TTCTGAAAGCGCAACCCCTGAATCCGGTCCAGGTACC TCTACTCCGGAAAGCGGTTCCGCATCTCCAGGTTCTA CCAGCGAATCCCCGTCTGGCACCGCACCAGGTTCTAC TAGCTCTACTGCTGAATCTCCGGGCCCAGGTACTTCT GAAAGCGCTACTCCGGAGTCCGGTCCAGGTACCTCTA CCGAACCGTCCGAAGGCAGCGCTCCAGGTACTTCTAC TGAACCTTCTGAGGGTAGCGCTCCA	GSEPATSGSETPGSPA GSPTSTEEGTSESATP ESGPGTSTEPSEGSAP GTSTEPSEGSAPGTSE SATPESGPGTSTPESG SASPGSTSESPSGTAP GSTSSTAESPGPGTSE SATPESGPGTSTEPSE GSAPGTSTEPSEGSAP
GGTACTTCTACCGAACCGTCCGAGGGCAGCGCTCCAG GTACTTCTACTGAACCTTCTGAAGGCAGCGCTCCAGG TACTTCTACTGAACCTTCCGAAGGTAGCGCACCAGGT ACCTCTACCGAACCGTCTGAAGGTAGCGCACCAGGTA CCTCTGAAAGCGCAACTCCTGAGTCCGGTCCAGGTAC TTCTGAAAGCGCAACCCCGGAGTCTGGCCCAGGTACT CCTGGCAGCGGTACCGCATCTTCCTCTCCAGGTAGCT CTACTCCGTCTGGTGCAACTGGTTCCCCAGGTGCTTCT CCGGGTACCAGCTCTACCGGTTCTCCAGGTTCCACCA GCTCTACTGCTGAATCTCCTGGTCCAGGTACCTCTCCT AGCGGTGAATCTTCTACTGCTCCAGGTACTTCTACTCC TGAAAGCGGCTCTGCTTCTCCA	GTSTEPSEGSAPGTST EPSEGSAPGTSTEPSE GSAPGTSTEPSEGSAP GTSESATPESGPGTSE SATPESGPGTPGSGTA SSSPGSSTPSGATGSP GASPGTSSTGSPGSTS STAESPGPGTSPSGES STAPGTSTPESGSASP
GGTAGCCCGGCAGGCTCTCCGACCTCTACTGAGGAAG GTACTTCTGAAAGCGCAACCCCGGAGTCCGGCCCAGG TACCTCTACCGAACCGTCTGAGGGCAGCGCACCAGGT ACTTCTGAAAGCGCAACCCCTGAATCCGGTCCAGGTA GCGAACCGGCTACTTCTGGCTCTGAGACTCCAGGTAC TTCTACCGAACCGTCCGAAGGTAGCGCACCAGGTAGC CCGGCTGGTTCTCCGACTTCCACCGAGGAAGGTACCT CTACTGAACCTTCTGAGGGTAGCGCTCCAGGTACCTC TACTGAACCTTCCGAAGGCAGCGCTCCAGGTACTTCT ACCGAACCGTCCGAGGGCAGCGCTCCAGGTACTTCTA CTGAACCTTCTGAAGGCAGCGCTCCAGGTACTTCTAC TGAACCTTCCGAAGGTAGCGCACCA	GSPAGSPTSTEEGTSE SATPESGPGTSTEPSE GSAPGTSESATPESGP GSEPATSGSETPGTST EPSEGSAPGSPAGSPT STEEGTSTEPSEGSAP GTSTEPSEGSAPGTST EPSEGSAPGTSTEPSE GSAPGTSTEPSEGSAP
GGTACTTCTCCGAGCGGTGAATCTTCTACCGCACCAG GTTCTACTAGCTCTACCGCTGAATCTCCGGCCCCAGG TACTTCTCCGAGCGGTGAATCTTCTACTGCTCCAGGT ACCTCTGAAAGCGCTACTCCGGAGTCTGGCCCAGGTA CCTCTACTGAACCGTCTGAGGGTAGCGCTCCAGGTAC TTCTACTGAACCGTCCGAAGGTAGCGCACCAGGTTCT AGCCCTTCTGCATCTACTGGTACTGGCCCAGGTAGCT CTACTCCTTCTGGTGCTACCGGCTCTCCAGGTGCTTCT CCGGGTACTAGCTCTACCGGTTCTCCAGGTACTTCTA CTCCGGAAAGCGGTTCCGCATCTCCAGGTACTTCTCC TAGCGGTGAATCTTCTACTGCTCCAGGTACCTCTCCTA GCGGCGAATCTTCTACTGCTCCA	GTSPSGESSTAPGSTS STAESPGPGTSPSGES STAPGTSESATPESGP GTSTEPSEGSAPGTST EPSEGSAPGSSPSAST GTGPGSSTPSGATGSP GASPGTSSTGSPGTST PESGSASPGTSPSGES STAPGTSPSGESSTAP
GGTACCTCTGAAAGCGCTACTCCGGAGTCTGGCCCAG GTACCTCTACTGAACCGTCTGAGGGTAGCGCTCCAGG TACTTCTACTGAACCGTCCGAAGGTAGCGCACCAGGT AGCCCTGCTGGCTCTCCGACTTCTACTGAGGAAGGTA GCCCGGCTGGTTCTCCGACTTCTACTGAGGAAGGTAC TTCTACCGAACCTTCCGAAGGTAGCGCTCCAGGTTCT AGCCCTTCTGCTTCCACCGGTACTGGCCCAGGTAGCT CTACCCCTTCTGGTGCTACCGGCTCCCCAGGTAGCTCT ACTCCTTCTGGTGCAACTGGCTCTCCAGGTAGCGAAC CGGCAACTTCCGGCTCTGAAACCCCAGGTACTTCTGA AAGCGCTACTCCTGAGTCTGGCCCAGGTAGCGAACCT GCTACCTCTGGCTCTGAAACCCCA	GTSESATPESGPGTST EPSEGSAPGTSTEPSE GSAPGSPAGSPTSTEE GSPAGSPTSTEEGTST EPSEGSAPGSSPSAST GTGPGSSTPSGATGSP GSSTPSGATGSPGSEP ATSGSETPGTSESATP ESGPGSEPATSGSETP

Sequence Trimmed	Protein Sequence
GGTACTTCTACCGAACCGTCCGAAGGCAGCGCTCCAG GTACCTCTACTGAACCTTCCGAGGGCAGCGCTCCAGG TACCTCTACCGAACCTTCTGAAGGTAGCGCACCAGGT AGCTCTACCCCGTCTGGTGCTACCGGTTCCCCAGGTG CTTCTCCTGGTACTAGCTCTACCGGTTCTCCAGGTAGC TCTACCCCGTCTGGTGCTACTGGCTCTCCAGGTACTTC TGAAAGCGCAACCCCTGAATCCGGTCCAGGTAGCGA ACCGGCTACTTCTGGCTCTGAGACTCCAGGTACTTCT ACCGAACCGTCCGAAGGTAGCGCACCAGGTTCTACTA GCGAATCTCCTTCTGGCACTGCACCAGGTTCTACCAG CGAATCTCCGTCTGGCACTGCACCAGGTACCTCTACC CCTGAAAGCGGTTCCGCTTCTCCA	GTSTEPSEGSAPGTST EPSEGSAPGTSTEPSE GSAPGSSTPSGATGSP GASPGTSSTGSPGSST PSGATGSPGTSESATP ESGPGSEPATSGSETP GTSTEPSEGSAPGSTS ESPSGTAPGSTSESPS GTAPGTSTPESGSASP
GGTACCTCTACTCCTGAAAGCGGTTCTGCATCTCCAG GTTCCACTAGCTCTACCGCAGAATCTCCGGGCCCAGG TTCTACTAGCTCTACTGCTGAATCTCCTGGCCCAGGTT CTAGCCCTTCTGCATCTACTGGTACTGGCCCAGGTAG CTCTACTCCTTCTGGTGCTACCGGCTCTCCAGGTGCTT CTCCGGGTACTAGCTCTACCGGTTCTCCAGGTAGCGA ACCGGCAACCTCCGGCTCTGAAACCCCAGGTACCTCT GAAAGCGCTACTCCTGAATCCGGCCCAGGTAGCCCGG CAGGTTCTCCGACTTCCACTGAGGAAGGTTCTACTAG CGAATCTCCTTCTGGCACTGCACCAGGTTCTACCAGC GAATCTCCGTCTGGCACTGCACCAGGTACCTCTACCC CTGAAAGCGGTTCCGCTTCTCCC	GTSTPESGSASPGSTS STAESPGPGSTSSTAE SPGPGSSPSASTGTGP GSSTPSGATGSPGASP GTSSTGSPGSEPATSG SETPGTSESATPESGP GSPAGSPTSTEEGSTS ESPSGTAPGSTSESPS GTAPGTSTPESGSASP
GGTAGCCCGGCTGGCTCTCCTACCTCTACTGAGGAAG GTACTTCTGAAAGCGCTACTCCTGAGTCTGGTCCAGG TACCTCTACTGAACCGTCCGAAGGTAGCGCTCCAGGT TCTACCAGCGAATCTCCTTCTGGCACCGCTCCAGGTT CTACTAGCGAATCCCCGTCTGGTACCGCACCAGGTAC TTCTCCTAGCGGCGAATCTTCTACCGCACCAGGTACC TCTACCGAACCTTCCGAAGGTAGCGCTCCAGGTAGCC CGGCAGGTTCTCCTACTTCCACTGAGGAAGGTACTTC TACCGAACCTTCTGAGGGTAGCGCACCAGGTAGCGA ACCTGCAACCTCTGGCTCTGAAACCCCAGGTACCTCT GAAAGCGCTACTCCTGAATCTGGCCCAGGTACTTCTA CTGAACCGTCCGAGGGCAGCGCACCA	GSPAGSPTSTEEGTSE SATPESGPGTSTEPSE GSAPGSTSESPSGTAP GSTSESPSGTAPGTSP SGESSTAPGTSTEPSE GSAPGSPAGSPTSTEE GTSTEPSEGSAPGSEP ATSGSETPGTSESATP ESGPGTSTEPSEGSAP
GGTACCTCTACCGAACCTTCCGAAGGTAGCGCTCCAG GTAGCCCGGCAGGTTCTCCTACTTCCACTGAGGAAGG TACTTCTACCGAACCTTCTGAGGGTAGCGCACCAGGT ACCTCCCCTAGCGGCGAATCTTCTACTGCTCCAGGTA CCTCTCCTAGCGGCGAATCTTCTACCGCTCCAGGTAC CTCCCCTAGCGGTGAATCTTCTACCGCACCAGGTTCT ACCAGCGAATCCCCTTCTGGTACTGCTCCAGGTTCTA CCAGCGAATCCCCTTCTGGCACCGCACCAGGTACTTC TACCCCTGAAAGCGGCTCCGCTTCTCCAGGTAGCGAA CCTGCAACCTCTGGCTCTGAAACCCCAGGTACCTCTG AAAGCGCTACTCCTGAATCTGGCCCAGGTACTTCTAC TGAACCGTCCGAGGGCAGCGCACCA	GTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSE GSAPGTSPSGESSTAP GTSPSGESSTAPGTSP SGESSTAPGSTSESPS GTAPGSTSESPSGTAP GTSTPESGSASPGSEP ATSGSETPGTSESATP ESGPGTSTEPSEGSAP
GGTAGCGAACCGGCAACCTCTGGCTCTGAAACCCCAG GTACCTCTGAAAGCGCTACTCCGGAATCTGGTCCAGG TACTTCTGAAAGCGCTACTCCGGAATCCGGTCCAGGT ACTTCTCCGAGCGGTGAATCTTCTACCGCACCAGGTT CTACTAGCTCTACCGCTGAATCTCCGGGCCCAGGTAC TTCTCCGAGCGGTGAATCTTCTACTGCTCCAGGTTCTA CTAGCGAATCCCCGTCTGGTACTGCTCCAGGTACTTC CCCTAGCGGTGAATCTTCTACTGCTCCAGGTTCTACC AGCTCTACCGCAGAATCTCCGGGTCCAGGTAGCTCTA CTCCGTCTGGTGCAACCGGTTCCCCAGGTAGCTCTAC CCCTTCTGGTGCAACCGGCTCCCCAGGTAGCTCTACC CCTTCTGGTGCAAACTGGCTCTCC	GSEPATSGSETPGTSE SATPESGPGTSESATP ESGPGTSPSGESSTAP GSTSSTAESPGPGTSP SGESSTAPGSTSESPS GTAPGTSPSGESSTAP GSTSSTAESPGPGSST PSGATGSPGSSTPSGA TGSPGSSTPSGANWL S

Sequence Trimmed	Protein Sequence
GGTAGCCCTGCTGGCTCTCCGACTTCTACTGAGGAAG GTAGCCCGGCTGGTTCTCCGACTTCTACTGAGGAAGG TACTTCTACCGAACCTTCCGAAGGTAGCGCTCCAGGT ACCTCTACTGAACCTTCCGAAGGCAGCGCTCCAGGTA CCTCTACCGAACCGTCCGAGGGCAGCGCACCAGGTAC TTCTGAAAGCGCAACCCCTGAATCCGGTCCAGGTGCA TCTCCTGGTACCAGCTCTACCGGTTCTCCAGGTAGCTC TACTCCTTCTGGTGCTACTGGCTCTCCAGGTGCTTCCC CGGGTACCAGCTCTACCGGTTCTCCAGGTAGCTCTAC CCCGTCTGGTGCTACTGGTTCTCCAGGTACTCCGGGC AGCGGTACTGCTTCTTCCTCTCCAGGTAGCTCTACCCC TTCTGGTGCTACTGGCTCTCCA	GSPAGSPTSTEEGSPA GSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAP GTSTEPSEGSAPGTSE SATPESGPGASPGTSS TGSPGSSTPSGATGSP GASPGTSSTGSPGSST PSGATGSPGTPGSGT ASSSPGSSTPSGATGS P
GGTAGCTCTACCCCGTCTGGTGCTACTGGTTCTCCAG GTACTCCGGGCAGCGGTACTGCTTCTTCCTCTCCAGG TAGCTCTACCCCTTCTGGTGCTACTGGCTCTCCAGGTA GCCCGGCTGGCTCTCCTACCTCTACTGAGGAAGGTAC TTCTGAAAGCGCTACTCCTGAGTCTGGTCCAGGTACC TCTACTGAACCGTCCGAAGGTAGCGCTCCAGGTACCT CTGAAAGCGCAACTCCTGAGTCTGGCCCAGGTAGCGA ACCTGCTACCTCCGGCTCTGAGACTCCAGGTACCTCT GAAAGCGCAACCCCGGAATCTGGTCCAGGTACTTCTA CTGAACCGTCTGAAGGTAGCGCACCAGGTACTTCTGA AAGCGCAACCCCGGAATCCGGCCCAGGTACCTCTGA AAGCGCAACCCCGGAGTCCGGCCCA	GSSTPSGATGSPGTPG SGTASSSPGSSTPSGA TGSPGSPAGSPTSTEE GTSESATPESGPGTST EPSEGSAPGTSESATP ESGPGSEPATSGSETP GTSESATPESGPGTST EPSEGSAPGTSESATP ESGPGTSESATPESGP

We claim:

1. A physiologically acceptable or pharmaceutical grade composition comprising a cell-reactive compstatin analog.

2. The composition of claim 1, wherein the composition is physiologically acceptable.

3. The composition of claim 1, wherein the composition is a pharmaceutical grade composition.

4. The composition of claim 1, wherein the composition is pharmaceutically acceptable for administration to a human.

5. The composition of claim 1, wherein the cell-reactive compstatin analog comprises a cell-reactive functional group capable of binding covalently to a mammalian cell.

6. The composition of claim 1, wherein the cell-reactive compstatin analog comprises a cell-reactive moiety comprising a cell-reactive functional group.

7. The composition of claim 1, wherein the compstatin analog comprises a cell-reactive functional group that reacts with an sulfhydryl (SH) group to form a covalent bond.

8. The composition of claim 1, wherein the compstatin analog comprises a cell-reactive functional group that reacts with an amine group to form a covalent bond.

9. The composition of claim 1, wherein the cell-reactive compstatin analog comprises a maleimide group.

10. The composition of claim 1, wherein the cell-reactive compstatin analog comprises a cell-reactive moiety, wherein the cell-reactive moiety comprises a cell-reactive functional group.

11. The composition of claim 1, wherein the cell-reactive compstatin analog comprises a cell-reactive functional group, a compstatin analog moiety, and a linking moiety that separates the cell-reactive functional group from the compstatin analog moiety portion of the cell-reactive compstatin analog.

12. The composition of claim 1, wherein the cell-reactive compstatin analog comprises an amino acid whose side chain contains a group of formula (NH)—R, wherein R represents a moiety comprising a cell-reactive functional group.

13. The composition of claim 1, wherein the cell-reactive compstatin analog is a compound that comprises a cyclic peptide having a core sequence of X'aa-Gln - Asp – Xaa – Gly (SEQ ID NO: 3), where X'aa and Xaa are selected from Trp and analogs of Trp, and wherein the compound comprises a cell-reactive moiety.

14. The composition of claim 1, wherein the compstatin analog is a compound that comprises a cyclic peptide having a core sequence of X'aa-Gln - Asp – Xaa – Gly-X"aa (SEQ ID NO: 4), where X'aa and Xaa are each independently selected from Trp and analogs of Trp, and X"aa is selected from His, Ala, single methyl unbranched amino acids, Phe, Trp, and analogs of Trp, and wherein the compound comprises a cell-reactive moiety.

15. The composition of claim 1, wherein the peptide has a sequence of X'aa1 - X'aa2 - X'aa3 - X'aa-Gln-Asp-Xaa-Gly- X"aa- X"aa2- X"aa3- X"aa4- X"aa5 (SEQ ID NO: 5), and X'aa1, X'aa2, X'aa3, X"aa2, X"aa3- X"aa4, and X"aa5 are identical to the amino acids at the corresponding positions in compstatin, and wherein the compound comprises a cell-reactive moiety.

16. The composition of claim 1, wherein the compstatin analog is a compound that comprises a cyclic peptide having a sequence of X'aa1 - X'aa2 - X'aa3 - X'aa4 -Gln-Asp-Xaa-Gly- X"aa1- X"aa2- X"aa3- X"aa4- X"aa5 (SEQ ID NO: 5), where X'aa4 and Xaa are selected from Trp and analogs of Trp, wherein X'aa1, X'aa2, X'aa3, X"aa1, X"aa2, X"aa3, X"aa4, and X"aa5, are independently selected from among amino acids and amino acid analogs, wherein the peptide is cyclized via a bond between X'aa2 and X"aa4, and wherein the compound comprises a cell-reactive moiety.

17. The composition of claim 16, wherein X'aa2 and X"aa4 are Cys, and wherein X"aa1 is optionally Ala or a single methyl unbranched amino acid, and wherein the compound comprises a cell-reactive moiety.

18. The composition of claim 16, wherein X'aa2 and X"aa4 are Cys, wherein any one or more of X'aa1, X'aa2, X'aa3, X"aa2, X"aa3, X"aa4, and X"aa5 are identical to the

amino acids at the corresponding positions in compstatin, and X"aa1 is Ala or a single methyl unbranched amino acid, and wherein the compound comprises a cell-reactive moiety.

19. The composition of claim 1, wherein the compstatin analog is a compound that comprises a cyclic peptide having a sequence:
Xaa1 – Cys – Val – Xaa2 - Gln - Asp – Xaa2* - Gly – Xaa3 - His - Arg – Cys – Xaa4 (SEQ ID NO: 6); wherein:
Xaa1 is Ile, Val, Leu, B^1-Ile, B^1-Val, B^1-Leu or a dipeptide comprising Gly-Ile or B^1-Gly-Ile, and B^1 represents a first blocking moiety;
Xaa2 and Xaa2* are independently selected from Trp and analogs of Trp;
Xaa3 is His, Ala or an analog of Ala, Phe, Trp, or an analog of Trp;
Xaa4 is L-Thr, D-Thr, Ile, Val, Gly, a dipeptide selected from Thr-Ala and Thr-Asn, or a tripeptide comprising Thr-Ala-Asn, wherein a carboxy terminal –OH of any of the L-Thr, D-Thr, Ile, Val, Gly, Ala, or Asn optionally is replaced by a second blocking moiety B^2; and
the two Cys residues are joined by a disulfide bond, and wherein the compound comprises a cell-reactive moiety.

20. The composition of claim 19, wherein
Xaa1 is Ile, Val, Leu, Ac-Ile, Ac-Val, Ac-Leu or a dipeptide comprising Gly-Ile or Ac-Gly-Ile;
Xaa2 and Xaa2* are independently selected from Trp and analogs of Trp;
Xaa3 is His, Ala or an analog of Ala, Phe, Trp, or an analog of Trp;
Xaa4 is L-Thr, D-Thr, Ile, Val, Gly, a dipeptide selected from Thr-Ala and Thr-Asn, or a tripeptide comprising Thr-Ala-Asn, wherein a carboxy terminal –OH of any of the L-Thr, D-Thr, Ile, Val, Gly, Ala, or Asn optionally is replaced by $-NH_2$.

21. The composition of claim 19, wherein Xaa2 is an analog of Trp having increased hydrophobic character relative to Trp.

22. The composition of claim 19, wherein Xaa2 is an analog of Trp comprising a substituted or unsubstituted bicyclic aromatic ring component or two or more substituted or unsubstituted monocyclic aromatic ring components.

23. The composition of claim 19, wherein Xaa2* is an analog of Trp having an electronegative substituent on the indole ring and not having increased hydrophobic character relative to Trp.

24. The composition of claim 19, wherein Xaa2* is an analog of Trp comprising a lower alkoxy or lower alkyl substituent at the 1 or 5 position of tryptophan or a halogen substituent at the 5 or 6 position of tryptophan.

25. The composition of claim 19, wherein Xaa2* is an analog of Trp comprising a lower alkoxy or lower alkyl substituent at the 1 or 5 position of tryptophan or a halogen substituent at the 5 or 6 position of tryptophan and Xaa2* is Trp.

26. The composition of claim 19, wherein the compstatin analog comprises a cyclic peptide having a sequence selected from the group consisting of: SEQ ID NOs: 9-36 and comprises a cell-reactive moiety.

27. The composition of claim 19, wherein the compstatin analog comprises a cyclic peptide having a sequence selected from the group consisting of: SEQ ID NOs: 14, 21, 28, 29, 32, 33, 34, or 36 and comprises a cell-reactive moiety.

28. The composition of claim 19, wherein the compstatin analog comprises a cyclic peptide having sequence of SEQ ID NO: 28, 32, or 34.

29. The composition of claim 1, wherein the compstatin analog is a compound that comprises a cyclic peptide having a sequence of X'aa1 - X'aa2 - X'aa3 - X'aa4 -Gln-Asp-Xaa-Gly- X"aa1- X"aa2- X"aa3- X"aa4- X"aa5 (SEQ ID NO: 5), where X'aa4 and Xaa are selected from Trp and analogs of Trp, wherein X'aa1, X'aa2, X'aa3, X"aa1, X"aa2, X"aa3, X"aa4, and X"aa5 are independently selected from among amino acids and amino acid analogs, X'aa2 and X"aa4 are not Cys, and the peptide is cyclized via a bond between X'aa2 and X"aa4, and wherein the compound comprises a cell-reactive moiety.

30. The composition of claim 29, wherein any one or more of X'aa1, X'aa3, X"aa2, X"aa3, and X"aa5 are identical to the amino acids at the corresponding positions in the peptide of any of claims 18 – 28, and X"aa1 is Ala or a single methyl unbranched amino acid.

31. The composition of claim 29, wherein one of X'aa2 and X"aa4 is an amino acid or amino acid analog having a side chain that comprises a primary or secondary amine, the other one of X'aa2 and X"aa4 is an amino acid or amino acid analog having a side chain that comprises a carboxylic acid group, and the bond is an amide bond.

32. The composition of claim 29, wherein X'aa1, X'aa3, X"aa1, X"aa2, X"aa3, and X"aa5 are identical to the amino acids at the corresponding positions in the cyclic peptide of any of claims 18 – 28, and wherein, optionally, one of X'aa2 and X"aa4 is an amino acid or amino acid analog having a side chain that comprises a primary or secondary amine, the other one of X'aa2 and X"aa4 is an amino acid or amino acid analog having a side chain that comprises a carboxylic acid group, and the bond is an amide bond.

33. The composition of any of claims 10 -32, wherein the cyclic peptide is acetylated at the N-terminus, amidated at the C-terminus, or both acetylated at the N-terminus and amidated at the C-terminus.

34. An isolated cell or organ having a compstatin analog covalently bonded thereto.

35. The isolated cell or organ of claim 34, wherein the isolated cell or organ is a human cell or organ.

36. The isolated cell or organ of claim 34, wherein the isolated cell is a blood cell or the isolated organ is a heart, kidney, liver, lung, or pancreas.

37. A method of reducing the sensitivity of a cell or organ to complement-dependent damage, the method comprising contacting the cell with a cell-reactive compstatin analog, wherein the cell-reactive compstatin analog binds covalently to the cell or organ.

38. The method of claim 37, wherein the cell or organ is a human cell or organ.

39. The method of claim 37, wherein the cell is a blood cell or the organ is a heart, kidney, liver, lung, or pancreas.

40. The method of claim 37, wherein the cell or organ is an isolated cell or organ to be transplanted into a subject, and the method comprises contacting the cell or organ with the cell-reactive compstatin analog prior to transplantation.

41. The method of claim 37, wherein the cell or organ has been transplanted into a subject, and the method comprises contacting the organ with the cell-reactive compstatin analog after transplantation.

42. The method of claim 37, wherein the method comprises perfusing the organ with a fluid comprising the cell-reactive compstatin analog.

43. The method of claim 37, wherein the method comprises contacting the cell or organ with the cell-reactive compstatin analog during transplantation of the organ into a subject.

44. The method of claim 37, wherein the the method comprises administering the cell-reactive compstatin analog to a subject after transplantation of the cell or organ into the subject.

45. The method of claim 37, wherein the the method comprises administering the cell-reactive compstatin analog to a subject after transplantation of the organ into the subject, wherein the cell-reactive compstatin analog is administered locally to the transplanted organ.

46. The method of claim 37, wherein the cell or organ has been or is to be transplanted into a subject who is at high risk of developing hyperacute or acute complement-mediated transfusion reaction or organ rejection.

47. A method of treating a subject in need of treatment for a complement-mediated disorder, the method comprising administering a cell-reactive compstatin analog to the subject.

48. The method of claim 47, wherein the cell-reactive compstatin analog is administered locally to a site at risk of or experiencing complement-mediated damage.

49. The method of claim 47, wherein the disorder results in complement-mediated damage to red blood cells, and the cell-reactive compstatin analog is administered intravascularly.

50. The method of claim 47, wherein the subject has a defect in complement regulation.

51. The method of claim 47, wherein the subject is in need of treatment for transplant rejection.

52. The method of claim 47, wherein the subject is in need of treatment for ischemia/reperfusion injury.

53. The method of claim 47, wherein the subject is in need of treatment for a hemolytic anemia.

54. A long-acting compstatin analog comprising a clearance reducing moiety (CRM) that has a molecular weight of at least 20 kilodaltons (kD).

55. The long-acting compstatin analog of claim 54 comprising a clearance-reducing moiety (CRM) having a molecular weight of between 20 kilodaltons (kD) and 100 kD.

56. The long-acting compstatin analog of claim 54 comprising a clearance-reducing moiety having a molecular weight of at least 30 kilodaltons (kD).

57. The long-acting compstatin analog of claim 54 comprising a clearance-reducing moiety having a molecular weight of at least 40 kilodaltons (kD).

58. The long-acting compstatin analog of claim 54, wherein the clearance reducing moiety comprises polyethylene glycol (PEG).

59. The long-acting compstatin analog of claim 54 , wherein the clearance reducing moiety comprises a linear PEG.

60. The long-acting compstatin analog of claim 54, wherein the clearance reducing moiety comprises a branched PEG.

61. The long-acting compstatin analog of claim 54 , wherein the clearance reducing moiety comprises a linear PEG and comprises a compstatin analog moiety attached at each end of the compound.

62. The long-acting compstatin analog of claim 54, wherein the clearance reducing moiety comprises a branched PEG having 3 to 10 branches.

63. The long-acting compstatin analog of claim 54 , wherein the clearance reducing moiety comprises a branched PEG having 3 to 10 branches and at least about 50% of said branches have a compstatin analog moiety attached thereto.

64. The long-acting compstatin analog of claim 54 , wherein the clearance reducing moiety comprises a branched PEG having 3 to 10 branches and at least about 75% of said branches have a compstatin analog moiety attached thereto.

65. The long-acting compstatin analog of claim 54 , wherein the clearance reducing moiety comprises human serum albumin.

66. The long acting-compstatin analog of claim 54 comprising between 2 and 10 compstatin analog moieties.

67. The long acting-compstatin analog of claim 54 comprising between 2 and 100 compstatin analog moieties.

68. The long-acting compstatin analog of claim 54, having a plasma half-life of at least 2 days when injected intravenously into a primate.

69. The long-acting compstatin analog of claim 54, having a plasma half-life of at least 3 days when injected intravenously into a primate

70. The long-acting compstatin analog of claim 54, having a plasma half-life of at least 4 days when injected intravenously into a primate.

71. The long-acting compstatin analog of claim 54 comprising a clearance reducing moiety and having a molar activity of at least about 20% of the activity of a corresponding compstatin analog that has the same amino acid sequence but does not comprise said clearance reducing moiety.

72. The long-acting compstatin analog of claim 54 comprising a clearance reducing moiety and having a molar activity of at least about 30% of the activity of a corresponding compstatin analog that does not comprise said clearance reducing moiety.

73. The long-acting compstatin analog of claim 54 comprising multiple compstatin analog moieties and having a molar activity that is at least about 10% of the sum of the activities of said compstatin analog moieties.

74. The long-acting compstatin analog of claim 54, comprising a clearance reducing moiety and having a terminal half-life at least 5-fold higher than that of a corresponding compstatin analog that does not comprise said CRM at a comparable dose.

75. The long-acting compstatin analog of claim 54, having a Cmax at least 10-fold higher than than of a compstatin analog that comprises the same amino acid sequence but does not comprise said clearance-reducing moiety, at a comparable dose.

76. A long-acting compstatin analog having a plasma half-life of at least 2 days when injected intravenously into a primate.

77. The long-acting compstatin analog of claim 76 having a plasma half-life of at least 3 days when injected intravenously into a primate.

78. The long-acting compstatin analog of claim 76 having a plasma half-life of at least 4 days when injected intravenously into a primate.

79. The long-acting compstatin analog of claim 76 comprising a clearance-reducing moiety and having a molar activity of at least about 20% of the activity of a corresponding compstatin analog that has the same amino acid sequence but does not comprise a clearance-reducing moiety.

80. The long-acting compstatin analog of claim 76 comprising a clearance-reducing moiety and having a molar activity of at least about 30% of the activity of a compstatin analog that comprises the same amino acid sequence but does not comprise said clearance-reducing moiety.

81. The long-acting compstatin analog of claim 76 comprising multiple compstatin analog moieties and having a molar activity that is at least about 10% of the sum of the activities of said compstatin analog moieties.

82. The long-acting compstatin analog of claim 76, having a Cmax at least 10-fold higher than than of a corresponding compstatin analog that does not comprise said clearance-reducing moiety, at a comparable dose.

83. A long-acting compstatin analog comprising a clearance-reducing moiety and having at least about 30% of the activity and a Cmax at least 10-fold higher than than of a corresponding compstatin analog that does not comprise said clearance-reducing moiety, at a comparable dose, and having a plasma half-life of at least 3 days.

84. The long-acting compstatin analog of any of claims 54- 83, wherein the long-acting compstatin analog is a compound that comprises a cyclic peptide having a core sequence of X'aa-Gln - Asp – Xaa – Gly (SEQ ID NO: 3), where X'aa and Xaa are selected from Trp and analogs of Trp, and wherein the compound comprises a clearance reducing moiety.

85. The long-acting compstatin analog of any of claims 54- 83, wherein the compstatin analog is a compound that comprises a cyclic peptide having a core sequence of X'aa-Gln - Asp – Xaa – Gly-X"aa (SEQ ID NO: 4), where X'aa and Xaa are each independently selected from Trp and analogs of Trp, and X"aa is selected from His, Ala, single methyl unbranched amino acids, Phe, Trp, and analogs of Trp, and wherein the compound comprises a clearance reducing moiety.

86. The long-acting compstatin analog of any of claims 54- 83, wherein the peptide has a sequence of X'aa1 - X'aa2 - X'aa3 - X'aa-Gln-Asp-Xaa-Gly- X"aa- X"aa2- X"aa3- X"aa4- X"aa5 (SEQ ID NO: 5), and X'aa1, X'aa2, X'aa3, X"aa2, X"aa3- X"aa4, and X"aa5 are identical to the amino acids at the corresponding positions in compstatin, and wherein the compound comprises a clearance reducing moiety.

87. The long-acting compstatin analog of any of claims 54- 83, wherein the compstatin analog is a compound that comprises a cyclic peptide having a sequence of X'aa1 - X'aa2 - X'aa3 - X'aa4 -Gln-Asp-Xaa-Gly- X"aa1- X"aa2- X"aa3- X"aa4- X"aa5 (SEQ ID NO: 5), where X'aa4 and Xaa are selected from Trp and analogs of Trp, wherein

X'aa1, X'aa2, X'aa3, X"aa1, X"aa2, X"aa3, X"aa4, and X"aa5, are independently selected from among amino acids and amino acid analogs, wherein the peptide is cyclized via a bond between X'aa2 and X"aa4, and wherein the compound comprises a clearance reducing moiety.

88. The long-acting compstatin analog of any of claims 54- 83, wherein X'aa2 and X"aa4 are Cys, and wherein X"aa1 is optionally Ala or a single methyl unbranched amino acid, and wherein the compound comprises a clearance reducing moiety.

89. The long-acting compstatin analog of any of claims 54- 83, wherein X'aa2 and X"aa4 are Cys, wherein any one or more of X'aa1, X'aa2, X'aa3, X"aa2, X"aa3, X"aa4, and X"aa5 are identical to the amino acids at the corresponding positions in compstatin, and X"aa1 is Ala or a single methyl unbranched amino acid, and wherein the compound comprises a clearance reducing moiety.

90. The long-acting compstatin analog of any of claims 54- 83, wherein the compstatin analog is a compound that comprises a cyclic peptide having a sequence:
Xaa1 – Cys – Val – Xaa2 - Gln - Asp – Xaa2* - Gly – Xaa3 - His - Arg – Cys – Xaa4 (SEQ ID NO: 6); wherein:
Xaa1 is Ile, Val, Leu, B^1-Ile, B^1-Val, B^1-Leu or a dipeptide comprising Gly-Ile or B^1-Gly-Ile, and B^1 represents a first blocking moiety;
Xaa2 and Xaa2* are independently selected from Trp and analogs of Trp;
Xaa3 is His, Ala or an analog of Ala, Phe, Trp, or an analog of Trp;
Xaa4 is L-Thr, D-Thr, Ile, Val, Gly, a dipeptide selected from Thr-Ala and Thr-Asn, or a tripeptide comprising Thr-Ala-Asn, wherein a carboxy terminal –OH of any of the L-Thr, D-Thr, Ile, Val, Gly, Ala, or Asn optionally is replaced by a second blocking moiety B^2; and
the two Cys residues are joined by a disulfide bond, and wherein the compound comprises a clearance reducing moiety.

91. The long-acting compstatin analog of claim 90, wherein
Xaa1 is Ile, Val, Leu, Ac-Ile, Ac-Val, Ac-Leu or a dipeptide comprising Gly-Ile or Ac-Gly-Ile;
Xaa2 and Xaa2* are independently selected from Trp and analogs of Trp;
Xaa3 is His, Ala or an analog of Ala, Phe, Trp, or an analog of Trp;

Xaa4 is L-Thr, D-Thr, Ile, Val, Gly, a dipeptide selected from Thr-Ala and Thr-Asn, or a tripeptide comprising Thr-Ala-Asn, wherein a carboxy terminal –OH of any of the L-Thr, D-Thr, Ile, Val, Gly, Ala, or Asn optionally is replaced by $-NH_2$.

92. The long-acting compstatin analog of claim 90, wherein Xaa2 is an analog of Trp having increased hydrophobic character relative to Trp.

93. The long-acting compstatin analog of claim 90, wherein Xaa2 is an analog of Trp comprising a substituted or unsubstituted bicyclic aromatic ring component or two or more substituted or unsubstituted monocyclic aromatic ring components.

94. The long-acting compstatin analog of claim 90, wherein Xaa2* is an analog of Trp having an electronegative substituent on the indole ring and not having increased hydrophobic character relative to Trp.

95. The long-acting compstatin analog of claim 90, wherein Xaa2* is an analog of Trp comprising a lower alkoxy or lower alkyl substituent at the 1 or 5 position of tryptophan or a halogen substituent at the 5 or 6 position of tryptophan.

96. The long-acting compstatin analog of claim 90, wherein Xaa2* is an analog of Trp comprising a lower alkoxy or lower alkyl substituent at the 1 or 5 position of tryptophan or a halogen substituent at the 5 or 6 position of tryptophan and Xaa2* is Trp.

97. The long-acting compstatin analog of claim 90 comprising a cyclic peptide having a sequence selected from the group consisting of: SEQ ID NOs: 9-36.

98. The long-acting compstatin analog of claim 90, wherein the compstatin analog comprises a cyclic peptide having a sequence selected from the group consisting of: SEQ ID NOs: 14, 21, 28, 29, 32, 33, 34, or 36.

99. The long-acting compstatin analog of claim 90, wherein the compstatin analog comprises a cyclic peptide having sequence of SEQ ID NO: 28, 32, or 34.

100. A long-acting compstain analog, wherein the long-acting compstatin analog is a compound that comprises a cyclic peptide having a sequence of X'aa1 - X'aa2 - X'aa3

- X'aa4 -Gln-Asp-Xaa-Gly- X"aa1- X"aa2- X"aa3- X"aa4- X"aa5 (SEQ ID NO: 5), where X'aa4 and Xaa are selected from Trp and analogs of Trp, wherein X'aa1, X'aa2, X'aa3, X"aa1, X"aa2, X"aa3, X"aa4, and X"aa5 are independently selected from among amino acids and amino acid analogs, X'aa2 and X"aa4 are not Cys, and the peptide is cyclized via a bond between X'aa2 and X"aa4.

101. The long-acting compstatin analog of claim 100, wherein any one or more of X'aa1, X'aa3, X"aa2, X"aa3, and X"aa5 are identical to the amino acids at the corresponding positions in the peptide of any of claims 18 – 28, and X"aa1 is Ala or a single methyl unbranched amino acid.

102. The long-acting compstatin analog of claim 100, wherein one of X'aa2 and X"aa4 is an amino acid or amino acid analog having a side chain that comprises a primary or secondary amine, the other one of X'aa2 and X"aa4 is an amino acid or amino acid analog having a side chain that comprises a carboxylic acid group, and the bond is an amide bond.

103. The long-acting compstatin analog of claim 100, wherein X'aa1, X'aa3, X"aa1, X"aa2, X"aa3, and X"aa5 are identical to the amino acids at the corresponding positions in the cyclic peptide of any of claims 18 – 28, and wherein, optionally, one of X'aa2 and X"aa4 is an amino acid or amino acid analog having a side chain that comprises a primary or secondary amine, the other one of X'aa2 and X"aa4 is an amino acid or amino acid analog having a side chain that comprises a carboxylic acid group, and the bond is an amide bond.

104. The long-acting compstatin analog of any of claims 84 - 103, wherein the cyclic peptide is acetylated at the N-terminus, amidated at the C-terminus, or both acetylated at the N-terminus and amidated at the C-terminus.

105. A long-acting compstatin analog comprising a compound that comprises a cyclic peptide having a sequence selected from the group consisting of: SEQ ID NOs: 14, 21, 28, 29, 32, 33, 34, or 36.

106. The long-acting compstatin analog of any of claims 84 - 105 comprising a compound in which at least one NHS ester of the compound of any of Formulae I – XVI or

Formulae A-H has reacted with an amino group in a side chain or terminus of a compstatin analog moiety.

107. A method of making a long-acting compstatin analog, comprising reacting the compound of any of Formulae I – XVI or Formulae A-H with a compstatin analog moiety.

108. A method of making a long-acting compstatin analog, comprising reacting the compound of any of Formulae I – XVI or Formulae A-H with a compstatin analog moiety comprising the amino acid sequence of any of SEQ ID NOs: 3-36, 37, 37A, 38A, 39A, 40a, or 41A.

109. A long-acting compstatin analog prepared according to the method of claim 107 – 108 or comprising the structure of said long-acting compstatin analog.

110. The long-acting compstatin analog of any of claims 84 - 106 or 109 or 145-164, further comprising a targeting moiety.

111. A composition comprising the long-acting compstatin analog of any of claims 54 – 105 or 109 or 145-164, and a pharmaceutically acceptable carrier.

112. A pharmaceutical grade composition comprising the long-acting compstatin analog of any of claims 54 - 106 or 109 or 145-164.

113. A pharmaceutical grade composition comprising the long-acting compstatin analog of any of claims 54 - 106 or 109 or 145-164, and a pharmaceutically acceptable carrier.

114. A method of reducing the sensitivity of a cell or organ to complement-dependent damage, the method comprising contacting the cell with a long-acting compstatin analog or composition of any of claims 54 - 106 or 109 – 113 or 144 - 165.

115. The method of claim 114, wherein the cell or organ is a human cell or organ.

116. The method of claim 114, wherein the cell is a blood cell or the organ is a heart, kidney, liver, lung, or pancreas.

117. The method of claim 114, wherein the the method comprises administering the long-acting compstatin analog or composition to a subject.

118. A method of treating a subject in need of treatment for a complement-mediated disorder, the method comprising administering a long-acting compstatin analog or composition of any of claims 54 - 106 or 109 - 113 or claims 144-165 to the subject.

119. The method of claim 118 wherein the long-acting compstatin analog is administered locally to a site at risk of or experiencing complement-mediated damage.

120. The method of claim118, wherein the disorder results in complement-mediated damage to red blood cells.

121. The method of claim 118, wherein the disorder results in complement-mediated damage to red blood cells, and the long-acting compstatin analog is administered intravascularly or subcutaneously.

122. The method of claim 118, wherein the subject has a defect in complement regulation.

123. The method of claim 118, wherein the subject is in need of treatment for transplant rejection.

124. The method of claim 118, wherein the subject is in need of treatment for ischemia/reperfusion injury.

125. The method of claim 118, wherein the subject is in need of treatment for a hemolytic anemia.

126. The method of claim 118, wherein the subject is in need of treatment for an autoimmune disease.

127. The method of claim 118, wherein the subject is in need of treatment for neuropathic pain.

128. The method of claim 118, wherein the subject is in need of treatment for MPGN.

129. The method of claim 118, wherein the subject is in need of treatment for neuromyelitis optica.

130. The method of claim 118, wherein the subject is in need of treatment for spinal cord injury.

131. The method of claim 118, wherein the subject is in need of treatment for asthma, COPD.

132. The method of claim 118, wherein the long-acting compstatin analog is administered intravascularly or subcutaneously.

133. The method of claim 118, wherein the long-acting compstatin analog is administered intravascularly or subcutaneously.

134. The method of claim 118, wherein the long-acting compstatin analog is administered intravenously.

135. The method of claim 118, wherein the long-acting compstatin analog is administered subcutaneously.

136. The method of claim 118, wherein the long-acting compstatin analog is administered intramuscularly.

137. The method of claim 118, wherein the long-acting compstatin analog is administered subcutaneously one or two times per day.

138. The method of claim 107 or claim 108, wherein the compound of any of Formulae I – XVI or Formulae A-H and the compstatin analog moiety each comprise a click functionality, and the method comprises performing a click chemistry reaction.

139. A compstatin analog comprising a click chemistry group.

140. The compstatin analog of claim 139, wherein the compstatin analog comprises a compound comprising any of SEQ ID NOs: 3-36, 37, 37A, 38A, 39A, 40A, or 41A.

141. The compstatin analog of claim 139 or claim 140, wherein the click chemistry group comprises an azide, an alkyne, an octyne, a dibenzoaryl cyclooctyne.

142. The compstatin analog of claim 140, wherein the click chemistry group is DBCO, DIBO, DIFO, BARAC, or BCN.

143. The compstatin analog of claim 140, wherein the click chemistry group is suitable for a copper-free click chemistry reaction.

144. A composition comprising a compstatin analog of any of claims 139 – 143 and a CRM.

145. A conjugate formed by reaction of a compstatin analog of any of claims 139 – 143 with a CRM comprising a complementary click chemistry group.

146. A long-acting compstatin analog comprising a compstatin analog moiety and a CRM conjugated via a click chemistry bond.

147. A long-acting compstatin analog comprising a compstatin analog moiety and a CRM, wherein the CRM comprises a POZ.

148. A long-acting compstatin analog comprising at least two compstatin analog moieties and a CRM.

149. The long-acting compstatin analog of claim 148, wherein the long-acting compstatin analog has at least 90% of the activity, or at least 100% of the activity, on a molar basis as a compstatin analog having the same sequence as the compstatin analog moiety.

150. The long-acting compstatin analog of claim 148, wherein the CRM comprises a PEG.

151. The long-acting compstatin analog of claim 148, wherein the CRM comprises a POZ.

152. The long-acting compstatin analog of claim 148, wherein the CRM comprises a polypeptide.

153. The long-acting compstatin analog of claim 148, comprising two compstatin analog moieties.

154. The long-acting compstatin analog of claim 148, comprising three compstatin analog moieties.

155. The long-acting compstatin analog of claim 148, comprising two to eight compstatin analog moieties.

156. The long-acting compstatin analog of claim 148, comprising two compstatin analog moieties and a PEG.

157. The long-acting compstatin analog of claim 148, comprising three compstatin analog moieties and a PEG.

158. The long-acting compstatin analog of claim 148, comprising two to eight compstatin analog moieties and a PEG.

159. The long-acting compstatin analog of claim 148, comprising two compstatin analog moieties and a POZ.

160. The long-acting compstatin analog of claim 148, comprising three compstatin analog moieties and a POZ.

161. The long-acting compstatin analog of claim 148, comprising two to eight compstatin analog moieties and a POZ.

162. The long-acting compstatin analog of claim 148, comprising two compstatin analog moieties and a polypeptide.

163. The long-acting compstatin analog of claim 148, comprising three compstatin analog moieties and a polypeptide.

164. The long-acting compstatin analog of claim 148, comprising two to eight compstatin analog moieties and a polypeptide.

165. A composition comprising a compstatin analog of any of claims 139 – 164, wherein the composition is optionally a pharmaceutical composition.

166. The method of claim 118, wherein the long-acting compstatin analog is administered to treat a Th17-associated disease.

167. A long-acting compstatin analog comprising a CRM and a compstatin analog moiety, wherein the long-acting compstatin analog has a molecular weight of at least about 30 kD, a terminal half-life of at least about 3 days when administered to a primate, and an activity on a molar basis of at least 80% of that of a compstatin analog comprising the same compstatin analog sequence as the compstatin analog moiety.but not linked to the CRM.

168. The long-acting compstatin analog of claim 167, having a molecular weight of at least at least about 40kD.

169. The long-acting compstatin analog of claim 167, having a terminal half-life of at least about 4 days.

170. The long-acting compstatin analog of claim 167, having a terminal half-life of at least about 5 days.

171. The long-acting compstatin analog of claim 167, having an activity on a molar basis of at least 90% of that of a compstatin analog comprising the same compstatin analog sequence as the compstatin analog moiety.but not linked to the CRM.

172. The long-acting compstatin analog of claim 167, having an activity on a molar basis at least equal to that of a compstatin analog comprising the same compstatin analog sequence as the compstatin analog moiety.but not linked to the CRM.

173. The long-acting compstatin analog of claim 167, wherein the compstatin analog moiety comprises any of SEQ ID NOs: 3-36, 37, 37A, 38A, 39A, 40A, or 41A.

174. The long-acting compstatin analog of any of claims 167-171, wherein the CRM comprises a PEG, POZ, or polypeptide.

175. A long-acting compstatin analog comprising at least one compstatin analog moiety and an unstructured polypeptide.

176. The long-acting compstatin analog of claim 175, wherein , the compstatin analog moiety comprises any of SEQ ID NOs: 3-36, 37, 37A, 38A, 39A, 40A, or 41A.

177. A method of treating a complement-mediated disorder comprising administering the long-acting compstatin analog of any of claims 167-176 to a subject in need thereof.

Figure 1

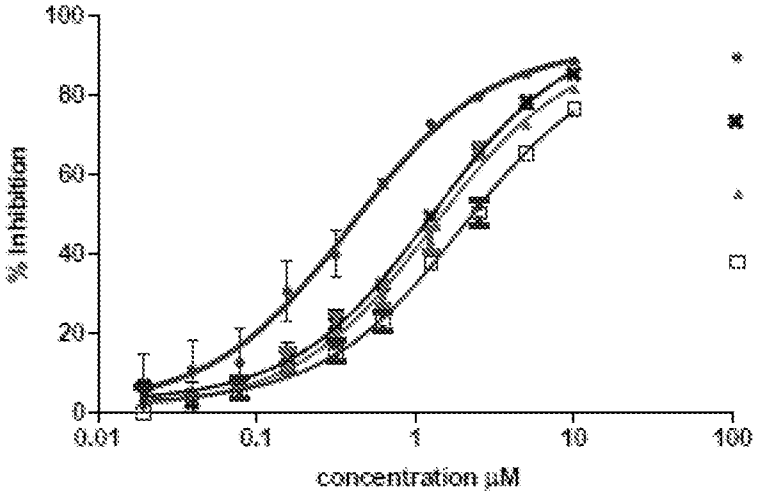


Figure 2

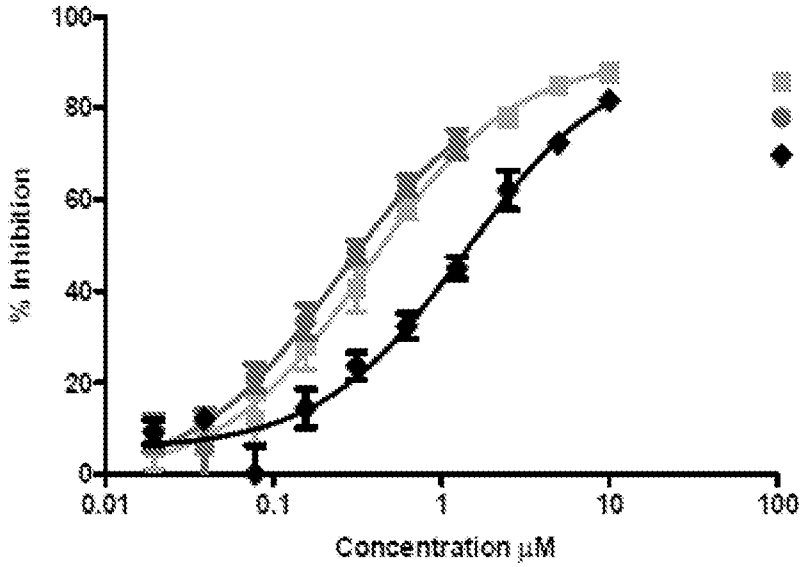


Figure 3

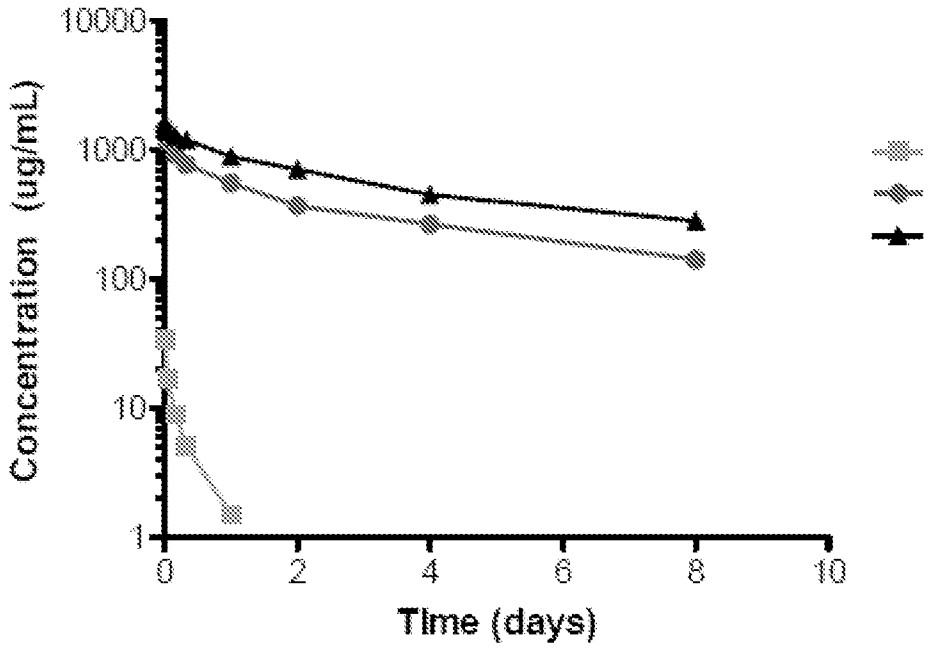


Figure 4

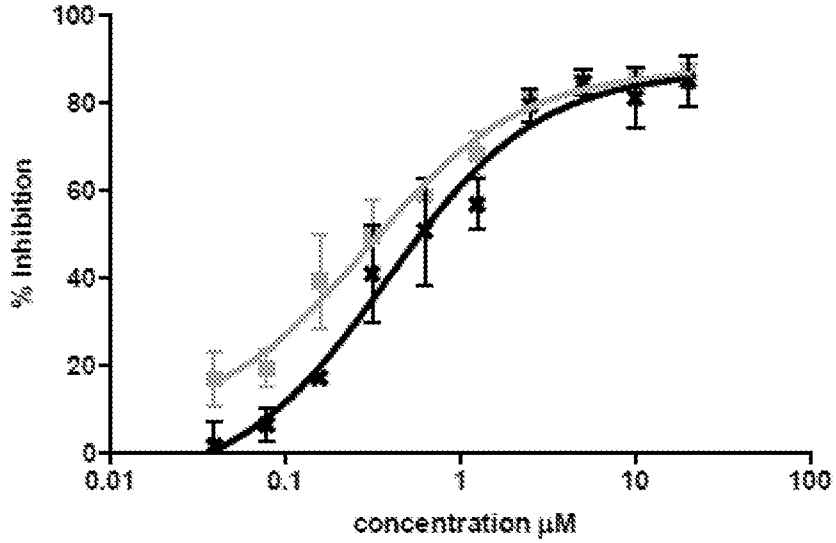


Figure 5

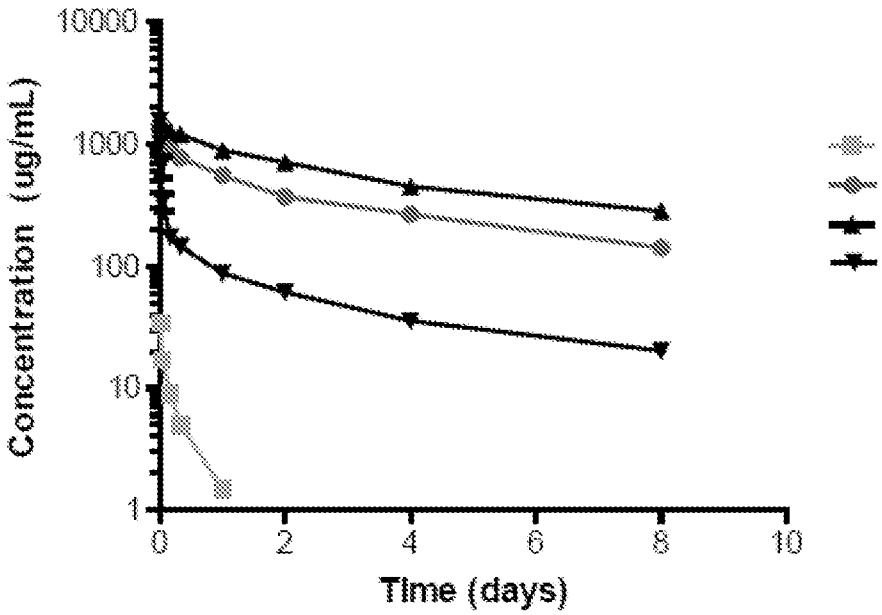


Figure 6

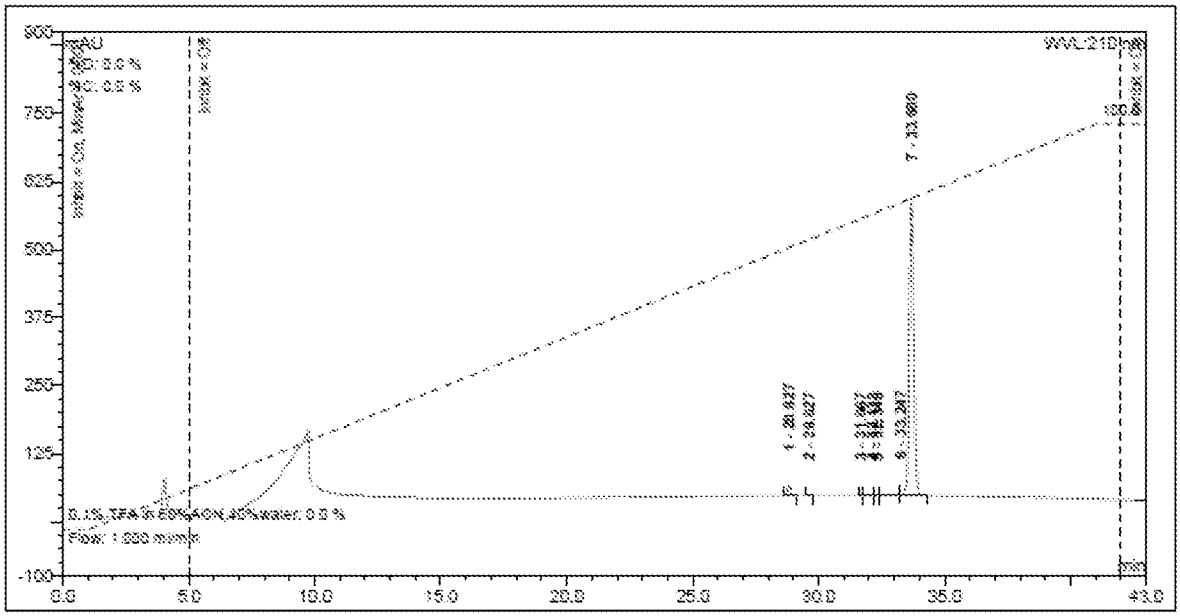

mAU
WVL:210
3 - 33.880
1 - 28.827
2 - 28.827
4 - 31.467
5 - 31.467
6 - 33.267
Flow: 1.000 ml/min
900
750
625
500
375
250
125
-100
0.0
5.0
10.0
15.0
20.0
25.0
30.0
35.0
40.0

Figure 7

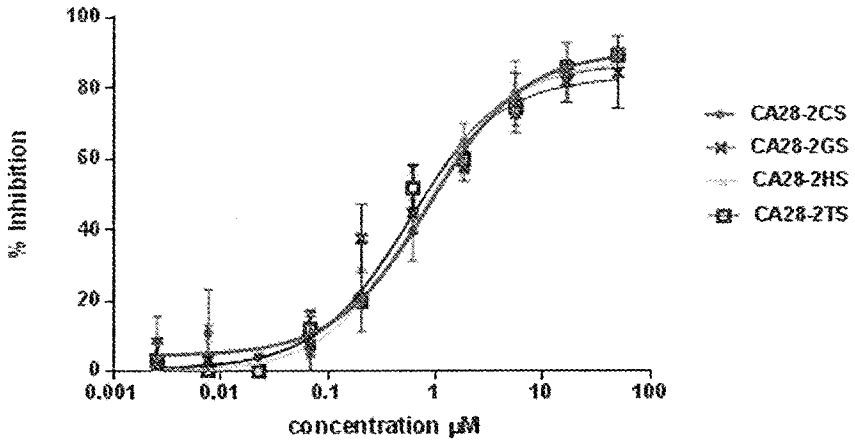

% Inhibition
100
80
60
40
20
0
0.001
0.01
0.1
1
10
100
concentration µM
CA28-2CS
CA28-2GS
CA28-2HS
CA28-2TS

Figure 8

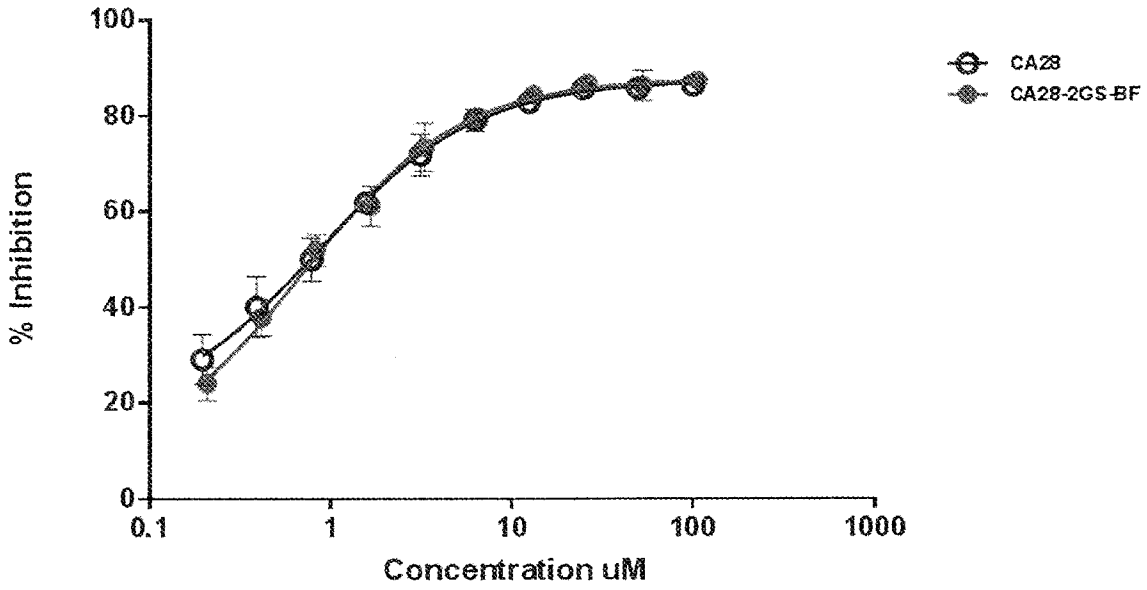
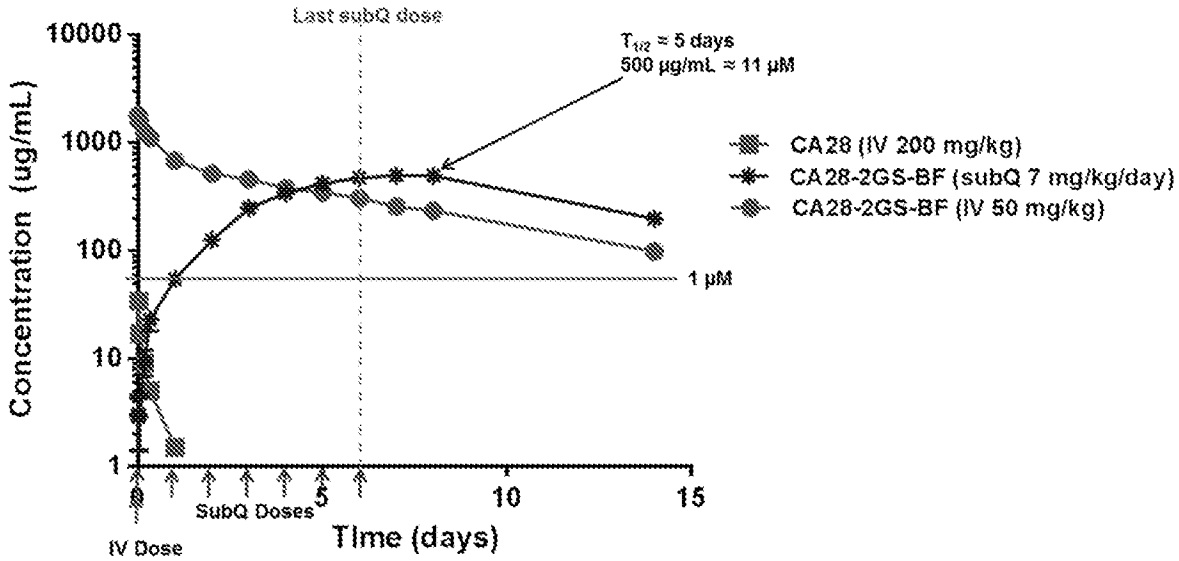

100
80
60
40
20
0
% Inhibition
0.1
1
10
100
1000
Concentration uM
CA28
CA28-2GS-BF

Figure 9


Last subQ dose
$T_{1/2}$ ≈ 5 days
500 µg/mL ≈ 11 µM
CA28 (IV 200 mg/kg)
CA28-2GS-BF (subQ 7 mg/kg/day)
CA28-2GS-BF (IV 50 mg/kg)
1 µM
Concentration (ug/mL)
10000
1000
100
10
1
0
5
10
15
SubQ Doses
IV Dose
Time (days)

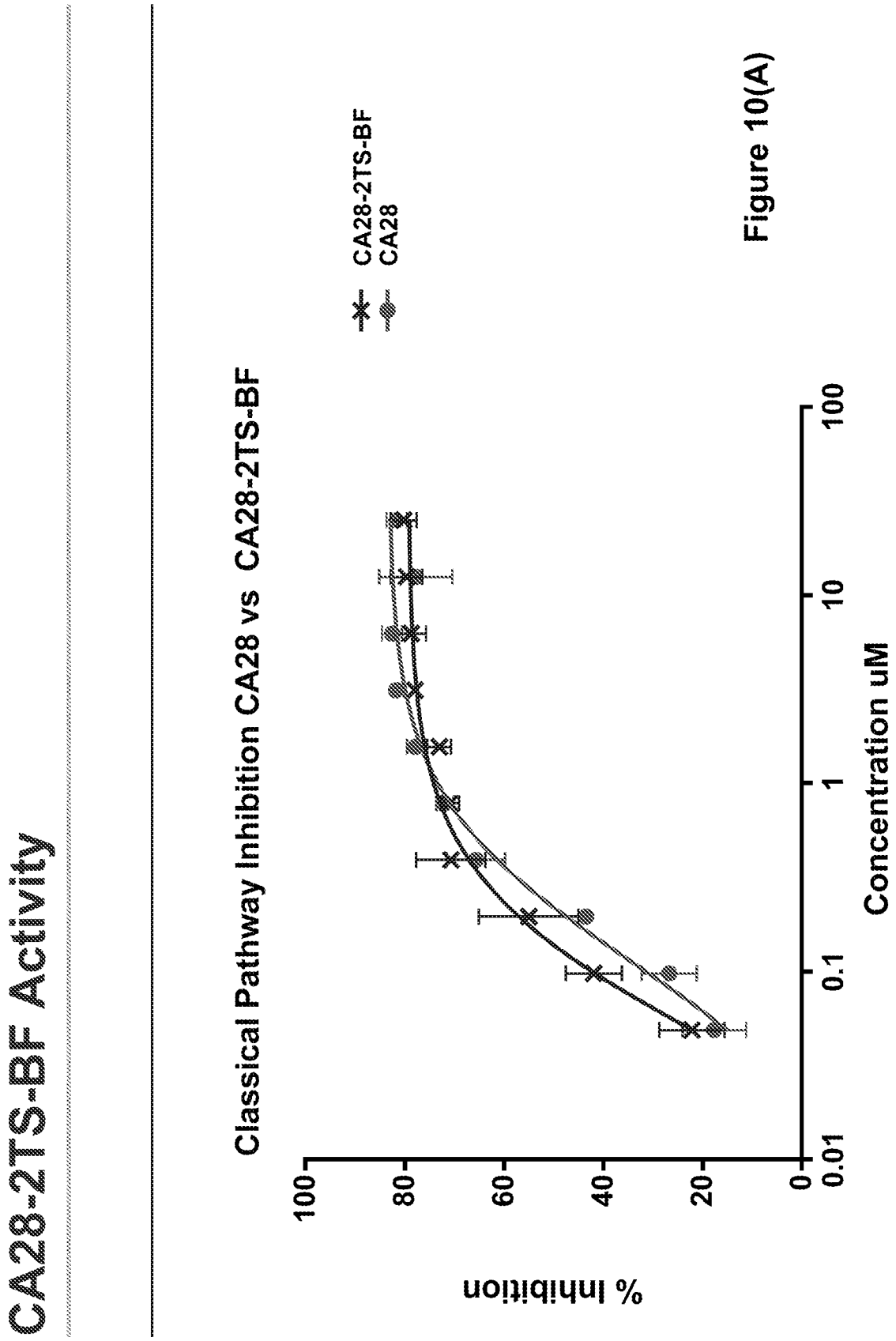

CA28-2TS-BF Activity
Classical Pathway Inhibition CA28 vs CA28-2TS-BF
CA28-2TS-BF
CA28
% Inhibition
100
80
60
40
20
0
0.01
0.1
1
10
100
Concentration uM

Figure 10(A)

CA28-2TS-BF Activity

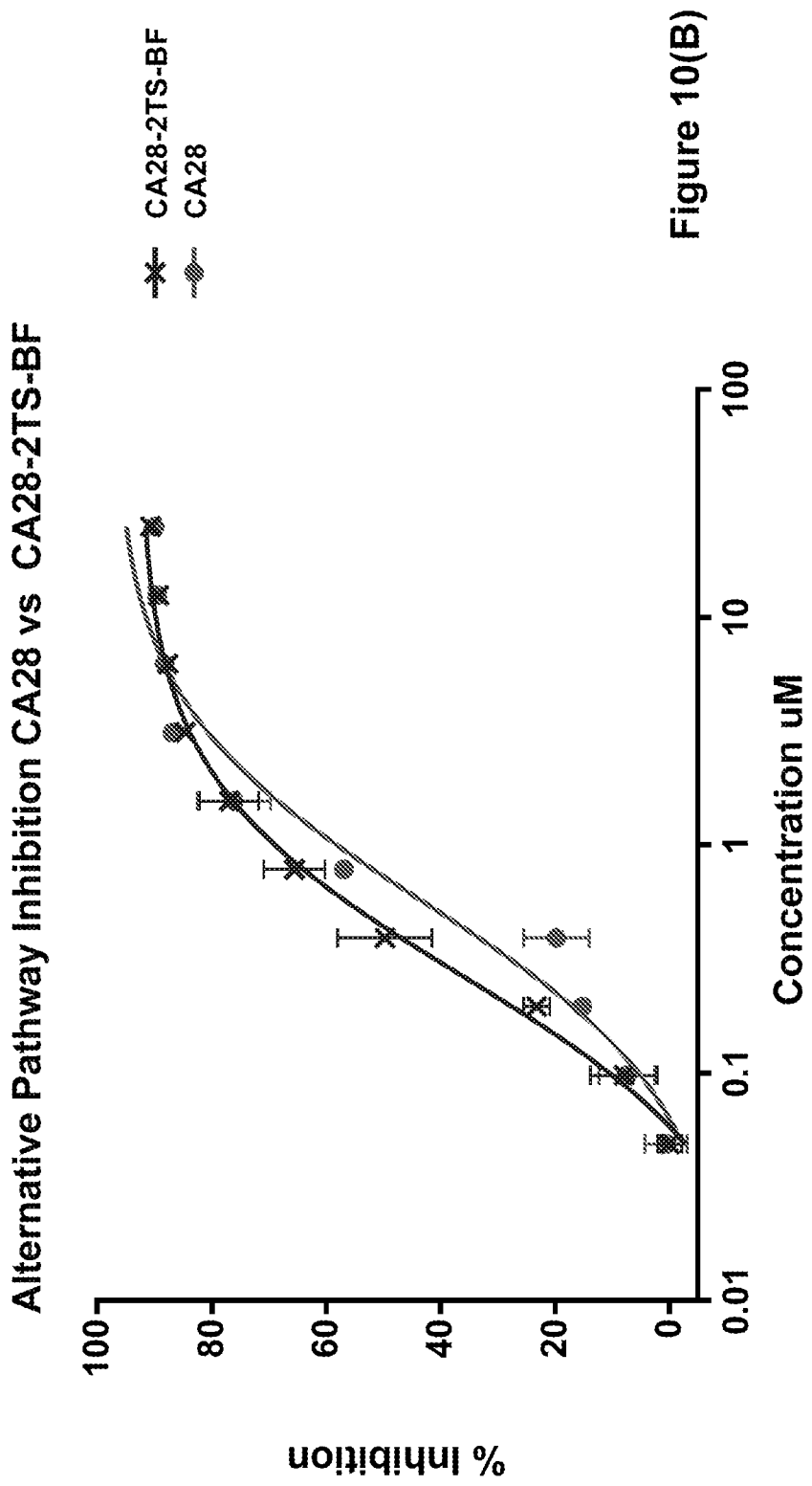


Figure 10(B)

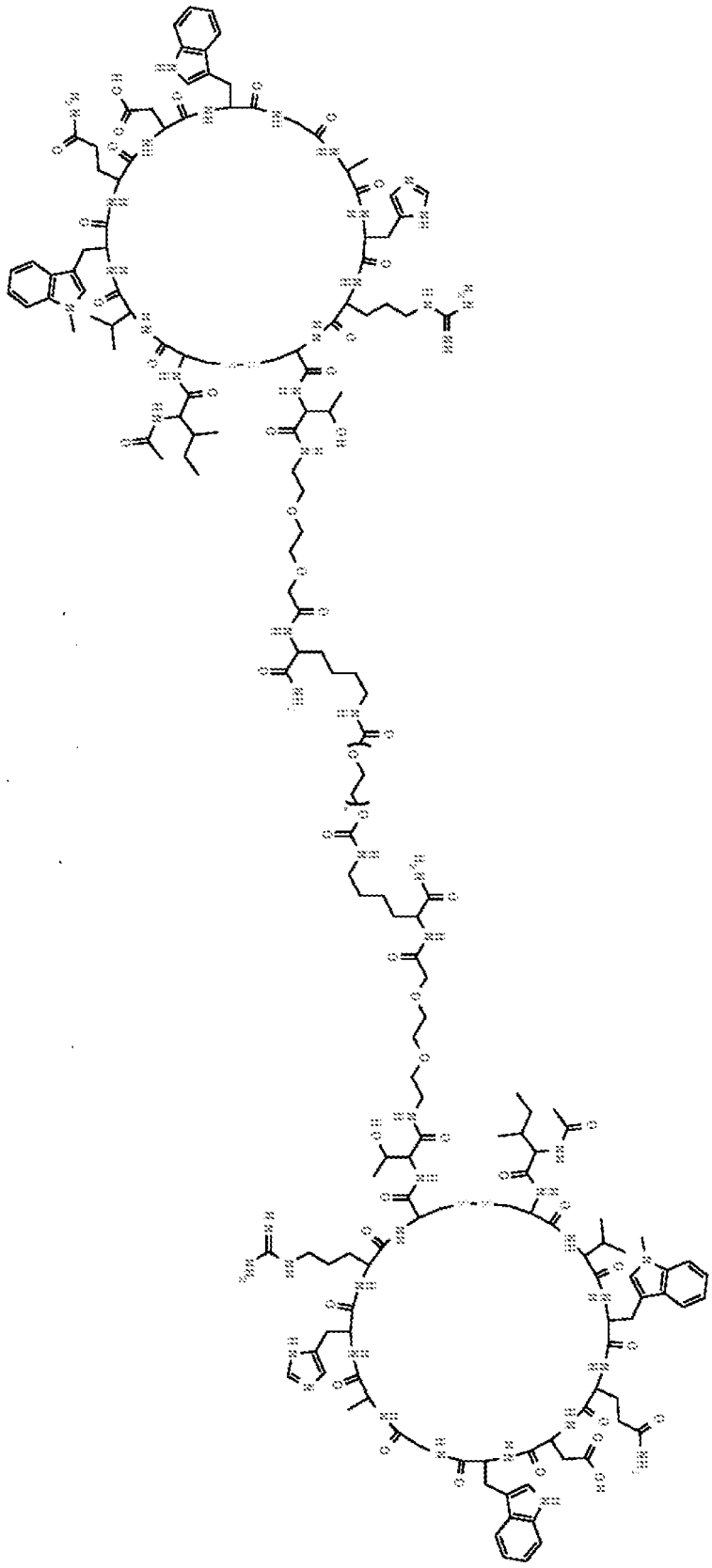


Figure 10(C)

CA28-2TS-BF IV vs CA28-2TS-BF SC vs CA28 PK (µg/mL)

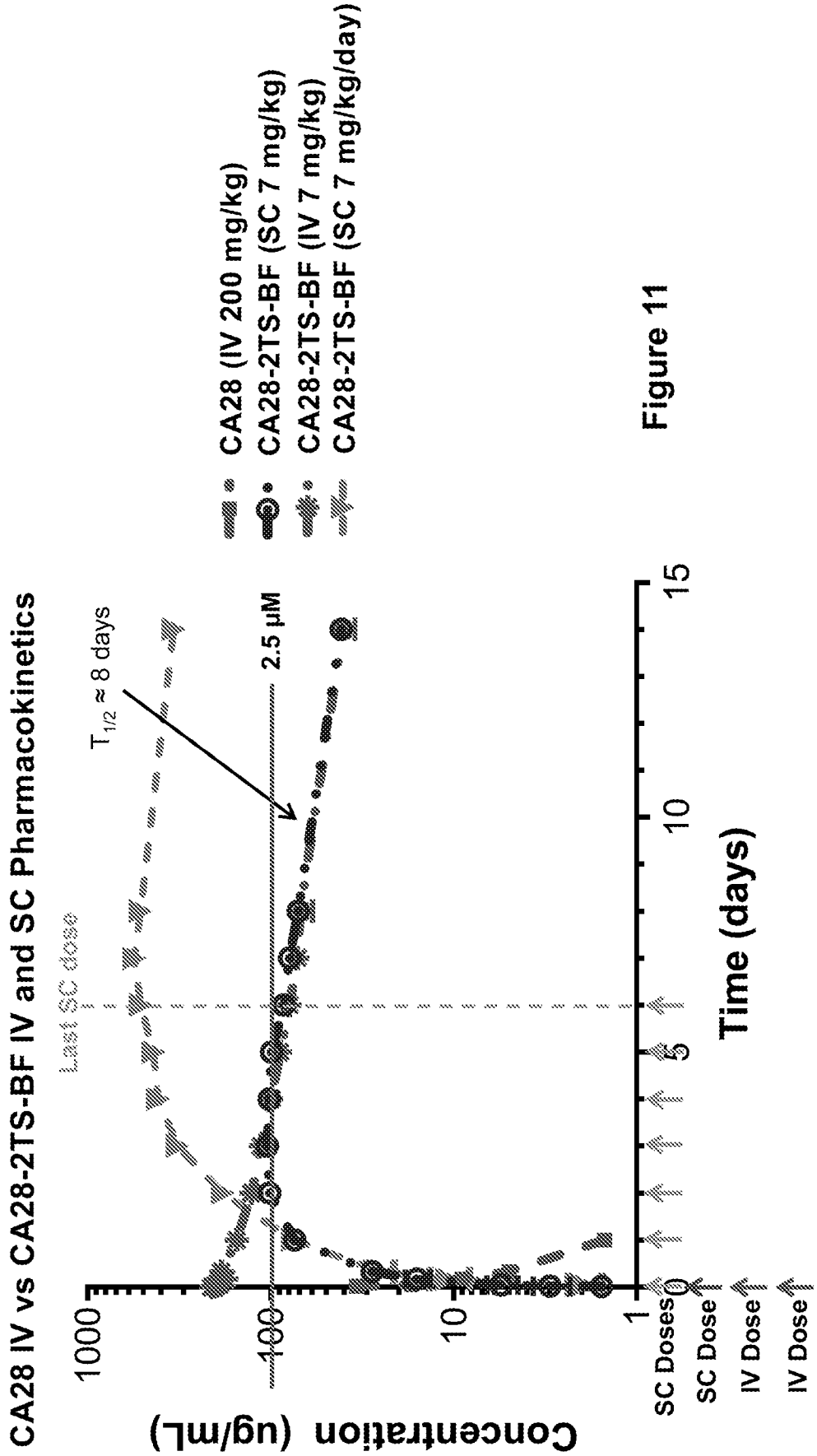


Figure 11

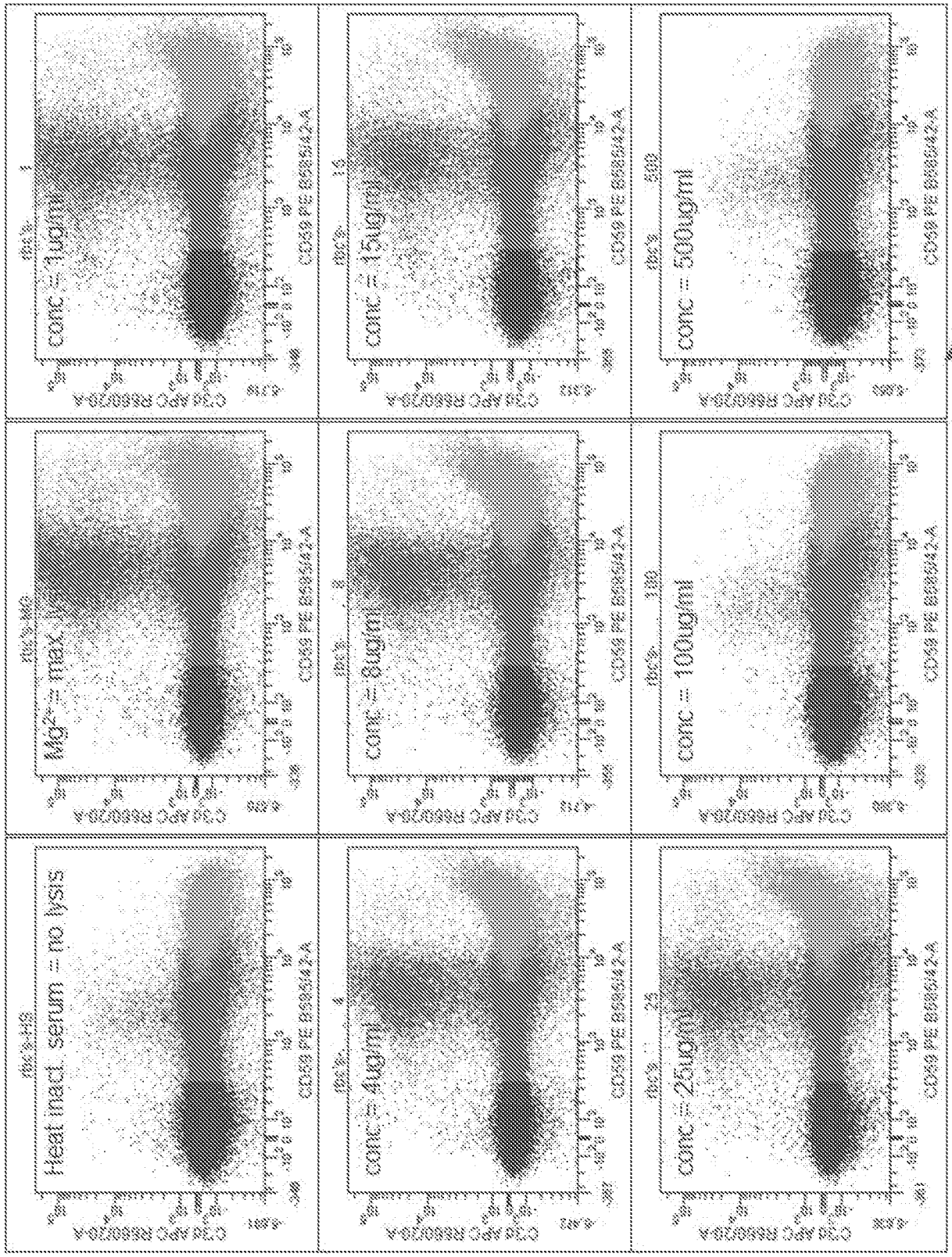

Dilution experiment – effect of CA28 on C3 loading
Heat inact. serum = no lysis
Mg2+ = max. lysis
conc = 1ug/ml
conc = 4ug/ml
conc = 8ug/ml
conc = 15ug/ml
conc = 25ug/ml
conc = 100ug/ml
conc = 500ug/ml
C3-loading
CD59 expression

Figure 12(A)

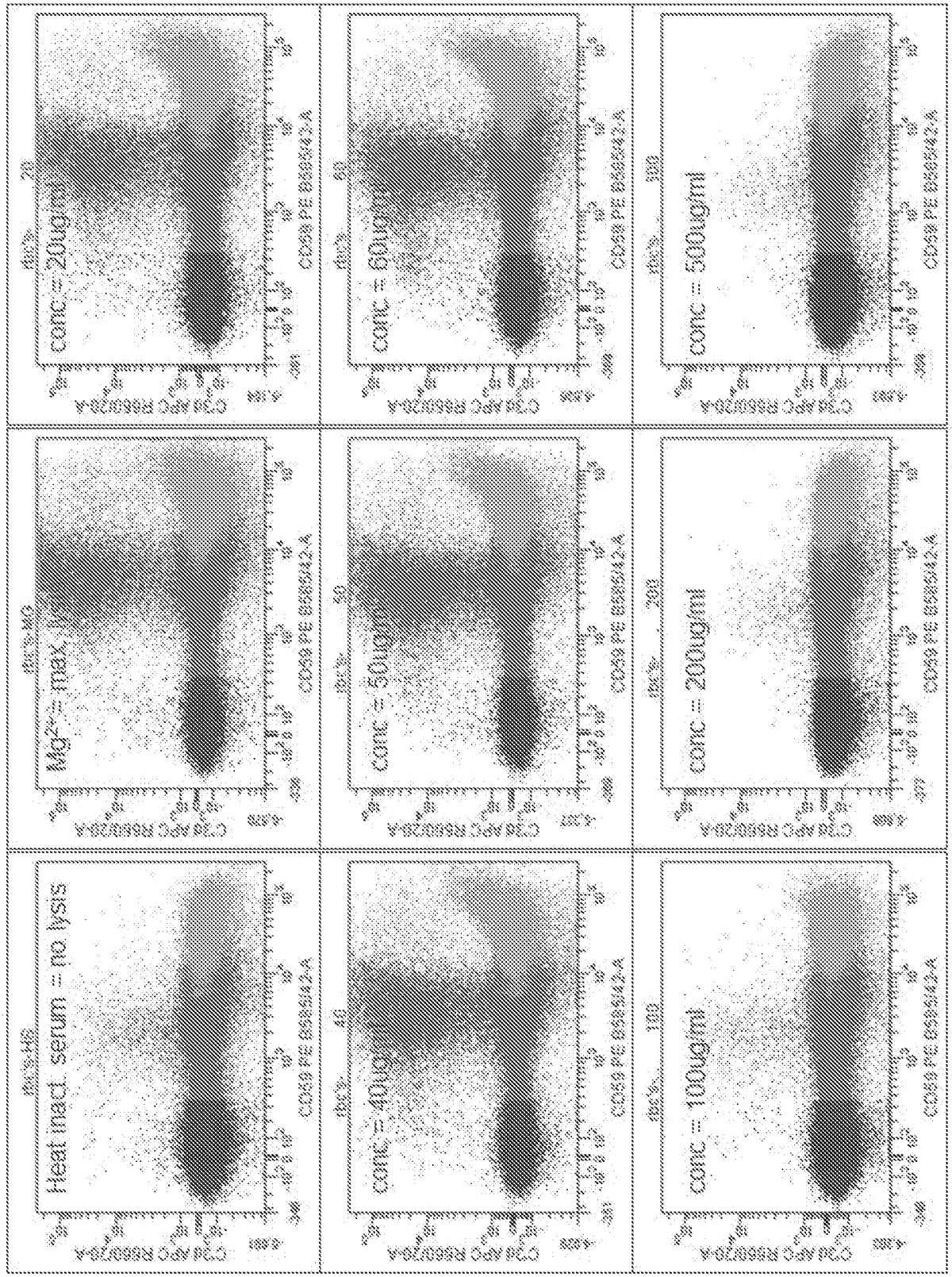

Dilution experiment – effect of CA28-2GS-BF on C3 loading
Heat inact. serum = no lysis
Mg2+ = max. lysis
conc = 20ug/ml
conc = 40ug/ml
conc = 50ug/ml
conc = 60ug/ml
conc = 100ug/ml
conc = 200ug/ml
conc = 500ug/ml
C3-loading
CD59 expression

Figure 12(B)

C3 loading PNH Red Cells during a modified Ham Test

Figure 13

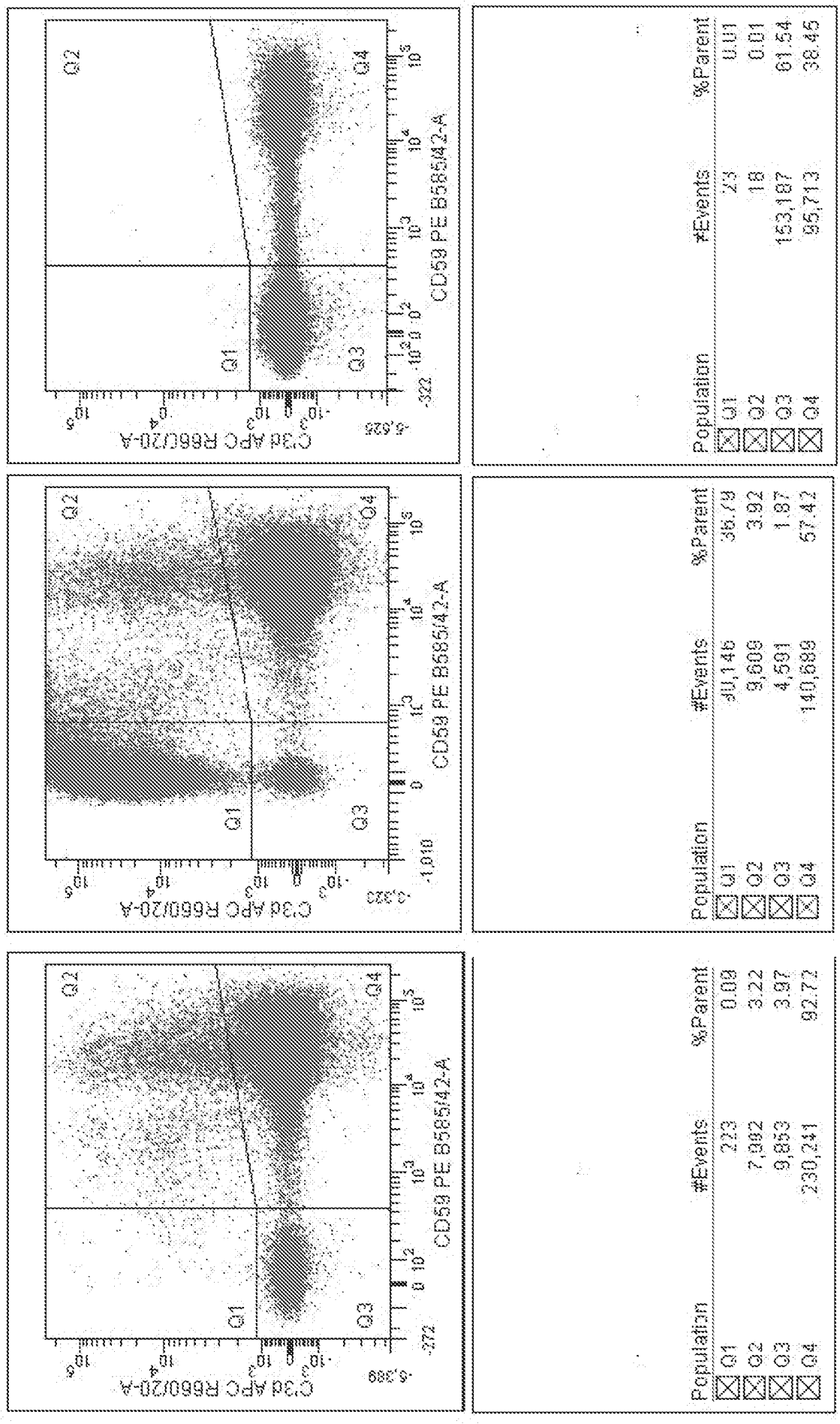


Figure 14

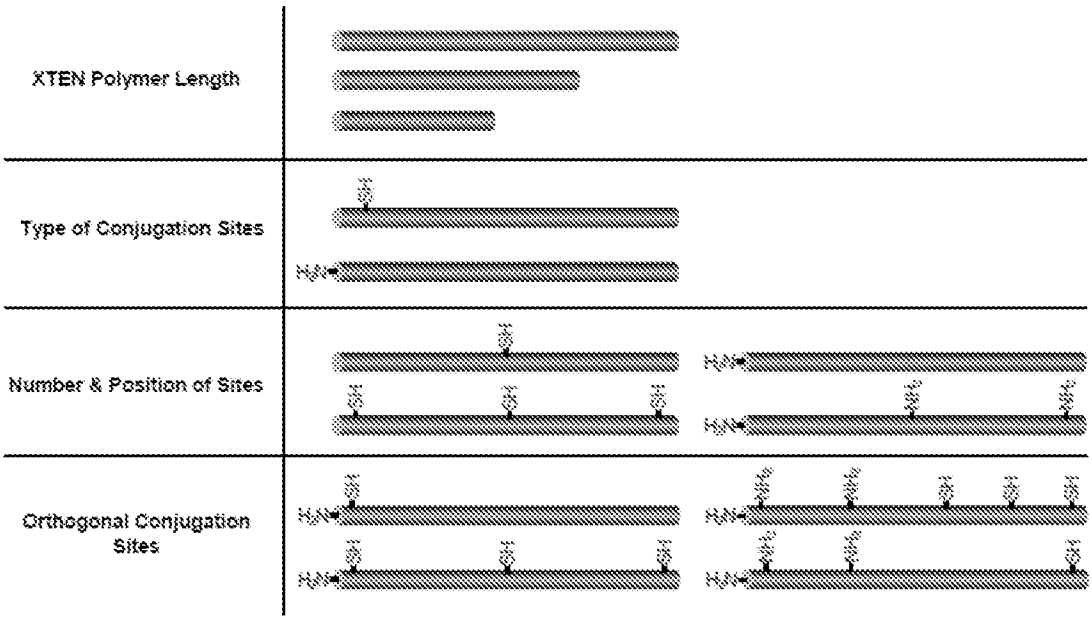
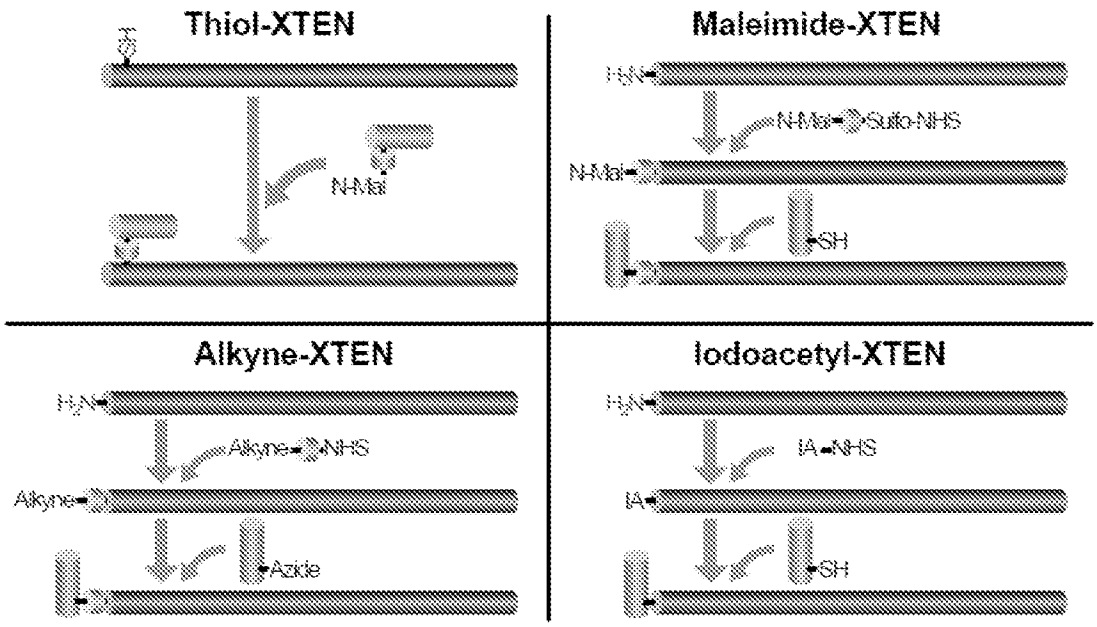


Figure 15



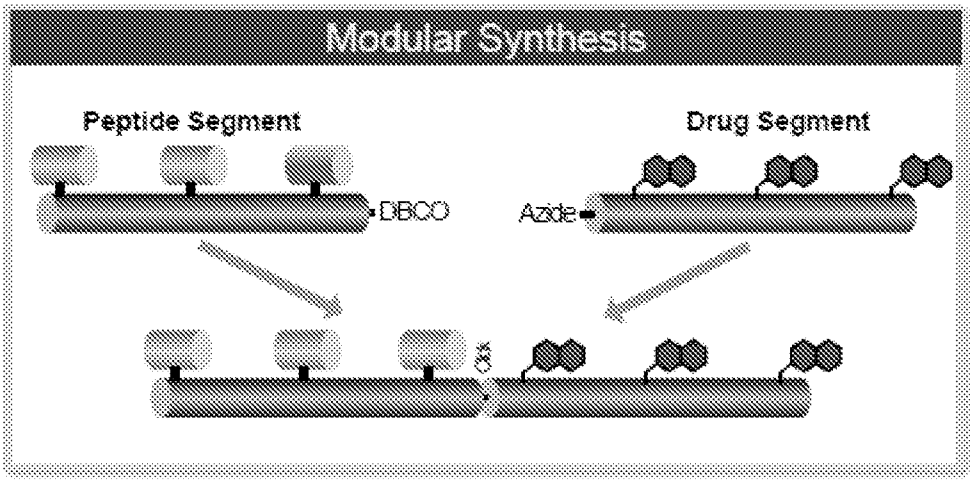
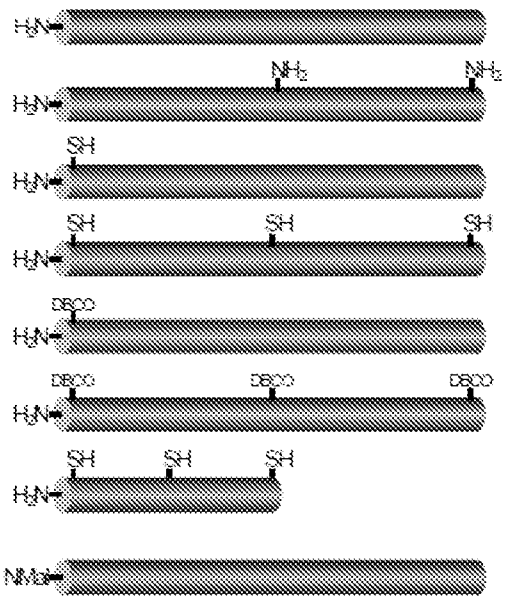


Figure 16

Figure 17



Product Number	Product Name
AXX-1	1xAmino-XTEN869
AXX-2	3xAmino-XTEN883
AXX-3	1xThiol, 1xAmino-XTEN880
AXX-4	3xThiol, 1xAmino-XTEN905
AXX-5	1xDBCO, 1xAmino-XTEN880
AXX-6	3xDBCO, 1xAmino-XTEN905
AXX-7	3xThiol, 1xAmino-XTEN432
AXX-12	1xNMal-XTEN869