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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 September 2013 (06.09.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/130684 A1

(51) International Patent Classification:
C07K 14/00 (2006.01)

(21) International Application Number:
PCT/US2013/028117

(22) International Filing Date:
27 February 2013 (27.02.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/634,312	27 February 2012 (27.02.2012)	US
61/690,187	18 June 2012 (18.06.2012)	US
61/709,942	4 October 2012 (04.10.2012)	US

(71) Applicant *(for all designated States except US)*: **AMUNIX OPERATING INC.** [US/US]; 500 Ellis Street, Mountain View, CA 94043-2206 (US).

(72) Inventors; and

(71) Applicants *(for US only)*: **SCHELLENBERGER, Volker** [DE/US]; 914 Moreno Avenue, Palo Alto, CA 94303 (US). **PODUST, Vladimir** [US/US]; 5805 Shadow Ridge Drive, Castro Valley, CA 94552 (US). **WANG, Chia-wei** [—/US]; 1449 David Lane, Milpitas, CA 95035 (US). **MCLAUGHLIN, Bryant** [US/US]; 190 Broadway, Millbrae, CA 94030 (US). **SIM, Bee-Cheng** [SG/US]; 2255 Showers Drive, Apt. #211, Mountain View, CA 94040 (US). **DING, Sheng** [CN/US]; 36 Duane Street, Apt. #1, Redwood City, CA 94062 (US). **GU, Chen** [US/US]; 2521 Parkside Drive, Fremont, CA 94536 (US).

(74) Agents: **ORBAI, Lucian** et al.; Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, CA 94304-1050 (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, IS, JP, KE, KG, KM, 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, 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,

[Continued on next page]

(54) Title: XTEN-FOLATE CONJUGATE COMPOSITIONS AND METHODS OF MAKING SAME

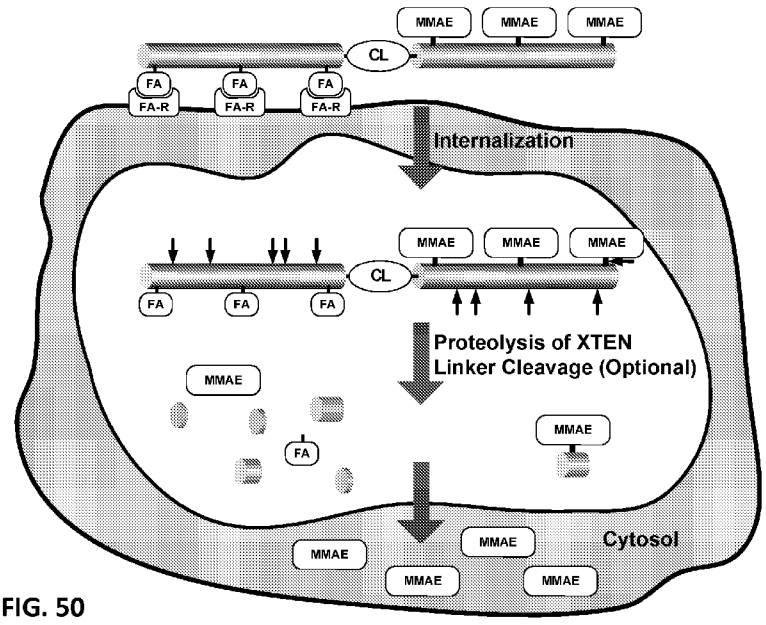


FIG. 50

(57) Abstract: The present invention relates to binding fusion protein compositions comprising targeting moieties linked to extended recombinant polypeptide (XTEN), binding fusion protein-drug conjugate compositions, and XTEN-drug conjugate compositions, isolated nucleic acids encoding the compositions and vectors and host cells containing the same, and methods of using such compositions in treatment of diseases, disorders, and conditions.

WO 2013/130684 A1

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

— *with sequence listing part of description (Rule 5.2(a))*

Published:

— *with international search report (Art. 21(3))*

XTEN-FOLATE CONJUGATE COMPOSITIONS AND METHODS OF MAKING SAME

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application claims priority benefit to U.S. Provisional Application Serial No. 61/634,312 filed February 27, 2012, U.S. Provisional Application Serial No. 61/690,187 filed June 18, 2012, and U.S. Provisional Application Serial No. 61/709,942 filed October 4, 2012, which applications are incorporated herein by reference.

FIELD OF THE INVENTION

[0002] This application relates to a medicament and its use in methods of treatment, diagnosis and imaging. In particular, it relates to the treatment of cancer with a cytotoxic agent linked to a targeting agent.

BACKGROUND OF THE INVENTION

[0003] It is generally accepted in the field of oncology that the main disadvantage of many chemotherapeutic agents are their inherent lack of specificity. Auristatins are highly potent antimitotic agents related in structure to the marine natural product, dolastatin 10. (Francisco, JA et al. Blood 2003 102: 1458-1465) The naturally occurring antimitotic pentapeptide dolastatin 10 and its synthetic analogue monomethyl auristatin E (MMAE) possess subnanomolar cytotoxicity against many human cancer cell lines and are over a hundred- to a thousand-times more potent than many pharmaceuticals, including taxol and doxorubicin, respectively, which are currently used in clinic. Dolastatins and auristatins inhibit tubulin polymerization in dividing cells, whereas taxol promotes it and stabilizes the microtubules. In the end, both dolastatins or auristatins and taxol inhibit cell proliferation. Additionally, auristatins function as the vascular disrupting agents and damage the established tumor vessels, thereby likely causing a more pronounced effect than taxol in vivo. Nonetheless, the therapeutic efficacies of the auristatins and dolastatins are poor as they also cause nonselective toxicity to normal cells, giving rise to significant systemic effects (Bajjuri, KM et al. ChemMedChem. (2011) 6(1): 54–59). Attempts have been made to reduce the non-selective toxicity by conjugating them to tumor-targeting antibodies in the Fc region through a linker. Upon uptake of the conjugates in tumor cells, free drugs are released by lysosomal protease-catalyzed hydrolysis of the linker. Antibody-drug conjugates (ADCs) consist of potent anticancer drugs covalently linked to monoclonal antibodies (mAb) that bind to tumor-associated antigens, and were introduced as a more selective way to direct highly toxic drugs to cells bearing such antigens. Although ADCs offer the opportunity to widen the therapeutic window of cancer therapies by increasing the amount of drug that gets to the tumor and sparing normal tissues from

systemic drug exposure, only a small percentage of administered mAb will localize within tumors (Wu AM and Senter PD (2005) Arming antibodies: prospects and challenges for immunoconjugates. Nat Biotechnol 23: 1137-1146). The rest is catabolized in the liver and several other organs, such as intestine, muscle, and skin (Henderson LA, Baynes JW, and Thorpe SR (1982) Identification of the sites of IgG catabolism in the rat. Arch Biochem Biophys 215: 1-11; Ferl GZ, Kenanova V, Wu AM, and DiStefano JJ III (2006) A two-tiered physiologically based model for dually labeled single-chain Fv-Fc antibody fragments. Mol Cancer Ther 5: 1550-1558.). For an ADC to be effective and relatively nontoxic, the chemotherapeutic drug should remain in stable linkage with the mAb while in circulation but should be efficiently released on internalization into the targeted cells. Thus, there remains a need to overcome this unspecific nature of many chemotherapeutics so that efficacy can be increased and side effects and tolerance in the patient can be reduced. However, it has been reported that folic acid binds to the folate receptor at the cell surface with very high affinity, and is internalized by receptor-mediated endocytosis. Furthermore, the folate receptor is overexpressed by carcinomas of the kidney, brain, lung, esophagus, and breast but has very restricted expression on most normal tissues. (Wu, W. et al. Int J Nanomedicine. 2012; 7: 3487–3502). As such, conjugates bearing folic acid may provide a mechanism to target toxic chemotherapeutics that do not have the deficiencies of ADCs.

[0004] The present invention addresses this need by the creation of extended recombinant polypeptide (XTEN) conjugates that can be purified to homogeneity and that are amenable to chemical conjugation with folic acid and toxins such as MMAE using a wide diversity of conjugation methods. The use of the purified XTEN reagents results in high-yield monodispersed products that have good solubility, that remain stable, and result in enhanced terminal half-life compared to unconjugated products, while the use of folic acid targeting moieties permits more specific delivery of anti-neoplastic agents such as MMAE and related analogs, resulting in increased efficacy and reduced non-specific toxicity.

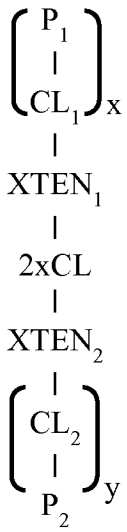
SUMMARY OF THE INVENTION

[0005] The present invention relates, in part, to novel anti-cancer medicaments that have enhanced properties of being specifically targeted to cancer cells by being linked to folate receptor ligands, that have enhanced pharmacokinetic properties that require less frequent dosing, and that are designed to release toxic payload drugs when internalized by the targeted cells. There is provided, in accordance with the present invention, a monodispersed macromolecular conjugate for targeted delivery of a therapeutic agent having a dimeric XTEN polypeptide carrier conjugated with: (a) one or more molecules of a folate targeting agent and (b) one or more molecules of a therapeutic agents, wherein each of the polypeptide carriers includes one or more orthogonal pendant reactive groups for specific attachment of the predetermined folate targeting and/or therapeutic agent attached to the polypeptide carrier, resulting in the XTEN-folate conjugate.

[0006] In one aspect, the invention provides conjugates comprising XTEN engineered for covalent linking to the one or more molecules of a first and a second payload using orthogonal pendant cross-linkers with reactive groups, resulting in XTEN-folate conjugates that comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more molecules of the two types of payloads. It is an object of the present invention to provide such engineered XTEN polypeptides for use in creating conjugates with the payloads as compositions with enhanced pharmaceutical properties, including enhanced pharmacokinetic properties. The invention provides XTEN that are substantially homogeneous in length and sequence that are useful for preparing the conjugates comprising the XTEN linked to one or more payloads such that the resulting XTEN-folate conjugates have a high degree of purity. Such conjugates are useful in preparing monodispersed pharmaceutical compositions that are used in the treatment of cancer subjects for which the one or more payloads have utility in the treatment of such cancer.

[0007] In some embodiments, the invention provides conjugates comprising a first and a second extended recombinant polypeptide (XTEN), wherein each of said first and second XTEN is independently selected from the group consisting of the XTEN sequences set forth in Table 1 and wherein: a) the first and the second XTEN are conjugated to each other by a reaction product of an alkyne and an azide reactant, one of said alkyne and azide reactant being located at the N-terminus of the first XTEN and the other of said alkyne and azide reactant being linked to the N-terminus of the second XTEN, wherein the alkyne and azide reactants are selected from the group consisting of the reactants set forth in Table 9; b)each of the first and the second XTEN comprises one or more cysteine residues and further comprises a first cross-linker conjugated to each cysteine residue of the first XTEN and a second cross-linker conjugated to each cysteine residue of the second XTEN, wherein the first and the second cross-linkers are independently selected from the group consisting of the cross-linkers set forth in Table 8; c) the conjugate comprises a first payload conjugated to each first cross-linker of the first XTEN, wherein the first payload is folic acid; and d) the conjugate comprises a second payload different from the first payload wherein the second payload is conjugated to each second cross-linker of the second XTEN, wherein the second payload is selected from the group consisting of auristatin E, auristatin F, monomethyl auristatin E, monomethyl auristatin F, valine-citrulline-p-aminobenzyloxycarbonyl-monomethyl auristatin E, and valine-citrulline-p-aminobenzyloxycarbonyl-monomethyl auristatin F.

[0008] . In one embodiment of the conjugate composition, the conjugate has the configuration of formula I



I

wherein independently for each occurrence P_1 is the first payload, P_2 is the second payload; CL_1 is the first cross-linker; x is an integer from 1 to about 100, or 1 to about 50, or 1 to about 40, or1 to about 20, or 1 to about 10, or 1 to about 5, or is 3, or is 2, or is 1; CL_2 is the cross-linker; y is an integer from 1 to about 100, or 1 to about 50, or 1 to about 40, or1 to about 20, or 1 to about 10, or 1 to about 5, or is 3, or is 2, or is 1 with the proviso that x + y is $\geq$ 2; 2xCL is the reaction product of the alkyne and azide reactants selected from the group consisting of the reactants set forth in Table 9; $XTEN_1$ is the first XTEN; and $XTEN_2$ is the second XTEN. In one embodiment of the conjugate the first XTEN and the second XTEN each independently have 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% sequence identity or is identical to a sequence selected from the group of sequences set forth in Table 1, when optimally aligned. In another embodiment, the first XTEN of the conjugate has 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% sequence identity or is identical, when optimally aligned, to a sequence selected from the group of sequences consisting of Seg 174, Seg 175, Seg 176, Seg 177, Seg 186, Seg 187, Seg 188, Seg 189, Seg 190, Seg 191, Seg 192, Seg 193, Seg 194, Seg 195, Seg 196, Seg 197, Seg 198, and Seg 199 set forth in Table 1 and the second XTEN has 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% sequence identity or is identical, when optimally aligned, to a different sequence selected from the group consisting of Seg 174, Seg 175, Seg 176, Seg 177, Seg 186, Seg 187, Seg 188, Seg 189, Seg 190, Seg 191, Seg 192, Seg 193, Seg 194, Seg 195, Seg 196, Seg 197, Seg 198, and Seg 199 set forth in Table 1. In another embodiment, the first XTEN and the second XTEN of the conjugate are the identical are each has 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% sequence identity or is identical, when optimally aligned, to a sequence selected from the group consisting of Seg 174, Seg 175, Seg 176, Seg 177, Seg 186, Seg 187, Seg 188, Seg 189, Seg 190, Seg 191, Seg 192, Seg 193, Seg 194, Seg 195, Seg 196, Seg 197, Seg 198, and Seg 199 set forth in Table 1. In another embodiment of the conjugate, the first cross-linker and the second cross linker are independently selected from the group consisting of N-maleimide, N-(5-aminopentyl)maleimide, 6-maleimidocaproic acid, iodoacetyl, pyridyl disulfide and vinyl sulfone, 3-propargyloxypropanoic acid, $(oxyethyl)_n$-acetylene where n is 1-10, dibenzylcyclooctyne (DBCO), cyclooctyne (COT), 3-azide-propionic acid, 6-azide-hexanoic acid, and $(oxyethyl)_n$-azide where n is 1-10. In a preferred embodiment of the conjugate, the first cross-linker is N-(5-aminopentyl)maleimide and the second cross linker is 6-maleimidocaproic acid. In another embodiment of the conjugate the alkyne reactant is selected from the group consisting of 3-propargyloxypropanoic acid NHS ester, acetylene-$(oxyethyl)_n$ -NHS ester where n is 1-10, dibenzylcyclooctyne (DBCO)-NHS ester, DBCO-$(oxyethyl)_n$ -NHS ester where n is 1-10, cyclooctyne (COT)-NHS ester, COT-$(oxyethyl)_n$ -NHS ester where n is 1-10, COT-$(oxyethyl)_n$ -pentafluorophenyl (PFP) ester where n is 1-10, BCOT-NHS ester, BCOT-$(oxyethyl)_n$ -NHS ester where n is 1-10, BCOT-$(oxyethyl)_n$ -pentafluorophenyl (PFP) ester where n is 1-10, acetylene-$(oxyethyl)_n$-maleimide where n is 1-10, DBCO-maleimide, COT-maleimide, BCOT-maleimide and the azide reactant is selected from the group consisting of 3-azide-propionic acid NHS ester, 6-azide-hexanoic acid, NHS ester, 3-azide-propionic acid PFP ester, 6-azide-hexanoic acid PFP ester, azide-$(oxyethyl)_n$ NHS ester where n is 1-10, azide-$(oxyethyl)_n$ -PFP ester where n is 1-10, azide-$(oxyethyl)_n$-maleimide where n is 1-10. In a preferred embodiment of the conjugate, the alkyne reactant is 6-(11,12-didehydrodibenzo[b,f]azocin-5(6H)-yl)-6-oxohexanoic acid N-hydroxysulfosuccinimide ester (DBCO-sulfo-NHS ester) and the azide reactant 1-azido-3,6,9,12-tetraoxapentadecan-15-oic acid N-hydroxysuccinimide ester (azide-(oxyethyl)4-NHS ester).

[0009] In the embodiments of the XTEN-folate conjugate, at least 90%, 91%, 92%, 93%, 94%, or 95% of individual first XTEN molecules in said conjugate have an identical sequence length and at least 90%, 91%, 92%, 93%, 94%, or 95% of individual second XTEN molecules in said conjugate have an identical sequence length.

[0010] In one embodiment of the XTEN-folate conjugate configured as formula I, the first XTEN and the second XTEN are identical and the sequence is Seg 176 set forth in Table 1, and x is 3 and y is 3. In another embodiment of the XTEN-folate conjugate configured as formula I, the first XTEN and the second XTEN are identical, the XTEN sequences are Seg 177 set forth in Table 1, and x is 9 and y is 9.

[0011] In one embodiment of the XTEN-folate conjugates, the second payload is valine-citrulline-*p*-aminobenzyloxycarbonyl-monomethylauristatin E. In another embodiment of the XTEN-folate conjugate, the second payload is valine-citrulline-*p*-aminobenzyloxycarbonyl-monomethylauristatin F.

[0012] In one embodiment, the invention provides an XTEN-folate conjugate having the structure set forth in FIG. 49.

[0013] In another aspect, the invention relates, in part, to XTEN-folate conjugates that are targeted and have enhance toxicity to cells bearing folate receptors. In some embodiments, the invention provides XTEN-payload conjugate wherein treatment of a folate receptor-bearing cell with a conjugate results in greater cytotoxicity compared to treatment with a comparable conjugate lacking folate. In one embodiment of the foregoing, the cytotoxicity is measured in an *in vitro* assay using a folate receptor-bearing cell selected from the group consisting of the cells set forth in Table 24 and the greater cytotoxicity is a reduction in IC50 by at least 100 nM, 200 nM, 300 nM, 400 nM, 500 nM, 600 nM, 700 nM, 800 nM, 900 nM or 1000 nM. In another embodiment, the cytotoxicity is measured in an *in vitro* assay using a KM folate receptor-bearing cell and the greater cytotoxicity is a reduction in IC50 by at least 100 nM, 200 nM, 300 nM, 400 nM, 500 nM, 600 nM, 700 nM, 800 nM, 900 nM or 1000 nM.

[0014] In another aspect, the invention relates, in part, to XTEN-folate conjugates that, when administered to a subject, exhibit a prolonged terminal half-life compared to the payloads not conjugated to XTEN. In one embodiment, administration of a single dose of the XTEN-folate conjugate described herein to a subject results in an increase in the terminal half-life of at least 2-fold, 4-fold, 8-fold, or 10-fold greater compared to the second payload administered singly to the subject using a comparable molar dose. In another embodiment, administration of a single dose of the XTEN-folate conjugate described herein to a subject results in a terminal half-life of at least 24 h, 48 h, 72 h, 96 h, 120 h, 144 h, or 7 days, or 10 days, or 14 days.

[0015] In another aspect, the invention relates, in part, to XTEN-folate conjugates in which the XTEN component is substantially homogeneous in length. In some embodiment, the first XTEN and the second XTEN utilized in the conjugates have the same XTEN sequence and are obtained from a substantially homogenous population of polypeptides prepared by the process comprising: culturing a host cell that comprises a vector encoding the polypeptide in a fermentation reaction under conditions effective to express the polypeptide as a component of a crude expression product of the host cell, wherein the encoded polypeptide comprises the XTEN, a first cleavage sequence, a first affinity tag with binding affinity to a first chromatography substrate, a second cleavage sequence, and a second affinity tag wherein the first and the second cleavage sequences are capable of being cleaved by the same protease and wherein the second affinity tag has binding affinity to a second, different chromatography substrate than the first affinity tag; adsorbing the

polypeptide of the crude expression product onto a first chromatography substrate under conditions effective to capture the first affinity tag onto the first chromatography substrate; eluting the polypeptide; adsorbing the polypeptide onto a second chromatography substrate under conditions effective to capture the second affinity tag onto the second chromatography substrate; eluting the polypeptide; and recovering the polypeptide wherein at least 90%, 91%, 92%, 93%, 94%, or 95% of the polypeptides of the population have an identical sequence length. In one embodiment of the foregoing, the polypeptide has a sequence selected from the group consisting of the sequences set forth in Table 35. In another embodiment of the foregoing, the first chromatography substrate is selected from the group consisting of a HIC substrate, a cation exchange substrate, an anion exchange substrate, and an IMAC substrate. In the embodiments of this paragraph, the first and the second affinity tags are independently selected from the group consisting of the affinity tags of Table 5. In another embodiment of the foregoing, the first chromatography substrate is a cation exchange substrate and the first affinity tag comprises a sequence selected from RPRPRPRPRPRPR and RPRPRPRPRPRPRPRPRPRPR. In another embodiment of the foregoing, the second chromatography substrate is an IMAC substrate and the second affinity tag comprises the sequence selected from HHHHHH and HHHHHHHH. In another embodiment of the foregoing, the process further comprises treating the polypeptide with a protease under conditions effective to cleave the cleavage sequences, thereby releasing the XTEN from the polypeptide; adsorbing the XTEN onto an anion chromatography substrate under conditions effective to capture the XTEN; eluting the XTEN; and recovering the XTEN wherein at least 90%, or at least 91%, or at least 92%, or at least 93%, or at least 94%, or at least 95% of the individual XTEN molecules have an identical sequence length. The anion chromatography substrate can be selected from the group consisting of macrocap Q, capto Q, superQ-650M, and poros D. In the foregoing embodiments of this paragraph, the cleavage sequences are capable of being cleaved by trypsin wherein the cleavage sequence is selected from the group consisting of the sequences set forth in Table 8, and the protease is trypsin. In another embodiment of the foregoing, the process further comprises treating the polypeptide with a protease such as trypsin under conditions effective to cleave the cleavage sequence(s), thereby releasing the XTEN from the polypeptide; adsorbing the protease onto a chromatography substrate under conditions effective to capture the protease but not the XTEN; and recovering the XTEN in the eluate wherein at least 90%, 91%, 92%, 93%, 94%, or 95% of the XTEN have an identical sequence length. In the foregoing embodiment, the chromatography substrate is one or more of a HIC substrate, a cation exchange substrate, and an IMAC substrate.

[0016] In another aspect, the invention provides pharmaceutical compositions comprising the XTEN-folate conjugate composition of any one of the preceding embodiments and a pharmaceutically acceptable carrier. In one embodiment, the pharmaceutical composition has

utility for treatment of a cancer. In another embodiment, the pharmaceutical composition is utilized for use in a pharmaceutical regimen for treatment of a subject with a cancer, said regimen comprising the pharmaceutical composition. In another embodiment, the pharmaceutical regimen further comprises the step of determining the amount of pharmaceutical composition needed to achieve an improvement in a parameter associated with the cancer. In another embodiment, the pharmaceutical regimen further comprises administering the pharmaceutical composition in an effective amount in two or more successive doses to the subject at an effective amount, wherein the administration results in at least a 10%, or 20%, or 30%, or 40%, or 50%, or 60%, or 70%, or 80%, or 90% greater improvement of at least one, two, or three parameters associated with the cancer compared to an untreated subject, wherein the parameters are selected from the group consisting of response rate as defined by the Response Evaluation Criteria in Solid Tumors (RECIST), time-to-progression of the cancer (relapse), discovery of local recurrence, discovery of regional metastasis, discovery of distant metastasis, onset of symptoms, hospitalization, increase in pain medication requirement, requirement of salvage chemotherapy, requirement of salvage surgery, requirement of salvage radiotherapy, time-to-treatment failure, and increased time of survival.

[0017] In another embodiment, the invention provides a XTEN-folate conjugate for use in the preparation of a medicament for treatment of mammalian cells or treatment of a cancer. In another embodiment, the invention provides an article of manufacture comprising a XTEN-folate conjugate as described herein, a container, and a package insert or label indicating that the compound can be used to treat cancer.

[0018] In another embodiment, the invention provides a kit comprising a pharmaceutical composition of the XTEN-folate conjugate, a container, and a package insert or label indicating that the compound can be used to treat cancer.

[0019] In another aspect, the invention relates, in part, to methods of treatment utilizing the XTEN-folate conjugate described herein. The method of the aspect of the invention can be used to kill cancer cells in vitro or in vivo. In one embodiment, the invention provides a method of treating a cancer cell in vitro, comprising administering to a culture of a cancer cell a composition comprising an effective amount of an XTEN-folate conjugate. In another embodiment, the invention provides a method of treating a cancer in a subject, comprising administering to the subject a pharmaceutical composition comprising an effective amount of an XTEN-folate conjugate. In one embodiment of the method, the pharmaceutical composition comprises the conjugate having the structure set forth in FIG. 49. In another embodiment of the method, the cancer is selected from the group consisting of non-small cell lung cancer, mesothelioma, platinum-resistant ovarian cancer, endometrial cancer, adenocarcinoma of the lung, and refractory advanced tumors. In another embodiment of the method, the administration results in at least a

10%, or 20%, or 30%, or 40%, or 50%, or 60%, or 70%, or 80%, or 90% greater improvement of at least one, two, or three parameters associated with the cancer compared to an untreated subject wherein the parameters are selected from the group consisting of response rate as defined by the Response Evaluation Criteria in Solid Tumors (RECIST), time-to-progression of the cancer (relapse), discovery of local recurrence, discovery of regional metastasis, discovery of distant metastasis, onset of symptoms, hospitalization, increase in pain medication requirement, requirement of salvage chemotherapy, requirement of salvage surgery, requirement of salvage radiotherapy, time-to-treatment failure, and increased time of survival.

[0020] It is specifically contemplated that the inventive XTEN-folate conjugates can exhibit one or more or any combination of the improved properties disclosed herein.

INCORPORATION BY REFERENCE

[0021] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0022] The features and advantages of the invention may be further explained by reference to the following detailed description and accompanying drawings that sets forth illustrative embodiments.

[0023] FIG. 1 shows schematics of XTEN suitable for conjugation with payloads. FIG. 1A shows unmodified XTEN. FIG. 1B shows a cysteine-engineered XTEN with an internal cysteine with a thiol side chain; below is an XTEN with an a reactive N-terminal amino group; below is an XTEN with an N-terminal cysteine with a thiol reactive group. FIG. 1C shows cysteine-engineered XTEN with multiple internal cysteines. FIG. 1D shows two variations of a cysteine-engineered XTEN with an internal cysteine with a thiol side chains and a reactive N-terminal amino group and, at the bottom, a shows a cysteine- and lysine-engineered XTEN with internal cysteines and internal lysines.

[0024] FIG. 2 shows a conjugation reaction utilizing NHS-esters and their water soluble analogs sulfo-NHS-esters) reacting with a primary amino group to yield a stable amide XTEN-payload product.

[0025] FIG. 3 shows a conjugation reaction utilizing thiol groups and an N-maleimide. The maleimide group reacts specifically with sulfhydryl groups when the pH of the reaction mixture is between pH 6.5 and 7.5, forming a stable thioether linkage that is not reversible.

[0026] FIG. 4 shows a conjugation reaction utilizing haloacetyls. The most commonly used haloacetyl reagents contain an iodoacetyl group that reacts with sulfhydryl groups at physiological

pH. The reaction of the iodoacetyl group with a sulfhydryl proceeds by nucleophilic substitution of iodine with a thiol producing a stable thioether linkage in the XTEN-payload.

[0027] FIG. 5 shows a conjugation reaction utilizing pyridyl disulfides. Pyridyl disulfides react with sulfhydryl groups over a broad pH range (the optimal pH is 4-5) to form disulfide bonds linking XTEN to payloads.

[0028] FIG. 6 shows a conjugation reaction utilizing zero-length cross-linkers wherein the cross-linkers are used to directly conjugate carboxyl functional groups of one molecule (such as a payload) to the primary amine of another molecule (such as an XTEN).

[0029] FIG. 7 shows a conjugation reaction utilizing the Huisgen 1,3-dipolar cycloaddition of alkynes to azides to form 1,4-disubsituted-1,2,3-triazoles, as shown.

[0030] FIG. 8 shows a conjugation reaction using thio-ene based click chemistry that may proceed by free radical reaction, termed thiol-ene reaction, or anionic reaction, termed thiol Michael addition.

[0031] FIG. 9 shows a conjugation reaction utilizing click chemistry based on reactions between hydrazides and aldehydes, resulting in the illustrated hydrazone linkage in the XTEN-payload.

[0032] FIG. 10 shows a conjugation reaction utilizing traceless Staudinger ligation, like Native Chemical Ligation (NCL), resulting in a native amide bond at the ligation site

[0033] FIG. 11 shows a conjugation reaction utilizing enzymatic ligation. Transglutaminases are enzymes that catalyze the formation of an isopeptide bond between the γ-carboxamide group of glutamine of a payload peptide or protein and the ε-amino group of a lysine in a lysine-engineered XTEN (or an N-terminal amino group), thereby creating inter- or intramolecular cross-links between the XTEN and payload.

[0034] FIG. 12 shows various XTEN-cross-linker precursor segments that are used as reactants to link to payloads or to other XTEN reactants. FIG. 12A is intended to show that the 1B represents the remaining reactive group of the precursors on the right. FIG. 12B shows similar reactive precursors with either multiple (left) or single (right) payload A molecules conjugated to the XTEN.

[0035] FIG. 13 shows exemplary permutations of XTEN-cross-linker precursor segments with two reactive groups of cross-linkers or reactive groups of an incorporated amino acid that are used as reactants to link to payloads or to other XTEN reactants. The 1B and 2B represent reactive groups that will, in other figures, react with a like-numbered reactive group; 1 with 1 and 2 with 2, etc.

[0036] FIG. 14 illustrates the creation of various XTEN precursor segments. FIG. 14A shows the steps of making an XTEN polypeptide, followed by reaction of the N-terminus with the cross-linker with 2B-1A reactive groups, with the 1A reacting with the N-terminal 1B (e.g., an alpha

amino acid) to create the XTEN precursor 2 with the reactive group 2B. FIG. 14B shows the sequential addition of two cross-linkers with 2A reactive groups to 2B reactive groups of the XTEN, resulting in XTEN precursor 4, which is then reacted with a cross-linker at the N-terminus between a reactive 1B and the 1A of a cross-linker, resulting in XTEN precursor 5, with reactive groups 4B and 3B. In such case, the XTEN-precursor 5 then could serve as a reactant with two different payloads or XTEN.

[0037] FIG. 15 illustrates various configurations of bispecific conjugates with two payloads. FIG. 15A illustrates configurations with one molecule each of two payloads, while FIG. 15B illustrates various configurations with multiple copies of one or both payloads.

[0038] FIG. 16 illustrates various examples of conjugates with high valency. Conjugations sites of payloads can grouped (FIG. 26A) or interspersed (FIG. 26B).

[0039] FIG. 17 illustrates the preparation of bispecific conjugates from an XTEN precursor carrying both amino and thiol groups in which many chemistries can be used and the order of payload addition can vary. One can generate linker-conjugates as precursors. FIG. 17A shows the creation of a single XTEN precursor to which two different payloads are attached. FIG. 17B shows a segment approach starting from two XTEN precursor molecules. This approach allows one to conjugate both payloads to XTEN using the same type of linker chemistry. In this case, the figure shows thiol as the group to which payloads are conjugated, and then the N-terminus of each segment is modified with a cross-linker to enable head-to-head segment conjugation, resulting in a dimeric, bispecific conjugate final product.

[0040] FIG. 18 is a schematic flowchart of representative steps in the assembly, production and the evaluation of a XTEN.

[0041] FIG. 19 is a schematic flowchart of representative steps in the assembly of an XTEN polynucleotide construct encoding a fusion protein. Individual oligonucleotides **501** are annealed into sequence motifs **502** such as a 12 amino acid motif ("12-mer"), which is ligated to additional sequence motifs from a library to create a pool that encompasses the desired length of the XTEN **504**, as well as ligated to a smaller concentration of an oligo containing BbsI, and KpnI restriction sites **503.** The resulting pool of ligation products is gel-purified and the band with the desired length of XTEN is cut, resulting in an isolated XTEN gene with a stopper sequence **505**. The XTEN gene is cloned into a stuffer vector. In this case, the vector encodes an optional CBD sequence **506** and a GFP gene **508.** Digestion is then performed with BbsI/HindIII to remove **507** and **508** and place the stop codon. The resulting product is then cloned into a BsaI/HindIII digested vector, resulting in gene **500** encoding an XTEN.

[0042] FIG. 20 is a schematic flowchart of representative steps in the assembly of a gene encoding XTEN, its expression, conjugation with a payload and recovery as an XTEN-payload, and its evaluation as a candidate product.

[0043] FIG. 21 shows generalized XTEN with either N- or C-terminal tags or N- and C-terminal sequences optimized for purification using methods illustrated in FIGS. 22.

[0044] FIG. 22 shows a generalized scheme for purification of XTEN with, in this illusrtrative embodiment, two tags in which a two-step purification method to capture first one tag and then the second can be utilized to remove truncated XTEN from fermentation, resulting in the highly purified target XTEN entity.

[0045] FIG. 23 shows an SDS-PAGE analysis of XTEN constructs with experimental tags after expression in *E.coli* as described in Example 6. Soluble lysates were loaded on the 4-12% Bis-Tris polyacrylamide gel, with amounts loaded per lane equivalent to 36 µl of cell culture suspension. The gel was stained with Coomassie Blue stain using standard methods.

[0046] FIG. 24 shows an SDS-PAGE analysis of the RP11-XTEN-His8 construct expressed in E.coli, as described in Example 6. Heat-treated soluble lysates were loaded on the 4-12% Bis-Tris polyacrylamide gel with amounts equivalent to 1 or 2 µl of cell culture suspension, respectively. The gel was stained with Coomassie Blue stain. The gel demonstrates that essentially all the expressed RP11-XTEN-His8 protein was found in the pelleted fraction.

[0047] FIG. 25 shows an SDS-PAGE analysis of the MacroCap SP purification of RP11-XTEN-His8 polypeptide described in Example 1. Fractions were analyzed by 4-12% SDS-PAGE followed by Coomassie staining.

[0048] FIG. 26 shows an SDS-PAGE analysis of the IMAC purification of the RP11-XTEN-His8 polypeptide described in Example 1. Fractions were analyzed by 4-12% SDS-PAGE followed by Coomassie staining.

[0049] FIG. 27 shows an SDS-PAGE analysis of the trypsin digestion of RP11-XTEN-His8 protein purified by two chromatographic steps (SP + IMAC) described in Example 1. Preparations were analyzed by 4-12% SDS-PAGE followed by Coomassie staining (FIG. 27A) and silver staining (FIG. 27B).

[0050] FIG. 28 shows the pharmacokinetic profile (plasma concentrations) in cynomolgus monkeys after single doses of different compositions of GFP linked to unstructured polypeptides of varying length, administered either subcutaneously or intravenously, as described in Example 35. The compositions were GFP-L288, GFP-L576, GFP-XTEN_AF576, GFP-Y576 and XTEN_AD836-GFP. Blood samples were analyzed at various times after injection and the concentration of GFP in plasma was measured by ELISA using a polyclonal antibody against GFP for capture and a biotinylated preparation of the same polyclonal antibody for detection. Results are presented as the plasma concentration versus time (h) after dosing and show, in particular, a considerable increase in half-life for the XTEN_AD836-GFP, the composition with the longest sequence length of XTEN. The construct with the shortest sequence length, the GFP-L288 had the shortest half-life.

[0051] FIG. 29 shows an SDS-PAGE gel of samples from a stability study of the fusion protein of XTEN_AE864 fused to the N-terminus of GFP. The GFP-XTEN was incubated in cynomolgus plasma and rat kidney lysate for up to 7 days at 37°C, as described in Example 28. In addition, GFP-XTEN administered to cynomolgus monkeys was also assessed. Samples were withdrawn at 0, 1 and 7 days and analyzed by SDS PAGE followed by detection using Western analysis and detection with antibodies against GFP.

[0052] FIG. 30 shows the near UV circular dichroism spectrum of Ex4-XTEN_AE864, performed as described in Example 29.

[0053] FIG. 31 shows results of a size exclusion chromatography analysis of glucagon-XTEN construct samples measured against protein standards of known molecular weight, with the graph output as absorbance versus retention volume, as described in Example 31. The glucagon-XTEN constructs are 1) glucagon-Y288; 2) glucagonY-144; 3) glucagon-Y72; and 4) glucagon-Y36.

[0054] FIG. 32 is a schematic of the logic flow chart of the algorithm SegScore (Example 33). In the figure the following legend applies: i, j - counters used in the control loops that run through the entire sequence; HitCount- this variable is a counter that keeps track of how many times a subsequence encounters an identical subsequence in a block; SubSeqX - this variable holds the subsequence that is being checked for redundancy; SubSeqY - this variable holds the subsequence that the SubSeqX is checked against; BlockLen - this variable holds the user determined length of the block; SegLen - this variable holds the length of a segment. The program is hardcoded to generate scores for subsequences of lengths 3, 4, 5, 6, 7, 8, 9, and 10; Block - this variable holds a string of length BlockLen. The string is composed of letters from an input XTEN sequence and is determined by the position of the i counter; SubSeqList - this is a list that holds all of the generated subsequence scores.

[0055] FIG. 33 depicts the application of the algorithm SegScore to a hypothetical XTEN of 11 amino acids in order to determine the repetitiveness. An XTEN sequence consisting of N amino acids is divided into N-S+1 subsequences of length S (S=3 in this case). A pair-wise comparison of all subsequences is performed and the average number of identical subsequences is calculated to result, in this case, in a subsequence score of 1.89.

[0056] FIG. 34 provides the results of the assay to measure the fluorescence signal of RP11 clones pSD0107 to pSD0118), as described in Example 6. One positive control (pLCW970) and two negative controls (pBr322 and pLCW970+10 mM phosphate) were included. The GFP expression level was measured using samples from 2-3 shake flasks per construct.

[0057] FIG. 35 shows the screening results of libraries LCW1157-1159. FIG. 35A-C provides the fluorescence histograms of LCW1157-1159, showing the number of colonies identified for each fluorescence signal region, as described in Example 6. The average fluorescence reading of the negative control (black arrow) and positive pSD0116 (white arrow) are marked in the figures.

FIG. 35D-F provides the correlation between the fluorescence reading in the original test and the retest of the select clones.

[0058] FIG. 36 shows results of the SDS-PAGE analysis of the top 8 expression construct products and controls under unreduced conditions, as described in Example 6. The desired full length protein end product RP11-XTEN-GFP is indicated by an arrow, and the higher band is the dimer of the protein. Lanes: 1-8: top 8 expression constructs (expression level from high to low, based on fluorescence reading of the retests), 1. LCW1159.004, 2. LCW1159.006, 3. LCW1158.004, 4. LCW1157.040, 5. LCW1158.003, 6. LCW1157.039, 7. LCW1157.025, 8. LCW1157.038; C1-C3: Controls: C1. pSD0114, C2. pSD0116, C3. pCW1146 (Negative control).

[0059] FIG. 37 shows the SDS-PAGE evaluation of the MacroCap SP capture efficiency for the top 4 expression construct products under non-reducing conditions, as described in Example 6. Lanes 1-4: load, flow through, wash and elution of LCW1159.004, 2. Lanes 5-8: load, flow through, wash and elution of LCW1159.006. Lanes 9-12: load, flow through, wash and elution of LCW1158.004. 13-16: load, flow through, wash and elution of LCW1157.040. Lanes 17-20 1-4: load, flow through, wash and elution of negative control. Unmarked lanes are molecular weight standards.

[0060] FIG. 38 shows the summary of library LCW1163 screening results with a comparison of the fluorescence signal of the top 4 expression products and the controls in the retests, as described in Example 6. Each sample had 4 replicates, represented by 4 individual dots in the figure.

[0061] FIG. 39 shows the summary of library LCW1160 screening results, as described in Example 6. Fluorescence histogram of LCW1157-1159, showing the number of colonies identified for each fluorescence signal region; average fluorescence reading of negative control (black arrow), pSD0116 (white arrow), and LCW1159.004 (high expression candidates from screening LCW1157-1159, grey arrow) were marked in the figures.

[0062] FIG. 40 shows 4-12% SDS-PAGE/silver staining analysis of MacroCap Q fractions as described in Example 2. FIG. 40A: Batch 2, lane 1: molecular weight standard; lanes 2-5: MacroCap Q flow through fractions 1-4, respectively; lanes 6-16: MacroCap Q elution fractions 1-11, respectively. FIG. 40B: Batch 1, lane 1: molecular weight standard; lanes 2-6: MacroCap Q flow through fractions 1-5, respectively; lanes 7-16: MacroCap Q elution fractions 1-10, respectively.

[0063] FIG. 41 shows results of analyses of reaction mixtures from the preparation of conjugates to 1xAzide,3xMMAE-XTEN analyzed by C18-RP-HPLC and mass spectroscopy, as described in Example 5. FIG. 41A is analysis of the initial 1xAmino,3xThiol-XTEN reactant. FIG. 41B is analysis of the protein modification with MMAE-Maleimide, showing the mass increase corresponding to modifications of three cysteines with MMAE-Mal. FIG. 41C shows the

analysis of the protein modification with Azide-PEG4-NHS ester, with mass increases corresponding to the single addition of the azide-PEG4 moiety.

[0064] FIG. 42 shows the results of screening libraries LCW1171, 1172, 1203, and 1204, as described in Example 2. FIG. 42A-D: Fluorescence histogram of LCW1171, 1172, 1203, 1204, showing the number of colonies identified for each fluorescence signal region; average fluorescence reading of negative control (black arrow) and pSD0116 (white arrow) when screening LCW1171-1172 were marked in the FIGS. 94A and B; average fluorescence reading of negative control (black arrow), pSD0116 (white arrow), and CBD control (grey arrow) when screening LCW1203-1204 are marked in FIGS. 94C and D.

[0065] FIG. 43 shows the results of screening libraries LCW1208-1210, as described in Example 6. FIGS. 95A-C: Fluorescence histograms of LCW1208-1210, showing the number of colonies identified for each fluorescence signal region; average fluorescence reading of negative control (black arrow) and CBD control (grey arrow) are marked in the figures.

[0066] FIG. 44 illustrated the production of XTEN segments from a precursor that contains three repeat copies of XTEN of identical length and sequence. In FIG. 44A, the XTEN precursor comprises three identical copies of XTEN that are flanked by identical protease cleavage sites. In FIG. 44B, the XTEN precursor further comprises N- and C- terminal affinity purification tags to facilitate purification of full-length precursor molecules. Following purification of the precursor it is cleaved by protease that acts on all the incorporated cleavage sequences to release the tags from the XTEN, which is followed by purification to separate the individual units of XTEN, facilitating the high-yield production of XTENs with short and intermediate lengths from long-chain precursor molecules.

[0067] FIG. 45 shows results of analyses of reaction mixtures from the preparation of conjugates to 1xDBCO,3xFA(γ)-XTEN analyzed by C18-RP-HPLC and mass spectroscopy, as described in Example 18. FIG. 45A is analysis of the initial 1xAmino,3xThiol-XTEN reactant. FIG. 45B is analysis of the protein modification with Folate-gamma-Maleimide, showing the mass increase corresponding to modifications of three cysteines with FA(γ)-Mal. FIG. 45C shows the analysis of the protein modification with DBCO-sulfo-NHS ester, with mass increases corresponding to the single addition of the DBCO moiety.

[0068] FIG. 46 shows C4 RP-HPLC analyses of the click conjugate reactants and product 3xFA(γ),3xMMAE-XTEN, as described in Example 19. (1) 1xDBCO,3xFA(γ)-XTEN; (2) 1xAzide,3xMMAE-XTEN; (3) products of click chemistry reaction.

[0069] FIG. 47 shows analyses of final 3xFA(γ),3xMMAE-XTEN product purified by preparative RP-HPLC, as described in Example 19. FIG. 47A shows size exclusion chromatography analysis (Phenomenex BioSep-SEC-s4000 600 x 7.80mm column, 50mM Sodium Phosphate pH 6.5, 300mM NaCl buffer, flow rate 0.5ml/min, isocratic elution 70min).

FIG. 47B shows RP-HPLC analysis (Phenomenex Jupiter C18 5μM 300Å 150 x 4.60mm column, Buffer A: 0.1% TFA in H2O, Buffer B: 0.1% TFA in CAN, flow rate 1ml/min, gradient 5% to 50%B in 45min). FIG. 3C shows ESI-MS analysis (QSTAR-XL, calculated MW 85,085.4 Da, experimental MW 85,091 Da).

[0070] FIG. 48 shows results of a killing assay demonstrating selective cytotoxicity of 3xFA(γ),3xMMAE-XTEN on KB cells, as described in Example 69. The inhibitory dose response curves are shown for the groups of free MMAE (filled circles); 3xMMAE-XTEN (filled, inverted triangles) and 3xFA(γ),3xMMAE-XTEN in the presence (filled triangles) and absence (filled squares) of folic acid competitor on KB cells.

[0071] FIG. 49 shows the structure of the XTEN-payload conjugate 3xFA(γ),3xMMAE-XTEN. FIG. 49A shows the two XTEN linked by the reaction of the azide 1-azido-3,6,9,12-tetraoxapentadecan-15-oic acid, N-hydroxysuccinimide ester and the alkyne 6-(11,12-didehydrodibenzo[b,f]azocin-5(6H)-yl)-6-oxohexanoic acid, N-hydroxysuccinimide (or N-hydroxysulfosuccinimide) ester. FIG. 49B shows the X residue of Cys modified with folate-g-aminopentyl-maleimide. FIG. 49C shows the Z residue of Cys modified with maleimidocaproyl-valine-citrulline-*p*-aminobenzyloxycarbonyl-monomethylauristatin.

[0072] FIG. 50 shows an example of an XTEN-folate conjugate comprising folate targeting moieties and toxin payloads that exert selective action on a target cell, such as a tumor cell. The particular design of the dimeric XTEN conjugate comprises folate and MMAE. This conjugate binds to the folate-receptor on that is over-expressed on many cancer cells. Receptor binding results in internalization followed by proteolytic break down and the intracellular liberation of MMAE, which is toxic to the cell.

DETAILED DESCRIPTION OF THE INVENTION

[0073] Before the embodiments of the invention are described, it is to be understood that such embodiments are provided by way of example only, and that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention.

[0074] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. In case of conflict, the patent specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be

limiting. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention.

DEFINITIONS

[0075] As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

[0076] As used in the specification and claims, the singular forms "a", "an" and "the" include plural references unless the context clearly dictates otherwise. For example, the term "a cell" includes a plurality of cells, including mixtures thereof.

[0077] The terms "polypeptide", "peptide", and "protein" are used interchangeably herein to refer to polymers of amino acids of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non-amino acids. The terms also encompass an amino acid polymer that has been modified, for example, by disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation, such as conjugation with a labeling component.

[0078] As used herein the term "amino acid" refers to either natural amino acids, including L optical isomers. Standard single or three letter codes are used to designate amino acids.

[0079] The term "natural L-amino acid" means the L optical isomer forms of glycine (G), proline (P), alanine (A), valine (V), leucine (L), isoleucine (I), methionine (M), cysteine (C), phenylalanine (F), tyrosine (Y), tryptophan (W), histidine (H), lysine (K), arginine (R), glutamine (Q), asparagine (N), glutamic acid (E), aspartic acid (D), serine (S), and threonine (T).

[0080] The term "non-naturally occurring," as applied to sequences and as used herein, means polypeptide or polynucleotide sequences that do not have a counterpart to, are not complementary to, or do not have a high degree of homology with a wild-type or naturally-occurring sequence found in a mammal. For example, a non-naturally occurring polypeptide or fragment may share no more than 99%, 98%, 95%, 90%, 80%, 70%, 60%, 50% or even less amino acid sequence identity as compared to a natural sequence when suitably aligned.

[0081] The terms "hydrophilic" and "hydrophobic" refer to the degree of affinity that a substance has with water. 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 characterized based on their hydrophobicity. A number of scales have been developed. An example is a scale developed by Levitt, M, et al., J Mol Biol (1976) 104:59, which is listed in Hopp, TP, et al., Proc Natl Acad Sci U S A (1981) 78:3824. Examples of "hydrophilic amino acids" are arginine, lysine, threonine, alanine, asparagine, and glutamine. Of particular interest are the hydrophilic amino acids aspartate,

glutamate, and serine, and glycine. Examples of "hydrophobic amino acids" are tryptophan, tyrosine, phenylalanine, methionine, leucine, isoleucine, and valine.

[0082] A "fragment" when applied to a biologically active protein, is a truncated form of a the biologically active protein that retains at least a portion of the therapeutic and/or biological activity. A "variant," when applied to a biologically active protein is a protein with sequence homology to the native biologically active protein that retains at least a portion of the therapeutic and/or biological activity of the biologically active protein. For example, a variant protein may share at least 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% amino acid sequence identity compared with the reference biologically active protein. As used herein, the term "biologically active protein variant" includes proteins modified deliberately, as for example, by site directed mutagenesis, synthesis of the encoding gene, insertions, or accidentally through mutations and that retain activity.

[0083] The term "sequence variant" means polypeptides that have been modified compared to their native or original sequence by one or more amino acid insertions, deletions, or substitutions. Insertions may be located at either or both termini of the protein, and/or may be positioned within internal regions of the amino acid sequence. A non-limiting example is insertion of an XTEN sequence within the sequence of the biologically-active payload protein. Another non-limiting example is substitution of an amino acid in an XTEN with a different amino acid. In deletion variants, one or more amino acid residues in a polypeptide as described herein are removed. Deletion variants, therefore, include all fragments of a payload polypeptide sequence. In substitution variants, one or more amino acid residues of a polypeptide are removed and replaced with alternative residues. In one aspect, the substitutions are conservative in nature and conservative substitutions of this type are well known in the art.

[0084] The term "moiety" means a component of a larger composition or that is intended to be incorporated into a larger composition, such as a functional group of a drug molecule or a targeting peptide joined to a larger polypeptide.

[0085] The term "XTEN release site" refers to a cleavage sequence in XTEN-payload that can be recognized and cleaved by a protease, effecting release of an XTEN or a portion of an XTEN from the XTEN-payload polypeptide. As used herein, "mammalian protease" means a protease that normally exists in the body fluids, cells or tissues of a mammal. XTEN release sites can be engineered to be cleaved by various mammalian proteases (a.k.a. "XTEN release proteases") such as trypsin, FXIa, FXIIa, kallikrein, FVIIIa, FVIIIa, FXa, FIIa (thrombin), Elastase-2, MMP-12, MMP13, MMP-17, MMP-20, or any protease that is present in a subject. Other equivalent proteases (endogenous or exogenous) that are capable of recognizing a defined cleavage site can be utilized. The cleavage sites can be adjusted and tailored to the protease utilized.

[0086] "Activity" as applied to form(s) of a XTEN-folate composition provided herein, refers to an action or effect, including but not limited to receptor binding, antagonist activity, agonist activity, a cellular or physiologic response, or an effect generally known in the art for the payload, whether measured by an *in vitro*, *ex vivo* or *in vivo* assay or a clinical effect.

[0087] As used herein, the term "ELISA" refers to an enzyme-linked immunosorbent assay as described herein or as otherwise known in the art.

[0088] A "host cell" includes an individual cell or cell culture which can be or has been a recipient for the subject vectors such as those described herein. Host cells include progeny of a single host cell. The progeny may not necessarily be completely identical (in morphology or in genomic of total DNA complement) to the original parent cell due to natural, accidental, or deliberate mutation. A host cell includes cells transfected in vivo with a vector of this invention.

[0089] "Isolated" when used to describe the various polypeptides disclosed herein, means polypeptide that has been identified and separated and/or recovered from a component of its natural environment. Contaminant components of its natural environment are materials that would typically interfere with diagnostic or therapeutic uses for the polypeptide, and may include enzymes, hormones, and other proteinaceous or non-proteinaceous solutes. As is apparent to those of skill in the art, a non-naturally occurring polynucleotide, peptide, polypeptide, protein, antibody, or fragments thereof, does not require "isolation" to distinguish it from its naturally occurring counterpart. In addition, a "concentrated", "separated" or "diluted" polynucleotide, peptide, polypeptide, protein, antibody, or fragments thereof, is distinguishable from its naturally occurring counterpart in that the concentration or number of molecules per volume is generally greater than that of its naturally occurring counterpart. In general, a polypeptide made by recombinant means and expressed in a host cell is considered to be "isolated."

[0090] An "isolated" nucleic acid is a nucleic acid molecule that is identified and separated from at least one contaminant nucleic acid molecule with which it is ordinarily associated in the natural source of the polypeptide-encoding nucleic acid. For example, an isolated polypeptide-encoding nucleic acid molecule is other than in the form or setting in which it is found in nature. Isolated polypeptide-encoding nucleic acid molecules therefore are distinguished from the specific polypeptide-encoding nucleic acid molecule as it exists in natural cells. However, an isolated polypeptide-encoding nucleic acid molecule includes polypeptide-encoding nucleic acid molecules contained in cells that ordinarily express the polypeptide where, for example, the nucleic acid molecule is in a chromosomal or extra-chromosomal location different from that of natural cells.

[0091] A "chimeric" protein contains at least one fusion polypeptide comprising at least one region in a different position in the sequence than that which occurs in nature. The regions may normally exist in separate proteins and are brought together in the fusion polypeptide; or they may

normally exist in the same protein but are placed in a new arrangement in the fusion polypeptide. A chimeric protein may be created, for example, by chemical synthesis, or by creating and translating a polynucleotide in which the peptide regions are encoded in the desired relationship.

[0092] "Fused," and "fusion" are used interchangeably herein, and refers to the joining together of two or more peptide or polypeptide sequences by recombinant means.

[0093] "Operably linked" means that the DNA sequences being linked are contiguous, and in reading phase or in-frame. An "in-frame fusion" refers to the joining of two or more open reading frames (ORFs) to form a continuous longer ORF, in a manner that maintains the correct reading frame of the original ORFs. For example, a promoter or enhancer is operably linked to a coding sequence for a polypeptide if it affects the transcription of the polypeptide sequence. Thus, the resulting recombinant fusion protein is a single protein containing two or more segments that correspond to polypeptides encoded by the original ORFs (which segments are not normally so joined in nature).

[0094] "Crosslinking," "conjugating," "link," "linking," and "joined to" are used interchangeably herein, and refer to the covalent joining of two different molecules by a chemical reaction. The crosslinking can occur in one or more chemical reactions, as described more fully, below.

[0095] The term "conjugation partner" as used herein, refers to the individual components that can be linked or are linked in a conjugation reaction.

[0096] The term "conjugate" is intended to refer to the heterogeneous molecule formed as a result of covalent linking of conjugation partners one to another, e.g., a biologically active payload covalently linked to a XTEN molecule or a cross-linker covalently linked to a reactive XTEN.

[0097] "Cross-linker" and "linker" and "cross-linking agent" are used interchangeably and in their broadest context to mean a chemical entity used to covalently join two or more entities. For example, a cross-linker joins two, three, four or more XTEN, or joins a payload to an XTEN, as the entities are defined herein. A cross-linker includes, but is not limited to, the reaction product of small molecule zero-length, homo- or hetero-bifunctional, and multifunctional cross-linker compounds, the reaction product of two click-chemistry reactants. It will be understood by one of skill in the art that a cross-linker can refer to the covalently-bound reaction product remaining after the crosslinking of the reactants. The cross-linker can also comprise one or more reactants which have not yet reacted but which are capable to react with another entity.

[0098] In the context of polypeptides, a "linear sequence" or a "sequence" is an order of amino acids in a polypeptide in an amino to carboxyl terminus direction in which residues that neighbor each other in the sequence are contiguous in the primary structure of the polypeptide. A

"partial sequence" is a linear sequence of part of a polypeptide that is known to comprise additional residues in one or both directions.

[0099] "Heterologous" means derived from a genotypically distinct entity from the rest of the entity to which it is being compared. For example, a glycine rich sequence removed from its native coding sequence and operatively linked to a coding sequence other than the native sequence is a heterologous glycine rich sequence. The term "heterologous" as applied to a polynucleotide, a polypeptide, means that the polynucleotide or polypeptide is derived from a genotypically distinct entity from that of the rest of the entity to which it is being compared.

[00100] The terms "polynucleotides", "nucleic acids", "nucleotides" and "oligonucleotides" are used interchangeably. They refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three-dimensional structure, and may perform any function, known or unknown. The following are non-limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component.

[00101] The term "complement of a polynucleotide" denotes a polynucleotide molecule having a complementary base sequence and reverse orientation as compared to a reference sequence, such that it could hybridize with a reference sequence with complete fidelity.

[00102] "Recombinant" as applied to a polynucleotide means that the polynucleotide is the product of various combinations of recombination steps which may include cloning, restriction and/or ligation steps, and other procedures that result in expression of a recombinant protein in a host cell.

[00103] The terms "gene" and "gene fragment" are used interchangeably herein. They refer to a polynucleotide containing at least one open reading frame that is capable of encoding a particular protein after being transcribed and translated. A gene or gene fragment may be genomic or cDNA, as long as the polynucleotide contains at least one open reading frame, which may cover the entire coding region or a segment thereof. A "fusion gene" is a gene composed of at least two heterologous polynucleotides that are linked together.

[00104] "Homology" or "homologous" or "sequence identity" refers to sequence similarity or interchangeability between two or more polynucleotide sequences or between two or more

polypeptide sequences. When using a program such as BestFit to determine sequence identity, similarity or homology between two different amino acid sequences, the default settings may be used, or an appropriate scoring matrix, such as blosum45 or blosum80, may be selected to optimize identity, similarity or homology scores. Preferably, polynucleotides that are homologous are those which hybridize under stringent conditions as defined herein and have at least 70%, preferably at least 80%, more preferably at least 90%, more preferably 95%, more preferably 97%, more preferably 98%, and even more preferably 99% sequence identity compared to those sequences. Polypeptides that are homologous preferably have sequence identities that are at least 70%, preferably at least 80%, even more preferably at least 90%, even more preferably at least 95-99% identical.

[00105] "Ligation" as applied to polynucleic acids refers to the process of forming phosphodiester bonds between two nucleic acid fragments or genes, linking them together. To ligate the DNA fragments or genes together, the ends of the DNA must be compatible with each other. In some cases, the ends will be directly compatible after endonuclease digestion. However, it may be necessary to first convert the staggered ends commonly produced after endonuclease digestion to blunt ends to make them compatible for ligation.

[00106] The terms "stringent conditions" or "stringent hybridization conditions" includes reference to conditions under which a polynucleotide will hybridize to its target sequence, to a detectably greater degree than other sequences (e.g., at least 2-fold over background). Generally, stringency of hybridization is expressed, in part, with reference to the temperature and salt concentration under which the wash step is carried out. Typically, stringent conditions will be those in which the salt concentration is less than about 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short polynucleotides (e.g., 10 to 50 nucleotides) and at least about 60°C for long polynucleotides (e.g., greater than 50 nucleotides)—for example, "stringent conditions" can include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and three washes for 15 min each in 0.1×SSC/1% SDS at 60°C to 65°C. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2×SSC, with SDS being present at about 0.1%. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating Tm and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. *et al.*, "Molecular Cloning: A Laboratory Manual," 3rd edition, Cold Spring Harbor Laboratory Press, 2001. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml.

Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art.

[00107] The terms "percent identity," percentage of sequence identity," and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences. Percent identity may be measured over the length of an entire defined polynucleotide sequence, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polynucleotide sequence, for instance, a fragment of at least 45, at least 60, at least 90, at least 120, at least 150, at least 210 or at least 450 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured. The percentage of sequence identity is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of matched positions (at which identical residues occur in both polypeptide sequences), dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. When sequences of different length are to be compared, the shortest sequence defines the length of the window of comparison. Conservative substitutions are not considered when calculating sequence identity.

[00108] "Percent (%) sequence identity," with respect to the polypeptide sequences identified herein, is defined as the percentage of amino acid residues in a query sequence that are identical with the amino acid residues of a second, reference polypeptide sequence or a portion thereof, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity, thereby resulting in optimal alignment. Alignment for purposes of determining percent amino acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as BLAST, BLAST-2, ALIGN or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve optimal alignment over the full length of the sequences being compared. Percent identity may be measured over the length of an entire defined polypeptide sequence, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150

contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

[00109] "Repetitiveness" used in the context of polynucleotide sequences refers to the degree of internal homology in the sequence such as, for example, the frequency of identical nucleotide sequences of a given length. Repetitiveness can, for example, be measured by analyzing the frequency of identical sequences.

[00110] A "vector" is a nucleic acid molecule, preferably self-replicating in an appropriate host, which transfers an inserted nucleic acid molecule into and/or between host cells. The term includes vectors that function primarily for insertion of DNA or RNA into a cell, replication of vectors that function primarily for the replication of DNA or RNA, and expression vectors that function for transcription and/or translation of the DNA or RNA. Also included are vectors that provide more than one of the above functions. An "expression vector" is a polynucleotide which, when introduced into an appropriate host cell, can be transcribed and translated into a polypeptide(s). An "expression system" usually connotes a suitable host cell comprised of an expression vector that can function to yield a desired expression product.

[00111] "Serum degradation resistance," as applied to a polypeptide, refers to the ability of the polypeptides to withstand degradation in blood or components thereof, which typically involves proteases in the serum or plasma. The serum degradation resistance can be measured by combining the protein with human (or mouse, rat, monkey, as appropriate) serum or plasma, typically for a range of days (e.g. 0.25, 0.5, 1, 2, 4, 8, 16 days), typically at about 37°C. The samples for these time points can be run on a Western blot assay and the protein is detected with an antibody. The antibody can be to a tag in the protein. If the protein shows a single band on the western, where the protein's size is identical to that of the injected protein, then no degradation has occurred. In this exemplary method, the time point where 50% of the protein is degraded, as judged by Western blots or equivalent techniques, is the serum degradation half-life or "serum half-life" of the protein.

[00112] The terms "$t_{1/2}$", "half-life", "terminal half-life", "elimination half-life" and "circulating half-life" are used interchangeably herein and, as used herein means the terminal half-life calculated as $\ln(2)/K_{el}$. K_{el} is the terminal elimination rate constant calculated by linear regression of the terminal linear portion of the log concentration vs. time curve. Half-life typically refers to the time required for half the quantity of an administered substance deposited in a living organism to be metabolized or eliminated by normal biological processes.

[00113] "Active clearance" means the mechanisms by which a protein is removed from the circulation other than by filtration, and which includes removal from the circulation mediated by cells, receptors, metabolism, or degradation of the protein.

[00114] "Apparent molecular weight factor" and "apparent molecular weight" are related terms referring to a measure of the relative increase or decrease in apparent molecular weight exhibited by a particular amino acid or polypeptide sequence. The apparent molecular weight is determined using size exclusion chromatography (SEC) or similar methods by comparing to globular protein standards, and is measured in "apparent kD" units. The apparent molecular weight factor is the ratio between the apparent molecular weight and the actual molecular weight; the latter predicted by adding, based on amino acid composition, the calculated molecular weight of each type of amino acid in the composition or by estimation from comparison to molecular weight standards in an SDS electrophoresis gel. Determination of both the apparent molecular weight and apparent molecular weight factor for representative proteins is described in the Examples.

[00115] The terms "hydrodynamic radius" or "Stokes radius" is the effective radius (R_h 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 the embodiments of the invention, the hydrodynamic radius measurements of the XTEN polypeptides correlate with the "apparent molecular weight factor" which is a more intuitive measure. The "hydrodynamic radius" of a protein affects its rate of diffusion in aqueous solution as well as its ability to migrate in gels of macromolecules. The hydrodynamic radius of a protein is determined by its molecular weight as well as by its structure, including shape and compactness. 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. Patent Nos. 6,406,632 and 7,294,513. Most proteins have globular structure, which is the most compact three-dimensional structure a protein can have with the smallest hydrodynamic radius. Some proteins adopt a random and open, unstructured, or 'linear' conformation and as a result have a much larger hydrodynamic radius compared to typical globular proteins of similar molecular weight.

[00116] "Physiological conditions" refers to a set of conditions in a living host as well as *in vitro* conditions, including temperature, salt concentration, pH, that mimic those conditions of a living subject. A host of physiologically relevant conditions for use in *in vitro* assays have been established. Generally, a physiological buffer contains a physiological concentration of salt and is adjusted to a neutral pH ranging from about 6.5 to about 7.8, and preferably from about 7.0 to about 7.5. A variety of physiological buffers are listed in Sambrook et al. (2001). Physiologically relevant temperature ranges from about 25^0C to about 38^0C, and preferably from about 35^0C to about 37^0C.

[00117] A "single atom residue of a payload" means the atom of a payload that is chemically linked to XTEN after reaction with the subject XTEN or XTEN-linker compositions; typically a sulfur, an oxygen, a nitrogen, or a carbon atom. For example, an atom residue of a payload could be a sulfur residue of a cysteine thiol reactive group in a payload, a nitrogen molecule of an amino

reactive group of a peptide or polypeptide or small molecule payload, a carbon or oxygen residue or a reactive carboxyl or aldehyde group of a peptide, protein or a small molecule or synthetic, organic drug.

[00118] A "reactive group" is a chemical structure that can be coupled to a second reactive group. Examples of reactive groups are amino groups, carboxyl groups, sulfhydryl groups, hydroxyl groups, aldehyde groups, azide groups. Some reactive groups can be activated to facilitate conjugation with a second reactive group, either directly or through a cross-linker. As used herein, a reactive group can be a part of an XTEN, a cross-linker, an azide/alkyne click-chemistry reactant, or a payload so long as it has the ability to participate in a chemical reaction. Once reacted, a conjugation bond links the residues of the payload or cross-linker or XTEN reactants.

[00119] "Controlled release agent", "slow release agent", "depot formulation" and "sustained release agent" are used interchangeably to refer to an agent capable of extending the duration of release of a polypeptide of the invention relative to the duration of release when the polypeptide is administered in the absence of agent. Different embodiments of the present invention may have different release rates, resulting in different therapeutic amounts.

[00120] The term "payload" as used herein refers to a moiety as described herein which has a biological, pharmacological or therapeutic activity or beneficial effect when administered in a subject or that can be demonstrated *in vitro*. Payload also includes a molecule that can be used for imaging or *in vivo* diagnostic purposes.

[00121] The term "antagonist", as used herein, includes any molecule that partially or fully blocks, inhibits, or neutralizes a biological activity of a native polypeptide disclosed herein. Methods for identifying antagonists of a polypeptide may comprise contacting a native polypeptide with a candidate antagonist molecule and measuring a detectable change in one or more biological activities normally associated with the native polypeptide. In the context of the present invention, antagonists may include proteins, nucleic acids, carbohydrates, antibodies or any other molecules that decrease the effect of a biologically active protein.

[00122] The term "agonist" is used in the broadest sense and includes any molecule that mimics a biological activity of a native polypeptide disclosed herein. Suitable agonist molecules specifically include agonist antibodies or antibody fragments, fragments or amino acid sequence variants of native polypeptides, peptides, small organic molecules, etc. Methods for identifying agonists of a native polypeptide may comprise contacting a native polypeptide with a candidate agonist molecule and measuring a detectable change in one or more biological activities normally associated with the native polypeptide.

[00123] "Inhibition constant", or "Ki", are used interchangeably and mean the dissociation constant of the enzyme-inhibitor complex, or the reciprocal of the binding affinity of the inhibitor to the enzyme.

[00124] A "defined medium" refers to a medium comprising nutritional and hormonal requirements necessary for the survival and/or growth of the cells in culture such that the components of the medium are known. Traditionally, the defined medium has been formulated by the addition of nutritional and growth factors necessary for growth and/or survival. Typically, the defined medium provides at least one component from one or more of the following categories: a) all essential amino acids, and usually the basic set of twenty amino acids plus cysteine; b) an energy source, usually in the form of a carbohydrate such as glucose; c) vitamins and/or other organic compounds required at low concentrations; d) free fatty acids; and e) trace elements, where trace elements are defined as inorganic compounds or naturally occurring elements that are typically required at very low concentrations, usually in the micromolar range. The defined medium may also optionally be supplemented with one or more components from any of the following categories: a) one or more mitogenic agents; b) salts and buffers as, for example, calcium, magnesium, and phosphate; c) nucleosides and bases such as, for example, adenosine and thymidine, hypoxanthine; and d) protein and tissue hydrolysates.

[00125] As used herein, "treat" or "treating," or "palliating" or "ameliorating" are used interchangeably and mean administering a drug or a biologic to achieve a therapeutic benefit, to cure or reduce the severity of an existing condition, or to achieve a prophylactic benefit, prevent or reduce the likelihood of onset or severity the occurrence of a condition. By therapeutic benefit is meant eradication or amelioration of the underlying condition being treated or one or more of the physiological symptoms associated with the underlying condition such that an improvement is observed in the subject, notwithstanding that the subject may still be afflicted with the underlying condition.

[00126] A "therapeutic effect" or "therapeutic benefit," as used herein, refers to a physiologic effect, including but not limited to the mitigation, amelioration, or prevention of disease in humans or other animals, or to otherwise enhance physical or mental wellbeing of humans or animals, resulting from administration of a polypeptide of the invention other than the ability to induce the production of an antibody against an antigenic epitope possessed by the biologically active protein. For prophylactic benefit, the compositions may be administered to a subject at risk of recurrence of a particular cancer, even though a diagnosis of the recurrence may not have been made.

[00127] The term "therapeutically effective amount" refers to an amount of a drug effective to treat a disease or disorder in a mammal. In the case of cancer, the therapeutically effective amount of the drug may reduce the number of cancer cells; reduce the tumor size; inhibit (i.e., slow to

some extent and preferably stop) cancer cell infiltration into peripheral organs; inhibit (i.e., slow to some extent and preferably stop) tumor metastasis; inhibit, to some extent, tumor growth; and/or relieve to some extent one or more of the symptoms associated with the cancer. To the extent the drug may inhibit the growth of and/or kill existing cancer cells, it may be cytostatic and/or cytotoxic. For cancer therapy, efficacy can, for example, be measured by assessing the time to disease progression and/or determining the response rate. Such effect need not be absolute to be beneficial. Determination of a therapeutically effective amount is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein.

[00128] The term "therapeutically effective dose regimen", as used herein, refers to a schedule for consecutively administered multiple doses (i.e., at least two or more) of a biologically active protein, either alone or as a part of a polypeptide composition, wherein the doses are given in therapeutically effective amounts to result in sustained beneficial effect on any symptom, aspect, measured parameter or characteristics of a disease state or condition.

I). GENERAL TECHNIQUES

[00129] The practice of the present invention employs, unless otherwise indicated, conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA, which are within the skill of the art. *See* Sambrook, J. *et al.*, "Molecular Cloning: A Laboratory Manual," 3rd edition, Cold Spring Harbor Laboratory Press, 2001; "Current protocols in molecular biology", F. M. Ausubel, *et al.* eds.,1987; the series "Methods in Enzymology," Academic Press, San Diego, CA.; "PCR 2: a practical approach", M.J. MacPherson, B.D. Hames and G.R. Taylor eds., Oxford University Press, 1995; "Antibodies, a laboratory manual" Harlow, E. and Lane, D. eds., Cold Spring Harbor Laboratory,1988; "Goodman & Gilman's The Pharmacological Basis of Therapeutics," 11th Edition, McGraw-Hill, 2005; and Freshney, R.I., "Culture of Animal Cells: A Manual of Basic Technique," 4th edition, John Wiley & Sons, Somerset, NJ, 2000, the contents of which are incorporated in their entirety herein by reference.

[00130] Host cells can be cultured in a variety of media. Commercially available media such as Ham's F10 (Sigma), Minimal Essential Medium (MEM, Sigma), RPMI-1640 (Sigma), and Dulbecco's Modified Eagle's Medium (DMEM, Sigma) are suitable for culturing eukaryotic cells. In addition, animal cells can be grown in a defined medium that lacks serum but is supplemented with hormones, growth factors or any other factors necessary for the survival and/or growth of a particular cell type. Whereas a defined medium supporting cell survival maintains the viability, morphology, capacity to metabolize and potentially, capacity of the cell to differentiate, a defined medium promoting cell growth provides all chemicals necessary for cell proliferation or multiplication. The general parameters governing mammalian cell survival and growth in vitro are well established in the art. Physicochemical parameters which may be controlled in different cell

culture systems are, e.g., pH, pO_2, temperature, and osmolarity. The nutritional requirements of cells are usually provided in standard media formulations developed to provide an optimal environment. Nutrients can be divided into several categories: amino acids and their derivatives, carbohydrates, sugars, fatty acids, complex lipids, nucleic acid derivatives and vitamins. Apart from nutrients for maintaining cell metabolism, most cells also require one or more hormones from at least one of the following groups: steroids, prostaglandins, growth factors, pituitary hormones, and peptide hormones to proliferate in serum-free media (Sato, G. H., et al. in "Growth of Cells in Hormonally Defined Media", Cold Spring Harbor Press, N.Y., 1982). In addition to hormones, cells may require transport proteins such as transferrin (plasma iron transport protein), ceruloplasmin (a copper transport protein), and high-density lipoprotein (a lipid carrier) for survival and growth in vitro. The set of optimal hormones or transport proteins will vary for each cell type. Most of these hormones or transport proteins have been added exogenously or, in a rare case, a mutant cell line has been found which does not require a particular factor. Those skilled in the art will know of other factors required for maintaining a cell culture without undue experimentation. Growth media for growth of prokaryotic host cells include nutrient broths (liquid nutrient medium) or LB medium (Luria Bertani). Suitable media include defined and undefined media. In general, media contains a carbon source such as glucose needed for bacterial growth, water, and salts. Media may also include a source of amino acids and nitrogen, for example beef or yeast extract (in an undefined medium) or known quantities of amino acids (in a defined medium). In some embodiments, the growth medium is LB broth, for example LB Miller broth or LB Lennox broth. LB broth comprises peptone (enzymatic digestion product of casein), yeast extract and sodium chloride. In some embodiments, a selective medium is used which comprises an antibiotic. In this medium, only the desired cells possessing resistance to the antibiotic will grow.

II). XTEN PROTEIN POLYMER AND CONJUGATE COMPOSITIONS

[00131] The present invention relates, in part, to compositions in which one or more XTEN are chemically linked to one or more payloads, including combinations of different payloads, in defined numbers in either monomeric or multimeric configurations to provide compositions with enhanced pharmaceutical, pharmacokinetic, and pharmacologic properties. Such compositions linked to such payloads may have utility, when administered to a subject, in the prevention, treatment or amelioration of cancer for which a payload has a pharmacologic or biologic effect.

[00132] In a first aspect, the invention provides compositions in which XTEN with defined numbers of orthogonal pendant reactive groups are conjugated to one or more molecules of a targeting moiety that serves as a ligand to a cell-surface receptor and one or more molecules of an effector drug. The resulting compositions, when administered in vitro or in vivo, can selectively

deliver the effector drug to cells bearing such receptors, resulting in inhibition of cell division, growth, or metastasis.

[00133] In general, the targeting moieties of the subject compositions exhibit a binding specificity to a given target or another desired biological characteristic when used in vivo or when utilized in an in vitro assay. The targets to which the subject binding fusion protein compositions are generally associated with a disease, disorder or condition; e.g., a cancer. As used herein, "a target associated with a disease, disorder or condition" means that the target is either expressed or overexpressed by disease cells or tissues, the target causes or is a mediator the disease, disorder or condition, or the target is generally found in higher concentrations in a tissue, organ or a localized region within a tissue or in an organ in a subject.

1. XTEN: extended recombinant polypeptides

[00134] For the subject conjugates, the invention provides substantially homogeneous XTEN polypeptides that are useful as conjugation partners to link to one or more payloads via a cross-linker reactant resulting in an XTEN-payload conjugate.

[00135] XTEN are polypeptides with non-naturally occurring, substantially non-repetitive sequences having a low degree or no secondary or tertiary structure under physiologic conditions. XTEN typically have from about 36 to about 3000 amino acids, of which the majority or the entirety are small hydrophilic amino acids. As used herein, "XTEN" specifically excludes whole antibodies or antibody fragments (e.g. single-chain antibodies and Fc fragments). XTEN polypeptides have utility as a conjugation partners in that they serve in various roles, conferring certain desirable properties when linked to a payload, described more fully, below. The resulting XTEN-payload conjugates have enhanced properties, such as enhanced pharmacokinetic, physicochemical, pharmacologic, and pharmaceutical properties compared to the corresponding payload not linked to XTEN, making them useful in the treatment of certain conditions for which the payload is known in the art to be used.

[00136] The unstructured characteristic and physicochemical properties of the XTEN result, in part, from the overall amino acid composition that is disproportionately limited to 4-6 types of hydrophilic amino acids, the linking of the amino acids in a quantifiable non-repetitive design, and the length of the XTEN polypeptide. In an advantageous feature common to XTEN but uncommon to native polypeptides, the properties of XTEN disclosed herein are not tied to absolute primary amino acid sequences, as evidenced by the diversity of the exemplary sequences of Table 1 that, within varying ranges of length, possess similar properties, many of which are documented in the Examples. Accordingly, XTEN have properties more like non-proteinaceous, hydrophilic polymers than they do proteins. The XTEN of the present invention exhibit one or more of the following advantageous properties: conformational flexibility, reduced or lack of secondary structure, high degree of aqueous solubility, high degree of protease resistance, low

immunogenicity, low binding to mammalian receptors, a defined degree of charge, and increased hydrodynamic (or Stokes) radii; properties that are similar to certain hydrophilic polymers (e.g., polyethylene glycol) that make them particularly useful as conjugation partners.

[00137] The XTEN component(s) of the subject conjugates are designed to behave like denatured peptide sequences under physiological conditions, despite the extended length of the polymer. "Denatured" describes the state of a peptide in solution that is characterized by a large conformational freedom of the peptide backbone. Most peptides and proteins adopt a denatured conformation in the presence of high concentrations of denaturants or at elevated temperature. Peptides in denatured conformation have, for example, characteristic circular dichroism (CD) spectra and are characterized by a lack of long-range interactions as determined by NMR. "Denatured conformation" and "unstructured conformation" are used synonymously herein. In some embodiments, the invention provides XTEN sequences that, under physiologic conditions, resemble denatured sequences that are largely devoid of secondary structure. In other cases, the XTEN sequences are substantially devoid of secondary structure under physiologic conditions. "Largely devoid," as used in this context, means that less than 50% of the XTEN amino acid residues of the XTEN sequence contribute to secondary structure as measured or determined by the means described herein. "Substantially devoid," as used in this context, means that at least about 60%, or about 70%, or about 80%, or about 90%, or about 95%, or about 97%, or at least about 99% of the XTEN amino acid residues of the XTEN sequence do not contribute to secondary structure, as measured or determined by the methods described herein.

[00138] A variety of methods and assays are known in the art for determining the physicochemical properties of the subject XTEN. Such properties include but are not limited to secondary or tertiary structure, solubility, protein aggregation, stability, absolute and apparent molecular weight, purity and uniformity, melting properties, contamination and water content. The methods to measure such properties include analytical centrifugation, EPR, HPLC-ion exchange, HPLC-size exclusion chromatography (SEC), HPLC-reverse phase, light scattering, capillary electrophoresis, circular dichroism, differential scanning calorimetry, fluorescence, HPLC-ion exchange, HPLC-size exclusion, IR, NMR, Raman spectroscopy, refractometry, and UV/Visible spectroscopy. In particular, secondary structure can be measured spectrophotometrically, e.g., by circular dichroism spectroscopy in the "far-UV" spectral region (190-250 nm). Secondary structure elements, such as alpha-helix and beta-sheet, each give rise to a characteristic shape and magnitude of CD spectra, as does the lack of these structure elements. Secondary structure can also be predicted for a polypeptide sequence via certain computer programs or algorithms, such as the well-known Chou-Fasman algorithm (Chou, P. Y., *et al.* (1974) *Biochemistry*, 13: 222-45) and the Garnier-Osguthorpe-Robson algorithm ("Gor algorithm") (Garnier J, Gibrat JF, Robson B. (1996), GOR method for predicting protein

secondary structure from amino acid sequence. Methods Enzymol 266:540-553), as described in US Patent Application Publication No. 20030228309A1. For a given sequence, the algorithms can predict whether there exists some or no secondary structure at all, 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). Polypeptide sequences can be analyzed using the Chou-Fasman algorithm using sites on the world wide web at, for example, fasta.bioch.virginia.edu/fasta_www2/fasta_www.cgi?rm=misc1 and the Gor algorithm at npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html (both accessed on September 5, 2012). Additional methods are disclosed in Arnau, *et al.*, Prot Expr and Purif (2006) 48, 1-13.

[00139] In some embodiments, the XTEN sequences used in the subject conjugates have an alpha-helix percentage ranging from 0% to less than about 5% as determined by the Chou-Fasman algorithm. In another embodiment, the XTEN sequences have a beta-sheet percentage ranging from 0% to less than about 5% as determined by the Chou-Fasman algorithm. In one embodiment, the XTEN sequences of the conjugates have 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 the Chou-Fasman algorithm. In one embodiment, the XTEN sequences of the conjugates have an alpha-helix percentage less than about 2% and a beta-sheet percentage less than about 2%. The XTEN sequences of the conjugate compositions have a high degree of random coil percentage, as determined by the GOR algorithm. In some embodiments, an XTEN sequence has at least about 80%, more preferably at least about 90%, more preferably at least about 91%, more preferably at least about 92%, more preferably at least about 93%, more preferably at least about 94%, more preferably at least about 95%, more preferably at least about 96%, more preferably at least about 97%, more preferably at least about 98%, and most preferably at least about 99% random coil, as determined by the GOR algorithm. In one embodiment, the XTEN sequences of the conjugate compositions have 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 the Chou-Fasman algorithm and at least about 90% random coil, as determined by the GOR algorithm. In another embodiment, the XTEN sequences of the disclosed compositions have an alpha-helix percentage less than about 2% and a beta-sheet percentage less than about 2% at least about 90% random coil, as determined by the GOR algorithm. In another embodiment, the XTEN sequences of the compositions are substantially lacking secondary structure as measured by circular dichroism.

[00140] The selection criteria for the XTEN to be linked to the payload used to create the conjugate compositions generally relate to attributes of physicochemical properties and

conformational structure of the XTEN that is, in turn, used to confer enhanced pharmaceutical, pharmacologic, and pharmacokinetic properties to the compositions.

[00141] XTEN are used as a carrier in the compositions, the invention taking advantage of the discovery that increasing the length of the non-repetitive, unstructured polypeptides enhances the unstructured nature of the XTENs and correspondingly enhances the physicochemical and pharmacokinetic properties of constructs comprising the XTEN carrier. In general, XTEN as monomers or as multimers with cumulative lengths longer that about 400 residues incorporated into the conjugates result in longer half-life compared to shorter cumulative lengths, e.g., shorter than about 280 residues. As described more fully in the Examples, proportional increases in the length of the XTEN, even if created by a repeated order of single family sequence motifs result in a sequence with a higher percentage of random coil formation, as determined by GOR algorithm, or reduced content of alpha-helices or beta-sheets, as determined by Chou-Fasman algorithm, compared to shorter XTEN lengths. In addition, increasing the length of the unstructured polypeptide fusion partner, as described in the Examples, results in a construct with a disproportionate increase in terminal half-life compared to polypeptides with unstructured polypeptide partners with shorter sequence lengths. In some embodiments, where the XTEN serve primarily as a carrier, the invention encompasses XTEN conjugate compositions comprising two, three, four or more XTEN wherein the cumulative XTEN sequence length of the XTEN proteins is greater than about 100, 200, 400, 500, 600, 800, 900, or 1000 to about 3000 amino acid residues, wherein the construct exhibits enhanced pharmacokinetic properties when administered to a subject compared to a payload not linked to the XTEN and administered at a comparable dose. In one embodiment of the foregoing, the two or more XTEN sequences each exhibit at least about 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, or 98% or more identity to a sequence selected from Table 1, and the remainder, if any, of the carrier sequence(s) contains at least 90% hydrophilic amino acids and less than about 2% of the overall sequence consists of hydrophobic or aromatic amino acids. The enhanced pharmacokinetic properties of the XTEN-payload conjugate, in comparison to payload not linked to XTEN, are described more fully, below.

1. Non-repetitive Sequences

[00142] It is specifically contemplated that the subject XTEN sequences included in the subject XTEN-folate conjugates are substantially non-repetitive. In general, repetitive amino acid sequences have a tendency to aggregate or form higher order structures, as exemplified by natural repetitive sequences such as collagens and leucine zippers. These repetitive amino acids may also tend to form contacts resulting in crystalline or pseudocrystaline structures. In contrast, the low tendency of non-repetitive sequences to aggregate enables the design of long-sequence XTENs with a relatively low frequency of charged amino acids that would otherwise be likely to aggregate if the sequences were repetitive. The non-repetitiveness of a subject XTEN can be

observed by assessing one or more of the following features. In one embodiment, a substantially non-repetitive XTEN sequence has no three contiguous amino acids in the sequence that are identical amino acid types unless the amino acid is serine, in which case no more than three contiguous amino acids are serine residues. In another embodiment, as described more fully below, a substantially non-repetitive XTEN sequence in which 80-99% of the sequence is comprised of motifs of 9 to 14 amino acid residues wherein the motifs consist of 3, 4, 5 or 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.

[00143] 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. According to the current invention, algorithms to be used in calculating the degree of repetitiveness of a particular polypeptide, such as an XTEN, are disclosed herein, and examples of sequences analyzed by algorithms are provided (*see* Examples, below). In one embodiment, the repetitiveness of a polypeptide of a predetermined length can be calculated (hereinafter "subsequence score") according to the formula given by Equation I:

$$\text{Subsequence score} = \frac{\sum_{i=1}^{m} Count_i}{m} \qquad \text{I}$$

wherein: m = (amino acid length of polypeptide) – (amino acid length of subsequence) + 1; and $Count_i$ = cumulative number of occurrences of each unique subsequence within sequence

[00144] An algorithm termed "SegScore" was developed to apply the foregoing equation to quantitate repetitiveness of polypeptides, such as an XTEN, providing the subsequence score wherein sequences of a predetermined amino acid length "n" are analyzed for repetitiveness by determining the number of times (a "count") a unique subsequence of length "s" appears in the set length, divided by the absolute number of subsequences within the predetermined length of the sequence. FIG. 32 depicts a logic flowchart of the SegScore algorithm, while FIG. 33 portrays a schematic of how a subsequence score is derived for a fictitious XTEN with 11 amino acids and a subsequence length of 3 amino acid residues. For example, a predetermined polypeptide length of 200 amino acid residues has 192 overlapping 9-amino acid subsequences and 198 3-mer subsequences, but the subsequence score of any given polypeptide will depend on the absolute number of unique subsequences and how frequently each unique subsequence (meaning a different amino acid sequence) appears in the predetermined length of the sequence.

[00145] In the context of the present invention, "subsequence score" means the sum of occurrences of each unique 3-mer frame across 200 consecutive amino acids of the cumulative XTEN polypeptide divided by the absolute number of unique 3-mer subsequences within the 200 amino acid sequence. Examples of such subsequence scores derived from 200 consecutive amino acids of repetitive and non-repetitive polypeptides are presented in Example 45. In one embodiment, the invention provides a XTEN-folate comprising one XTEN in which the XTEN has a subsequence score 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. In another embodiment, the invention provides XTEN-cross-linker conjugates comprising an XTEN in which the XTEN have a subsequence score of 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. In another embodiment, the invention provides XTEN-click-chemistry conjugates comprising an XTEN in which the XTEN have a subsequence score of 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. In yet another embodiment, the invention provides XTEN conjugate compositions comprising at least two linked XTEN in which each individual XTEN 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, or less. In yet another embodiment, the invention provides XTEN conjugate compositions comprising at least three linked XTEN in which each individual XTEN 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, or less. In the embodiments of the XTEN compositions described herein, an XTEN with a subsequence score of 10 or less (i.e., 9, 8, 7, etc.) is characterized as substantially non-repetitive.

[00146] It was discovered that the non-repetitive characteristic of XTEN together with the particular types of amino acids that predominate in the XTEN, rather than the absolute primary sequence, confers one or more of the enhanced physicochemical and biological properties of the XTEN and the resulting XTEN-folate conjugates in which they are utilized. These enhanced properties include a higher degree of expression of the XTEN protein in the host cell, greater genetic stability of the gene encoding XTEN, a greater degree of solubility, less tendency to aggregate, and enhanced pharmacokinetics of the resulting conjugate compared to payloads not conjugated to XTEN. These enhanced properties permit more efficient manufacturing, greater uniformity of the final product, lower cost of goods, and facilitate the formulation of pharmaceutical preparations containing extremely high protein concentrations, in some cases exceeding 100 mg/ml. In addition, the XTEN polypeptide sequences of the conjugates are designed to have a low degree of internal repetitiveness in order to reduce or substantially eliminate immunogenicity when administered to a mammal.

[00147] The present invention encompasses XTEN used as conjugation partners that comprise multiple units of shorter sequences, or motifs, in which the amino acid sequences of the motifs are substantially non-repetitive. The non-repetitive property can be met even using a "building block" approach using a library of sequence motifs that are multimerized to create the XTEN sequences, as shown in FIGS. 18-19. While an XTEN sequence may 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 XTEN sequence is designed to render the sequence substantially non-repetitive. Examples of amino acids that are included in XTEN are, e.g., arginine, lysine, threonine, alanine, asparagine, glutamine, aspartate, glutamate, serine, and glycine. In one embodiment, XTEN sequences have predominately four to six types of amino acids selected from glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) or proline (P) that are arranged in a substantially non-repetitive sequence that is about 36 to about 3000, or about 100 to about 2000, or about 144 to about 1000 amino acid residues in length.

[00148] In some embodiments wherein the XTEN has less than 100% of its amino acids consisting of 4, 5, or 6 types of amino acid selected from glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P), the other amino acid residues of the XTEN are selected from any of the other 14 natural L-amino acids, but are preferentially selected from hydrophilic amino acids such that the XTEN sequence contains at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or at least about 99% hydrophilic amino acids. An individual amino acid or a short sequence of amino acids other than glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P) may be incorporated into the XTEN to achieve a needed property, such as to permit incorporation of a restriction site by the encoding nucleotides, or to facilitate linking to a payload component, or incorporation of a cleavage sequence. As one exemplary embodiment, described more fully below, the invention provides XTEN that incorporates from 1 to about 20, or 1 to about 15, or 1 to about 10, or 1 to 5 lysine residues wherein the reactive lysines are utilized for linking to cross-linkers or payloads, as described herein. In another embodiment, described more fully below, the XTEN incorporates from 1 to about 20, or 1 to about 15, or 1 to about 10, or 1 to 5 cysteine residues wherein the reactive cysteines are utilized for linking to cross-linkers or payloads, as described herein. In another embodiment, the XTEN incorporates from 1 to about 20 cysteine and lysine residues wherein the lysines and cysteines are utilized for linking to different cross-linkers or payloads, as described herein. In another embodiment, the XTEN incorporations 1, 2, 3, 4, 5 or more arginine residues that are not followed by proline residues to provide cleavage sequences that can be cleaved by trypsin to create XTEN segments, described more fully herein, below. The XTEN amino acids that are not glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) and proline (P) are either interspersed throughout the XTEN sequence, are located within or between the sequence

motifs, or are concentrated in one or more short stretches of the XTEN sequence such as at or near the N- or C-terminus. As hydrophobic amino acids impart structure to a polypeptide, the invention provides that the content of hydrophobic amino acids in the XTEN utilized in the conjugation constructs will typically be less than 5%, or less than 2%, or less than 1% hydrophobic amino acid content. Hydrophobic residues that are less favored in construction of XTEN include tryptophan, phenylalanine, tyrosine, leucine, isoleucine, valine, and methionine. Additionally, one can design the XTEN sequences to contain less than 5% or less than 4% or less than 3% or less than 2% or less than 1% or none of the following amino acids: methionine (to avoid oxidation), asparagine and glutamine (to avoid desamidation). In other embodiments, the amino acid content of methionine and tryptophan in the XTEN component used in the conjugation constructs is typically less than 5%, or less than 2%, and most preferably less than 1%. In other embodiments, the XTEN of the subject XTEN conjugates will have 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, the sum of methionine and tryptophan residues will be less than 2%, and the sum of asparagine and glutamine residues will be less than 5% of the total XTEN sequence.

2. Cysteine- and Lysine-Engineered XTEN Sequences

[00149] In another aspect, the invention provides XTEN-folate conjugates comprising XTEN with defined numbers of incorporated cysteine or lysine residues; "cysteine-engineered XTEN" and "lysine-engineered XTEN", respectively. It is an object of the invention to provide XTEN with defined numbers of cysteine and/or lysine residues to permit conjugation between the thiol group of the cysteine or the epsilon amino group of the lysine and a reactive group on a payload or a cross-linker to be conjugated to the XTEN backbone. In one embodiment of the foregoing, the XTEN of the invention 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, or about 1 to about 5 lysine residues, or 1 to about 3 lysine residues, or alternatively only a single lysine residue. In another embodiment of the foregoing, the XTEN of the invention has between about 1 to about 100 cysteine residues, or about 1 to about 70 cysteine residues, or about 1 to about 50 cysteine residues, or about 1 to about 30 cysteine residues, or about 1 to about 20 cysteine residues, or about 1 to about 10 cysteine residues, or about 1 to about 5 cysteine residues, or 1 to about 3 cysteine residues, or alternatively only a single cysteine residue. In another embodiment of the foregoing, the XTEN of the invention has about 1 to about 10 lysine residues and about 1 to about 10 cysteine residues. Using the foregoing lysine- and/or cysteine-containing XTEN, conjugates can be constructed that comprise XTEN, a cross-linker, plus one or more types of payloads useful in the treatment of a cancer in a subject wherein the maximum number of molecules of the

payload agent linked to the XTEN component is determined by the numbers of lysines or cysteines with a reactive side group (i.e., a terminal amino or a thiol) incorporated into the XTEN.

[00150] In one embodiment, the invention provides polynucleotides encoding cysteine-engineered XTEN wherein nucleotides encoding one or more amino acids of an XTEN are replaced with nucleotides encoding a cysteine amino acid to create a gene encoding the cysteine-engineered XTEN. In another embodiment, the invention provides polynucleotides encoding cysteine-engineered XTEN where nucleotides encoding one or more cysteine amino acids are inserted into an-XTEN encoding gene to create the cysteine-engineered XTEN gene. In other cases, oligonucleotides encoding one or more motifs of about 9 to about 14 amino acids comprising codons encoding one or more cysteines are linked in frame with other oligos encoding XTEN motifs or full-length XTEN to create the cysteine-engineered XTEN gene. In one embodiment of the foregoing, where the one or more cysteines are inserted into an XTEN sequence during the creation of the XTEN gene, nucleotides encoding cysteine can be linked to codons encoding amino acids used in XTEN to create a cysteine-XTEN motif with the cysteine(s) at a defined position using the methods described herein (see Example 2 and FIG. 1), or by standard molecular biology techniques, and the motifs subsequently assembled into the gene encoding the full-length cysteine-engineered XTEN. In such cases, where, for example, nucleotides encoding a single cysteine are added to the DNA encoding a 12 amino acid motif, the resulting motif would have 13 amino acids, while incorporating two cysteines would result in a motif having 14 amino acids, etc. In other cases, a cysteine-motif can be created *de novo* and be of a pre-defined length and number of cysteine amino acids by linking nucleotides encoding cysteine to nucleotides encoding one or more amino acid residues used in XTEN (e.g., G, S, T, E, P, A) at a defined position, and the encoding motifs subsequently assembled by annealing with other XTEN-encoding motif sequences into the gene encoding the full-length XTEN, as described herein and illustrated in FIGS. 7-8. In cases where a lysine-engineered XTEN is utilized to make the conjugates of the invention, the approaches described above would be performed with codons encoding lysine instead of cysteine. Thus, by the foregoing, a new XTEN motif can be created that could comprise about 9-14 amino acid residues and have one or more reactive amino acids; i.e., cysteine or lysine. Non-limiting examples of motifs suitable for use in an engineered XTEN that contain a single cysteine or lysine are:

GGSPAGSCTSP
GASASCAPSTG
TAEAAGCGTAEAA
GPEPTCPAPSG
GGSPAGSKTSP
GASASKAPSTG

However, the invention contemplates motifs of different lengths, such as those of Table 3, for incorporation into XTEN.

[00151] In such cases where a gene encoding an XTEN with one or more cysteine and/or lysine motifs is to be constructed from existing XTEN motifs or segments, the gene can be designed and built by linking existing "building block" polynucleotides encoding both short- and long-length XTENs; e.g., AE48, AE144, AE288, AE432, AE576, AE864, AM48, AM875, AE912, AG864, which can be fused in frame with the nucleotides encoding the cysteine- and/or lysine-containing motifs or, alternatively, the cysteine- and/or lysine-encoding nucleotides can be PCR'ed into an existing XTEN sequence (as described more fully below and in the Examples) using, for example, nucleotides encoding the islands of Tables 4 and 5 to build an engineered XTEN in which the reactive cysteine and/or lysines are placed in one or more designed locations in the sequence in the desired quantity. Non-limiting examples of such engineered XTEN are provided in Table 1. Thus, in one embodiment, the invention provides an XTEN sequence having at least about 80% sequence identity, 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% sequence identity, or is identical to a sequence or a fragment of a sequence selected from of Table 1, when optimally aligned. However, application of the cysteine- or lysine-engineered methodology to create XTEN encompassing cysteine or lysine residues is not meant to be constrained to the precise compositions or range of composition identities of the foregoing embodiments. As will be appreciated by those skilled in the art, the precise location and numbers of incorporated cysteine or lysine residues in an XTEN can be varied without departing from the invention as described.

Table 1: Cysteine- and lysine-engineered XTEN

Name	Amino Acid Sequence
Seg 1	AGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSE SATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGS APGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPTAEAAG CGTAEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTST EEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEP ATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESAT PESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTST EEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPR
Seg 2	ATAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGS PAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSG SETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTE EGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGT STEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATP ESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTE EGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGS

Name	Amino Acid Sequence
	EPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATP ESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPTA EAAGCGTAEAAR
Seg 3	ATAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGS PAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSG SETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTE EGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGT STEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPT STEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSET PGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT SESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTE PSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSG SETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESG PGTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 4	ATAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGS PAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSG SETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPTAEAAGCGTA EAAGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSA PGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGS PAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPT STEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAGTSESATP ESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSA PGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGS PAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPA TSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSG SETPGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 5	ATAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGS PAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAGTSES ATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATP ESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSA PGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGS PAGSPTSTEEGTSTEPSEGSAPTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGT SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAG SPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATP ESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSA PTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESG PGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT SESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTE PSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 6	ATAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGS PAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE TAEAAGCGTAEAAPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSES ATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATP ESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGST AEAAGCGTAEAAAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSA PGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGT

Name	Amino Acid Sequence
	SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSTAE AAGCGTAEAAPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEP SEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPE SGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATTAEA AGCGTAEAAPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGP GTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTS TEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPTAEAA GCGTAEAAR
Seg 7	ATAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGS PAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAGTSTEPSEGSAPGS PAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSES ATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATP ESGPTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP ESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSA PGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPTAEAAGCGTAEAAGTSESATPESG PGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGT SESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTE PSEGSAPTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSG SETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEETAEAAGCGTAEAAGTSTEPSEG SAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETP GSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTS TEPSEGSAPTAEAAGCGTAEAAR
Seg 8	ATAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGS PAGSPTSTEEGTSTAEAAGCGTAEAAESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGS PAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSES ATPESGPGSEPATSGSETPGTSTEPSTAEAAGCGTAEAAEGSAPGTSTEPSEGSAPGTSES ATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP ESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSTAEAAGCGTAEAAAPGTSTEPSE GSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESG PGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTTAEAAGCGTAE AASTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTS ESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPAT AEAAGCGTAEAATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESA TPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEG SAPGTSESATPETAEAAGCGTAEAASGPGTSESATPESGPGTSESATPESGPGSEPATSGS ETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETP GTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 9	ATAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGS PAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSG SETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTE EGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGT STEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATP ESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTE EGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGS EPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATP ESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPR
Seg 10	AGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSE SATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGS

Name	Amino Acid Sequence
	APGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGS APGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPES GPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 11	ATAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGS PAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSG SETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTE EGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGT STEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPT STEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSET PGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT SESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTE PSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSG SETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESG PGTSTEPSEGSAPR
Seg 12	AGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSE SATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGS APGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTTAEAAGCGTAEAAEPSEGSAPGSPAGSPTSTEEGTST EPSEGSAPGTSESATPESGPGSEPATTAEAAGCGTAEAASGSETPGTSESATPESGPGSEP ATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEG TSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTST EPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATS GSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPR
Seg 13	AGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSE SATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGS APGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPTAEAAG CGTAEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTST EEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEP ATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESAT PESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTST EEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPT AEAAGCGTAEAAR
Seg 14	AGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSE SATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS

Name	Amino Acid Sequence
	EGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGS APGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGGKPGG TSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTST EPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESAT PESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPG TSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSE SATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPS EGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPR
Seg 15	AGGKPGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTE EGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGT STEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTE PSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATP ESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSA PGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGT SESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTE PSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATP ESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSET PGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGT SESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSES ATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGGKPGR
Seg 16	AGGKPGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTE EGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGT STEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTE PSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATP ESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSA PGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGG KPGGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGP GTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTS ESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPAT SGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGP GTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTS TEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGGKPG R
Seg 17	AGGKPGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTE EGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGT STEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTE PSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGGKPGGSPAGSPTSTEEGT SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTE PSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESG PGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGT SESATPESGPGTSTEPSEGSAPGGKPGGTSESATPESGPGSEPATSGSETPGTSESATPESG PGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGS EPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSES ATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPT STEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSA PGGKPGR
Seg 18	AGGKPGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTE EGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGT

Name	Amino Acid Sequence
	STEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGGKPGGTSESATPESGPGSEPATSGSET PGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGT SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTE PSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGGKPGGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPT STEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESG PGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGGKPGGSPAGSPTSTEEGTSESATP ESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSA PGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGS PAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTE PSEGSAPGGKPGR
Seg 19	AGGKPGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTE EGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEGGKPGPSE GSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSET PGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGT SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSGGKPGAPGTSTEPSEGSA PGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGT STEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGTSTEPSEGSAPGTSESATPESGPGSGGKPGPAGSPTSTEEGSPAGSPTSTEEGS PAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATP ESGPGSEPATSGSETPGTSESATGGKPGPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTE PSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSG SETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESG PGTSTEPSEGSAPGGKPGR
Seg 20	AGGKPGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTE EGTSESATPESGPGTSTEPSEGSAPGGKPGGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSET PGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGGKPGGSPAGSPT STEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSA PGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGT STEPSEGSAPGGKPGGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSET PGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGS PAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGGKPGGTSESATPESGPGSEPATSGSET PGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT SESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGGKP GGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGS EPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSES ATPESGPGTSTEPSEGSAPGGKPGR
Seg 21	AGGKPGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTE EGTSGGKPGESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSET PGTSTEPSGGKPGEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPT STEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSA PGTSTEPSEGSGGKPGAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPT STEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSET PGTSESATPESGPGTGGKPGSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPT STEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSET PGTSESATPESGPGSEPAGGKPGTSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPT STEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTE EGTSTEPSEGSAPGTSESATPEGGKPGSGPGTSESATPESGPGTSESATPESGPGSEPATSG SETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSET PGTSESATPESGPGTSTEPSEGSAPGGKPGR
Seg 22	AGGKPGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE

Name	Amino Acid Sequence
	GSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTE EGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGT STEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTE PSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATP ESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSA PGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGT SESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTE PSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATP ESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSET PGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGT SESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSES ATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPR
Seg 23	AGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSE SATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGS APGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGS APGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPES GPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGGKPGR
Seg 24	AGGKPGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTE EGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGT STEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTE PSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATP ESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSA PGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGG KPGGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGP GTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTS ESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPAT SGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGP GTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTS TEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPR
Seg 25	AGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSE SATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGS APGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTGGKPGEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG TSESATPESGPGSEPATGGKPGSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEG TSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEP ATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATS GSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPES GPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEG TSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPR
Seg 26	AGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSE SATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS

Name	Amino Acid Sequence
	EGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGS APGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGGKPGG TSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTST EPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESAT PESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPG TSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSE SATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPS EGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGGKPGR
Seg 27	AEATTAAGGAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTST EPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSP TSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGS APGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPG TSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSE SATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPS EGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGS APTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEG SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGS APGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPG SPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTST EPSEGSAPR
Seg 28	AEATTAAGGATAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPG SPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEP ATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPES GPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPG SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEG SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGS APGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPG SPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTST EPSEGSAPTAEAAGCGTAEAAR
Seg 29	AEATTAAGGATAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPG SPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEP ATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPES GPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPG SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTST EEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPG SEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPA GSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSE TPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 30	AEATTAAGGATAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPG SPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEP

Name	Amino Acid Sequence
	ATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPES GPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPT AEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPG TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTST EEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEA AGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPG TSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPA GSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESAT PESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGS APGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 31	ATGTATSEGSPEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT STEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAG SPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSET PGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGT SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTE PSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSG SETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTE EGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGT SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSE GSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSET PGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGT STEPSEGSAPR
Seg 32	ATGTATSEGSPETAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP GSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSE PATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGS PTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPE SGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGP GSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGS PTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEE GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTS ESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESA TPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEG SAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETP GSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTS TEPSEGSAPTAEAAGCGTAEAAR
Seg 33	ATGTATSEGSPETAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP GSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSE PATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGS PTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPE SGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGP GSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGS PTSTEEGTSTEPSEGSAPTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTS TEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGP GSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSP AGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGS PTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGS ETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETP GTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAR

Name	Amino Acid Sequence
Seg 34	ATGTATSEGSPETAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP GSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSE PATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGS PTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPE SGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGP TAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGP GTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTS TEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTS TEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAE AAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGP GTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSP AGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESA TPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEG SAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 35	EPTAATTGESAGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT STEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAG SPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSET PGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGT SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTE PSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSG SETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTE EGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGT SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSE GSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSET PGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGT STEPSEGSAPR
Seg 36	EPTAATTGESAGTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP GSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSE PATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGS PTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPE SGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGP GSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGS PTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEE GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTS ESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESA TPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEG SAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETP GSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTS TEPSEGSAPTAEAAGCGTAEAAR
Seg 37	EPTAATTGESAGTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP GSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSE PATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGS PTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPE SGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGP GSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGS PTSTEEGTSTEPSEGSAPTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTS TEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGP GSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSP AGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGS PTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGS

Name	Amino Acid Sequence
	ETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETP GTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 38	EPTAATTGESAGTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP GSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSE PATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGS PTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPE SGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGP TAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGP GTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTS TEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTS TEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAE AAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGP GTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSP AGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESA TPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEG SAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 39	AEATTAAGGAEEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEE GTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPAT SGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTS TEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTS ESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEP SEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGS ETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGP GTSTEPSEGSAPR
Seg 40	AEATTAAGGAEEETAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPG SEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPES GPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATS GSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTST EEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPG TSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSE SATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPS EGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSE TPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPG TSTEPSEGSAPTAEAAGCGTAEAAR
Seg 41	AEATTAAGGAEEETAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPG SEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPES GPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPT STEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESG PGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGS

Name	Amino Acid Sequence
	PAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAG SPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSG SETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSET PGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 42	AEATTAAGGAEEETAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPG SEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPES GPTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPG TSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSP TSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPTAEAAGCGT AEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPG SPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSE SATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 43	ATGTATSEGSPEEEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGS PAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSG SETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTE EGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGT STEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPT STEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSET PGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT SESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTE PSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSG SETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESG PGTSTEPSEGSAPR
Seg 44	ATGTATSEGSPEEEETAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEG SAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETP GSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSP AGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESA TPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPE SGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP GTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSP AGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPAT SGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTS TEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTS ESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEP SEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGS ETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGP GTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 45	ATGTATSEGSPEEEETAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEG SAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETP GSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSP AGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESA TPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPE SGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP GTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSP AGSPTSTEEGTSTEPSEGSAPTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTS ESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGS

Name	Amino Acid Sequence
	PTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPE SGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP GSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSP AGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPAT SGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGS ETPGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 46	ATGTATSEGSPEEEETAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEG SAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETP GSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSP AGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESA TPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPE SGPTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPE SGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAP GTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTS ESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGS PTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPTAEAAGCG TAEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPE SGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGP GSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTS ESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEP SEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 47	AEATTAAGGAEEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGS ETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEE GTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPAT SGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTS TEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTS ESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEP SEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGS ETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGP GTSTEPSEGSAPRPRPRPRP
Seg 48	AEATTAAGGAEEETAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPG SEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPES GPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATS GSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTST EEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPG TSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSE SATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPS EGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSE TPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPG TSTEPSEGSAPTAEAAGCGTAEAARPRPRPRP
Seg 49	AEATTAAGGAEEETAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPG SEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPES GPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPA

Name	Amino Acid Sequence
	GSPTSTEEGTSTEPSEGSAPTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPT STEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESG PGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGS PAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAG SPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSG SETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSET PGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAARPRPRPRP
Seg 50	AEATTAAGGAEEETAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPG SEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPES GPTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPG TSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSP TSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPTAEAAGCGT AEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPG SPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSE SATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAARPRPRPR P
Seg 51	AEATTAAGGAEEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGS ETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEE GTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPAT SGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTS TEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTS ESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEP SEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGS ETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGP GTSTEPSEGSAPR
Seg 52	AEATTAAGGAEEETAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPG SEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPES GPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATS GSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTST EEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPG TSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSE SATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPS EGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSE TPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPG TSTEPSEGSAPTAEAAGCGTAEAAR
Seg 53	AEATTAAGGAEEETAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPG SEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESAT

Name	Amino Acid Sequence
	PESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPES GPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPT STEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESG PGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGS PAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAG SPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSG SETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSET PGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 54	AEATTAAGGAEEETAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPG SEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPES GPTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPG TSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSP TSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPTAEAAGCGT AEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPG SPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSE SATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 55	AEATTAAGGAEEETAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPG SEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPES GPGGKPGGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPS EGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPA GSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSE GSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTE EGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGS EPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPA TSGSETPGTSESATPESGPGTSTEPSEGSAPGGKPGR
Seg 56	AEATTAAGGAEEETAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPG SEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPES GPTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPG TSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSP TSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPTAEAAGCGT AEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPG SPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSE SATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGGKPGR
Seg 57	AEATTAAGGAEEEGGKPGGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSE

Name	Amino Acid Sequence
	TPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEG TSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEP ATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGGKPGG SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEG SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGGKPGGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEG TSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTST EPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATS GSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 58	AEATTAAGGAEEETAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPG SEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPES GPGGKPGGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPS EGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGS APGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPA GSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSE GSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTE EGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGS EPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPA TSGSETPGTSESATPESGPGTSTEPSEGSAPGGKPGR
Seg 59	AEATTAAGGAEEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGS ETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEE GTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPE SGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEE GTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 60	AEATTAAGGAEEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGS ETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEE GTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPTAEAAGCGTAEAAR
Seg 61	AEATTAAGGAEEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGS ETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTAEAAGCGTAE AAR
Seg 62	AEATTAAGGAEEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 63	AEATTAAGGAEEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP

Name	Amino Acid Sequence
	SEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGS ETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEE GTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPE SGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEE GTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 64	AEATTAAGGAEEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGS ETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEE GTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPTAEAAGCGTAEAAR
Seg 65	AEATTAAGGAEEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGS ETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPTAEAAGCGTAE AAR
Seg 66	AEATTAAGGAEEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAR
Seg 67	AEATTAAGGAEEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGS ETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEE GTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPE SGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEE GTSESATPESGPGTSTEPSEGSAPGGKPGR
Seg 68	AEATTAAGGAEEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGS ETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEE GTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGGKPGR
Seg 69	AEATTAAGGAEEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGS ETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGGKPGR
Seg 70	AEATTAAGGAEEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGGKPGR
Seg 71	AEATTAAGGAEEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGS ETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEE GTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPE

Name	Amino Acid Sequence
	SGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEE GTSESATPESGPGTSTEPSEGSAPGGKPGR
Seg 72	AEATTAAGGAEEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGGKPGR
Seg 73	AEATTAAGGAEEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGS ETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGGKPGR
Seg 74	AEATTAAGGAEEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGGKPGR
Seg 75	AGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPG SSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGAS PGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGS GTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSG ATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSST GSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGS PGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPT AEAAGCGTAEAAGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPG SSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGAS PGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGS GTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSAST GTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGT GPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGP GASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGA SPGTSSTGSPR
Seg 76	ATAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSS PGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPG TPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGAS PGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPS GATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTA SSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSS PGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPG ASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSP SASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGT SSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTA SSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTG PGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPG SSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGAS PGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPG TSSTGSPTAEAAGCGTAEAAR
Seg 77	ATAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSS PGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPG TPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGAS PGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPS GATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTA SSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSS PGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPG ASPGTSSTGSPTAEAAGCGTAEAAGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPG SSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGAS PGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPS

Name	Amino Acid Sequence
	GATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGA TGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATG SPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSP GSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGS STPSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 78	ATAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSS PGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPG TPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGAS PGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPS GATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPTAEAAGC GTAEAAGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSAST GTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATG SPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGA SPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPTAEAAGCGTAEAAGT PGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTP SGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGT ASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSST GSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSS PGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 79	ATAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSS PGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPG TPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGAS PGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPTAEAAGCGTAEAAGSS PSASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPG TSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSAST GTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATG SPGSSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAGTPGSGTASSSPGSSTPSGATG SPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSP GTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGT PGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTP SGATGSPTAEAAGCGTAEAAGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPG TSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSS TGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGT GPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAE AAR
Seg 80	ATAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSS PGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPG TPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGAS PGTAEAAGCGTAEAATSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSS PSASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPG TSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSAST GTTAEAAGCGTAEAAGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSS TGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATG SPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSP GTTAEAAGCGTAEAAPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSP GASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASP GTSTAEAAGCGTAEAASTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSP SASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGT SSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSST GSPTAEAAGCGTAEAAR
Seg 81	ATAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSS PGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPG TPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGAS PGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPS GATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSST GSPTAEAAGCGTAEAAGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGT

Name	Amino Acid Sequence
	SSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSST GSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAGTPGSGTAS SSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSP GASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGA SPGTSSTGSPTAEAAGCGTAEAAGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGS STPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPS ASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPTAEAAGCGTAEAAGSSPS ASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSAS TGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGAT GSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 82	ATAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSS PGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPG TPGSGTASSSPGASTAEAAGCGTAEAAPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPG SSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSP SASTGTGPGSSTPSGATGSPGSSTPSGTAEAAGCGTAEAAATGSPGASPGTSSTGSPGASP GTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGT SSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGTAEAAGCGTAEAASPGASPGT SSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTA SSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGATAEAAGCG TAEAASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGS PGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPG SSTPTAEAAGCGTAEAASGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPG SSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSP SASTGTGPGSSPSASTGTAEAAGCGTAEAATGPGASPGTSSTGSPGASPGTSSTGSPGSST PSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPS GATGSPGSSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 83	ATAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSS PGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPG TPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGAS PGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPS GATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTA SSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSS PGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPG ASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSP SASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGT SSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTA SSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTG PGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPG SSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGAS PGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPG TSSTGSPR
Seg 84	AGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPG SSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGAS PGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGS GTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSG ATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSST GSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGS PGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPG TPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGAS PGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASPG TSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGA TGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGT GPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGP GASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGS SPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPTAEA AGCGTAEAAR
Seg 85	ATAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSS PGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPG

Name	Amino Acid Sequence
	TPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGAS PGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPS GATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTA SSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSS PGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPG ASPGTSSTGSPTAEAAGCGTAEAAGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPG SSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGAS PGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPS GATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGA TGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATG SPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSP GSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGS STPSGATGSPGASPGTSSTGSPR
Seg 86	AGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPG SSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGAS PGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGS GTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSG ATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSST GSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGS PGASPGTSSTGSPGASPGTSSTGSPGSSTTAEAAGCGTAEAAPSGATGSPGSSTPSGATGS PGASPGTSSTGSPGTPGSGTASSSPGSSTPSTAEAAGCGTAEAAGATGSPGSSTPSGATGS PGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPG ASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSST PSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPS GATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGA TGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATG SPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSP GSSTPSGATGSPGASPGTSSTGSPR
Seg 87	AGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPG SSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGAS PGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGS GTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSG ATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSST GSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGS PGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPT AEAAGCGTAEAAGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPG SSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGAS PGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGS GTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSAST GTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGT GPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGP GASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGA SPGTSSTGSPTAEAAGCGTAEAAR
Seg 88	AGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPG SSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGAS PGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGS GTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSG ATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSST GSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGS PGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPG GKPGGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGT GPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSP GASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGS STPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPS ASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSAS TGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSST GSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGS PR

Name	Amino Acid Sequence
Seg 89	AGGKPGGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGA TGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASS SPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSP GTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGS STPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASP GTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGT SSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSST GSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTG PGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPG ASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSST PSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSA STGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSAST GTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTG SPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSP GGKPGR
Seg 90	AGGKPGGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGA TGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASS SPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSP GTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGS STPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASP GTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGT SSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSST GSPGGKPGGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPSA STGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSS TGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASS SPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGP GSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGS SPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASP GTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGT SSTGSPGGKPGR
Seg 91	AGGKPGGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGA TGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASS SPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSP GTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGS STPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGGKPGGTPGSGTASSS PGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPG ASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGAS PGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPSA STGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSS TGSPGASPGTSSTGSPGASPGTSSTGSPGGKPGGTPGSGTASSSPGSSTPSGATGSPGTPGS GTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSAST GTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGT GPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGP GASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGA SPGTSSTGSPGGKPGR
Seg 92	AGGKPGGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGA TGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASS SPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSP GTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGGKPGGSSPSASTGTGPGSSTPSGAT GSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSS PGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPG ASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGAS PGTSSTGSPGGKPGGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSP GSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGA SPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGS GTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSPS ASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSAS

Name	Amino Acid Sequence
	TGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGAT GSPGASPGTSSTGSPGGKPGR
Seg 93	AGGKPGGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGA TGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASS SPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGGGKPGTS STGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGAT GSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSS PGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGGKPGGPGTPGSGT ASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGAT GSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGS PGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTGGKPGPGSGTASSSPGASPGTSS TGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATG SPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSP GSSPSASTGTGPGSSPSASTGTGPGASPGTSGGKPGSTGSPGTPGSGTASSSPGSSTPSGAT GSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGS PGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPG SSTPSGATGSPGASPGTSSTGSPGGKPGR
Seg 94	AGGKPGGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGA TGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASS SPGASPGTSSTGSPGASPGTSSTGSPGGKPGGTPGSGTASSSPGSSTPSGATGSPGASPGTS STGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGAT GSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGGKPGGTPGSG TASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTA SSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGS PGASPGTSSTGSPGGKPGGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGA TGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTG SPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGGKPGGTPGSGTASSSPGSSTPSG ATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGAT GSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGS PGGKPGGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGA TGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATG SPGSSTPSGATGSPGASPGTSSTGSPGGKPGR
Seg 95	AGGKPGGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGA TGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASS SPGASGGKPGPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTS STGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGAT GSPGSSTPSGGGKPGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSG TASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTA SSSPGASPGTSSTGGGKPGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTP SGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSG ATGSPGSSPSASTGTGPGAGGKPGSPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASP GTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSG ATGSPGTPGSGTASSSPGSSTPGGKPGSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSST PSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPS GATGSPGSSPSASTGTGPGSSPSASTGGGKPGTGPGASPGTSSTGSPGASPGTSSTGSPGSS TPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPS GATGSPGSSTPSGATGSPGASPGTSSTGSPGGKPGR
Seg 96	AGGKPGGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGA TGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASS SPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSP GTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGS STPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASP GTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGT SSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSST GSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTG PGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPG ASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSST PSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSA

Name	Amino Acid Sequence
	STGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSAST GTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTG SPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSP R
Seg 97	AGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPG SSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGAS PGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGS GTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSG ATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSST GSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGS PGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPG TPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGAS PGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPG TSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGA TGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGT GPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGP GASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGS SPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGGKP GR
Seg 98	AGGKPGGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGA TGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASS SPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSP GTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGS STPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASP GTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGT SSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSST GSPGGKPGGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPSA STGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSS TGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASS SPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGP GSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGS SPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASP GTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGT SSTGSPR
Seg 99	AGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPG SSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGAS PGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGS GTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSG ATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSST GSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGS PGASPGTSSTGSPGASPGTSSTGSPGSSTGGKPGPSGATGSPGSSTPSGATGSPGASPGTSS TGSPGTPGSGTASSSPGSSTPSGGKPGGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPS ASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTS STGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTAS SSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGP GSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGS SPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASP GTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGT SSTGSPR
Seg 100	AGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPG SSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGAS PGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGS GTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSG ATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSST GSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGS PGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPG GKPGGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGT GPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSP

Name	Amino Acid Sequence
	GASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGS STPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPS ASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSAS TGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSST GSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGS PGGKPGR
Seg 101	AEATTAAGGAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSS TPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGS GTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTS STGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGAT GSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSS PGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPG ASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGAS PGTSSTGSPTAEAAGCGTAEAAGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSS TPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPG TSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGA TGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATG SPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSP GSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGS SPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTP SGATGSPGASPGTSSTGSPR
Seg 102	AEATTAAGGATAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPG TPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSST PSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPS GATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSAST GTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTG SPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGP GTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGS STPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTP SGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTS STGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGAT GSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGS PGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPG SSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSP SASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPS GATGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 103	AEATTAAGGATAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPG TPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSST PSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPS GATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSAST GTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTG SPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGP GTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGS STPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSP SASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPS GATGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 104	AEATTAAGGATAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPG TPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSST PSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPS GATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSAST GTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTG SPTAEAAGCGTAEAAGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTG SPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSP

Name	Amino Acid Sequence
	GSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGS STPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPG SGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPTAEAAG CGTAEAAGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSG TASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSST GSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGS PGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPG TPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 105	ATGTATSEGSPEGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGS STPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPG SGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGT SSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGAT GSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSS PGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPG ASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGAS PGTSSTGSPTAEAAGCGTAEAAGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSS TPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPG TSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGA TGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATG SPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSP GSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGS SPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTP SGATGSPGASPGTSSTGSPR
Seg 106	ATGTATSEGSPETAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGP GTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGS STPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTP SGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSAS TGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSST GSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTG PGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPG SSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSST PSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGT SSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGAT GSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGS PGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPG SSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSP SASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPS GATGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 107	ATGTATSEGSPETAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGP GTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGS STPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTP SGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSAS TGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSST GSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTG PGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPG SSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAGTPGSGTASSSPGSSTPSGATGSPG SSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTP GSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGS GTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSG ATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTAS SSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGS STPSGATGSPGSSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 108	ATGTATSEGSPETAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGP GTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGS STPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTP SGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSAS TGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSST

Name	Amino Acid Sequence
	GSPTAEAAGCGTAEAAGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSST GSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGS PGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPG SSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTP GSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPTAEAA GCGTAEAAGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGS GTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTS STGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSST GSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTG PGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAEA AR
Seg 109	EPTAATTGESAGGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGS STPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPG SGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGT SSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGAT GSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSS PGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPG ASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGAS PGTSSTGSPTAEAAGCGTAEAAGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSS TPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPG TSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGA TGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATG SPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSP GSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGS SPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTP SGATGSPGASPGTSSTGSPR
Seg 110	EPTAATTGESAGTAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTG PGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPG SSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSST PSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSA STGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSS TGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGT GPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSP GSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGS STPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASP GTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSG ATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGAT GSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGS PGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPG SSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSST PSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 111	EPTAATTGESAGTAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTG PGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPG SSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSST PSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSA STGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSS TGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGT GPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSP GSSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAGTPGSGTASSSPGSSTPSGATGSP GSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPG SGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSG ATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTAS SSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGS STPSGATGSPGSSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 112	EPTAATTGESAGTAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTG PGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPG

Name	Amino Acid Sequence
	SSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSST PSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSA STGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSS TGSPTAEAAGCGTAEAAGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSS TGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTG SPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSP GSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPTAEA AGCGTAEAAGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPG SGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGT SSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSST GSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTG PGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAEA AR
Seg 113	AEATTAAGGAEEEGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASP GTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSG ATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTAS SSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSP GASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGA SPGTSSTGSPTAEAAGCGTAEAAGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGS STPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASP GTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSG ATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGAT GSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGS PGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPG SSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSST PSGATGSPGASPGTSSTGSPR
Seg 114	AEATTAAGGAEEETAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGT GPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSP GSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGS STPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPS ASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTS STGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGT GPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSP GSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGS STPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASP GTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSG ATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGAT GSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGS PGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPG SSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSST PSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 115	AEATTAAGGAEEETAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGT GPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSP GSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGS STPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPS ASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTS STGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGT GPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSP GSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGS STPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASP GTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSG ATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGAT GSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGS PGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPG SSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSST PSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAR

Name	Amino Acid Sequence
	STPSGATGSPGSSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 116	AEATTAAGGAEEETAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGT GPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSP GSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGS STPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPS ASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTS STGSPTAEAAGCGTAEAAGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTS STGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSST GSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGS PGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPG TPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPTAE AAGCGTAEAAGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTP GSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPG TSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSS TGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGT GPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAE AAR
Seg 117	ATGTATSEGSPEEEEGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSS PGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPG TPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGAS PGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPS GATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTA SSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSS PGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPG ASPGTSSTGSPTAEAAGCGTAEAAGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPG SSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGAS PGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPS GATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGA TGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATG SPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSP GSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGS STPSGATGSPGASPGTSSTGSPR
Seg 118	ATGTATSEGSPEEEETAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTG TGPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSS PGSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPG SSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSP SASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGT SSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTG TGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPG SSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGAS PGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPS GATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGA TGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATG SPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSP GSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGS STPSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 119	ATGTATSEGSPEEEETAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTG TGPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSS PGSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPG SSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSP SASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGT SSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTG TGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPG SSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGAS PGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPS GATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGA TGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATG SPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSP GSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGS STPSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAGTPGSGTASSSPGSSTPSGATGS PGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPG TPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTP GSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPS

Name	Amino Acid Sequence
	GATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTA SSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGS PGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPG SSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 120	ATGTATSEGSPEEEETAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTG TGPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSS PGSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPG SSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSP SASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGT SSTGSPTAEAAGCGTAEAAGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGT SSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSST GSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGS PGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPG TPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPTAE AAGCGTAEAAGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTP GSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPG TSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSS TGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGT GPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAE AAR
Seg 121	AEATTAAGGAEEEGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASP GTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSG ATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTAS SSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSP GASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGA SPGTSSTGSPTAEAAGCGTAEAAGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGS STPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASP GTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSG ATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGAT GSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGS PGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPG SSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSST PSGATGSPGASPGTSSTGSPRPRPRPRP
Seg 122	AEATTAAGGAEEETAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGT GPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSP GSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGS STPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPS ASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTS STGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGT GPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSP GSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGS STPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASP GTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSG ATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGAT GSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGS PGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPG SSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSST PSGATGSPGASPGTSSTGSPTAEAAGCGTAEAARPRPRPRP
Seg 123	AEATTAAGGAEEETAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGT GPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSP GSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGS STPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPS ASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTS STGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGT GPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSP GSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGS STPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASP GTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSG ATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGAT GSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGS PGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPG SSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSST PSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAGTPGSGTASSSPGSSTPSGATGSP

Name	Amino Acid Sequence
	GSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPG SGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSG ATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTAS SSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGS STPSGATGSPGSSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAEAARPRPRPRP
Seg 124	AEATTAAGGAEEETAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGT GPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSP GSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGS STPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPS ASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTS STGSPTAEAAGCGTAEAAGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTS STGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSST GSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGS PGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPG TPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPTAE AAGCGTAEAAGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTP GSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPG TSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSS TGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGT GPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAE AARPRPRPRP
Seg 125	AEATTAAGGAEEEGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASP GTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSG ATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTAS SSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSP GASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGA SPGTSSTGSPTAEAAGCGTAEAAGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGS STPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASP GTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSG ATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGAT GSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGS PGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPG SSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSST PSGATGSPGASPGTSSTGSPR
Seg 126	AEATTAAGGAEEETAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGT GPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSP GSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGS STPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPS ASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTS STGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGT GPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSP GSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGS STPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASP GTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSG ATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGAT GSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGS PGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPG SSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSST PSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 127	AEATTAAGGAEEETAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGT GPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSP GSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGS STPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPS ASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTS

Name	Amino Acid Sequence
	STGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGT GPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSP GSSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAGTPGSGTASSSPGSSTPSGATGSP GSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPG SGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSG ATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTAS SSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGS STPSGATGSPGSSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 128	AEATTAAGGAEEETAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGT GPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSP GSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGS STPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPS ASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTS STGSPTAEAAGCGTAEAAGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTS STGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSST GSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGS PGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPG TPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPTAE AAGCGTAEAAGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTP GSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPG TSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSS TGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGT GPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAE AAR
Seg 129	AEATTAAGGAEEETAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGT GPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSP GSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGS STPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPS ASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTS STGSPGGKPGGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSP SASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPS GATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGA TGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASS SPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPTAEAAGCGTAE AAGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSP GSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGT PGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASP GTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSG TASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGGKPGR
Seg 130	AEATTAAGGAEEETAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGT GPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSP GSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGS STPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPS ASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTS STGSPTAEAAGCGTAEAAGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTS STGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSST GSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGS PGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPG TPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPTAE AAGCGTAEAAGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTP GSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPG TSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSS TGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGT GPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGGKPGR
Seg 131	AEATTAAGGAEEEGGKPGGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSG TASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGAT

Name	Amino Acid Sequence
	GSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGS PGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPG SSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGGK PGGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGP GTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGS STPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTP SGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTS STGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGGKPGGTPGSGTASSSPGSST PSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPS GATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGA TGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATG SPGSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSP GSSTPSGATGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 132	AEATTAAGGAEEETAEAAGCGTAEAAGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGT GPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSP GSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGS STPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPS ASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTS STGSPGGKPGGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSP SASTGTGPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPS GATGSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGA TGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASS SPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPTAEAAGCGTAE AAGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSP GSSTPSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGT PGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASP GTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPS GATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGSGSSPSG TASSSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGGKPGR
Seg 133	AEATTAAGGAEEEGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASP GTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSG ATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTAS SSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSP GASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGA SPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPS ASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTS STGSPGASPGTSSTGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 134	AEATTAAGGAEEEGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASP GTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSG ATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTAS SSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSP GASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGA SPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPS ASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTS STGSPGASPGTSSTGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 135	AEATTAAGGAEEEGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASP GTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSG ATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPTAEAAGCG TAEAAR
Seg 136	AEATTAAGGAEEEGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 137	AEATTAAGGAEEEGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASP

Name	Amino Acid Sequence
	GTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSG ATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTAS SSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSP GASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGA SPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPS ASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTS STGSPGASPGTSSTGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 138	AEATTAAGGAEEEGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASP GTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSG ATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTAS SSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSP GASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGA SPGTSSTGSPTAEAAGCGTAEAAR
Seg 139	AEATTAAGGAEEEGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASP GTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSG ATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPTAEAAGCG TAEAAR
Seg 140	AEATTAAGGAEEEGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPTAEAAGCGTAEAAR
Seg 141	AEATTAAGGAEEEGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASP GTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSG ATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTAS SSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSP GASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGA SPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPS ASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTS STGSPGASPGTSSTGSPGASPGTSSTGSPGGKPGR
Seg 142	AEATTAAGGAEEEGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASP GTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSG ATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTAS SSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSP GASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGA SPGTSSTGSPGGKPGR
Seg 143	AEATTAAGGAEEEGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASP GTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSG ATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGGKPGR
Seg 144	AEATTAAGGAEEEGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGGKPGR
Seg 145	AEATTAAGGAEEEGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASP GTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSG ATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTAS SSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSP GASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGA SPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPS

Name	Amino Acid Sequence
	ASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTS STGSPGASPGTSSTGSPGASPGTSSTGSPGGKPGR
Seg 146	AEATTAAGGAEEEGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASP GTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSG ATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTAS SSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSP GASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGA SPGTSSTGSPGGKPGR
Seg 147	AEATTAAGGAEEEGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASP GTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSG ATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGGKPGR
Seg 148	AEATTAAGGAEEEGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSP GSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGGKPGR
Seg 149	GSEPATSGSETPGTSESATPESGPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGSE PATSGSETPGSEPATSGSETPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGSEPAT SGSETPGTSTEPSEGSAPGGGSPAGSCTSPGGSPAGSPTSTEEGTSESATPESGPGTSTEPS EGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSE TPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPG SPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSE SATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESAT PESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGS APGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPG SPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEP ATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPG
Seg 150	MAEPAGSPTSTEEGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGSPAGSPTSTEEG TSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSE SATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPS EGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPES GPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPG TSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPS EGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPG SPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSE SATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPS EGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTST EEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPG SEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEP ATSGSETPGTSESATPESGPGTSTEPSEGSAPGGASASCAPSTGGGSEPATSGSETPGTSES ATPESGPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGSEPATSG SETPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSA PG
Seg 151	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTS TEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESA TPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEG SAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSE PATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPE SGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTS TEPSEGSAPGGGSPAGSCTSPGGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEP

Name	Amino Acid Sequence
	ATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATS GSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPES GPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEG TSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG
Seg 152	MAEPAGSPTSTEEGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGGPEPTCPAPSGG GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTS TEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESA TPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEG SAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSE PATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPE SGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTS TEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPE SGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGP GTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTS TEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG
Seg 153	GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTS TEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGS PTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPE SGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAP GSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGGGSPAGSCTSPGGTSESATPESGPG SEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPA GSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSE TPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPGGGSPAGSCTSPGGTSESATPESGPGSEPATSGSETPGT SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSE GSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSET PGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGT STEPSEGSAPG
Seg 154	MAEPAGSPTSTEEGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGGASASCAPSTG GGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSE SATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGS APGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGS APGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPG TSTEPSEGSAPGGASASCAPSTGGGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGS PAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSETPGSEPATSGSETPGTSTE PSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPG
Seg 155	MAEPAGSPTSTEEGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGGGSPAGSCTSP GGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGSTSSTAESPGPGTSTPESGSASPGS TSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSTPESGSASPGSEPATSGSETPGTSES ATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSTEPSE GSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESG PGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGS PAGSPTSTEEGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTSTEPSEGSAPGTSTE PSEGSAPGSEPATSGSETPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGASASGA PSTGGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSTSSTAESPGPGSTSESPSGTA PGTSPSGESSTAPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSEPATSGSETPGT SESATPESGPGSEPATSGSETPGSTSSTAESPGPGSTSSTAESPGPGTSPSGESSTAPGSEPA

Name	Amino Acid Sequence
	TSGSETPGSEPATSGSETPGTSTEPSEGSAPGSTSSTAESPGPGTSTPESGSASPGSTSESPS GTAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSSTPSGATGSPGSSPSASTGT GPGASPGTSSTGSPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSSTPSGATGSP GSSPSASTGTGPGASPGTSSTGSPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPG
Seg 156	MAEPAGSPTSTEEGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGGASASCAPSTG GGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGSTSSTAESPGPGTSTPESGSASPGS TSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSTPESGSASPGSEPATSGSETPGTSES ATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSTEPSE GSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESG PGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGS PAGSPTSTEEGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTSTEPSEGSAPGTSTE PSEGSAPGSEPATSGSETPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGPEPTGPA PSGGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSPAGSPTSTEE GSPAGSPTSTEEGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSTSSTAESPGPGST SESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGSTSESPSGTAPGTSPSGESSTAPGTSTEP SEGSAPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSESATPE SGPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSPSGESSTAPGTSPSGESSTAP GTSPSGESSTAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGSSPSASTGTGPGSS TPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASAS GAPSTGGTSPSGESSTAPGSTSSTAESPGPGTSPSGESSTAPGTSESATPESGPGTSTEPSEG SAPGTSTEPSEGSAPGSSPSASTGTGPGSSTPSGATGSPGASPGTSSTGSPGTSTPESGSASP GTSPSGESSTAPGTSPSGESSTAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGST SESPSGTAPGSTSESPSGTAPGTSTPESGSASPGSPAGSPTSTEEGTSESATPESGPGTSTEP SEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGSSTPSGATGSPGASPGTSST GSPGSSTPSGATGSPGSTSESPSGTAPGTSPSGESSTAPGSTSSTAESPGPGSSTPSGATGSP GASPGTSSTGSPGTPGSGTASSSPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPG
Seg 157	MAEPAGSPTSTEEGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGGPEPTCPAPSGG GTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGSTSSTAESPGPGTSTPESGSASPGST SESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSTPESGSASPGSEPATSGSETPGTSESA TPESGPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEG SAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGP GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGSP AGSPTSTEEGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTSTEPSEGSAPGTSTEP SEGSAPGSEPATSGSETPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGASASGAP STGGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSTSSTAESPGPGSTSESPSGTAP GTSPSGESSTAPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSEPATSGSETPGTS ESATPESGPGSEPATSGSETPGSTSSTAESPGPGSTSSTAESPGPGTSPSGESSTAPGSEPAT SGSETPGSEPATSGSETPGTSTEPSEGSAPGSTSSTAESPGPGTSTPESGSASPGSTSESPSG TAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSSTPSGATGSPGSSPSASTGTGP GASPGTSSTGSPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSSTPSGATGSPGSS PSASTGTGPGASPGTSSTGSPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPG
Seg 158	MAEPAGSPTSTEEGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGGPEPTCPAPSGG GTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGSTSSTAESPGPGTSTPESGSASPGST SESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSTPESGSASPGSEPATSGSETPGTSESA TPESGPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEG SAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGP GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGSP AGSPTSTEEGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTSTEPSEGSAPGTSTEP SEGSAPGSEPATSGSETPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGASASGAP STGGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSTSSTAESPGPGSTSESPSGTAP GTSPSGESSTAPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSEPATSGSETPGTS ESATPESGPGSEPATSGSETPGSTSSTAESPGPGSTSSTAESPGPGTSPSGESSTAPGSEPAT SGSETPGSEPATSGSETPGTSTEPSEGSAPGSTSSTAESPGPGTSTPESGSASPGSTSESPSG TAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSSTPSGATGSPGSSPSASTGTGP GASPGTSSTGSPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSSTPSGATGSPGSS PSASTGTGPGASPGTSSTGSPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGGPEP TCPAPSGGMAEPAGSPTSTEEGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPG
Seg 159	GSEPATSGSETPGTSESATPESGPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGSE

Name	Amino Acid Sequence
	PATSGSETPGSEPATSGSETPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGSEPAT SGSETPGTSTEPSEGSAPGGGSPAGSKTSPGGSPAGSPTSTEEGTSESATPESGPGTSTEPS EGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSE TPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPG SPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSE SATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESAT PESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGS APGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPG SPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEP ATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPG
Seg 160	MAEPAGSPTSTEEGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGSPAGSPTSTEEG TSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSE SATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPS EGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPES GPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPG TSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPS EGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPG SPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSE SATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPS EGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTST EEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPG SEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEP ATSGSETPGTSESATPESGPGTSTEPSEGSAPGGASASKAPSTGGGSEPATSGSETPGTSES ATPESGPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGSEPATSG SETPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSA PG
Seg 161	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTS TEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESA TPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEG SAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSE PATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPE SGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTS TEPSEGSAPGGGSPAGSKTSPGGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEP ATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATS GSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPES GPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEG TSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG
Seg 162	MAEPAGSPTSTEEGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGGASASKAPSTG GGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSE SATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGS APGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGS APGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPES GPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG

Name	Amino Acid Sequence
Seg 163	GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTS TEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGS PTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPE SGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAP GSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGGGSPAGSKTSPGGTSESATPESGPG SEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPA GSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSE TPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPGGGSPAGSKTSPGGTSESATPESGPGSEPATSGSETPGT SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSE GSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSET PGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGT STEPSEGSAPG
Seg 164	MAEPAGSPTSTEEGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGGGSPAGSKTSP GGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSE SATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGS APGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGS APGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPG TSTEPSEGSAPGGGSPAGSKTSPGGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGS PAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSETPGSEPATSGSETPGTSTE PSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPG
Seg 165	MAEPAGSPTSTEEGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGGGSPAGSKTSP GGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGSTSSTAESPGPGTSTPESGSASPGS TSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSTPESGSASPGSEPATSGSETPGTSES ATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSTEPSE GSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESG PGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGS PAGSPTSTEEGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTSTEPSEGSAPGTSTE PSEGSAPGSEPATSGSETPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGASASGA PSTGGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSTSSTAESPGPGSTSESPSGTA PGTSPSGESSTAPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSEPATSGSETPGT SESATPESGPGSEPATSGSETPGSTSSTAESPGPGSTSSTAESPGPGTSPSGESSTAPGSEPA TSGSETPGSEPATSGSETPGTSTEPSEGSAPGSTSSTAESPGPGTSTPESGSASPGSTSESPS GTAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSSTPSGATGSPGSSPSASTGT GPGASPGTSSTGSPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSSTPSGATGSP GSSPSASTGTGPGASPGTSSTGSPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPG
Seg 166	MAEPAGSPTSTEEGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGGASASKAPSTG GGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGSTSSTAESPGPGTSTPESGSASPGS TSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSTPESGSASPGSEPATSGSETPGTSES ATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSTEPSE GSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESG PGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGS PAGSPTSTEEGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTSTEPSEGSAPGTSTE PSEGSAPGSEPATSGSETPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGPEPTGPA PSGGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSPAGSPTSTEE GSPAGSPTSTEEGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSTSSTAESPGPGST SESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGSTSESPSGTAPGTSPSGESSTAPGTSTEP SEGSAPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSESATPE SGPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSPSGESSTAPGTSPSGESSTAP GTSPSGESSTAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGSSPSASTGTGPGSS TPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASAS

Name	Amino Acid Sequence
	GAPSTGGTSPSGESSTAPGSTSSTAESPGPGTSPSGESSTAPGTSESATPESGPGTSTEPSEG SAPGTSTEPSEGSAPGSSPSASTGTGPGSSTPSGATGSPGASPGTSSTGSPGTSTPESGSASP GTSPSGESSTAPGTSPSGESSTAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGST SESPSGTAPGSTSESPSGTAPGTSTPESGSASPGSPAGSPTSTEEGTSESATPESGPGTSTEP SEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGSSTPSGATGSPGASPGTSST GSPGSSTPSGATGSPGSTSESPSGTAPGTSPSGESSTAPGSTSSTAESPGPGSSTPSGATGSP GASPGTSSTGSPGTPGSGTASSSPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPG
Seg 167	MAEPAGSPTSTEEGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGGASASKAPSTG GGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGSTSSTAESPGPGTSTPESGSASPGS TSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSTPESGSASPGSEPATSGSETPGTSES ATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSTEPSE GSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESG PGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGS PAGSPTSTEEGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTSTEPSEGSAPGTSTE PSEGSAPGSEPATSGSETPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGASASGA PSTGGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSTSSTAESPGPGSTSESPSGTA PGTSPSGESSTAPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSEPATSGSETPGT SESATPESGPGSEPATSGSETPGSTSSTAESPGPGSTSSTAESPGPGTSPSGESSTAPGSEPA TSGSETPGSEPATSGSETPGTSTEPSEGSAPGSTSSTAESPGPGTSTPESGSASPGSTSESPS GTAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSSTPSGATGSPGSSPSASTGT GPGASPGTSSTGSPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSSTPSGATGSP GSSPSASTGTGPGASPGTSSTGSPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPG
Seg 168	MAEPAGSPTSTEEGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGGASASKAPSTG GGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGSTSSTAESPGPGTSTPESGSASPGS TSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSTPESGSASPGSEPATSGSETPGTSES ATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSTEPSE GSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESG PGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGS PAGSPTSTEEGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTSTEPSEGSAPGTSTE PSEGSAPGSEPATSGSETPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGASASGA PSTGGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSTSSTAESPGPGSTSESPSGTA PGTSPSGESSTAPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSEPATSGSETPGT SESATPESGPGSEPATSGSETPGSTSSTAESPGPGSTSSTAESPGPGTSPSGESSTAPGSEPA TSGSETPGSEPATSGSETPGTSTEPSEGSAPGSTSSTAESPGPGTSTPESGSASPGSTSESPS GTAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSSTPSGATGSPGSSPSASTGT GPGASPGTSSTGSPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSSTPSGATGSP GSSPSASTGTGPGASPGTSSTGSPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGG ASASKAPSTGGMAEPAGSPTSTEEGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPG
Seg 169	GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTS TEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGS PTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPE SGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAP GSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGGGSPAGSCTSPGGTSESATPESGPG SEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPA GSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSE TPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPGGGSPAGSKTSPGGTSESATPESGPGSEPATSGSETPGT SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSE GSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSET PGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGT STEPSEGSAPG
Seg 170	MAEPAGSPTSTEEGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGGGSPAGSCTSP GGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGS EPATSGSETPGSEPATSGSETPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGSEPA TSGSETPGTSTEPSEGSAPGGGSPAGSCTSPGSEPATSGSETPGTSESATPESGPGSEPATS GSETPGSPAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSETPGSEPATSGSE

Name	Amino Acid Sequence
	TPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGGGSPAGSCTSP GGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGS EPATSGSETPGSEPATSGSETPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGSEPA TSGSETPGTSTEPSEGSAPGGGSPAGSCTSPGGSEPATSGSETPGTSESATPESGPGSEPAT SGSETPGSPAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSETPGSEPATSGS ETPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPG
Seg 171	MAEPAGSPTSTEEGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGGGSPAGSKTSP GGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGS EPATSGSETPGSEPATSGSETPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGSEPA TSGSETPGTSTEPSEGSAPGGGSPAGSKTSPGSEPATSGSETPGTSESATPESGPGSEPATS GSETPGSPAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSETPGSEPATSGSE TPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGGGSPAGSKTSP GGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGS EPATSGSETPGSEPATSGSETPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGSEPA TSGSETPGTSTEPSEGSAPGGGSPAGSKTSPGGSEPATSGSETPGTSESATPESGPGSEPAT SGSETPGSPAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSETPGSEPATSGS ETPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPG
Seg 172	SAGSPTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPS EGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPES GPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSE SATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPTAEAAGCGTAEAASASR
Seg 173	SAGSPGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGT STEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSES ATPESGPGTSTEPSEGSAGPTKPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTS TEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPTKPGTSPTSTEEGTSESATPE SGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAP GTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPTAE AAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTS ESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEP SEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPE SGPGTSTEPSEGSAPTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGP GTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTS TEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPTAEAAGCGTAEAASASR
Seg 174	SAGSPTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSE TPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEG TSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEP ATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSP TSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGS APGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEG TSTEPSEGSAPTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPA GSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATS GSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTST EEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEG TSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEP ATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESAT PESGPGTSTEPSEGSAPTAEAAGCGTAEAASASR
Seg 175	SAGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGS APGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEG TSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTST EPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPS EGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPES

Name	Amino Acid Sequence
	GPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPTAEA AGCGTAEAAPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTAEAAGCGTAEAA STEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPATAEAAGCGTAEAASPTSTEEGTSE SATPESGPGTSTEPSEGSAPGTSESATTAEAAGCGTAEAASETPGTSESATPESGPGSEPA TSGSETPGTSESATPESGTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGSEPAT SGSETPGTTAEAAGCGTAEAAAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESTA EAAGCGTAEAATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGTAEAAGCGTAE AATEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPTAEAAGCGTAEAASASR
Seg 176	SAGSPTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPS EGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPES GPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSE SATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPTAEAAGCGTAEAASASR
Seg 177	SAGSPTAEAAGCGTAEAAPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTAEAA GCGTAEAASTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPATAEAAGCGTAEAASP TSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATTAEAAGCGTAEAASETPGTSESATPE SGPGSEPATSGSETPGTSESATPESGTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPES GPGSEPATSGSETPGTTAEAAGCGTAEAAAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSA PGTSESTAEAAGCGTAEAATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGTAEA AGCGTAEAATEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPTAEAAGCGTAEA ASASR
Seg 178	SAGSPTGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPS EGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTST EEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPG SEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEP ATSGSETPGTSESATPESGPGTSTEPSEGSAPSASR
Seg 179	GSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTS ESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEP SEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEG SAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGP GSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESA TPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEG SAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGP GTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTS ESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPE SGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAP GTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGR
Seg 180	GSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTS ESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEP SEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEG SAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGP GSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPTAEAA GKPGTAEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPT STEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESG PGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGS EPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSES ATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPT STEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSA PGK

Name	Amino Acid Sequence
Seg 181	SAGSPTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSE TPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEG TSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEP ATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSP TSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGS APGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSP TSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPES GPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSE SATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGS APSASR
Seg 182	CGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSA PGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGT SESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTE PSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSE GSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESG PGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGT STEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSE GSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESG PGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGT SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATP ESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSA PGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGR
Seg 183	MKNPEQAEEQAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPTGPGSEPATSGS ETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTS TEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPAT SGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPE SGPGTSTEPSEGSAPSASRSAHHHHHHHHH
Seg 184	MKNPEQAEEQAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPTAEAAGCGTAE AAPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTAEAAGCGTAEAASTEPSEGS APGTSESATPESGPGSPAGSPTSTEEGSPATAEAAGCGTAEAASPTSTEEGTSESATPESG PGTSTEPSEGSAPGTSESATTAEAAGCGTAEAASETPGTSESATPESGPGSEPATSGSETP GTSESATPESGTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPG TTAEAAGCGTAEAAAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESTAEAAGCG TAEAATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGTAEAAGCGTAEAATEEGT STEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPTAEAAGCGTAEAASASRSAHHHHHHH H
Seg 185	MKNPEQAEEQAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPGSPAGSPTSTEE GTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTS ESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEP SEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPE SGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAP GTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTS ESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPTAEAAGCGTAEAAPGSEPA TSGSETPGTSESATPESGPGSEPATSGSETPGTAEAAGCGTAEAASTEPSEGSAPGTSESA TPESGPGSPAGSPTSTEEGSPATAEAAGCGTAEAASPTSTEEGTSESATPESGPGTSTEPSE GSAPGTSESATTAEAAGCGTAEAASETPGTSESATPESGPGSEPATSGSETPGTSESATPE SGTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTTAEAAGC GTAEAAAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESTAEAAGCGTAEAATPES GPGTSESATPESGPGSEPATSGSETPGSEPATSGTAEAAGCGTAEAATEEGTSTEPSEGSA

Name	Amino Acid Sequence
	PGTSTEPSEGSAPGSEPATSGSETPTAEAAGCGTAEAASASRSAHHHHHHHHH
Seg 186	GSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTS ESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEP SEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEG SAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGP GSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTS TEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPTAEAA GKPGTAEAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPT STEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESG PGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGS EPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSES ATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPT STEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSA PGK
Seg 187	SAGSPTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSE TPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEG TSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEP ATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSP TSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGS APGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSP TSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPES GPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSE SATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGS APSASR
Seg 188	SAGSPTEGTSTEPSEGSAPGTSESTAEAAGCGTAEAATPESGPGTSESATPESGPGSEPAT SGSETPGSEPATSGTAEAAGCGTAEAATEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATS GSETPTAEAAGCGTAEAASASR
Seg 189	SAGSPTPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATP ESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTE EGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGS EPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATP ESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTE EGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPTAEAAGCGTAEAASASR
Seg 190	SAGSPTGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEG SAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGP GTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTS ESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGSE PATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSEST AEAAGCGTAEAATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPTAEAAGCGTAEAASASR
Seg 191	SAGSPTPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATP ESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTE EGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGS EPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESTAE AAGCGTAEAATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGTAEAAGCGTAEA ATEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPTAEAAGCGTAEAASASR
Seg 192	SAGSPGSTSSTAESPGPGSTSSTAESPGPGCTSESPSGTAPGSTSSTAESPGPGSTSSTAESP GPGTSTPESGSASPGSTSCSPSGEAPGTSPSGESSTAPGSTSESPSGTAPGSTSESPSGTAPE TSPSGESCTAPGSTSASR

Name	Amino Acid Sequence
Seg 193	SAGSPGTPGSGTASSSPGSSTPSGATGSPGCAGSGTASSSPGSSTPSGATGSPGTPGSGTAS SSPGSSTPSGATGSPGSSTCSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGS PGTPGSGTACSSPGSSSASR
Seg 194	SAGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGS APGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEG TSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTST EPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPS EGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPES GPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSE SATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPS EGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPES GPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESAT PESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGS APGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTESASK
Seg 195	SAGSPTGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPS EGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTST EEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPG SEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEP ATSGSETPGTSESATPESGPGTSTACSEGSAPSASR
Seg 196	SAGSPTGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPS EGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTST EEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPG SEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEP ATSGSETPGTSESATPESGPGTSTEPSEGSCASASR
Seg 197	SAGSPGSCAGSPTSTEEGTSESACPESGPGTSTEPSEGSCPGSPAGSPTSTEEGTCTEPSEG SAPGTSTEPCSGSAPGTSESATPESCPGSEPATSGSETPGSCPATSGSETPGSPAGSCTSTEE GTSESATPESCPGTESASR
Seg 198	SAGSPTGCGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPS EGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTST EEGSPAGSPTSTEEGTSCTPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPG SEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEP ATSGSETPGTSESATPESGPGTSTEPSCGSAPSASR
Seg 199	SAGSPTGCGSEPATSGSETPGTSESATPESGPGSEPATSGSCTPGTSESATPESGPGTSTEPS EGSAPGSPAGSPCSTEEGTSESATPESGPGSEPATSGSETPGTSESCTPESGPGSPAGSPTST EEGSPAGSPTSTEEGTSCTPSEGSAPGTSESATPESGPGTSESATPESGPGCSESATPESGP GSEPATSGSETPGSEPATSGSETCGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGCAPGS EPATSGSETPGTSESATPESGPGTSTEPSCGSAPSASR

[00152] In another embodiment, the invention provides inserts of cysteine as part of a longer sequence defined as a cysteine island. Examples of cysteine island are shown in Table 3. Examples of lysine island are shown in Table 2. The benefit of flanking all cysteine residues in an XTEN with similar or identical sequence is that it results in a more uniform chemical reactivity for each cysteine. Another benefit results from the ability to perform peptide mapping to measure the degree of payload conjugation. Examples include islands I_C4, I_C7, I_C8, and I_C9 of Table 3. These islands comprise glutamate residues that facilitate peptide mapping using GluC protease. The islands can be inserted into constructs encoding the existing XTEN by conventional PCR methods, as described

above and in the Examples. Oligonucleotides encoding the islands can be inserted into constructs encoding the existing XTEN by conventional PCR methods. For example, in one embodiment, where an existing full-length XTEN gene is to be modified with nucleotides encoding one or more reactive cysteine or lysine residues, an oligonucleotide can be created that encodes a cysteine or lysine and that exhibits partial homology to and can hybridize with one or more short sequences of the XTEN, resulting in a recombination event and substitution of a cysteine or the lysine codon for an existing codon of the XTEN gene (see, e.g., Examples 6 and 7 for a description of the general methods). In one exemplary embodiment, the recombination results in a replacement with the amino acid sequence GGSPAGSCTSP of the I_C1 island. However, the oligonucleotides can be designed to place the cysteine (or lysine) in a different location in the motif or to include a second cysteine (or lysine) in the motif. The cysteine- or lysine-encoding oligonucleotides can be designed to hybridize with a given sequence segment at different points along the known XTEN sequence to permit their insertion into an XTEN-encoding gene. Thus, the invention contemplates that multiple XTEN gene constructs can be created with cysteines or lysines inserted at different locations within the XTEN sequence by the selection of restriction sites within the XTEN sequence and the design of oligonucleotides appropriate for the given location and that encode a cysteine or lysine, including use of designed oligonucleotides that result in multiple insertions in the same XTEN sequence. By the design and selection of one or more such oligonucleotides in consideration of the known sequence of the XTEN, and the appropriate use of the methods of the invention, the potential number of substituted reactive cysteine or lysine residues inserted into the full-length XTEN can be estimated and then confirmed by sequencing the resulting XTEN gene.

Table 2: Examples of lysine islands

Designator	Amino Acid Sequence
I_L1	GGSPAGSKPTSP
I_L2	GASASKPAPSTG
I_L3	PKP
I_L4	PPKPP
I_L5	GGKPG
I_L6	EGGKPGES
I_L7	EGGSPAGSKPTSPE
I_L8	EGASASKPAPSTGE

Table 3: Examples of cysteine islands

Designator	Sequence
I_C1	GGSPAGSCTSP
I_C2	GASASCAPSTG
I_C3	GPEPTCPAPSG
I_C4	TAEAAGCGTAEAA
I_C5	GECEP
I_C6	GRPCRP
I_C7	GETSPAGSCTSPTET
I_C8	TESGRPCRPSET
I_C9	GPEPTCPAPSEG

[00153] XTEN can be designed to comprise both lysine and cysteine residues for conjugation as illustrated in Figure 1D. This enables one to conjugate two different payloads to the same XTEN polymer using conjugations methods tailored to react with the functional group or linker attached to the cysteine or lysine. Such mixed payloads can have additive and/or synergistic pharmacologic effects when administered to a subject in a single composition. Alternatively, the mixed payloads can be a combination of a targeting moiety and an active payload in order to deliver the pharmacophore to a desired location in the subject. By controlling the number and position of lysine and cysteine residues one can control the number and position of conjugated payloads. This enables one to adjust the relative potency or selectivity of the payloads in the resulting XTEN-folate conjugate.

[00154] The design, selection, and preparation methods of the invention enable the creation of engineered XTEN that are reactive with electrophilic functionality. The methods to make the subject conjugates provided herein enable the creation of XTEN-folate conjugates, XTEN-cross-linker conjugates, and XTEN-azide/alkyne reactant conjugates with the linker or payload molecules added in a quantified fashion at designated sites, as illustrated schematically in FIG. 1. Payloads, cross-linkers, and azide/alkyne reactants may be site-specifically and efficiently linked to the N- or C-terminus of XTEN, to cysteine-engineered XTEN with a thiol-reactive reagent, or to lysine-engineered XTEN of the invention with an amine-reactive reagent, and to an alpha amino group at the N-terminus of XTEN, as described more fully, below, and then are purified and characterized as shown schematically in FIG. 40 using, for example, the non-limiting methods described more specifically in the Examples.

3. XTEN Segments from XTEN Precursors

[00155] In another aspect, the invention provides XTEN with cleavage sequences incorporated internal to the sequence at defined intervals such that the XTEN can be processed by cleavage into 2, 3, 4, 5, or 6 shorter XTEN of uniform lengths. As illustrated in FIG. 44A, a monomeric XTEN is designed with two internal cleavage sequences that, when treated with a protease under conditions effective to result in the cleavage of all cleavage sequences, results in three XTEN segments of uniform length. In addition, the XTEN are designed with a sequence such that the resulting XTEN segments also have the identical amino acid sequence, inclusive of the residual cleavage sequence. In one embodiment, the invention provides an XTEN with a defined, sequence comprising 1, 2, 3, 4, or 5 arginine (R) residues internal to the XTEN sequence and spaced at uniform intervals along the XTEN sequence bridging identical XTEN segments wherein treatment with trypsin results in cleavage of the XTEN into XTEN segments to having an identical length and sequence. In the foregoing embodiment, the arginine residue does not have a proline residue at the adjacent P1' position. Thus, by treatment of the foregoing with trypsin, an XTEN with 1 internal arginine would result in 2 identical XTEN segments, an XTEN with 2 internal argninines would result in 3 identical XTEN segments, etc. In another embodiment, each arginine of the foregoing embodiments is replaced with lysine residues. In another embodiment, the invention provides an XTEN with a defined sequence comprising 1, 2, 3, 4, or 5 cleavage sequences internal to the XTEN sequence and spaced at uniform intervals along the XTEN sequence, wherein each cleavage sequence is SASRSA, and wherein treatment with trypsin results in cleavage of the XTEN into XTEN segments to having an identical length and sequence. In another embodiment, the invention provides an XTEN with 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% sequence identity to a sequence selected from the group of sequences set forth in Table 4. In another embodiment, the invention provides an XTEN with 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% sequence identity to a sequence selected from the group of sequences set forth in Table 4, wherein the XTEN further comprises a first and a second affinity tag wherein each affinity tags are linked to the XTEN by a cleavage sequence at

the N- and C-termini of the XTEN, respectively, wherein each cleavage sequence is capable of being cleaved by trypsin, and wherein the first affinity tag is different from the second affinity tag and each is independently selected from the group consisting of the affinity tags set forth in Table 5. The foregoing embodiment is illustrated in FIG. 44B, wherein the treatment with protease of the XTEN with two internal cleavage sequences and an N-terminal and a C-terminal affinity tag each linked to the XTEN by cleavage sequence results in cleavage of the construct into three XTEN segments of uniform length and liberation of the two affinity tags, the resulting preparation of which can be subsequently processed into substantially homogeneous XTEN as described herein, below. Variations of XTEN comprising such uniform cleavage sequences and their distribution in the sequence are contemplated by the invention.

Table 4: Precursor XTEN with Internal Cleavage Sequences

XTEN Name	Amino Acid Sequence
AE864_R2 (2x AE432_R1)	SAGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPS EGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGS PTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPAT SGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGS PTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAG SPTSTEEGTESASRSAGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEP ATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSE SATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSE SATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTST EPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGSPAGSPTSTEEGTESASR
AE864_R3 (3x AE288_R1)	SAGSPTGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTST EPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPA GSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSE SATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTST EPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPSASRSAGSPTGPGSEP ATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPA GSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPA GSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEP ATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEP ATSGSETPGTSESATPESGPGTSTEPSEGSAPSASRSAGSPTGPGSEPATSGSETPGTS ESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTS ESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTS TEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSE PATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTS ESATPESGPGTSTEPSEGSAPSASR
AE864_R6 (6x AE144_R1)	SAGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPS EGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGS PTSTEEGTSESATPESGPGTESASRSAGSPGSPAGSPTSTEEGTSESATPESGPGTSTEP SEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPAT SGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTESASRSAGSPGSPAG SPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTE

XTEN Name	Amino Acid Sequence
	PSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSES ATPESGPGTESASRSAGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEP ATSGSETPGSPAGSPTSTEEGTSESATPESGPGTESASRSAGSPGSPAGSPTSTEEGTS ESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTS ESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTE SASRSAGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTS TEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPTSTEEGTSESATPESGPGTESASR
Seg 200 (3x Seg 195)	SAGSPTGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTST EPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPA GSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSE SATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTST EPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTACSEGSAPSASRSAGSPTGPGSE PATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSP AGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSP AGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSE PATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSE PATSGSETPGTSESATPESGPGTSTACSEGSAPSASRSAGSPTGPGSEPATSGSETPGT SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT SESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGT STEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGS EPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGT SESATPESGPGTSTACSEGSAPSASR
Seg 201 (3x Seg 196)	SAGSPTGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTST EPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPA GSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSE SATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTST EPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSCASASRSAGSPTGPGSE PATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSP AGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSP AGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSE PATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSE PATSGSETPGTSESATPESGPGTSTEPSEGSCASASRSAGSPTGPGSEPATSGSETPGT SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT SESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGT STEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGS EPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGT SESATPESGPGTSTEPSEGSCASASR

4. Net charge

[00156] In other embodiments, the XTEN polypeptides have an unstructured characteristic imparted by incorporation of amino acid residues with a net charge and containing a low percentage or no hydrophobic amino acids in the XTEN sequence. The overall net charge and net charge density is controlled by modifying the content of charged amino acids in the XTEN sequences, either positive or negative, with the net charge typically represented as the percentage of amino acids in the polypeptide contributing to a charged state beyond those residues that are cancelled by a residue with an opposing charge. In some embodiments, the net charge density of the XTEN of the conjugates 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. In other embodiments, the net charge of an XTEN can 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. Based on the net charge, some XTENs 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 even 6.5. In one embodiment, the XTEN will have an isoelectric point between 1.5 and 4.5 and carry a net negative charge under physiologic conditions.

[00157] Since most tissues and surfaces in a human or animal have a net negative charge, in some embodiments the XTEN sequences are designed to have a net negative charge to minimize non-specific interactions between the XTEN containing compositions and various surfaces such as blood vessels, healthy tissues, or various receptors. Not to be bound by a particular theory, an XTEN can adopt open conformations due to electrostatic repulsion between individual amino acids of the XTEN polypeptide that individually carry a net negative charge and that are distributed across the sequence of the XTEN polypeptide. In some embodiments, the XTEN sequence is designed with at least 90% to 95% of the charged residues separated by other non-charged residues such as serine, alanine, threonine, proline or glycine, which leads to a more uniform distribution of charge, better expression or purification behavior. Such a uniform distribution of net negative charge in the extended sequence lengths of XTEN also contributes to the unstructured conformation of the polymer that, in turn, can result in an effective increase in hydrodynamic radius. In preferred embodiments, the negative charge of the subject XTEN is conferred by incorporation of glutamic acid residues. Generally, the glutamic residues are spaced uniformly across the XTEN sequence. In some cases, the XTEN can contain about 10-80, or about 15-60, or about 20-50 glutamic residues per 20kDa of XTEN that can result in an XTEN with charged residues that would have very similar pKa, which can increase the charge homogeneity of the product and sharpen its isoelectric point, enhance the physicochemical properties of the resulting XTEN-folate conjugate for, and hence, simplifying purification procedures. Accordingly, in one embodiment the invention provides XTEN in which the XTEN sequences contain about 1%, 2%, 4%, 8%, 10%, 15%, 17%, 20%, 25%, or even about 30% glutamic acid. In some cases, the XTEN can contain about 10-80, or about 15-60, or about 20-50 glutamic residues per 20kDa of XTEN that can result in an XTEN with charged residues that would have very similar pKa, which can increase the charge homogeneity of the product and sharpen its isoelectric point, enhance the physicochemical properties of the resulting XTEN conjugate, and hence, simplifying purification procedures.

5. Low immunogenicity

[00158] In another aspect, the invention provides XTEN-folate conjugates having a low degree of immunogenicity or are substantially non-immunogenic. Several factors can contribute to the low immunogenicity of XTEN, e.g., the non-repetitive sequence, the unstructured conformation,

the high degree of solubility, the low degree or lack of self-aggregation, the low degree or lack of proteolytic sites within the sequence, and the low degree or lack of epitopes in the XTEN sequence.

[00159] Conformational epitopes are 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 the latter case, the immune response to a protein in an individual is heavily influenced by T-cell epitope recognition that is a function of the peptide binding specificity of that individual's HLA-DR allotype. Engagement of a MHC Class II peptide complex by a cognate T-cell receptor on the surface of the T-cell, together with the cross-binding of certain other co-receptors such as the CD4 molecule, can induce an activated state within the T-cell. Activation leads to the release of cytokines further activating other lymphocytes such as B cells to produce antibodies or activating T killer cells as a full cellular immune response.

[00160] The ability of a peptide to bind a given MHC Class II molecule for presentation on the surface of an APC (antigen presenting cell) is dependent on a number of factors; most notably its primary sequence. In one embodiment, a lower degree of immunogenicity is achieved by designing XTEN sequences that resist antigen processing in antigen presenting cells, and/or choosing sequences that do not bind MHC receptors well. The invention provides XTEN-folate conjugates with substantially non-repetitive XTEN polypeptides designed to reduce binding with MHC II receptors, as well as avoiding formation of epitopes for T-cell receptor or antibody binding, resulting in a low degree of immunogenicity. Avoidance of immunogenicity can attribute to, at least in part, a result of the conformational flexibility of XTEN sequences; i.e., the lack of secondary structure due to the selection and order of amino acid residues. For example, of particular interest are sequences having a low tendency to adapt compactly folded conformations in aqueous solution or under physiologic conditions that could result in conformational epitopes. The administration of compositions comprising XTEN, using conventional therapeutic practices and dosing, would generally not result in the formation of neutralizing antibodies to the XTEN sequence, and also reduce the immunogenicity of the payload in the conjugates.

[00161] In one embodiment, the XTEN sequences utilized in the subject polypeptides can be substantially free of epitopes recognized by human T cells. The elimination of such epitopes for the purpose of generating less immunogenic proteins has been disclosed previously; see for example WO 98/52976, WO 02/079232, and WO 00/3317 which are incorporated by reference herein. Assays for human T cell epitopes have been described (Stickler, M., *et al.* (2003) *J Immunol Methods*, 281: 95-108). Of particular interest are peptide sequences that can be

oligomerized without generating T cell epitopes or non-human sequences. This is achieved by testing direct repeats of these sequences for the presence of T-cell epitopes and for the occurrence of 6 to 15-mer and, in particular, 9-mer sequences that are not human, and then altering the design of the XTEN sequence to eliminate or disrupt the epitope sequence. In some embodiments, the XTEN sequences are substantially non-immunogenic by the restriction of the numbers of epitopes of the XTEN predicted to bind MHC receptors. With a reduction in the numbers of epitopes capable of binding to MHC receptors, there is a concomitant reduction in the potential for T cell activation as well as T cell helper function, reduced B cell activation or upregulation and reduced antibody production. The low 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), as shown in Example 35. 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 (corresponding to binding constraints of $10e^{10}$ K_d to $10e^{-10}$ K_d), 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, an XTEN component incorporated into either a XTEN-folate, XTEN-cross-linker, or XTEN-click-chemistry reactant conjugate does not have a predicted T-cell epitope at a TEPITOPE threshold score of about -5, or -6, or -7, or -8, or -9, or at a TEPITOPE score of -10. As used herein, a score of "-9" is a more stringent TEPITOPE threshold than a score of -5.

6. Increased Hydrodynamic radius

[00162] In another aspect, a subject XTEN useful as a conjugation partner has a high hydrodynamic radius; a property that confers a corresponding increased apparent molecular weight to the XTEN-folate conjugates compared to the payload without the XTEN. As detailed in Example 31, the linking of XTEN to therapeutic protein sequences results in compositions that can have increased hydrodynamic radii, increased apparent molecular weight, and increased apparent molecular weight factor compared to a therapeutic protein not linked to an XTEN. For example, in therapeutic applications in which prolonged half-life is desired, conjugates in which one or more XTEN with a high hydrodynamic radius are conjugated to a payload can effectively enlarge the hydrodynamic radius of the conjugate beyond the glomerular pore size of approximately 3-5 nm (corresponding to an apparent molecular weight of about 70 kDa) (Caliceti. 2003. Pharmacokinetic and biodistribution properties of poly(ethylene glycol)-protein conjugates. Adv Drug Deliv Rev 55:1261-1277), resulting in reduced renal clearance of circulating proteins with a corresponding increase in terminal half-life and other enhanced pharmacokinetic properties. The hydrodynamic radius of a protein is conferred by its molecular weight as well as

by its structure, including shape or compactness. The open, extended and unstructured conformation of the XTEN polypeptide has a greater proportional hydrodynamic radius compared to polypeptides of a comparable sequence length and/or molecular weight that have secondary 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. Patent Nos. 6,406,632 and 7,294,513. Example 31 demonstrates that increases in XTEN length result in proportional increase in the hydrodynamic radius, apparent molecular weight, and/or apparent molecular weight factor, and thus permit the tailoring of an XTEN-folate conjugate composition to desired cut-off values of apparent molecular weights or hydrodynamic radii. Accordingly, in certain embodiments, the XTEN-folate conjugate composition can be configured with an XTEN such that the resulting conjugate can have a hydrodynamic radius of at least about 5 nm, or at least about 8 nm, or at least about 10 nm, or about 12 nm, or about 15 nm, or about 20 nm, or about 30 nm or more. In the foregoing embodiments, the large hydrodynamic radius conferred by the XTEN in a XTEN-folate conjugate can lead to reduced clearance of the resulting conjugate, an increase in terminal half-life, and an increase in mean residence time. As described in the Examples, when the molecular weights of the XTEN-containing conjugates are derived from size exclusion chromatography analyses, the open conformation of the XTEN due to the low degree of secondary structure results in an increase in the apparent molecular weight of the conjugates into which they are incorporated. In one embodiment, the present invention makes use of the discovery that the increase in apparent molecular weight can be accomplished by the linking not only of a single XTEN of a given length, but also by the linking of two XTEN of proportionally shorter lengths, either in linear fashion or as a dimeric, branched configuration, as described more fully, below. In some embodiments, the XTEN comprising a payload and one or more XTEN exhibits an apparent molecular weight of at least about 400 kD, or at least about 500 kD, or at least about 700 kD, or at least about 1000 kD, or at least about 1400 kD, or at least about 1600 kD, or at least about 1800kD, or at least about 2000 kD. Accordingly, the XTEN-folate conjugate exhibits an apparent molecular weight that is about 1.3-fold greater, or about 2-fold greater, or about 3-fold greater or about 4-fold greater, or about 8-fold greater, or about 10-fold greater, or about 12-fold greater, or about 15-fold, or about 20-fold greater than the actual molecular weight of the conjugate. In one embodiment, the isolated XTEN-folate conjugate of any of the embodiments disclosed herein exhibit an apparent molecular weight factor under physiologic conditions that is greater than about 1.3, or about 2, or about 3, or about 4, or about 5, or about 6, or about 7, or about 8, or about 10, or greater than about 15. In another embodiment, the XTEN-folate conjugate composition has, under physiologic conditions, an apparent molecular weight factor that is about 3 to about 20, or is about 5 to about 15, or is about 8 to about 12, or is about 9 to about 10 relative to the actual molecular weight of the conjugate. Generally, the

increased apparent molecular weight of the subject XTEN-folate conjugates enhances the pharmacokinetic properties of the composition by a combination of factors, which include reduced active clearance, reduced renal clearance, and reduced loss through capillary and venous junctions.

7. Compositions and methods of purifying XTEN as substantially homogeneous preparations

[00163] It is an object of the invention to provide compositions of XTEN and methods of making preparations comprising XTEN with a high level of purity and uniformity in the length and composition of the XTEN described herein, for use of conjugation partners to make the subject XTEN-folate conjugates.

[00164] The expression of recombinant XTEN protein or a recombinant fusion protein comprising XTEN in a host cell normally, like any globular protein, results in a mixture of different compounds in which a portion are truncated versions of the desired protein length. The truncation can be the result of early termination of translation, mRNA instability, or proteolysis in the host cell. Because globular proteins generally have efficient or complete folding into their three-dimensional structure while truncated versions do not, typical purification and recovery processes can successfully separate and remove the truncated versions such that a high level of product homogeneity is achieved in a given preparation of globular proteins. However, protein polymers such as XTEN are unique in that, given their unstructured nature, they don't fold into three-dimensional structures. As a consequence, until now, it has not been possible to obtain a substantially homogeneous preparation of full-length sequences of XTEN from a mixture of XTEN of different lengths in a crude expression product. This is because incomplete or truncated XTEN chains differ only slightly in their physicochemical properties from the desired full-length sequences such that traditional processes that would be sufficient for purification of globular proteins are not effective in the removal of truncated XTEN from the expression product in order to obtain a substantially homogeneous preparation of full-length sequences. While the subject XTEN of the invention, including XTEN linked to payload, can be purified to a moderate degree of homogeneity by conventional means used for proteins, such as salt fractionation, ion exchange chromatography, size exclusion chromatography, hydroxyapatite adsorption chromatography, hydrophobic interaction chromatography or gel electrophoresis, these methods alone do not result in preparations wherein the XTEN are substantially homogeneous in sequence length.

[00165] It has now been discovered that the XTEN derived from complex fermentation mixtures utilized in the preparation of the subject compositions can be purified from the heterogeneous population of XTEN truncation products such that they are substantially homogeneous with respect to sequence length if the XTEN are designed to further comprise, as a fusion protein, affinity tags located at either or both of the N- and C-termini of the XTEN such

that the expressed product can be subject to purification methods to selectively capture the full-length expressed polypeptide, thereby removing truncated XTEN by-products (see FIGS. 41-42). Non-limiting examples of affinity tags that can be added to the termini of XTEN are presented in Table 5. Non-limiting examples of methods of the design, expression, and purification methods to achieve substantially homogeneous XTEN are described in the Examples.

[00166] In some embodiments, the invention provides substantially homogeneous polypeptide compositions with XTEN fused directly to one affinity tag (such as, but not limited to the tags of Table 5) linked to either the N- or C-terminus of the XTEN. In other embodiments, the invention provides substantially homogeneous polypeptide compositions with XTEN fused to one affinity tag (such as, but not limited to the tags of Table 5) by a cleavage sequence linked to either the N- or C-terminus of the XTEN. In other embodiments, the invention provides substantially homogeneous polypeptide compositions with XTEN fused directly to one or two different affinity tags (such as, but not limited to the tags of Table 5) linked to the N- and/or C-termini of the XTEN, as shown in FIG. 21. In other embodiments, the invention provides substantially homogeneous compositions with XTEN fused to one or two cleavage sequences (such as, but not limited to the cleavage sequences of Table 8 or Table 9) which, in turn, are each fused to different affinity tags (such as, but not limited to the tags of Table 5) linked to the N- or C-termini or both the N- and C-termini of the XTEN, as shown in FIG. 21. In yet other embodiments, the invention provides substantially homogeneous polypeptide compositions with XTEN fused directly to one or two different affinity tags (such as, but not limited to the tags of Table 5) linked to the N- and/or C-termini of the XTEN that further comprise a helper sequence (such as, but not limited to the sequences of Table 12) fused to the N-terminus of the protein. As used in the context of the proteins described herein, "substantially homogeneous" means that at least about 85%, or 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% of the polypeptide sequences of the preparation have an identical sequence length. The percent values are based on the area percentage of the chromatogram of the preparation analyzed by HPLC or the area percentage of the scan of the preparation analyzed by SDS-PAGE, or by other such standard procedures known in the art for assessing the purity of proteins.

[00167] The polypeptide constructs comprising affinity tags have the advantageous property, compared to XTEN not linked to affinity tags, of being able to be purified to substantially homogeneous length by use of chromatography substrates to which the affinity tags will bind. In some embodiments, the categories of chromatography substrates used in the method of purification are selected from the chromatography substrates set forth in Table 5, which are utilized for the purification of XTEN linked to the corresponding the indicated affinity tag in the tables. As will be appreciated by one of skill in the art, the categories of chromatography

substrate can encompass different chemical groups linked to different matrices or resins; e.g., anion exchange substrates include quaternary trimethylammonium and diethylaminoethyl bound to resins, cation exchange substrates include sulfo or sulfopropyl or carboxymethyl or phosphate groups bound to resins, HIC substrates include ethyl, isopropyl, butyl, phenyl or octyl groups bound to resins, and IMAC substrates include iminodiacetic acid and nitriloacetic acid groups bound to resins. The foregoing substrates are listed for illustrative purposes and are not intended to limit the scope of substrates that can be employed to practice the invention.

[00168] In some embodiments, the invention provides substantially homogeneous XTEN prepared from the a polypeptide comprising an XTEN fused to a first or a first and a second affinity tag by cleavage sequences (such as, but not limited to the cleavage sequences of Table 7) capable of being cleaved by a protease, wherein the preparation is treated with the protease to cleave the cleavage sequences to release the XTEN from the polypeptide, followed by a chromatography step to bind and then elute, and then recover the substantially homogeneous XTEN. In one embodiment of the foregoing, the protease is trypsin and the cleavage sequences are capable of being cleaved by trypsin, non-limiting examples of which are listed in Tables 6 and 7. In another embodiment of the foregoing, the protease is TEV and the cleavage sequences are capable of being cleaved by TEV. In another embodiment of the foregoing, the cleaved XTEN is purified by binding to MacoCap SP chromatography substrate followed by elution with a salt or buffer solution such as, but not limited to, sodium phosphate/NaCl, resulting in the substantially homogenous XTEN. As used in the context of XTEN and/or polypeptides comprising XTEN, a preparation that is "substantially purified" means that at least about 85%, and more preferably at least about 90%, and more preferably at least about 91%, and more preferably at least about 92%, and more preferably at least about 93%, and more preferably at least about 94%, and more preferably at least about 95% or more of the individual molecules of a given preparation have an identical sequence length; in other words, the same number of amino acids. The methods that can be utilized to assay for homogeneity of length include mass spectroscopy, size exclusion chromatography/HPLC, or SDS-PAGE followed by silver staining; the methods can be used individually or collectively to quantitate the degree of homogeneity. A generalized scheme for purification of polypeptides comprising XTEN with affinity tag sequences optimized for purification is shown in FIGS. 41-42. After purification the tags can be proteolytically cleaved (FIG. 21B) or retained (FIG. 21C). The XTEN can be purified from contaminants due to the unique amino acid composition of XTEN, as illustrated in FIG. 42.

[00169] In one embodiment of XTEN compositions with affinity tags, the first and second affinity tags are selected from the group of affinity tags set forth in Table 5. In one embodiment, the first affinity tag linked to the XTEN as a fusion protein comprises the sequence RPRPRPRPRPRPR and the chromatography substrate used to bind the polypeptide is MacroCap

SP. In another embodiment, the first affinity tag linked to a first terminus of the XTEN as a fusion protein comprises the sequence RPRPRPRPRPRPRPRPRPRPR, the second affinity tag linked to a second terminus of the XTEN comprises the sequence HHHHHHHH, the first chromatography substrate used to bind the polypeptide is MacroCap SP, and the second chromatography substrate used to bind the polypeptide is a immobilized metal on affinity (IMAC) substrate. In another embodiment, the first affinity tag fused to a cleavage sequence fused to a first terminus of the XTEN as a fusion protein comprises the sequence RPRPRPRPRPRPR or RPRPRPRPRPRPRPRPRPRPRPR, the second affinity tag fused to a cleavage sequence to a second terminus of the XTEN comprises the sequence HHHHHH or HHHHHHHH, the first chromatography substrate used to bind the polypeptide is MacroCap SP, the second chromatography substrate used to bind the polypeptide is a immobilized metal on affinity (IMAC) substrate, the cleavage sequences comprise an arginine or lysine (including, but not limited to the sequences of Tables 6 and 7) and are cleaved by trypsin, and Macrocap Q is the chromatography substrate used to bind the XTEN freed from the affinity tags or, in the alternative, the freed XTEN is captures as flow-through by passing the protease-treated preparation through one or more of cation exchange, HIC and/or IMAC to capture the cleavage products and protease, leaving the XTEN in the flow-through, which is then recovered as a substantially homogeneous preparation.

[00170] It will be further appreciated by one of skill in the art that once the XTEN polypeptides comprising cleavage sequences and affinity tags are purified, the recovered polypeptide can be subsequently treated by proteolysis to release the one or two affinity tags, followed by passing the treated XTEN through a chromatography substrate to recover the XTEN without linked affinity tags. A schematic of the method is illustrated in FIG. 42 and exemplary methodologies are described in the Examples. Many different proteases can be utilized for the release of terminal purification tags, depending on the sequence linking the affinity tag to the XTEN, including but not limited to a protease selected from Table 7. In one embodiment, the protease is selected from the group consisting of trypsin, chymotrypsin, tobacco etch mosaic virus protease (TEV), FXa, and enterokinase. In another embodiment, the cleavage sequence incorporated into the polypeptide comprises an arginine residue that can be cleaved and the affinity tag removed by treatment with trypsin, thereby releasing the XTEN that is subsequently recovered in substantially purified form by chromatography such as, by capture using anion exchange (including but not limited to, MacroCap Q) or recovered as flow-through wherein the non-XTEN cleavage products and protease are captured by one or more of HIC, cation exchange, or IMAC chromatography, leaving substantially homogeneous XTEN in the flow-through.

Table 5: Affinity Tags and Chromatography Substrate Categories that Bind Affinity Tags

Affinity Tag Amino Acid Sequence	Chromatography Substrate
LYPYPYP, LYYYPP, WPWP, FPFPFP	HIC
$(Y)_n$, $(W)_n$, $(YP)_n$, $(WP)_n$, $(FP)_n$, $(LP)_n$ with n=3-20	HIC
$(RP)_n$, $(KP)_n$, $(HP)_n$, $(H)_n$, $(R)_n$ with n=3-20	Cation exchange, IMAC
$(E)_n$, $(D)_n$, $(ED)_n$, $(EP)_n$, $(DP)_n$ with n=3-20	Anion exchange
RPRPRPRPRP	Cation exchange
RPRPRPRPRPGR	Cation exchange
RPRPRPRPRPRPRP	Cation exchange
RPRPRPRPRPRPRPGR	Cation exchange
KPKPKPKPKP	Cation exchange
KPKPKPKPKPGR	Cation exchange
RPRPRPRPRPRPRPRPRP	Cation exchange
RPRPRPRPRPRPRPRPRPGR	Cation exchange
RPRPRPRPRPRPRPRPRPRPRP	Cation exchange
RPRPRPRPRPRPRPRPRPRPRPGR	Cation exchange
RPRPRPRPRPRPRPRPRPRPRPRPRP	Cation exchange
RPRPKPRPKPRPKPRPKP	Cation exchange
PRPKPRPKPRPKPRPKPGR	Cation exchange
RPRPKPRPKPRPKPRPKPRPKP	Cation exchange
RPRPKPRPKPRPKPRPKPRPKPGR	Cation exchange
GSPYGYPYSYS, GSPWGSPTSTE, GSPAGSPTSTE,	HIC
GSPXGXPXSXS, GSPSGXPXSXS, GSPSGTPXSXS where X = Ile, Leu, Val, Phe, Trp, or Tyr	HIC
GSPXGXPXSXS, GSPSGXPXSXS, GSPSGTPXSXS where X = Arg, Lys, or His	Cation exchange, IMAC
HHHHHH, HHHHHHHH	IMAC
STRPSRRSRRG, STRRGTRRGTRRG,	Cation exchange
STRPSRGRARG, STRPSRRARG, STRPSRRRRG,	Cation exchange
STEPSEESEEG, STEEGTEEGTEEG,	Anion exchange
STEPSEGEAEG, STEPSEEAEG, STEPSEEEEG,	Anion exchange

Table 6: Trypsin Cleavage Sequences

P4	P3	P2	P1	P1'	P2'
S	A	S	R	S	A
S	A	S	K	S	A
G	S	G	R	A	T
E	A	A	R	H	H
A	P	G	R	H	H
G	S	G	R	G	S
			R	X*	
			K	X*	

* X = any L-amino acid other than proline

Table 7: Proteases and Protease Cleavage Sequences

Protease Acting Upon Sequence	Exemplary Cleavage Sequence	Cleavage Sequences*
FXIa	KLTR↓AET	KD/FL/T/R/VA/VE/GT/GV
FXIa	DFTR↓VVG	KD/FL/T/R/VA/VE/GT/GV

FXIIa	TMTR↓IVGG	
Kallikrein	SPFR↓STGG	-/-/FL/RY/SR/RT/-/-
FVIIa	LQVR↓IVGG	
FIXa	PLGR↓IVGG	-/-/G/R/-/-/-/-
FXa	IEGR↓TVGG	IA/E/GFP/R/STI/VFS/-/G
FIIa (thrombin)	LTPR↓SLLV	-/-/PLA/R/SAG/-/-/-
Elastase-2	LGPV↓SGVP	-/-/-/VIAT/-/-/-/-
Granzyme-B	VAGD↓SLEE	V/-/-/D/-/-/-/-
MMP-12	GPAG↓LGGA	G/PA/-/G/L/-/G/-
MMP-13	GPAG↓LRGA	G/P/-/G/L/-/GA/-
MMP-17	APLG↓LRLR	-/PS/-/-/LQ/-/LT/-
MMP-20	PALP↓LVAQ	
TEV	ENLYFQ↓G	ENLYFQ/G/S
Enterokinase	DDDK↓IVGG	DDDK/IVGG
Protease 3C (PreScission™)	LEVLFQ↓GP	LEVLFQ/GP
Sortase A	LPKT↓GSES	L/P/KEAD/T/G/-/EKS/S
Trypsin	K↓X** or R↓X	K/X or R/X
Trypsin	R↓X**	SASRSA

↓indicates cleavage site
* the listing of multiple amino acids before, between, or after a slash indicate alternative amino acids that can be substituted at the position; "-" indicates that any amino acid may be substituted for the corresponding amino acid indicated in the middle column
** x is any L-amino acid other than proline

III). PAYLOADS

[00171] The present invention relates in part to XTEN conjugates linked to one or more payload molecules. The invention addresses a long-felt need in increasing the terminal half-life of exogenously administered therapeutic payloads to a subject in need thereof, as well as combinations of payloads that may include a therapeutic component and a targeting component to increase efficacy and reduce systemic toxicity.

[00172] It is contemplated that XTEN can be linked to combinations of targeting and drug payload molecules, including pharmacologically active small-molecules and targeting small molecules payloads, as well as combinations of these types of payloads resulting in compositions with two types of payloads. For the XTEN-folate conjugation compositions, it is specifically contemplated that a payload can be a pharmacologically active agent that possesses a suitably reactive functional group, including, but not limited to a native amino group, a sulfydryl group, a carboxyl group, an aldehyde group, a ketone group, an alkene group, an alkyne group, an azide group, an alcohol group, a heterocycle, or, alternatively, is modified to contain at least one of the foregoing reactive groups suitable for coupling to either an XTEN, XTEN-cross-linker, or XTEN-

click-chemistry reactant of the invention using any of the conjugation methods described herein or are otherwise known to be useful in the art for conjugating such reactive groups. Specific functional moieties and their reactivities are described in Organic Chemistry, 2nd Ed. Thomas Sorrell, University Science Books, Herndon, VA (2005). Further, it will be understood that any payload containing a reactive group or that is modified to contain a reactive group will also contain a residue after conjugation to which either the XTEN, the XTEN-cross-linker, or the XTEN-click-chemistry reactant is linked.

1. Drugs as payloads

[00173] In one aspect, the invention provides compositions comprising XTEN covalently linked to a small molecule payload drug. In some embodiments, the drug payload for conjugation to either the subject XTEN, the XTEN-cross-linkers, or the XTEN-click-chemistry reactants described herein is one or more analogs of auristatin, or a pharmaceutically acceptable salt, acid or derivative or agonist thereof.

[00174] In one embodiment, the payload is one or more molecules of auristatin E. In another embodiment, the payload is one or more molecules of auristatin F. In another embodiment, the payload is one or more molecules of monomethyl auristatin E. In another embodiment, the payload is one or more molecules of monomethyl auristatin F. In another embodiment, the monomethyl auristatin E drug is derivatized to introduce a reactive group for conjugation to the subject XTEN, the XTEN-cross-linkers, or the XTEN-click-chemistry reactants described herein. In another embodiment, the monomethyl auristatin F drug is derivatized to introduce a reactive group for conjugation to the subject XTEN, the XTEN-cross-linkers, or the XTEN-click-chemistry reactants described herein.

[00175] "Auristatin F" or "AF" is N,N-dimethylvaline-valine-dolaisoleuine(dil)-dolaproine(dap)-phenylalanine. "MMAF" is N-methylvaline-valine-dolaisoleuine(dil)-dolaproine(dap)-phenylalanine (MW 731.5).

[00176] Monomethyl auristatin E (MMAE, vedotin) is antimitotic agent which inhibits cell division by blocking the polymerisation of tubulin. Because of its toxicity, it cannot be used as a drug itself, but can be linked to targeting agent such as a monoclonal antibody or a receptor ligand that directs it to the cancer cells. When conjugated to the XTEN carrier with a cleavable linker, it is cleaved by cathepsin once the conjugate has entered a tumor cell, thus activating the anti-mitotic mechanism. The chemical name for MMAE is *(S)*-*N*-((*3R*,*4S*,*5S*)-1-((*S*)-2-((*1R*,*2R*)-3-(((*1S*,*2R*)-1-hydroxy-1-phenylpropan-2-yl)amino)-1-methoxy-2-methyl-3-oxopropyl)pyrrolidin-1-yl)-3-methoxy-5-methyl-1-oxoheptan-4-yl)-*N*,3-dimethyl-2-(*(S)*-3-methyl-2-(methylamino)butanamido)butanamide.

[00177] Monomethyl auristatin F (MMAF) is an anti-mitotic agent that inhibits cell division by blocking the polymerization of tubulin. It is an auristatin derivative with a charged C-terminal

phenylalanine residue that attenuates its cytotoxic activity compared to its uncharged counterpart MMAE. Because of its toxicity, it cannot be used as a drug itself, but can be linked to targeting agent such as a monoclonal antibody or a receptor ligand that directs it to the cancer cells. When conjugated to the XTEN carrier with a cleavable linker, it is cleaved by cathepsin once the conjugate has entered a tumor cell, thus activating the anti-mitotic mechanism. The chemical name is (*S*)-2-((2*R*,3*R*)-3-((*S*)-1-((3*R*,4*S*,5*S*)-4-((*S*)-*N*,3-dimethyl-2-((*S*)-3-methyl-2-(methylamino)butanamido)butanamido)-3-methoxy-5-methylheptanoyl)pyrrolidin-2-yl)-3-methoxy-2-methylpropanamido)-3-phenylpropanoic acid.

2. Targeting moieties as payloads

[00178] In another aspect, the invention provides compositions comprising XTEN covalently linked to a folate molecule that serves as a targeting moiety. In one embodiment, the small molecule targeting moiety is a ligand to a cell surface receptor expressed on a cancer cell, such as the folate receptor.

[00179] "Folate" and "folic acid" are used interchangeably herein to mean the chemical also known as pteroyl-L-glutamic acid, vitamin B9, folacin, and (2S)-2-[(4-{[(2-amino-4-hydroxypteridin-6-yl)methyl]amino}phenyl)formamido]pentanedioic acid. Folate is a ligand for the cell receptor known as folate receptor.

[00180] Folate receptor alpha is a protein that in humans is encoded by the FOLR1 gene (Campbell IG, et al. (1991). Folate-binding protein is a marker for ovarian cancer (Cancer Res 51 (19): 5329–5338). Many cancer cells have a high requirement for folic acid and overexpress the folate receptor. The folate receptor encoded by this gene is a member of the folate receptor (FOLR) family, and members have a high affinity for folic acid and for several reduced folic acid derivatives, and mediate delivery of 5-methyltetrahydrofolate to the interior of cells. Folate receptor can be overexpressed by a number of tumors including ovarian, breast, renal, lung, colorectal, and brain.

3. Cross-linker and azide/alkyne click-chemistry reactants for conjugation

[00181] In another aspect, the invention relates to XTEN conjugated to cross-linkers, resulting in XTEN-cross-linker conjugation partners that can be utilized to prepare XTEN–payload conjugation compositions. In particular, the herein-described XTEN-cross-linker conjugate partners are useful for conjugation to payload agents or surfaces bearing at least one thiol, amino, carboxyl, aldehyde or alcohol or any other reactive group available and suitable, as known in the art, for reaction between the components described herein.

[00182] In another aspect, the invention relates to methods of making conjugates of XTEN-cross-linker reactants and XTEN-click-chemistry azide/alkyne reactants, resulting in conjugates that can be utilized to prepare the subject XTEN–payload compositions. In particular, the herein-described methods for making XTEN-cross-linkers and XTEN-azide/alkyne reactants are useful

wherein the payload agent or a reaction surface bears at least one thiol, amino, carboxyl, aldehyde, alkene, alkyne, heterocycle, alcohol, or other reactive group available for reaction.

[00183] Exemplary embodiments of XTEN have been described above, including preparations of substantially homogeneous XTEN. The invention provides XTEN that further serve as a platform to which payloads can be conjugated, such that they serve as a "carrier", conferring certain desirable pharmacokinetic, chemical and pharmaceutical properties to the compositions, amongst other properties described below. In other embodiments, the invention provides polynucleotides that encode XTEN that can be linked to genes encoding peptide or polypeptide payloads that can be incorporated into expression vectors and incorporated into suitable hosts for the expression and recovery of the subject XTEN.

[00184] In some embodiments, the XTEN components as described herein, above, are engineered to incorporate a defined number of reactive amino acid residues that can be reacted with cross-linking agents or can further contain reactive groups that can be used to conjugate to payloads. In one embodiment, the invention provides cysteine-engineered XTEN wherein the cysteine, each of which contains a reactive thiol group, are conjugated to a cross-linker, resulting in an XTEN-cross-linker conjugate. In another embodiment, invention provides lysine-engineered XTEN wherein lysine, each of which contains a positively charged hydrophilic ε-amino group, are conjugated to a cross-linker, resulting in an XTEN-cross-linker conjugate. In the embodiments of cysteine-engineered XTEN, each comprises about 1 to about 100 cysteine amino acids, or from 1 to about 50 cysteine amino acids, or from 1 to about 40 cysteine amino acids, or from 1 to about 20 cysteine amino acids, or from 1 to about 10 cysteine amino acids, or from 1 to about 5 cysteine amino acids, or 9 cysteines, or 3 cysteines, or a single cysteine amino acid that is available for conjugation. In the embodiments of lysine-engineered XTEN, each comprises about 1 to about 100 lysine amino acids, or from 1 to about 50 lysine amino acids, or from 1 to about 40 lysine engineered amino acids, or from 1 to about 20 lysine engineered amino acids, or from 1 to about 10 lysine engineered amino acids, or from 1 to about 5 lysine engineered amino acids, or 9 cysteines, or 3 cysteines, or a single lysine that is available for conjugation. In another embodiment, the engineered XTEN comprises both cysteine and lysine residues of the foregoing ranges or numbers.

[00185] Generally, XTEN cysteine thiol groups are more reactive, i.e., more nucleophilic, towards electrophilic conjugation reagents than amine or hydroxyl groups. In addition, cysteine residues are generally found in smaller numbers in a given protein; thus are less likely to result in multiple conjugations within the same protein. Cysteine residues have been introduced into proteins by genetic engineering techniques to form covalent attachments to ligands or to form new intramolecular disulfide bonds (Better et al (1994) J. Biol. Chem. 13:9644-9650; Bernhard et al (1994) Bioconjugate Chem. 5:126-132; Greenwood et al (1994) Therapeutic Immunology 1:247-

255; Tu et al (1999) Proc. Natl. Acad. Sci USA 96:4862-4867; Kanno et al (2000) J. of Biotechnology, 76:207-214; Chmura et al (2001) Proc. Nat. Acad. Sci. USA 98(15):8480-8484; U.S. Pat. No. 6,248,564).

[00186] In one embodiment, the invention provides an isolated composition comprising a cysteine-engineered XTEN conjugated to a cross-linker, wherein the cross-linker is selected from sulfhydryl-reactive homobifunctional or heterobifunctional cross-linkers. In another embodiment, the invention provides an isolated composition comprising a lysine-engineered XTEN conjugated by a cross-linker, wherein the cross-linker is selected from amine-reactive homobifunctional or heterobifunctional cross-linkers. Cross-linking is the process of chemically linking two or more molecules by a covalent bond. The process is also called conjugation or bioconjugation with reference to its use with proteins and other biomolecules. For example, proteins can be modified to alter N- and C-termini, and amino acid side chains on proteins and peptides in order to block or expose reactive binding sites, inactivate functions, or change functional groups to create new targets for cross-linking

[00187] In one aspect, the invention provides methods for the site-specific conjugation to XTEN polymer, accomplished using chemically-active amino acid residues or their derivatives (e.g., the N-terminal α-amine group, the ε-amine group of lysine, the thiol group of cysteine, the C-terminal carboxyl group, carboxyl groups of glutamic acid and aspartic acid. Functional groups suitable for reactions with primary α- and ε-amino groups are chlorocyanurates, dichlorotreazines, trezylates, benzotriazole carbonates, p-nitrophenyl carbonates, trichlorophenyl carbonates, aldehydes, mixed anhydrides, carbonylimidazoles, imidoesters, N-hydroxysuccinimide esters, N-hydroxysulfosuccinimide esters (Harris, J. M., Herati, R. S. *Polym. Prepr.* (*Am. Chem. Soc., Div. Polym. Chem*), 32(1), 154-155 (1991); Herman, S., et al. *Macromol. Chem. Phys.* 195, 203-209 (1994); Roberts, M. J. et. al. *Advanced Drug Delivery Reviews,* 54, 459-476 (2002)). N-hydroxysuccinimide esters (NHS-esters and their water soluble analogs sulfo-NHS-esters) are commonly used for protein conjugation (see FIG. 2). NHS-esters yield stable amide products upon reaction with primary amines with relatively efficient coupling at physiological pH. The conjugation reactions are typically performed in 50-200 mM phosphate, bicarbonate/carbonate, HEPES or borate buffers (pH between 7 and 9) at 4°C to room temperature from 0.5 to 2 hrs. NHS-esters are usually used at two- to 50-fold molar excess to protein. Typically, the concentration of the reagent can vary from 0.1-10 mM, while the optimal protein concentration is 50-100 μM.

[00188] In another method, given that XTEN polypeptides possess only a single N-terminal α-amino group, the XTEN can be engineered to contain additional ε-amino group(s) of intentionally incorporated lysine residues; exemplary sequences of which are provided in Table 1. The α-and ε-amino groups have different pKa values: approximately 7.6 to 8.0 for the α-amino group of the N-

terminal amino acid, and approximately 10-10.5 for the ε-amino group of lysine. Such a significant difference in pKa values can be used for selective modification of amino groups. Deprotonation of all primary amines occurs at pH above pH 8.0. In this environment, the nucleophilic properties of different amines determine their reactivity. When deprotonated, the more nucleophilic ε-amino groups of lysines are generally more reactive toward electrophiles than α-amino groups. On the other hand, at a lower pH (for example pH 6), the more acidic α-amino groups are generally more deprotonated than ε-amino groups, and the order of reactivity is inverted. For example, the FDA-approved drug Neulasta (pegfilgranstim) is granulocyte colony-stimulating factor (G-CSF) modified by covalent attachment of 20 kDa PEG-aldehyde. Specific modification of the protein's N-terminal amino acid was accomplished by exploiting the lower pKa of α-amino group as compared to ε-amino groups of internal lysines (Molineaux, G. *Curr. Pharm. Des.* 10, 1235-1244 (2004), US Patent 5,824,784).

[00189] The XTEN polypeptides comprising cysteine residues can be genetically engineered using recombinant methods described herein (see, e.g., Examples) or by standard methods known in the art. Conjugation to thiol groups can be carried using highly specific reactions, leading to the formation of single conjugate species joined by cross-linking agents. Functional groups suitable for reactions with cysteine thiol-groups are N-maleimides, haloacetyls, and pyridyl disulfides. The maleimide group reacts specifically with sulfhydryl groups when the pH of the reaction mixture is between pH 6.5 and 7.5, forming a stable thioether linkage that is not reversible (see FIG. 3). At neutral pH, maleimides react with sulfhydryls 1,000-fold faster than with amines, but when the pH is raised to greater than 8.5, the reaction favors primary amines. Maleimides do not react with tyrosines, histidines or methionines. For reaction solutions, thiols must be excluded from reaction buffers used with maleimides as they will compete for coupling sites. Excess maleimides in the reaction can be quenched at the end of a reaction by adding free thiols, while EDTA can be included in the coupling buffer to minimize oxidation of sulfhydryls.

[00190] In another embodiment, the invention contemplates use of haloacetyl reagents that are useful for cross-linking sulfhydryls groups of XTEN or payloads to prepare the subject conjugates. The most commonly used haloacetyl reagents contain an iodoacetyl group that reacts with sulfhydryl groups at physiological pH. The reaction of the iodoacetyl group with a sulfhydryl proceeds by nucleophilic substitution of iodine with a thiol producing a stable thioether linkage (see FIG. 4). Using a slight excess of the iodoacetyl group over the number of sulfhydryl groups at pH 8.3 ensures sulfhydryl selectivity. If a large excess of iodoacetyl group is used, the iodoacetyl group can react with other amino acids. Imidazoles can react with iodoacetyl groups at pH 6.9-7.0, but the incubation must proceed for longer than one week. Histidyl side chains and amino groups react in the unprotonated form with iodoacetyl groups above pH 5 and pH 7, respectively. In another embodiment, cross-linkers useful for sulfhydryls groups are pyridyl disulfides. Pyridyl

disulfides react with sulfhydryl groups over a broad pH range (the optimal pH is 4-5) to form disulfide bonds linking XTEN to payloads (see FIG. 5). As a disulfide, conjugates prepared using these reagents are cleavable. During the reaction, a disulfide exchange occurs between the molecule's –SH group and the 2-pyridyldithiol group. As a result, pyridine-2-thione is released. These reagents can be used as crosslinkers and to introduce sulfhydryl groups into proteins. The disulfide exchange can be performed at physiological pH, although the reaction rate is slower.

[00191] Since cysteines are generally less abundant in natural peptide and protein sequences than lysines, the use of cysteines as a site for conjugation reduces the likelihood of multiple conjugations to XTEN-cross-linker molecules in a reaction. It also reduces the likelihood of peptide/protein deactivation upon conjugation. Moreover, conjugation to cysteine sites can often be carried out in a well-defined manner, leading to the formation of single species XTEN polymer-peptide or XTEN polymer–polypeptide conjugates. In some cases cysteine may be absent in the amino acid sequence of the peptide to be conjugated. In such a case, cysteine residue can be added to the N- or C-terminus of the peptide either recombinantly or synthetically using standard methods. Alternatively, a selected amino acid can be chemically or genetically modified to cysteine. As one example, serine modification to cysteine is considered a conservative mutation. Another approach to introduce a thiol group in cysteine-lacking peptides is chemical modification of the lysine ε-amino group using thiolating reagents such as 2-iminothiolane (Traut's reagent), SATA (N-succinimidyl S-acetylthioacetate), SATP (N-succinimidyl S-acetylthiopropionate), SAT-PEO_4-Ac (N-Succinimidyl S-acetyl(thiotetraethylene glycol)), SPDP (*N*-Succinimidyl 3-(2-pyridyldithio)propionate), LC-SPDP (Succinimidyl 6-(3'-[2-pyridyldithio]propionamido)hexanoate) (described more fully, below). Once a unique thiol group is introduced in the peptide, it can be selectively modified by compounds containing sufhydryl-reactive such as N-maleimides, haloacetyls, and pyridyl disulfides, as described above.

[00192] The conjugation between the XTEN polypeptide and a peptide, protein or small molecule drug payload may be achieved by a variety of linkage chemistries, including commercially available zero-length, homo- or hetero-bifunctional, and multifunctional cross-linker compounds, according to methods known and available in the art, such as those described, for example, in R. F. Taylor (1991) "Protein immobilization. Fundamentals and Applications", Marcel Dekker Inc., N.Y.; G. T. Hermanson et al. (1992) "Immobilized Affinity Ligand Techniques", Academic Press, San Diego; G. T. Hermanson (2008) "Bioconjugate Techniques", 2nd. ed. Elsevier, Inc., S. S. Wong (1991) "Chemistry of Protein Conjugation and Crosslinking", CRC Press, Boca Raton. Suitable cross-linking agents for use in preparing the conjugates of the disclosure are commercially-available from companies like Sigma-Aldrich, Thermo Fisher Scientific (Pierce Protein Research Products), Invitrogen, ProteoChem, G-Biosciences. Preferred

embodiments of cross-linkers comprise a thiol-reactive functional group or an amino-reactive functional group. A list of exemplary cross-linkers is provided in Table 8.

Table 8: Exemplary cross-linkers

Cross-linker
maleimides, haloacetyls, pyridyl disulfides, haloacetyls, pyridyl disulfides, ABH (*p*-Azidobenzoyl hydrazide), AMAS (*N*-(α-Maleimidoacetoxy)-succinimide ester), ANB-NOS (*N*-5-Azido-2-nitrobenzyloxy-succinimide), APDP (*N*-(4-[*p*-Azidosalicylamido]butyl)-3'-(2'-pyridyldithio) propionamide), ASBA (4-(*p*-Azidosalicylamido)-butylamine), BASED (*Bis* (β-[4-azidosalicylamido]ethyl) disulfide), BMB (1,4-*Bis*-Maleimidobutane), BMDB (1,4 Bismaleimidyl-2,3-dihydroxybutane), BMH (*Bis*-Maleimidohexane), BMOE (*Bis*-Maleimidoethane), BMPH (*N*-(β-Maleimidopropionic acid)hydrazide), BMPS (*N*-(β-Maleimidopropyloxy)succinimide ester), $BM(PEG)_2$ (1,8-*Bis*-Maleimidodiethylene-glycol), $BM(PEG)_3$ (1,11-*Bis*-Maleimidotriethyleneglycol), BS^2G (*Bis* (sulfosuccinimidyl)glutarate), BS^3 (Sulfo-DSS) (*Bis* (sulfosuccinimidyl)suberate), $BS[PEG]_5$ (*Bis* (NHS)PEG5), $BS(PEG)_9$ (*Bis* (NHS)PEG9), BSOCOES (Bis(2-[succinimidoxycarbonyloxy]ethyl)sulfone), C6-SANH (C6-Succinimidyl 4-hydrazinonicotinate acetone hydrazone), C6-SFB (C6-Succinimidyl 4-formylbenzoate), DCC (*N,N*-Dicyclohexylcarbodiimide), DFDNB (1-5-Difluoro-2,4-dinitrobenzene), DMA (Dimethyl adipimidate), DMP (Dimethyl pimelimidate), DMS (Dimethyl suberimidate), DPDPB (1,4-Di-(3'-[2'pyridyldithio]propionamido) butane), DSG (Disuccinimidyl glutarate), DSP (Dithiobis(succimidylpropionate), Lomant's Reagent), DSS (Disuccinimidyl suberate), DST (Disuccinimidyl tartarate), DTBP (Dimethyl 3,3'-dithiobispropionimidate), DTME (Dithiobis-maleimidoethane), DTSSP (Sulfo-DSP) (3,3'-Dithiobis (sulfosuccinimidylpropionate)), EDC (1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride), EGS (Ethylene glycol *bis*(succinimidylsuccinate)), EMCA (*N*-ε-Maleimidocaproic acid), EMCH (*N*-(ε-Maleimidocaproic acid)hydrazide), EMCS (*N*-(ε-Maleimidocaproyloxy)succinimide ester), GMBS (*N*-(γ-Maleimidobutyryloxy)succinimide ester), KMUA (*N*-κ-Maleimidoundecanoic acid), KMUH (*N*-(κ-Maleimidoundecanoic acid)hydrazide), LC-SDA (NHS-LC-Diazirine), LC-SMCC (Succinimidyl 4-(*N*-maleimidomethyl) cyclohexane-1-carboxy-(6-amidocaproate)), LC-SPDP (Succinimidyl 6-(3'-[2-pyridyldithio]propionamido)hexanoate), MBS (*m*-Maleimidobenzoyl-*N*-hydroxysuccinimide ester), MPBH (4-(4-N-Maleimidophenyl)-butyric acid hydrazide), NHS-ASA (*N*-Hydroxysuccinimidyl-4-azidosalicylic acid), PDPH (3-(2-Pyridyldithio)propionylhydrazide), PMPI (*N*-(*p*-Maleimidophenyl)isocyanate), SADP (Succinimidyl (4-azidophenyl dithio) propionate), SAED (Succimidyl 2-[7-azido-4-methylcoumarin-3-acetamido]ethyl-1,3'-dithiopropionate), SAND (Succinimidyl-2-(m-azido-*o*-nitrobenzamido)ethyl 1,3'-dithiopropionate), SANH (Succinimidyl 4-hydrazinonicotinate acetone hydrazone), SANPAH (*N*-Succinimidyl 6-(4'-azido-2'-nitrophenylamino)hexanoate), SASD (Succinimidyl-2-(p-azidosalycylamido)ethyl-1,3-dithiopropionate), SBAP (Succinimdyl 3-(bromoacetamido)propionate), SDA (NHS-Diazirine), SDAD (NHS-SS-Diazirine), SFAD (Succinimidyl(perfluoroazidobenzamido)ethyl 1,3'-dithiopropionate), SFB (Succinimidyl 4-formylbenzoate), SHTH (Succinimidyl 4-hydrazidoterephthalate), SIA (*N*-succinimidyl iodoacetate), SIAB (*N*-Succinimidyl(4-iodoacetyl)aminobenzoate), SMPB (Succinimidyl 4-(p-maleimidophenyl) butyrate), SMCC (Succinimidyl 4-(*N*-maleimido-methyl)cyclohexane-1-carboxylate), $SM[PEG]_2$ (NHS-PEG2-Maliemide), $SM[PEG]_4$ (NHS-PEG4-Maliemide), $SM(PEG)_6$ (NHS-PEG6-Maleimide), $SM[PEG]_8$ (NHS-PEG8-Maliemide), $SM[PEG]_{12}$ (NHS-PEG12-Maliemide), $SM(PEG)_{24}$ (NHS-PEG24-Maleimide), SMPB (Succinimidyl 4-(*p*-maleimido-phenyl)butyrate), SMPH (Succinimidyl-6-(β-maleimidopropionamido)hexanoate), SMPT (4-Succinimidyloxycarbonyl-methyl-α-(2-pyridyldithio)toluene), SPB (Succinimidyl-(4-psoralen-8-yloxy)butyrate), SPDP (*N*-Succinimidyl 3-(2-pyridyldithio)propionate), Sulfo-DST (Sulfodisuccinimidyl tartrate), Sulfo-EGS (Ethylene glycol bis (sulfo-succinimidyl succinate)), Sulfo-EMCS (*N*-(ε-Maleimidocaproyloxy)sulfosuccinimide ester), Sulfo-GMBS (*N*-(γ-Maleimidobutryloxy)sulfosuccinimide ester), Sulfo-HSAB (*N*-Hydroxysulfosuccinimidyl-4-azidobenzoate), Sulfo-KMUS (*N*-(κ-Maleimidoundecanoyloxy)sulfosuccinimide ester), Sulfo-LC-SDA (Sulfo-NHS-LC-Diazirine), Sulfo-LC-SMPT (Sulfosuccinimidyl 6-(α-methyl-α-[2-

Cross-linker
pyridyldithio]-toluamido)hexanoate), Sulfo-LC-SPDP (Sulfosuccinimidyl 6-(3'-[2-pyridyldithio]propionamido)hexanoate), Sulfo-MBS (*m*-Maleimidobenzoyl-*N*-hydroxysulfosuccinimide ester), Sulfo-NHS-LC-ASA (Sulfosuccinimidyl(4-azido-salicylamido) hexanoate), Sulfo-SADP (Sulfosuccinimidyl (4-azidophenyl dithio) propionate), Sulfo-SAED (Sulfosuccimidyl 2-[7-azido-4-methylcoumarin-3-acetamido]ethyl-1,3'-dithiopropionate), Sulfo-SAND (Sulfosuccinimidyl-2-(m-azido-*o*-nitrobenzamido)ethyl 1,3'-dithiopropionate), Sulfo-SANPAH (Sulfosuccinimidyl 6-(4'-azido-2'-nitrophenylamino)hexanoate), Sulfo-SASD (Sulfosuccinimidyl-2-(p-azidosalycylamido)ethyl-1,3-dithiopropionate), Sulfo-SDA (Sulfo-NHS-Diazirine), Sulfo-SDAD (Sulfo-NHS-SS-Diazirine), Sulfo-SFAD (Sulfosuccinimidyl(perfluoroazidobenzamido)ethyl 1,3'-dithiopropionate), Sulfo-SIAB (Sulfosuccinimidyl(4-iodo-acetyl)aminobenzoate), Sulfo-SMCC (Sulfosuccinimidyl 4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate), Sulfo-SMPB (Sulfosuccinimidyl 4-(*p*-maleimidophenyl)butyrate), THPP (β-(Tris[hydroxymethyl]phosphine)propionic acid (betaine)), TMEA (*Tris*-(2-Maleimidoethyl)amine), TSAT (*Tris*-(succinimidyl aminotriacetate)), 3-propargyloxypropanoic acid,NHS ester, acetylene-PEG-NHS ester, dibenzylcyclooctyne, (DBCO)-NHS ester, DBCO-PEG-NHS ester, cyclooctyne (COT)-NHS ester, COT-PEG-NHS ester, COT-PEG-pentafluorophenyl (PFP) ester, BCOT-NHS ester, BCOT-PEG-NHS ester, BCOT-PEG-pentafluorophenyl (PFP) ester, Acetylene-PEG4-maleimide, DBCO-maleimide, COT-maleimide, BCOT-maleimide, 3-azide-propionic acid, NHS ester, 6-azide-hexanoic acid, NHS ester, 3-azide-propionic acid, PFP ester, 6-azide-hexanoic acid, PFP ester, azide-PEG-NHS ester, azide-PEG-PFP ester, azide-PEG-maleimide, N-(5-Aminopentyl)maleimide, aminopentyl-maleimide

[00193] Non-limiting examples of cross-linkers are ABH (*p*-Azidobenzoyl hydrazide), AMAS (*N*-(α-Maleimidoacetoxy)-succinimide ester), ANB-NOS (*N*-5-Azido-2-nitrobenzyloxy-succinimide), APDP (*N*-(4-[*p*-Azidosalicylamido]butyl)-3'-(2'-pyridyldithio) propionamide), ASBA (4-(*p*-Azidosalicylamido)-butylamine), BASED (*Bis* (β-[4-azidosalicylamido]ethyl) disulfide), BMB (1,4-*Bis*-Maleimidobutane), BMDB (1,4 Bismaleimidyl-2,3-dihydroxybutane), BMH (*Bis*-Maleimidohexane), BMOE (*Bis*-Maleimidoethane), BMPH (*N*-(β-Maleimidopropionic acid)hydrazide), BMPS (*N*-(β-Maleimidopropyloxy)succinimide ester), $BM(PEG)_2$ (1,8-*Bis*-Maleimidodiethylene-glycol), $BM(PEG)_3$ (1,11-*Bis*-Maleimidotriethyleneglycol), BS^2G (*Bis* (sulfosuccinimidyl)glutarate), BS^3 (Sulfo-DSS) (*Bis* (sulfosuccinimidyl)suberate), $BS[PEG]_5$ (*Bis* (NHS)PEG5), $BS(PEG)_9$ (*Bis* (NHS)PEG9), BSOCOES (Bis(2-[succinimidoxycarbonyloxy]ethyl)sulfone), C6-SANH (C6-Succinimidyl 4-hydrazinonicotinate acetone hydrazone), C6-SFB (C6-Succinimidyl 4-formylbenzoate), DCC (*N,N*-Dicyclohexylcarbodiimide), DFDNB (1-5-Difluoro-2,4-dinitrobenzene), DMA (Dimethyl adipimidate), DMP (Dimethyl pimelimidate), DMS (Dimethyl suberimidate), DPDPB (1,4-Di-(3'-[2'pyridyldithio]propionamido) butane), DSG (Disuccinimidyl glutarate), DSP (Dithiobis(succimidylpropionate), Lomant's Reagent), DSS (Disuccinimidyl suberate), DST (Disuccinimidyl tartarate), DTBP (Dimethyl 3,3'-dithiobispropionimidate), DTME (Dithiobis-maleimidoethane), DTSSP (Sulfo-DSP) (3,3'-Dithiobis (sulfosuccinimidylpropionate)), EDC (1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride), EGS (Ethylene glycol *bis*(succinimidylsuccinate)), EMCA (*N*-ε-Maleimidocaproic acid), EMCH (*N*-(ε-

Maleimidocaproic acid)hydrazide), EMCS (*N*-(ε-Maleimidocaproyloxy)succinimide ester), GMBS (*N*-(γ-Maleimidobutyryloxy)succinimide ester), KMUA (*N*-κ-Maleimidoundecanoic acid), KMUH (*N*-(κ-Maleimidoundecanoic acid)hydrazide), LC-SDA (NHS-LC-Diazirine), LC-SMCC (Succinimidyl 4-(*N*-maleimidomethyl) cyclohexane-1-carboxy-(6-amidocaproate)), LC-SPDP (Succinimidyl 6-(3'-[2-pyridyldithio]propionamido)hexanoate), MBS (*m*-Maleimidobenzoyl-*N*-hydroxysuccinimide ester), MPBH (4-(4-N-Maleimidophenyl)-butyric acid hydrazide), NHS-ASA (*N*-Hydroxysuccinimidyl-4-azidosalicylic acid), PDPH (3-(2-Pyridyldithio)propionylhydrazide), PMPI (*N*-(*p*-Maleimidophenyl)isocyanate), SADP (Succinimidyl (4-azidophenyl dithio) propionate), SAED (Succimidyl 2-[7-azido-4-methylcoumarin-3-acetamido]ethyl-1,3'-dithiopropionate), SAND (Succinimidyl-2-(m-azido-*o*-nitrobenzamido)ethyl 1,3'-dithiopropionate), SANH (Succinimidyl 4-hydrazinonicotinate acetone hydrazone), SANPAH (*N*-Succinimidyl 6-(4'-azido-2'-nitrophenylamino)hexanoate), SASD (Succinimidyl-2-(p-azidosalycylamido)ethyl-1,3-dithiopropionate), SBAP (Succinimdyl 3-(bromoacetamido)propionate), SDA (NHS-Diazirine), SDAD (NHS-SS-Diazirine), SFAD (Succinimidyl(perfluoroazidobenzamido)ethyl 1,3'-dithiopropionate), SFB (Succinimidyl 4-formylbenzoate), SHTH (Succinimidyl 4-hydrazidoterephthalate), SIA (*N*-succinimidyl iodoacetate), SIAB (*N*-Succinimidyl(4-iodoacetyl)aminobenzoate), SMPB (Succinimidyl 4-(p-maleimidophenyl) butyrate), SMCC (Succinimidyl 4-(*N*-maleimido-methyl)cyclohexane-1-carboxylate), SM[PEG]$_2$ (NHS-PEG2-Maliemide), SM[PEG]$_4$ (NHS-PEG4-Maliemide), SM(PEG)$_6$ (NHS-PEG6-Maleimide), SM[PEG]$_8$ (NHS-PEG8-Maliemide), SM[PEG]$_{12}$ (NHS-PEG12-Maliemide), SM(PEG)$_{24}$ (NHS-PEG24-Maleimide), SMPB (Succinimidyl 4-(*p*-maleimido-phenyl)butyrate), SMPH (Succinimidyl-6-(β-maleimidopropionamido)hexanoate), SMPT (4-Succinimidyloxycarbonyl-methyl-α-(2-pyridyldithio)toluene), SPB (Succinimidyl-(4-psoralen-8-yloxy)butyrate), SPDP (*N*-Succinimidyl 3-(2-pyridyldithio)propionate), Sulfo-DST (Sulfodisuccinimidyl tartrate), Sulfo-EGS (Ethylene glycol bis (sulfo-succinimidyl succinate)), Sulfo-EMCS (*N*-(ε-Maleimidocaproyloxy)sulfosuccinimide ester), Sulfo-GMBS (*N*-(γ-Maleimidobutryloxy)sulfosuccinimide ester), Sulfo-HSAB (*N*-Hydroxysulfosuccinimidyl-4-azidobenzoate), Sulfo-KMUS (*N*-(κ-Maleimidoundecanoyloxy)sulfosuccinimide ester), Sulfo-LC-SDA (Sulfo-NHS-LC-Diazirine), Sulfo-LC-SMPT (Sulfosuccinimidyl 6-(α-methyl-α-[2-pyridyldithio]-toluamido)hexanoate), Sulfo-LC-SPDP (Sulfosuccinimidyl 6-(3'-[2-pyridyldithio]propionamido)hexanoate), Sulfo-MBS (*m*-Maleimidobenzoyl-*N*-hydroxysulfosuccinimide ester), Sulfo-NHS-LC-ASA (Sulfosuccinimidyl(4-azido-salicylamido) hexanoate), Sulfo-SADP (Sulfosuccinimidyl (4-azidophenyl dithio) propionate), Sulfo-SAED (Sulfosuccimidyl 2-[7-azido-4-methylcoumarin-3-acetamido]ethyl-1,3'-dithiopropionate), Sulfo-SAND (Sulfosuccinimidyl-2-(m-azido-*o*-nitrobenzamido)ethyl 1,3'-dithiopropionate), Sulfo-SANPAH (Sulfosuccinimidyl 6-(4'-azido-2'-nitrophenylamino)hexanoate), Sulfo-SASD

(Sulfosuccinimidyl-2-(p-azidosalycylamido)ethyl-1,3-dithiopropionate), Sulfo-SDA (Sulfo-NHS-Diazirine), Sulfo-SDAD (Sulfo-NHS-SS-Diazirine), Sulfo-SFAD (Sulfosuccinimidyl(perfluoroazidobenzamido)ethyl 1,3'-dithiopropionate), Sulfo-SIAB (Sulfosuccinimidyl(4-iodo-acetyl)aminobenzoate), Sulfo-SMCC (Sulfosuccinimidyl 4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate), Sulfo-SMPB (Sulfosuccinimidyl 4-(*p*-maleimidophenyl)butyrate), THPP (β-(Tris[hydroxymethyl]phosphine)propionic acid (betaine)), TMEA (*Tris*-(2-Maleimidoethyl)amine), TSAT (*Tris*-(succinimidyl aminotriacetate)).

[00194] In some embodiments, XTEN-folate conjugates using cross-linking reagents introduce non-natural spacer arms. However, in cases where a native peptide bond is preferred, the invention provides that a reaction can be carried out using zero-length cross-linkers that act via activation of a carboxylate group. In the embodiments thereof, in order to achieve reaction selectivity, the first polypeptide has to contain only a free C-terminal carboxyl group while all lysine, glutamic acid and aspartic acid side chains are protected and the second peptide/protein N-terminal α-amine has to be the only available unprotected amino group (requiring that any lysines, asparagines or glutamines be protected). In such cases, use of XTEN sequences that are without glutamic acid as the first polypeptide in the XTEN-payload or XTEN-cross-linker is preferred. Accordingly, in one embodiment, the invention provides XTEN-cross-linker and XTEN-folate sequences wherein the compositions are conjugated to payloads using a zero-length cross-linkers. Exemplary zero-length cross-linkers utilized in the embodiment include but are not limited to DCC (N,N-Dicyclohexylcarbodiimide) and EDC (1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride) wherein the cross-linikers are used to directly conjugate carboxyl functional groups of one molecule (such as a payload) to the primary amine of another molecule, such as a payload with that functional group (see FIG. 6). Sulfo-NHS (N-hydroxysulfosuccinimide) and NHS (N-hydroxysuccinimide) are used as catalysts for conjugation, increasing reaction efficiency (Grabarek Z, Gergely J. Zero-length crosslinking procedure with the use of active esters. (1990) Anal. Biochem. 185(1), 131-135). EDC reacts with carboxylic acid group and activates the carboxyl group to form an active O-acylisourea intermediate, allowing it to be coupled to the amino group in the reaction mixture. The O-acylisourea intermediate is unstable in aqueous solutions, making it ineffective in two-step conjugation procedures without increasing the stability of the intermediate using N-hydroxysuccinimide. This intermediate reacts with a primary amine to form an amide derivative. The crosslinking reaction is usually performed between pH 4.5 to 5 and requires only a few minutes for many applications. However, the yield of the reaction is similar at pH from 4.5 to 7.5. The hydrolysis of EDC is a competing reaction during coupling and is dependent on temperature, pH and buffer composition. 4-Morpholinoethanesulfonic acid (MES) is an effective carbodiimide reaction buffer. Phosphate buffers reduce the reaction efficiency of the EDC, but increasing the

amount of EDC can compensate for the reduced efficiency. Tris, glycine and acetate buffers may not be used as conjugation buffers.

[00195] In other embodiments, XTEN and payloads can be conjugated using a broad group of cross-linkers, including those consisting of a spacer arm (linear or branched) and two or more reactive ends that are capable of attaching to specific functional groups (e.g., primary amines, sulfhydryls, etc.) on proteins or other molecules. Linear cross-linkers can be homobifunctional or heterobifunctional. Homobifunctional cross-linkers have two identical reactive groups which are used to cross-link proteins in one step reaction procedure. Non-limiting examples of amine-reactive homobifunctional cross-linkers are BS2G (Bis (sulfosuccinimidyl)glutarate), BS3 (Sulfo-DSS) (Bis (sulfosuccinimidyl)suberate), BS[PEG]5 (Bis (NHS)PEG5), BS(PEG)9 (Bis (NHS)PEG9), BSOCOES (Bis(2-[succinimidoxycarbonyloxy]ethyl)sulfone), DFDNB (1-5-Difluoro-2,4-dinitrobenzene), DMA (Dimethyl adipimidate), DMP (Dimethyl pimelimidate), DMS (Dimethyl suberimidate), DSG (Disuccinimidyl glutarate), DSP (Dithiobis(succimidylpropionate) (Lomant's Reagent), DSS (Disuccinimidyl suberate), DST (Disuccinimidyl tartarate), DTBP (Dimethyl 3,3'-dithiobispropionimidate), DTSSP (Sulfo-DSP) (3,3'-Dithiobis (sulfosuccinimidylpropionate)), EGS (Ethylene glycol bis(succinimidylsuccinate)), Sulfo-EGS (Ethylene glycol bis (sulfo-succinimidyl succinate)).

[00196] Additionally, examples of homobifunctional cross-linkers employed in the compositions and in the methods to create the XTEN-folate conjugate composition and/or XTEN-cross-linker compositions are sulfhydryl–reactive agents such as BMB (1,4-Bis-Maleimidobutane), BMH (Bis-Maleimidohexane), BMDB (1,4 Bismaleimidyl-2,3-dihydroxybutane), BMOE (Bis-Maleimidoethane), BM(PEG)2 (1,8-Bis-Maleimidodiethylene-glycol), BM(PEG)3 (1,11-Bis-Maleimidotriethyleneglycol), DPDPB (1,4-Di-(3'-[2'pyridyldithio]propionamido) butane), DTME (Dithiobis-maleimidoethane).

[00197] For the creation of XTEN-cross-linker conjugates for subsequent conjugation to payloads, as well as the creation of XTEN-folate conjugates, heterobifunctional cross-linkers are preferred as the sequential reactions can be controlled. As heterobifunctional cross-linkers possess two different reactive groups, their use in the compositions allows for sequential two-step conjugation. A heterobifunctional reagent is reacted with a first protein using the more labile group. In one embodiment, the conjugation of the heterobifunctional cross-linker to a reactive group in an XTEN results in an XTEN-cross-linker conjugate. After completing the reaction and removing excess unreacted cross-linker, the modified protein (such as the XTEN-cross-linker) can be added to the payload which interacts with a second reactive group of the cross-linker, resulting in an XTEN-payload conjugate. Most commonly used heterobifunctional cross-linkers contain an amine-reactive group at one end and a sulfhydryl–reactive group at the other end. Accordingly, these cross-linkers are suitable for use with cysteine- or lysine-engineered XTEN, or with the

alpha-amino group of the N-terminus of the XTEN. Non-limiting examples of heterobifunctional cross-linkers are AMAS (*N*-(α-Maleimidoacetoxy)-succinimide ester), BMPS (*N*-(β-Maleimidopropyloxy)succinimide ester), EMCS (*N*-(ε-Maleimidocaproyloxy)succinimide ester), GMBS (*N*-(γ-Maleimidobutyryloxy)succinimide ester), LC-SMCC (Succinimidyl 4-(*N*-maleimidomethyl) cyclohexane-1-carboxy-(6-amidocaproate)), LC-SPDP (Succinimidyl 6-(3'-[2-pyridyldithio]propionamido)hexanoate), MBS (*m*-Maleimidobenzoyl-*N*-hydroxysuccinimide ester), SBAP (Succinimdyl 3-(bromoacetamido)propionate), SIA (*N*-succinimidyl iodoacetate), SIAB (*N*-Succinimidyl(4-iodoacetyl)aminobenzoate), SMPB (Succinimidyl 4-(p-maleimidophenyl) butyrate), SMCC (Succinimidyl 4-(*N*-maleimido-methyl)cyclohexane-1-carboxylate), SM[PEG]$_2$ (NHS-PEG2-Maliemide), SM[PEG]$_4$ (NHS-PEG4-Maliemide), SM(PEG)$_6$ (NHS-PEG6-Maleimide), SM[PEG]$_8$ (NHS-PEG8-Maliemide), SM[PEG]$_{12}$ (NHS-PEG12-Maliemide), SM(PEG)$_{24}$ (NHS-PEG24-Maleimide), SMPB (Succinimidyl 4-(*p*-maleimido-phenyl)butyrate), SMPH (Succinimidyl-6-(β-maleimidopropionamido)hexanoate), SMPT (4-Succinimidyloxycarbonyl-methyl-α-(2-pyridyldithio)toluene), SPDP (*N*-Succinimidyl 3-(2-pyridyldithio)propionate), Sulfo-EMCS (*N*-(ε-Maleimidocaproyloxy)sulfosuccinimide ester), Sulfo-GMBS (*N*-(γ-Maleimidobutryloxy)sulfosuccinimide ester), Sulfo-KMUS (*N*-(κ-Maleimidoundecanoyloxy)sulfosuccinimide ester), Sulfo-LC-SMPT (Sulfosuccinimidyl 6-(α-methyl-α-[2-pyridyldithio]-toluamido)hexanoate), Sulfo-LC-SPDP (Sulfosuccinimidyl 6-(3'-[2-pyridyldithio]propionamido)hexanoate), Sulfo-MBS (*m*-Maleimidobenzoyl-*N*-hydroxysulfosuccinimide ester), Sulfo-SIAB (Sulfosuccinimidyl(4-iodo-acetyl)aminobenzoate), Sulfo-SMCC (Sulfosuccinimidyl 4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate), Sulfo-SMPB (Sulfosuccinimidyl 4-(*p*-maleimidophenyl)butyrate). An example of a heterobifunctional cross-linker that allows covalent conjugation of amine- and sulfhydryl-containing molecules is Sulfo-SMCC (SulfoSulfosuccinimidyl 4-(*N*-maleimidomethyl)cyclohexane-1-carboxylate). Sulfo-SMCC is a water soluble analog of SMCC that can be prepared in aqueous buffers up to 10 mM concentration. The cyclohexane ring in the spacer arm of this cross-linker decreases the rate of hydrolysis of the maleimide group compared to similar reagents not containing this ring. This feature enables XTEN that have been maleimide-activated with SMCC or Sulfo-SMCC to be lyophilized and stored for later conjugation to a sulfhydryl-containing molecule. Thus, in one embodiment, the invention provides an XTEN-cross-linker having an XTEN having at least about 80% sequence identity, 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% sequence identity, or is identical to a sequence or a fragment of a sequence selected from of Table 1, when optimally aligned, wherein XTEN-cross-linker has one or more cross-linkers of sulfo-SMCC linked to the α-amino group of the XTEN or the ε-amine of a lysine-engineered XTEN. In another embodiment, the invention provides an XTEN-cross-linker having an XTEN having at least about

80% sequence identity, 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% sequence identity, or is identical to a sequence or a fragment of a sequence selected from of Table 2, when optimally aligned., wherein the XTEN-cross-linker has one sulfo-SMCC linked to the amino group of the N-terminus of the XTEN. The foregoing described heterobifunctional cross-linkers conjugate two molecules via a single amine and a single cysteine. A special type of cross-linker was developed for site-specific conjugation to disulfide bridges in proteins (Balan S. et al. Site-specific PEGylation of protein disulfide bonds using a three-carbon bridge. (2007) Bioconjugate Chem. 18, 61-76; Brocchini S. et al. Disulfide bridge based PEGylation of proteins. (2008) Advanced Drug Delivery Reviews 60, 3-12). First, the linker is synthesized as an amine-specific 4-[2,2-bis[*p*-tolylsulfonyl)methyl]acetyl) benzoic acid–NHS ester. This molecule can be covalently attached to the amino group of XTEN yielding XTEN-Bis(sulfone). Incubation of the latter molecule in 50 mM sodium phosphate buffer, pH 7.8, will result in elimination of toluene sulfinic acid to generate XTEN-α,β-unsaturated β'-monosulfone. The resulting molecule will react with a disulfide bridge-containing payload protein in a site-specific manner. In a first step the disulfide bridge is converted into two thiols by reduction. In a second step, the XTEN-monosulfone bis-alkylates two cysteines resulting in a chemically-stable three-carbon bridge. The same α,β-unsaturated β'-monosulfone can be used not only for conjugation to two thiol groups derived from a disulfide bridge but also for conjugation to polyhistidine tags (Cong Y. et al. Site-specific PEGylation at histidine tags. (2012) Bioconjugate Chem. 23, 248-263).

[00198] Conjugation using XTEN-cross-linker compositions with the sulfo-SMCC is usually performed in a two-step process. In one embodiment, the amine-containing protein is prepared in conjugation buffer of, e.g., phosphate-buffered saline (PBS = 100mM sodium phosphate, 150mM sodium chloride, pH 7.2) or a comparable amine- and sulfhydryl-free buffer at pH 6.5-7.5. The addition of EDTA to 1-5mM helps to chelate divalent metals, thereby reducing disulfide formation in the sulfhydryl-containing protein. The concentration of the amine-containing protein determines the cross-linker molar excess to be used. In general, in protein samples of < 1mg/ml utilize an 40-80-fold molar excess, protein samples of 1-4mg/ml utilize a 20-fold molar excess, and protein samples of 5-10mg/ml utilize a 5- to 10-fold molar excess of the cross-linker. The reaction mixture (amine-containing protein and cross-linker) is incubated for 30 minutes at room temperature or 2 hours at 4°C and then the excess cross-linker is removed using a desalting column equilibrated with conjugation buffer. In the case of preparing a XTEN-cross-linker, the composition would be held at that point. In embodiments wherein the XTEN-cross-linker is conjugated to a payload, the sulfhydryl-containing payload and the XTEN-cross-linker conjugate are mixed in a molar ratio corresponding to that desired for the final conjugate (taking into account the number of expected cross-linkers conjugated to one or more amino groups per

molecule of the XTEN) and consistent with the single sulfhydryl group that exists on the payload. The reaction mixture is incubated at room temperature for 30 minutes or 2 hours at 4°C. Conjugation efficiency can be estimated by SDS-PAGE followed by protein staining or by appropriate analytical chromatography technique such as reverse phase HPLC or cation/anion exchange chromatography.

[00199] Cross-linkers can be classified as either "homobifunctional" or "heterobifunctional" wherein homobifunctional cross-linkers have two or more identical reactive groups and are used in one-step reaction procedures to randomly link or polymerize molecules containing like functional groups, and heterobifunctional cross-linkers possess different reactive groups that allow for either single-step conjugation of molecules that have the respective target functional groups or allow for sequential (two-step) conjugations that minimize undesirable polymerization or self-conjugation. In a preferred embodiment, where XTEN-cross-linkers are prepared and isolated as compositions for subsequent reaction, the XTEN-cross-linker is linked to a heterbifunctional cross-linker and has at least one reactive group available for subsequent reaction.

[00200] In one embodiment, the invention provides XTEN-cross-linkers and XTEN-payloads that are conjugated utilizing cleavable cross-linkers with disulfide bonds. Typically, the cleavage is effected by disulfide bond reducing agents such as β-mercaptoethanol, DTT, TCEP, however it is specifically contemplated that such compositions would be cleavable endogenously in a slow-release fashion by conditions with endogenous reducing agents (such as cysteine and glutathione). The following are non-limiting examples of such cross-linkers: APDP (*N*-(4-[*p*-Azidosalicylamido]butyl)-3'-(2'-pyridyldithio) propionamide), BASED (*Bis* (β-[4-azidosalicylamido]ethyl) disulfide), DPDPB (1,4-Di-(3'-[2'pyridyldithio]propionamido) butane), DSP (Dithiobis(succimidylpropionate) (Lomant's Reagent), DTBP (Dimethyl 3,3'-dithiobispropionimidate), DTME (Dithiobis-maleimidoethane), DTSSP (Sulfo-DSP) (3,3'-Dithiobis (sulfosuccinimidylpropionate)), LC-SPDP (Succinimidyl 6-(3'-[2-pyridyldithio]propionamido)hexanoate), PDPH (3-(2-Pyridyldithio)propionylhydrazide), SDAD (NHS-SS-Diazirine), SMPT (4-Succinimidyloxycarbonyl-methyl-α-(2-pyridyldithio)toluene), SPDP (*N*-Succinimidyl 3-(2-pyridyldithio)propionate), Sulfo-LC-SMPT (Sulfosuccinimidyl 6-(α-methyl-α-[2-pyridyldithio]-toluamido)hexanoate), Sulfo-LC-SPDP (Sulfosuccinimidyl 6-(3'-[2-pyridyldithio]propionamido)hexanoate), Sulfo-SAED (Sulfosuccimidyl 2-[7-azido-4-methylcoumarin-3-acetamido]ethyl-1,3'-dithiopropionate), Sulfo-SAND (Sulfosuccinimidyl-2-(m-azido-*o*-nitrobenzamido)ethyl 1,3'-dithiopropionate), Sulfo-SDAD (Sulfo-NHS-SS-Diazirine), Sulfo-SFAD (Sulfosuccinimidyl(perfluoroazidobenzamido)ethyl 1,3'-dithiopropionate. In another embodiment, XTEN-folate conjugates comprising BSOCOES (Bis(2-[succinimidoxycarbonyloxy]ethyl)sulfone) can be cleaved under alkaline conditions. In another embodiment, XTEN-folate conjugates comprising DST (Disuccinimidyl tartarate) and BMDB

(1,4 Bismaleimidyl-2,3-dihydroxybutane) can be cleaved by periodate oxidation. EGS (Ethylene glycol *bis*(succinimidylsuccinate)) and Sulfo-EGS (Ethylene glycol bis (sulfo-succinimidyl succinate)) are cleaved by hydroxylamine but would be expected to be cleaved endogenously such that the active payload would be released from the conjugate.

[00201] In general, the conjugation reagents described above assume that a cross-linker is reactive with the otherwise stable and inert groups such as amines, sulfhydryls and carboxyls. In other embodiments, the invention provides a different approach of conjugation based on separate modifications of the XTEN and payload with two functional groups which are stable and inactive toward biopolymers in general yet highly reactive toward each other. Several orthogonal reactions have been grouped under the concept of click chemistry, which provides XTEN-azide/alkyne reactants that have good stability properties and are therefor particularly suited as reagents for subsequent conjugation with payloads in a separate reaction (Kolb H.C., Finn M.G., Sharpless K.B. Click chemistry: diverse chemical function from a few good reactions. (2001) *Angew. Chem. Int. Ed. Engl.* 40(11), 2004-2021). Generally, click chemistry is used as a reaction concept which embraces reactions involving (1) alkyne-azide; (2) "ene"-thiol, and (3) aldehyde-hydrazide, and the invention contemplates use of all three. One example is the Huisgen 1,3-dipolar cycloaddition of alkynes to azides to form 1,4-disubsituted-1,2,3-triazoles, shown in FIG. 8. Azide and alkyne moieties can be introduced into peptide/protein or drug payloads or into XTEN by chemical modification of N-terminal α-amino groups, ε-amino groups of lysine, and sulfhydryl groups of cysteine. Table 9 provides non-limiting examples of click chemistry reactants contemplated for use in the making of the conjugate compositions, wherein one component of the intended conjugate (and XTEN or a payload) is reacted with a reactant 1 of the Table 9 and the second component (a payload or an XTEN) is reacted with a azide reactant 2 of the Table 9. For example, one molecule is modified with an alkyne moiety using an amine-reactive alkyne, such as 3-propargyloxypropanoic acid, NHS ester, acetylene-PEG4-NHS ester, dibenzylcyclooctyne (DBCO)-NHS ester, DBCO-PEG4-NHS ester, cyclooctyne (COT)-PEG2-NHS ester, COT-PEG3-NHS ester, COT-PEG4-NHS ester, COT-PEG2-pentafluorophenyl (PFP) ester, COT-PEG3-PFP ester, COT-PEG4-PFP ester, BCOT-PEG2-NHS ester, BCOT-PEG3-NHS ester, BCOT-PEG4-NHS ester, BCOT-PEG2-PFP ester, BCOT-PEG3-PFP ester, BCOT-PEG4-PFP ester. Alternatively, the molecule is modified with a sulfhydryl-reactive alkyne such as acetylene-PEG4-Maleimide, DBCO-Maleimide, or DBCO-PEG4-Maleimide. The second molecule is modified with azide-PEG2-NHS ester, azide-PEG3-NHS ester , azide-PEG4-NHS ester , azide-PEG2-PFP ester, azide-PEG3-PFP ester , azide-PEG4-PFP ester or azide-PEG4-Maleimide. The azide and alkyne moieties can be used interchangeably; they are biologically unique, stable and inert towards biological molecules and aqueous environments. When mixed, the azide and alkyne reactants form an irreversible covalent bond without any side reactions (Moses J.E. and

Moorhouse A.D. The growing applications of click chemistry. (2007) Chem. Soc. Rev. 36, 1249–1262; Breinbauer R. and Köhn M. Azide-alkyne coupling: a powerful reaction for bioconjugate chemistry. (2003) ChemBioChem 4(11), 1147–1149; Rostovtsev V.V., Green L.G., Fokin V.V., Sharpless K.B. A stepwise Huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes.(2002) Angew Chem Int Ed Engl. 41(14), 2596-2599). In one embodiment, the invention provides a conjugate comprising a first XTEN conjugated to a second XTEN wherein the first XTEN is linked to a alkyne reactant 1 from Table 9, the second XTEN is linked to a azide reactant 2 from Table 9, and then the first XTEN and the second XTEN are linked under conditions effective to react the alkyne reactant 1 and the azide reactant 2, resulting in the XTEN-XTEN conjugate. In another embodiment, the invention provides a conjugate comprising a first XTEN conjugated to a payload wherein the XTEN is linked to a alkyne reactant 1 from Table 9, the payload is linked to a azide reactant 2 from Table 9, and then the XTEN and the payload are linked under conditions effective to react the alkyne reactant 1 and the azide reactant 2, resulting in the XTEN-folate conjugate. In another embodiment, the invention provides a conjugate comprising a first XTEN conjugated to a payload wherein the XTEN is linked to a azide reactant 2 from Table 9, the payload is linked to a alkyne reactant 1 from Table 9, and then the XTEN and the payload are linked under conditions effective to react the alkyne reactant 1 and the azide reactant 2, resulting in the XTEN-folate conjugate. In the foregoing embodiments, the conditions to effect the reactions are those described herein or are reaction conditions known in the art for the conjugation of such reactants. The invention also contemplates the various combinations of the foregoing conjugates; e.g., an XTEN-XTEN conjugate in which the XTEN are linked by click chemistry reactants and in which one XTEN further comprises one or more molecules of a payload conjugated to the XTEN using click chemistry, an XTEN-XTEN conjugate in which the XTEN are linked by click chemistry reactants in which one XTEN further comprises one or more molecules of a first payload conjugated to the XTEN using click chemistry and the second XTEN further comprises one or more molecules of a second payload conjugated to the XTEN using click chemistry. Additional variations on these combinations will be readily apparent to those of ordinary skill in the art.

Table 9: Alkyne and Azide Click-chemistry Reactants

Alkyne Reactant 1	Attached to:	Azide Reactant 2	Attached to:
3-propargyloxypropanoic acid, NHS ester*	Amine	3-azide-propionic acid, NHS ester*	Amine
acetylene-(oxyethyl)$_n$-NHS ester*, where n is 1-10	Amine	6-azide-hexanoic acid, NHS ester*	Amine
dibenzylcyclooctyne (DBCO)-NHS ester*	Amine	3-azide-propionic acid, PFP ester	Amine

DBCO-(oxyethyl)$_n$- NHS ester*, where n is 1-10	Amine	6-azide-hexanoic acid, PFP ester	Amine
cyclooctyne (COT)-NHS ester*	Amine	azide-(oxyethyl)$_n$NHS ester*, where n is 1-10	Amine
COT-(oxyethyl)$_n$- NHS ester*, where n is 1-10	Amine	azide-(oxyethyl)$_n$- PFP ester, where n is 1-10	Amine
COT-(oxyethyl)$_n$-pentafluorophenyl (PFP) ester, where n is 1-10	Amine	1-azido-3,6,9,12-tetraoxapentadecan-15-oic acid N-hydroxysuccinimide ester	Amine
BCOT-NHS ester*	Amine	azide-(oxyethyl)$_n$- maleimide, where n is 1-10	Thiol
BCOT-(oxyethyl)$_n$- NHS ester*, where n is 1-10	Amine		
BCOT-(oxyethyl)$_n$-pentafluorophenyl (PFP) ester, where n is 1-10	Amine		
6-(11,12-didehydrodibenzo[b,f]azocin-5(6H)-yl)-6-oxohexanoic acid N-hydroxysulfosuccinimide ester	Amine		
ccetylene-(oxyethyl)$_n$-maleimide, where n is 1-10	Thiol		
DBCO-maleimide	Thiol		
COT-maleimide	Thiol		
BCOT-maleimide	Thiol		

*could be either NHS ester or sulfo-NHS ester

[00202] In some embodiments, the XTEN-XTEN conjugates and the XTEN-folate conjugates are conjugated using thio-ene based click chemistry that proceeds by free radical reaction, termed thiol-ene reaction, or anionic reaction, termed thiol Michael addition (see FIG. 9) (Hoyle C. E. and Bowman C.N. Thiol-ene click chemistry. (2010) *Angew. Chem. Int. Ed.* 49, 1540-1573). It particular, is believed that thiol Michael addition is better suited for XTEN-payload conjugates wherein the payload is a protein (Pounder R. J. et. al. Metal free thiol-maleimide 'Click' reaction as a mild functionalisation strategy for degradable polymers. (2008) *Chem Commun (Camb).* 41, 5158-5160). As at least one molecule needs to contain a free thiol group, a cysteine-engineered XTEN can be utilized if the payload does not contain cysteine. Alternatively, the thiol group can be introduced by chemical modification of N-terminal α-amino group or the lysine ε-amino groups of either the XTEN or the payload peptide/protein using thiolating reagents such as 2-iminothiolane (Traut's reagent), SATA (N-succinimidyl S-acetylthioacetate), SATP (N-succinimidyl S-acetylthiopropionate), SAT-PEO$_4$-Ac (N-Succinimidyl S-acetyl(thiotetraethylene glycol)), SPDP (*N*-Succinimidyl 3-(2-pyridyldithio)propionate), LC-SPDP (Succinimidyl 6-(3'-[2-pyridyldithio]propionamido)hexanoate). Such methods are known in the art (Carlsson J. et al. (1978) *Biochem. J.* 173, 723-737; Wang D. et al. (1997) *Bioconjug. Chem.* 8, 878-884; Traut R.R. et al. (1973) *Biochemistry* 12(17), 3266-3273; Duncan, R.J.S. et.al. (1983) *Anal. Biochem.* 132.

68-73; U.S. Pat. No. 5,708,146). The second component of thiol-Michael addition reaction requires a reagent with electron-deficient carbon-carbon double bond, such as in (meth)acrylates, maleimides, α,β-unsaturated ketones, fumarate esters, acrylonitrile, cinnamates, and crotonates. The N-maleimides are commonly used as sulfhydryl-reactive functionalities and can be introduced into the payload protein or the XTEN molecule via N-terminal α-amino group or Lys ε-amino group modification using commercially available heterobifunctional cross-linkers such as AMAS (*N*-(α-Maleimidoacetoxy)-succinimide ester), BMPS (*N*-(β-Maleimidopropyloxy)succinimide ester) and others described above. The resulting two molecules containing free thiol and maleimide moieties, respectively, form a stable covalent bond under mild conditions, resulting in a XTEN-payload linked by maleimide.

[00203] In other embodiments, XTEN-XTEN conjugates and XTEN-folate conjugates are created utilizing click chemistry based on reactions between hydrazides and aldehydes, such as described by Ganguly et al. and as shown in FIG. 9 (Ganguly T. et al. The hydrazide/hydrazone click reaction as a biomolecule labeling strategy for M(CO)3 (M = Re, 99mTc) radiopharmaceuticals. (2011) *Chem. Commun.* 47, 12846-12848). For example, an XTEN can be modified to have a hydrazine or hydrazide that is mixed with a payload having an aldehyde group to yield the desired XTEN-folate conjugate. In one embodiment, the invention provides XTEN with at least one hydrazine or hydrazide introduced in either the α–N-terminal amino group or, alternatively one or more lysine ε-amino groups are modified to provide an XTEN suitable as a reagent for conjugation to a target payload as it is considered to be stable. The resulting bis-arylhydrazones formed from aromatic hydrazines and aromatic aldehydes are stable to 92°C and a wide range of pH values from 2.0-10.0 (Solulink, Inc., Protein-Protein Conjugation Kit, Technical Manual, Catalog # S-9010-1). The leaving group in the reaction is water and no reducing agents (e.g., sodium cyanoborohydride) are required to stabilize the bond. Molecules modified with either hydrazine/hydrazide or aldehyde moieties have good stability in aqueous environments and remain active without special handling requirements. The amino group(s) of the XTEN molecule are modified by NHS-ester/hydrazide, such as SANH (succinimidyl 4-hydrazinonicotinate acetone hydrazone), C6-SANH (C6-Succinimidyl 4-hydrazinonicotinate acetone hydrazone), SHTH (Succinimidyl 4-hydrazidoterephthalate hydrocholoride). In a typical reaction, a protein is prepared as 1-5 mg/ml solution in modification buffer (100 mM Phosphate, 150 mM NaCl, pH 7.4) and the modifying agent is added in a 5- to 20-fold molar excess and the reaction is carried out for 2 hrs at room temperature. Separately, the payload molecule is modified with NHS-ester/aldehyde SFB (succinimidyl 4-formylbenzoate) or C6-SFB (C6-Succinimidyl 4-formylbenzoate) under similar conditions. Both modified molecules are then desalted into conjugation buffer (100 mM phosphate, 150 mM NaCl, pH 6.0). The resulting components are mixed together using 1 mole equivalent of a limiting protein and 1.5-2 mole equivalents of a

protein that can be used in abundance. A catalyst buffer of 100 mM aniline in 100 mM phosphate, 150 mM NaCl, pH 6.0 is added to adjust the final concentration of aniline to 10 mM and the reaction is carried out for 2 hrs at room temperature.

[00204] In another embodiment, the XTEN-folate conjugate can be produced by reaction between an aldehyde and primary amino group followed by reduction of the formed Schiff base with sodium borohydride or cyanoborohydride. As a first step in the method, an XTEN molecule, such as XTEN with a primary α-amino group or Lys-containing XTEN with an ε-amino group, is modified by NHS-ester/aldehyde SFB (succinimidyl 4-formylbenzoate), C6-SFB (C6-succinimidyl 4-formylbenzoate) or SFPA (succinimidyl 4-formylphenoxyacetate) using typical amine-NHS chemistry in an amine-free coupling buffer such as 0.1M sodium phosphate, 0.15M NaCl, pH 7.2. The resulting modified aldehyde-XTEN can either be held at this point as an XTEN-cross-linker composition or can be used as a reagent to create an XTEN-folate conjugate. To make the XTEN-folate, the modified aldehyde-XTEN is mixed with a payload with a reactive amino-group and a mild reducing agent such as 20-100 mM sodium cyanoborohydride. The reaction mixture is incubated up to 6 hours at room temperature or overnight at 4°C. Unreacted aldehyde groups are then blocked with 50-500 mM Tris•HCl, pH 7.4 and 20-100 mM sodium cyanoborohydride, permitting separation of the conjugated purified XTEN-folate.

[00205] In other embodiments, the invention provides XTEN-folate conjugates comprising peptides or protein payloads wherein the payload is conjugated via chemical ligation based on the reactivity of the peptide/protein C-terminal acyl azide of the payload. As an example, when the peptide or protein is produced using solid-phase peptide synthesis (SPPS) with hydroxymethylbenzoic acid (HMBA) resin, the final peptide can be cleaved from the resin by a variety of nucleophilic reagents to give access to peptides with diverse C-terminal functionalities. In one embodiment, the method includes hydrazinolysis of the peptidyl/protein resins to yield peptide or protein hydrazides. Nitrosation of resulting acyl hydrazides with sodium nitrite or *tert*-butyl nitrite in dilute hydrochloric acid then results in formation of acyl azides. The resulting carbonyl azide (or acyl azide) is an activated carboxylate group (esters) that can react with a primary amine of an XTEN to form a stable amide bond, resulting in the XTEN-folate conjugate. In alternative embodiments, the primary amine could be the α-amine of the XTEN N-terminus or one or more ε-amine of engineered lysine residues in the XTEN sequence. In the conjugation reaction, the azide function is the leaving group, shown in FIG. 11. The conjugation reaction with the amine groups occurs by attack of the nucleophile at the electron-deficient carbonyl group (Meienhofer, J. (1979) The Peptides: Analysis, Synthesis, Biology. Vol. 1, Academic Press: N.Y.; ten Kortenaar P. B. W. et. al. Semisynthesis of horse heart cytochrome c analogues from two or three fragments. (1985) *Proc. Natl. Acad. Sci. USA* 82, 8279-8283).

[00206] In another embodiment, the invention provides XTEN-folate conjugates prepared by enzymatic ligation. Transglutaminases are enzymes that catalyze the formation of an isopeptide bond between the γ-carboxamide group of glutamine of a payload peptide or protein and the ε-amino group of a lysine in a lysine-engineered XTEN, thereby creating inter- or intramolecular cross-links between the XTEN and payload (see FIG. 11), resulting in the composition (Lorand L, Conrad S.M. Transglutaminases.(1984) Mol. Cell Biochem. 58(1-2), 9-35). Non-limiting examples of enzymes that have been successfully used for ligations are factor XIIIa (Schense J.C., Hubbell J.A. Cross-linking exogenous bifunctional peptides into fibrin gels with factor XIIIa. (1999) *Bioconjug. Chem.* 10(1):75-81) and tissue transglutaminase (Collier J.H., Messersmith P.B. Enzymatic modification of self-assembled peptide structures with tissue transglutaminase. (2003) *Bioconjug. Chem.* 14(4), 748-755; Davis N. E., Karfeld-Sulzer L. S., Ding S., Barron A. E. Synthesis and characterization of a new class of cationic protein polymers for multivalent display and biomaterial applications. (2009) *Biomacromolecules* 10 (5), 1125-1134). The glutamine substrate sequence GQQQL is known to have high specificity toward tissue transglutaminase (Hu B.H., Messersmith P.B. Rational design of transglutaminase substrate peptides for rapid enzymatic formation of hydrogels.(2003*) J. Am. Chem. Soc.* 125(47), 14298-14299). Tissue transglutaminase sequence specificity was less stringent for an acyl acceptor (lysine) than for acyl donor (glutamine) (Greenberg C. S., Birckbichler P. J., Rice R. H. Transglutaminases: multifunctional cross-linking enzymes that stabilize tissues. (1991) *FASEB J.* 1991, 5, 3071–3077).

[00207] While the various embodiments of conjugation chemistry have been described in terms of protein-protein conjugations, it is specifically intended that in practicing the invention, the payload moiety of the XTEN-folate conjugates can be a small molecule drug in those conjugation methods applicable to functional groups like amines, sulfhydryls, carboxyl that are present in the target small molecule drugs. It will be understood by one of ordinary skill in the art that one can apply even more broad chemical techniques compared to protein and peptides whose functionalities are usually limited to amino, sulfhydryl and carboxyl groups. Drug payloads can be conjugated to the XTEN through functional groups including, but not limited to, primary amino groups, aminoxy, hydrazide, hydroxyl, thiol, thiolate, succinate (SUC), succinimidyl succinate (SS), succinimidyl propionate (SPA), succinimidyl butanoate (SBA), succinimidyl carboxymethylate (SCM), benzotriazole carbonate (BTC), N-hydroxysuccinimide (NHS), p-nitrophenyl carbonate (NPC). Other suitable reactive functional groups of drug molecule payloads include acetal, aldehydes (e.g., acetaldehyde, propionaldehyde, and butyraldehyde), aldehyde hydrate, alkenyl, acrylate, methacrylate, acrylamide, active sulfone, acid halide, isocyanate, isothiocyanate, maleimide, vinylsulfone, dithiopyridine, vinylpyridine, iodoacetamide, epoxide, glyoxal, dione, mesylate, tosylate, and tresylate.

[00208] In another embodiment, the drug payloads can also be conjugated to XTEN-cross-linker conjugates using a heterocycle ring system in which one or more ring atoms is a heteroatom, e.g. a nitrogen, an oxygen, a phosphorus or a sulfur atom. The heterocycle group comprises at least 1 to as many as 20 carbon atoms and 1 to 3 heteroatoms selected from N, O, P, and S. In the embodiment, the heterocycle may be a monocycle having 3 to 7 ring members (2 to 6 carbon atoms and 1 to 3 heteroatoms selected from N, O, P, and S) or a bicycle having 7 to 10 ring members (4 to 9 carbon atoms and 1 to 3 heteroatoms selected from N, O, P, and S), for example: a bicyclo [4,5], [5,5], [5,6], or [6,6] system. Heterocycles are described in Paquette, Leo A. "Principles of Modern Heterocyclic Chemistry", W. A. Benjamin, New York, (1968); "The Chemistry of Heterocyclic Compounds, A series of Monographs" (John Wiley & Sons, New York, 1950 to present), in particular Volumes 13, 14, 16, 19, and 28. Non-limiting examples of heterocycles that may be found in drugs suitable for conjugation include pyridyl, dihydroypyridyl, tetrahydropyridyl (piperidyl), thiazolyl, tetrahydrothiophenyl, sulfur oxidized tetrahydrothiophenyl, pyrimidinyl, furanyl, thienyl, pyrrolyl, pyrazolyl, imidazolyl, tetrazolyl, benzofuranyl, thianaphthalenyl, indolyl, indolenyl, quinolinyl, isoquinolinyl, benzimidazolyl, piperidinyl, 4-piperidonyl, pyrrolidinyl, 2-pyrrolidonyl, pyrrolinyl, tetrahydrofuranyl, bis-tetrahydrofuranyl, tetrahydropyranyl, bis-tetrahydropyranyl, tetrahydroquinolinyl, tetrahydroisoquinolinyl, decahydroquinolinyl, octahydroisoquinolinyl, azocinyl, triazinyl, 6H-1,2,5-thiadiazinyl, 2H,6H-1,5,2-dithiazinyl, thienyl, thianthrenyl, pyranyl, isobenzofuranyl, chromenyl, xanthenyl, phenoxathinyl, 2H-pyrrolyl, isothiazolyl, isoxazolyl, pyrazinyl, pyridazinyl, indolizinyl, isoindolyl, 3H-indolyl, 1H-indazolyl, purinyl, 4H-quinolizinyl, phthalazinyl, naphthyridinyl, quinoxalinyl, quinazolinyl, cinnolinyl, pteridinyl, 4Ah-carbazolyl, carbazolyl, β-carbolinyl, phenanthridinyl, acridinyl, pyrimidinyl, phenanthrolinyl, phenazinyl, phenothiazinyl, furazanyl, phenoxazinyl, isochromanyl, chromanyl, imidazolidinyl, imidazolinyl, pyrazolidinyl, pyrazolinyl, piperazinyl, indolinyl, isoindolinyl, quinuclidinyl, morpholinyl, oxazolidinyl, benzotriazolyl, benzisoxazolyl, oxindolyl, benzoxazolinyl, and isatinoyl.

[00209] In some embodiments of the XTEN-folate conjugates with drugs as the payload, the drug molecules are attached to lysine- or cysteine engineered XTEN (such as the sequences of Table 1) by cross-linkers having two reactive sites for binding to the drug and the XTEN. Preferred cross-inker groups are those that are relatively stable to hydrolysis in the circulation, are biodegradable and are nontoxic when cleaved from the conjugate. In addition, the use of cross-linkers can provide the potential for conjugates with an even greater flexibility between the drug and the XTEN, or provide sufficient space between the drug and the XTEN such that the XTEN does not interfere with the binding between the pharmacophore and its binding site. In one embodiment, a cross-linker has a reactive site that has an electrophilic group that is reactive to a nucleophilic group present on an XTEN. Preferred nucleophiles include thiol, thiolate, and

primary amine. The heteroatom of the nucleophilic group of a lysine- or cysteine-engineered XTEN is reactive to an electrophilic group on a cross-linker and forms a covalent bond to the cross-linker unit, resulting in an XTEN-cross-linker conjugate. Useful electrophilic groups for cross-linkers include, but are not limited to, maleimide and haloacetamide groups, and provide a convenient site for attachment to the XTEN. In another embodiment, a cross-linker has a reactive site that has a nucleophilic group that is reactive to an electrophilic group present on a drug such that a conjugation can occur between the XTEN-cross-linker and the payload drug, resulting in an XTEN-drug conjugate. Useful electrophilic groups on a drug include, but are not limited to, hydroxyl, thiol, aldehyde, alkene, alkane, azide and ketone carbonyl groups. The heteroatom of a nucleophilic group of a cross-linker can react with an electrophilic group on a drug and form a covalent bond. Useful nucleophilic groups on a cross-linker include, but are not limited to, hydrazide, oxime, amino, hydrazine, thiosemicarbazone, hydrazine carboxylate, and arylhydrazide. The electrophilic group on a drug provides a convenient site for attachment to a cross-inker.

[00210] In a particular embodiment, the conjugation of drugs to the lysine epsilon amino group of a subject lysine-engineered XTEN makes use of a reactive drug-N-hydroxylsuccinimide reactant, or esters such as drug-succinimidyl propionate, or drug-succinimidyl butanoate or other drug-succinimide conjugates. Alternatively, lysine residues of the subject lysine-engineered XTEN may be used to introduce free sulfhydryl groups through reaction with 2-iminothiolane. Alternatively, targeting substance lysines of subject lysine-engineered XTEN may be linked to a heterobifunctional reagent having a free hydrazide or aldehyde group available for conjugation with an active drug agent. Reactive esters can conjugate at physiological pH, but less reactive derivatives typically require higher pH values. Low temperatures may also be employed if a labile protein payload is being used. Under low temperature conditions, a longer reaction time may be used for the conjugation reaction.

[00211] In another particular embodiment, the invention provides XTEN-folate conjugates with an amino group conjugation with lysine residues of a subject lysine-engineered XTEN wherein the conjugation is facilitated by the difference between the pKa values of the α-amino group of the N-terminal amino acid (approximately 7.6 to 8.0) and pKa of the ε-amino group of lysine (approximately 10). Conjugation of the terminal amino group often employs reactive drug-aldehydes (such as drug-propionaldehyde or drug-butylaldehyde), which are more selective for amines and thus are less likely to react with, for example, the imidazole group of histidine. In addition, amino residues are reacted with succinic or other carboxylic acid anhydrides, or with N,N′-Disuccinimidyl carbonate (DSC), N,N′-carbonyl diimidazole (CDI), or p-nitrophenyl chloroformate to yield the activated succinimidyl carbonate, imidazole carbamate or p-nitrophenyl carbonate, respectively. Derivatization with these agents has the effect of reversing the charge of

the lysinyl residues. Conjugation of a drug-aldehyde to the terminal amino group of a subject XTEN typically takes place in a suitable buffer performed at a pH which allows one to take advantage of the pKa differences between the ε-amino groups of the lysine residues and that of the α-amino group of the N-terminal residue of the protein. In the method of the embodiment, the reaction for coupling uses a pH in the range of from about pH 7 up to about 8. Useful methods for conjugation of the lysine epsilon amino group have been described in U.S. Pat. No. 4,904,584 and U.S. Pat. No. 6,048,720.

[00212] The person with ordinary skill in the art will be aware that the activation method and/or conjugation chemistry to be used in the creation of the XTEN-folate conjugates depends on the reactive groups of the XTEN polypeptide as well as the functional groups of the drug moiety (e.g., being amino, hydroxyl, carboxyl, aldehyde, sulfhydryl, alkene, alkane, azide, etc), the functional group of the drug-cross-linker reactant, or the functional group of the XTEN-cross-linker reactant. The drug conjugation may be directed towards conjugation to all available attachment groups on the engineered XTEN polypeptide such as the specific engineered attachment groups on the incorporated cysteine residues or lysine residues. In order to control the reactants such that the conjugation is directed to the appropriate reactive site, the invention contemplates the use of protective groups during the conjugation reaction. A "protecting group" is a moiety that prevents or blocks reaction of a particular chemically reactive functional group in a molecule under certain reaction conditions. The protecting group will vary depending upon the type of chemically reactive group being protected as well as the reaction conditions to be employed, as well as the presence of additional reactive groups in the molecule. Non-limiting examples of functional groups which may be protected include carboxylic acid groups, hydroxyl groups, amino groups, thiol groups, and carbonyl groups. Representative protecting groups for carboxylic acids and hydroxyls include esters (such as a p-methoxybenzyl ester), amides and hydrazides; for amino groups, carbamates (such as tert-butoxycarbonyl) and amides; for hydroxyl groups, ethers and esters; for thiol groups, thioethers and thioesters; for carbonyl groups, acetals and ketals; and the like. Such protecting groups are well-known to those skilled in the art and are described, for example, in T. W. Greene and G. M. Wuts, *Protecting Groups in Organic Synthesis* , Third Edition, Wiley, New York, 1999, and references cited therein. The conjugation may be achieved in one step or in a stepwise manner (e.g., as described in WO 99/55377), such as through addition of a reaction intermediate cross-linker, using the cross-linkers disclosed herein or those known in the art to be useful for conjugation to cysteine or lysine residues of polypeptides to be linked to reactive functional groups on drug molecules.

[00213] In some embodiments of the invention, the method for conjugating a cross-linker to a cysteine-engineered XTEN may provide that the XTEN is pre-treated with a reducing agent, such as dithiothreitol (DTT) to reduce any cysteine disulfide residues to form highly nucleophilic

cysteine thiol groups (—CH_2 SH). The reducing agent is subsequently removed by any conventional method, such as by desalting. The reduced XTEN thus reacts with drug-linker compounds, or cross-linker reagents, with electrophilic functional groups such as maleimide or α-halo carbonyl, according to, for example, the conjugation method of Klussman et al. (2004) *Bioconjugate Chemistry* 15(4), 765-773. Conjugation of a cross-linker or a drug to a cysteine residue typically takes place in a suitable buffer at pH 6-9 at temperatures varying from 4°C to 25°C for periods up to about 16 hours. Alternatively, the cysteine residues can be derivatized. Suitable derivatizing agents and methods are well known in the art. For example, cysteinyl residues most commonly are reacted with α-haloacetates (and corresponding amines), such as iodoacetic acid or iodoacetamide, to give carboxymethyl or carboxyamidomethyl derivatives. Cysteinyl residues also are derivatized by reaction with bromotrifluoroacetone, α-bromo-β-(4-imidozoyl)propionic acid, chloroacetyl phosphate, N-alkylmaleimides, 3-nitro-2-pyridyl disulfide, methyl 2-pyridyl disulfide, p-chloromercuribenzoate, 2-chloromercuri-4-nitrophenol, or chloro-7-nitrobenzo-2-oxa-1,3-diazole.

[00214] In some instances, the conjugation is performed under conditions aiming at reacting as many of the available XTEN attachment groups as possible with drug or drug-linker molecules. This is achieved by means of a suitable molar excess of the drug in relation to the polypeptide. Typical molar ratios of activated drug or drug-linker molecules to polypeptide are up to about 1000-1, such as up to about 200-1 or up to about 100-1. In some cases, the ratio may be somewhat lower, however, such as up to about 50-1, 10-1 or 5-1. Equimolar ratios also may be used.

[00215] In the embodiments, the XTEN-folate conjugates of the disclosure retain at least a portion of the pharmacologic activity compared to the corresponding payload not linked to XTEN. In one embodiment, the XTEN-folate retains at least about 1%, or at least about 5%, or at least about 10%, or at least about 20%, or at least about 30%, or at least about 40%, or at least about 50%, or at least about 60%, or at least about 70%, or at least about 80%, or at least about 90%, or at least about 95% of the pharmacologic activity of the payload not linked to XTEN.

[00216] In one embodiment, XTEN-folate conjugates can be designed to release the payload in the body by unspecific or enzymatic hydrolysis of the linker, including disulfide bond reduction, pH-dependent release, or by exogenous or endogenous proteases, including the proteases of Table 7. Macromolecules can be taken up by the cell either through receptor-mediated endocytosis, adsorptive endocytosis or fluid phase endocytosis (Jain R.K. Transport of molecules across tumor vasculature. (1987) *Cancer Metastasis Rev.* 6(4), 559-593; Jain R.K. Transport of molecules, particles, and cells in solid tumors. (1999) *Ann. Rev. Biomed. Eng.* 1, 241-263; Mukherjee S., Ghosh R.N., Maxfield F.R. Endocytosis. (1997) *Physiol. Rev.* 77(3), 759-803). Upon cellular uptake of XTEN-folate, the payload can be released by low pH values in endosomes (pH 5.0 – 6.5) and lysosomes (pH 4.5 – 5.0), as well as by lysosomal enzymes (e.g., esterases and

proteases). Example of acid-sensitive cross-linker is 6-maleimidodocaproyl hydrazone which can be coupled to thiol-bearing carriers. The hydrazone linker is rapidly cleaved at pH values < 5 allowing a release of the payload in the acidic pH of endosomes and lysosomes following internalization of the conjugate (Trail P.A. et al. Effect of linker variation on the stability, potency, and efficacy of carcinoma-reactive BR64-doxorubicin immunoconjugates. (1997) *Cancer Res.* 57(1), 100-105; Kratz F. et al. Acute and repeat-dose toxicity studies of the (6-maleimidocaproyl)hydrazone derivative of doxorubicin (DOXO-EMCH), an albumin-binding prodrug of the anticancer agent doxorubicin. (2007) *Hum. Exp. Toxicol.* 26(1), 19-35). Clinically approved mAb–drug conjugate, gemtuzumab ozogamicin (Mylotarg™) is a drug–antibody conjugate containing a humanized mAb P67.6 against CD33, linked chemically to the cytotoxic antibiotic agent calicheamicin. The linker between the antibody and the drug incorporates two labile bonds: a hydrazone and a sterically hindered disulfide. It has been shown that the acid-sensitive hydrazone bond is the actual cleavage site (Jaracz S., Chen J., Kuznetsova L.V., Ojima I. Recent advances in tumor-targeting anticancer drug conjugates. (2005) *Bioorg. Med. Chem.* 13(17), 5043-5054).

[00217] For those XTEN-folate conjugates in which the payload is linked by a disulfide bond, the payload can be released from XTEN by reduction of disulfide bond within the labile linker. For example, huN901–DM1 is a tumor-activated immunotherapeutic prodrug developed by ImmunoGen, Inc. for the treatment of small cell lung cancer. The prodrug consists of humanized anti-CD56 mAb (huN901) conjugated with microtubule inhibitor maytansinoid DM1. An average of 3.5 – 3.9 molecules of DM1 are bound to each antibody via hindered disulfide bonds. Although the disulfide link is stable in blood, it is cleaved rapidly on entering the cell targeted by huM901, thus releasing active DM1 (Smith S.V. Technology evaluation: huN901-DM1, ImmunoGen. (2005) *Curr. Opin. Mol. Ther.* 7(4), 394-401). DM1 has been also coupled to Millennium Pharmaceuticals MLN-591, an anti-prostate-specific membrane antigen mAb. DM1 is linked to the antibody via a hindered disulfide bond that provides serum stability at the same time as allowing intracellular drug release on internalization (Henry M.D. et al. A prostate-specific membrane antigen-targeted monoclonal antibody-chemotherapeutic conjugate designed for the treatment of prostate cancer. (2004) *Cancer Res.* 64(21), 7995-8001).

[00218] Release of the payload from the carrier XTEN can be achieved by creating compositions using short cleavable peptides as linkers between the payload and XTEN. Example of the conjugate assessed clinically is doxorubicin–HPMA (N-(2-hydroxypropyl)methacrylamide) conjugate in which doxorubicin is linked through its amino sugar to the HPMA copolymer via a tetrapeptide spacer GlyPheLeuGly that is cleaved by lysosomal proteases, such as cathepsin B (Vasey P.A. et al. Phase I clinical and pharmacokinetic study of PK1 [N-(2-hydroxypropyl)methacrylamide copolymer doxorubicin]: first member of a new class of

chemotherapeutic agents-drug-polymer conjugates. (1999) *Clin. Cancer Res.* 5(1), 83-94). Other examples of carrier-drug conjugates with peptide linkers that reached clinical stage of development are macromolecular platinum complexes. Two HPMA-based drug candidates consisted of a HPMA copolymer backbone to which the complexing aminomalonate platinum complexes were bound through cathepsin B-cleavable peptide spacer GlyPheLeuGly or tripeptide spacer GlyGlyGly (Rademaker-Lakhai J.M. et al. A Phase I and pharmacological study of the platinum polymer AP5280 given as an intravenous infusion once every 3 weeks in patients with solid tumors. (2004) *Clin. Cancer Res.* 10(10), 3386-3395; Sood P. et al. Synthesis and characterization of AP5346, a novel polymer-linked diaminocyclohexyl platinum chemotherapeutic agent. (2006) Bioconjugate Chem. 17(5), 1270-1279).

[00219] A highly selective method was developed to target prostate cancer via prostate-specific antigen (PSA) protease which is almost exclusively expressed in prostate tissue and prostate carcinomas. A novel albumin-binding prodrug of paclitaxel, EMC-ArgSerSerTyrTyrSerLeu-PABC-paclitaxel (EMC: ε-maleimidocaproyl; PABC: p-aminobenzyloxycarbonyl) was synthesized. This prodrug was water soluble and was bound to endogenous and exogenous albumin. Albumin-bound form of the prodrug was cleaved by PSA releasing the paclitaxel-dipeptide Ser-Leu-PABC-paclitaxel. Due to the incorporation of a PABC self-eliminating linker, this dipeptide was rapidly degraded to liberate paclitaxel as a final cleavage product (Elsadek B. et al. Development of a novel prodrug of paclitaxel that is cleaved by prostate-specific antigen: an in vitro and in vivo evaluation study. (2010) *Eur. J. Cancer* 46(18), 3434-3444).

[00220] Self-immolative spacers have gained significant interest due to their utility in prodrug delivery systems. Several reports described linear self-eliminating systems or dendrimeric structures which can release all of their units through a domino-like chain fragmentation, initiated by a single cleavage event (Haba K. et al. Single-triggered trimeric prodrugs. (2005) *Angew. Chem., Int. Ed.* 44, 716-720; Shabat D. Self-immolative dendrimers as novel drug delivery platforms. (2006) *J. Polym. Sci., Part A: Polym. Chem.* 44, 1569-1578.Warnecke A., Kratz F. 2,4-Bis(hydroxymethyl)aniline as a building block for oligomers with self-eliminating and multiple release properties. (2008) *J. Org. Chem.* 73, 1546-1552; Sagi A. et al. Self-immolative polymers. (2008) *J. Am. Chem. Soc.* 130, 5434–5435). In one study, a self-immolative dendritic prodrug with four molecules of the anticancer agent camptothecin and two molecules of PEG5000 was designed and synthesized. The prodrug was effectively activated by penicillin-G-amidase under physiological conditions and free camptothecin was released to the reaction media to cause cell-growth inhibition (Gopin A. et al. Enzymatic activation of second-generation dendritic prodrugs: conjugation of self-immolative dendrimers with poly(ethylene glycol) via click chemistr y. (2006) *Bioconjugate Chem.* 17, 1432-1440). Incorporation of a specific enzymatic substrate, cleaved by a protease that is overexpressed in tumor cells, could generate highly efficient cancer-cell-specific

dendritic prodrug activation systems. Non-limiting examples of spacer sequences that are cleavable by proteases are listed in Table 7.

[00221] In some embodiments, the invention provides XTEN-folate configurations, including dimeric conjugates in which the payload is attached to the XTEN using a labile linker as described herein, above, and as illustrated in FIG. 50. In one embodiment of the foregoing, the composition further includes a targeting component to deliver the composition to a ligand or receptor on a targeted cell. As illustrated in FIG. 50, once the XTEN-folate conjugate attaches to the folate receptor of the tumor cell, it is internalized and the toxin payload can be cleaved from the conjugate by intracellular proteases, releasing it in a more active form.

[00222] The unstructured characteristics and uniform composition and charge of XTEN result in properties that can be exploited for purification of XTEN-folate conjugates following a conjugation reaction. Of particular utility is the capture of XTEN conjugates by ion exchange, which allows the removal of un-reacted payload and payload derivatives. Of particular utility is the capture of conjugates by hydrophobic interaction chromatography (HIC). Due to their hydrophilic nature, most XTEN polypeptides show low binding to HIC resins, which facilitates the capture of XTEN-folate conjugates due to hydrophobic interactions between the payload and the column material, and their separation from un-conjugated XTEN that failed to conjugate to the payload during the conjugation process. The high purity of XTEN and XTEN-folate conjugates offers a significant benefit compared to most chemical or natural polymers, particularly pegylated payloads. Most chemical and natural poylmers are produced by random-or semi-random polymerization, which results in the generation of many homologs. Such polymers can be fractionated by various methods to increase fraction of the target entity in the product. However, even after enrichment most preparations of natural polymers and their payload conjugates contain less than 10% target entity. Examples of PEG conjugates with G-CSF have been described in [Bagal, D., et al. (2008) Anal Chem, 80: 2408-18]. This publication shows that even a PEG conjugate that is approved for therapeutic use contains more than 100 homologs that occur with a concentration of at least 10% of the target entity.

[00223] The complexity of random polymers, such as PEG, is a significant impediment for the monitoring and quality control during conjugation and purification. In contrast, XTEN purified by the methods described herein have high levels of purity and uniformity. In addition, the conjugates created as described herein routinely contain greater than about 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% of the intended target and in the intended configuration, resulting in easy to interpret mass spectra and chromatograms.

4. Dimeric Payload Configurations of XTEN-Payload Compositions

[00224] In another aspect, the invention provides conjugates in which two XTEN or XTEN-folate precursor segments are linked by a divalent cross-linker, resulting in a divalent

configuration, such as shown in FIG. 17B. In one embodiment, the invention provides dimeric conjugates containing two different payload molecules linked separately to two cysteine-engineered XTEN backbones, as illustrated in FIG. 17A, resulting in a dimeric conjugate. In one embodiment, the bivalent configuration conjugate is created by reacting the engineered XTEN, such as those specifically provided in Table 1, with a first XTEN-payload comprising a linker appropriate for reaction with the cysteine-engineered XTEN, followed by a second reaction with a second XTEN-payload comprising a linker appropriate for reaction with the lysine-engineered XTEN, resulting in the final product. The number and location of payloads is controlled by the design of the engineered XTEN, with the placement of the reactive thiol or amino group being determinative. In one embodiment, the bivalent conjugate comprises a single molecule of a first payload and a single molecule of a second payload linked to the cysteine-lysine-engineered XTEN by linkers. In another embodiment, the bivalent conjugate comprises one, or two, or three, or four, or five molecules of a first payload and a single molecule of a second payload linked to the cysteine-lysine-engineered XTEN by linkers. In another embodiment, the bivalent conjugate comprises one, or two, or three, or four, or five molecules of a first payload and one, or two, or three, or four, or five molecules of a second payload linked to the cysteine-lysine-engineered XTEN by linkers.

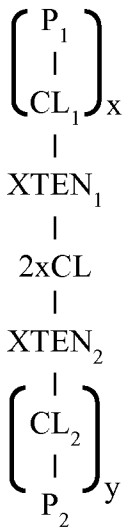
[00225] The central XTEN carrier can be used in the preparation of a monodisperse dimeric conjugate composition by conjugating the pendant functional groups either directly with reactive functional groups on an effector, therapeutic agent, or water-soluble polymer, or by first further functionalizing the pendant functional group to improve its reactivity with or selectivity for a functional group on an effector, therapeutic agent, or water-soluble polymer.

[00226] Some suitable reactive groups for use in the present invention include thiols, primary amines, carboxylic acids, alcohols, phenols, hydrazides, hydrazones, ketones, oximes, and aldehydes. Any of these reactive groups can be present or formed on the effector, therapeutic agent, water-soluble polymer, or peptidic carrier.

[00227] A method for preparing a monodisperse macromolecular conjugate composition for in-vivo delivery of a therapeutic agent is also provided, wherein the macromolecular conjugate comprises a cysteine-engineered XTEN carrier conjugated with one or more targeting moieties and one or more therapeutic agents, wherein the method comprises the steps of (a) providing a XTEN carrier comprising orthogonal pendant reactive groups linked to thiol groups of cysteine residues of the XTEN for specific attachment of a predetermined targeting moiety or therapeutic agent; (b) attaching an targeting moiety to a first XTEN carrier by means of a cross-linker such that each copy of the targeting moiety is attached to the same reactive group on each XTEN carrier in the composition; (c) attaching a drug payload to a second XTEN carrier by means of a cross-linker such that each copy of the drug is attached to the same reactive group on each XTEN

carrier in the composition; and (d) conjugating the first XTEN and the second XTEN by reacting an azide reactant linked to the N-terminus of one of the XTEN carriers and an alkyne reactant linked to the N-terminus of the other of the XTEN carriers, resulting in the final XTEN composition comprising the targeting moiety and the drug payloads.

[00228] In one embodiment, the invention provides a dimeric conjugate having the configuration of formula I



I

wherein independently for each occurrence P_1 is folic acid, P_2 is a second payload selected from the group of payloads consisting of auristatin E, auristatin F, monomethyl auristatin E, monomethyl auristatin F, and valine-citrulline-p-aminobenzyloxycarbonyl-monomethylauristatin E; CL_1 is a cross-linker selected from the group consisting of the cross-linkers of Table 8 and each CL_1 is linked to a cysteine thiol of the $XTEN_1$; x is an integer from 1 to about 20, or 1 to about 10, or 1 to about 5, or is 9, or is 3, or is 2, or is 1; CL_2 is a cross-linker selected from the group consisting of the cross-linkers of Table 8 and each CL_2 is linked to a cysteine thiol of $XTEN_2$; y is an integer from 1 to about 20, or 1 to about 10, or 1 to about 5, or is 9, or is 3, or is 2, or is 1, with the proviso that x + y is ≥ 2; 2xCL is alternatively a divalent cross-linker or the reaction product of a first and a second click chemistry reactant selected from Table 9 and is linked to the N-terminus of $XTEN_1$ and $XTEN_2$; $XTEN_1$ is a first substantially homogeneous XTEN having at least 80%, or 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%, or having 100% sequence identity to a sequence selected from the group of sequences set forth in Table 1; and $XTEN_2$ is a first substantially homogeneous having at least 80%, or 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%, or having 100% sequence identity to a sequence selected from the group of sequences set forth in Table 1. In another embodiment

of the conjugate of formula I, $XTEN_1$ and $XTEN_2$ are identical. In another embodiment of the conjugate of formula I, $XTEN_1$ and $XTEN_2$ are different. In another embodiment, the invention provides a preparation of the conjugate of formula I in which at least about 80%, or 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% of the respective $XTEN_1$ and $XTEN_2$ molecules of the preparation of the conjugate have an identical sequence length. In another embodiment, $XTEN_1$ and $XTEN_2$ are identical and have the sequence selected from the group consisting of Seg 174, Seg 175, Seg 176, Seg 177, Seg 186, Seg 187, Seg 188, Seg 189, Seg 190, Seg 191, Seg 192, Seg 193, Seg 194, Seg 195, Seg 196, Seg 197, Seg 198, and Seg 199 of Table 1. In another embodiment, $XTEN_1$ and $XTEN_2$ are identical and have the sequence of Seg 176 of Table 1. In another embodiment, $XTEN_1$ and $XTEN_2$ are different and each is independently selected from the group consisting of Seg 174, Seg 175, Seg 176, Seg 177, Seg 186, Seg 187, Seg 188, Seg 189, Seg 190, Seg 191, Seg 192, Seg 193, Seg 194, Seg 195, Seg 196, Seg 197, Seg 198, and Seg 199 of Table 1. In another embodiment, $XTEN_1$ has the sequence of Seg 176 and $XTEN_2$ has the sequence of Seg 177.

[00229] In one embodiment, XTEN-folate conjugates can be designed to release the payload in the body by unspecific or enzymatic hydrolysis of the linker, including disulfide bond reduction, pH-dependent release, or by exogenous or endogenous proteases, including the proteases of Table 8. In a particular embodiment, the drug for conjugation is derivatized to introduce a cleavable linker such as, but not limited to, valine-citrulline-PAB, wherein the self-immolative linker is capable of being cleaved by an intracellular protease after administration to a subject, such as cathepsin inside tumor cells, thereby freeing the drug from the conjugate. Another example of acid-sensitive cross-linker is 6-maleimidodocaproyl hydrazone which can be coupled to thiol-bearing carriers. The hydrazone linker is rapidly cleaved at pH values < 5 allowing a release of the payload in the acidic pH of endosomes and lysosomes following internalization of the conjugate (Trail P.A. et al. Effect of linker variation on the stability, potency, and efficacy of carcinoma-reactive BR64-doxorubicin immunoconjugates. (1997) Cancer Res. 57(1), 100-105; Kratz F. et al. Acute and repeat-dose toxicity studies of the (6-maleimidocaproyl)hydrazone derivative of doxorubicin (DOXO-EMCH), an albumin-binding prodrug of the anticancer agent doxorubicin. (2007) Hum. Exp. Toxicol. 26(1), 19-35). Clinically approved mAb–drug conjugate, gemtuzumab ozogamicin (Mylotarg™) is a drug–antibody conjugate containing a humanized mAb P67.6 against CD33, linked chemically to the cytotoxic antibiotic agent calicheamicin. The linker between the antibody and the drug incorporates two labile bonds; a hydrazone and a sterically hindered disulfide. It has been shown that the acid-sensitive hydrazone bond is the actual cleavage site (Jaracz S., Chen J., Kuznetsova L.V., Ojima I. Recent advances in tumor-targeting anticancer drug conjugates. (2005) Bioorg. Med. Chem. 13(17), 5043-5054). Another example is doxorubicin–HPMA (N-(2-hydroxypropyl)methacrylamide) in which doxorubicin is linked

through its amino sugar to the HPMA copolymer via a tetrapeptide spacer GlyPheLeuGly that is cleaved by lysosomal proteases, such as cathepsin B (Vasey P.A. et al. Phase I clinical and pharmacokinetic study of PK1 [N-(2-hydroxypropyl)methacrylamide copolymer doxorubicin]: first member of a new class of chemotherapeutic agents-drug-polymer conjugates. (1999) Clin. Cancer Res. 5(1), 83-94). Other examples of carrier-drug conjugates with peptide linkers that reached clinical stage of development are macromolecular platinum complexes. Two HPMA-based drug candidates consisted of a HPMA copolymer backbone to which the complexing aminomalonate platinum complexes were bound through cathepsin B-cleavable peptide spacer GlyPheLeuGly or tripeptide spacer GlyGlyGly (Rademaker-Lakhai J.M. et al. A Phase I and pharmacological study of the platinum polymer AP5280 given as an intravenous infusion once every 3 weeks in patients with solid tumors. (2004) Clin. Cancer Res. 10(10), 3386-3395; Sood P. et al. Synthesis and characterization of AP5346, a novel polymer-linked diaminocyclohexyl platinum chemotherapeutic agent. (2006) Bioconjugate Chem. 17(5), 1270-1279). The invention contemplates the incorporation of any one of the foregoing labile spacers of this paragraph into the conjugate embodiments between the payload and the XTEN to facilitate the release of the payload. In one embodiment, the XTEN-folate incorporates the labile spacer valine-citrulline-*p*-aminobenzyloxycarbonyl between the auristatin payload and the cross-linker.

IV). THE DNA SEQUENCES OF THE INVENTION

[00230] The present invention provides isolated polynucleic acids encoding XTEN and binding fusion protein chimeric polypeptides and sequences complementary to polynucleic acid molecules encoding XTEN and binding fusion protein chimeric polypeptides, including homologous variants. In another aspect, the invention encompasses methods to produce polynucleic acids encoding XTEN and binding fusion protein chimeric polypeptides and sequences complementary to polynucleic acid molecules encoding binding fusion protein chimeric polypeptides, including homologous variants. In general, and as illustrated in FIGS. 7-9, the methods of producing a polynucleotide sequence coding for an XTEN or a binding fusion protein and expressing the resulting gene product include assembling nucleotides encoding targeting moieties and XTEN (and any linker sequences, if any), linking the components in frame, incorporating the encoding gene into an appropriate expression vector, transforming an appropriate host cell with the expression vector, and causing the fusion protein to be expressed in the transformed host cell, thereby producing the biologically-active binding fusion protein. Standard recombinant techniques in molecular biology can be used to make the polynucleotides and expression vectors of the present invention.

[00231] In accordance with the invention,

[00232] separations are also described in Baron, *et al.*, Crit. Rev. Biotechnol. 10:179-90 (1990) and Below, *et al.*, J. Chromatogr. A. 679:67-83 (1994).

V). PHARMACEUTICAL COMPOSITIONS

[00233] The present invention provides pharmaceutical compositions comprising binding fusion proteins, or BFP-D conjugates. In one embodiment, the pharmaceutical composition comprises the binding fusion protein and at least one pharmaceutically acceptable carrier. In another embodiment, the pharmaceutical composition comprises the BFP-D and at least one pharmaceutically acceptable carrier. The pharmaceutical compositions of the present invention can be formulated according to known methods to prepare pharmaceutically useful compositions, whereby the polypeptide is combined in admixture with a pharmaceutically acceptable carrier vehicle, such as aqueous solutions or buffers, pharmaceutically acceptable suspensions and emulsions. Examples of non-aqueous solvents include propyl ethylene glycol, polyethylene glycol and vegetable oils. Therapeutic formulations are prepared for storage by mixing the active ingredient having the desired degree of purity with optional physiologically acceptable carriers, excipients or stabilizers, as described in Remington's Pharmaceutical Sciences 16th edition, Osol, A. Ed. (1980), in the form of lyophilized formulations or aqueous solutions. In addition, the pharmaceutical compositions can also contain other pharmaceutically active compounds or a plurality of compositions of the invention.

[00234] The pharmaceutical compositions may be administered for therapy by any suitable route including intravitreously, topically (including transdermal, aerosol, buccal and sublingual), and parenterally (including subcutaneous, subcutaneous by infusion pump, intramuscular, intravenous and intradermal). It will also be appreciated that the preferred route will vary with the condition and age of the recipient, and the disease being treated.

[00235] The compositions of the invention may be formulated using a variety of excipients. Suitable excipients include microcrystalline cellulose (e.g. Avicel PH102, Avicel PH101), polymethacrylate, poly(ethyl acrylate, methyl methacrylate, trimethylammonioethyl methacrylate chloride) (such as Eudragit RS-30D), hydroxypropyl methylcellulose (Methocel K100M, Premium CR Methocel K100M, Methocel E5, Opadry®), magnesium stearate, talc, triethyl citrate, aqueous ethylcellulose dispersion (Surelease®), and protamine sulfate. The slow release agent may also comprise a carrier, which can comprise, for example, solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents. Pharmaceutically acceptable salts can also be used in these slow release agents, for example, mineral salts such as hydrochlorides, hydrobromides, phosphates, or sulfates, as well as the salts of organic acids such as acetates, proprionates, malonates, or benzoates. The composition may also contain liquids, such as water, saline, glycerol, and ethanol, as well as substances such as

wetting agents, emulsifying agents, or pH buffering agents. Liposomes may also be used as a carrier.

[00236] For liquid formulations, a desired property is that the formulation be supplied in a form that can pass through a 25, 28, 30, 31, 32 gauge needle for intravenous, intramuscular, intraarticular, or subcutaneous administration. Syringe pumps may also be used to delivery the pharmaceutical compositions of the invention. Such devices are described in U.S. Pat. Nos. 4,976,696; 4,933,185; 5,017,378; 6,309,370; 6,254,573; 4,435,173; 4,398,908; 6,572,585; 5,298,022; 5,176,502; 5,492,534; 5,318,540; and 4,988,337, the contents of which are incorporated herein by reference. One skilled in the art, considering both the disclosure of this invention and the disclosures of these other patents could produce a syringe pump for the extended release of the compositions of the present invention.

VI). METHODS OF TREATMENT

[00237] In another aspect, the invention relates, in part, to methods of treatment utilizing the XTEN-folate conjugate composition described herein. The method of the aspect of the invention can be used to kill cancer cells in vitro or in vivo. In one embodiment, the invention provides a method of treating a cancer cell in vitro, comprising administering to a culture of a cancer cell a composition comprising an effective amount of an XTEN-folate composition. In another embodiment, the invention provides a method of treating a cancer in a subject, comprising administering to the subject a pharmaceutical composition comprising an effective amount of an XTEN-folate composition. In one embodiment of the method, the pharmaceutical composition comprises the composition having the structure set forth in FIG. 49. In another embodiment of the method, the cancer is selected from the group consisting of non-small cell lung cancer or mesothelioma. platinum-resistant ovarian cancer, endometrial cancer, adenocarcinoma of the lung, and refractory advanced tumors. . In another embodiment of the method, the administration results in at least a 10%, or 20%, or 30%, or 40%, or 50%, or 60%, or 70%, or 80%, or 90% greater improvement of at least one, two, or three parameters associated with the cancer compared to an untreated subject wherein the parameters are selected from the group consisting of response rate as defined by the Response Evaluation Criteria in Solid Tumors (RECIST), time-to-progression of the cancer (relapse), discovery of local recurrence, discovery of regional metastasis, discovery of distant metastasis, onset of symptoms, hospitalization, increase in pain medication requirement, requirement of salvage chemotherapy, requirement of salvage surgery, requirement of salvage radiotherapy, time-to-treatment failure, and increased time of survival.

VII). PHARMACEUTICAL KITS

[00238] In another aspect, the invention provides a kit to facilitate the use of the composition embodiments disclosed herein. In one embodiment, the kit comprises, in at least a first container: an amount of a conjugate composition drug sufficient to administer in treatment of a subject with

a disease; an amount of a pharmaceutically acceptable carrier; together in a formulation ready for injection or for reconstitution with sterile water, buffer, or dextrose; together with a label identifying the drug and storage and handling conditions, and/or a sheet of the approved indications for the drug and instructions for the reconstitution and/or administration of the drug for the use for the treatment of a approved indication, appropriate dosage and safety information, and information identifying the lot and expiration of the drug.

[00239] In another embodiment, the kit comprises, in at least a first container: a first container: an amount of a conjugate composition drug sufficient to administer in treatment of a subject with a disease; an amount of a pharmaceutically acceptable carrier; a second container that can carry a suitable diluent for the subject composition, which will provide the user with the appropriate concentration of the pharmaceutical composition to be delivered to the subject; together with a label identifying the drug and storage and handling conditions, and/or a sheet of the approved indications for the drug and instructions for the reconstitution and/or administration of the drug for the use for the treatment of a approved indication, appropriate dosage and safety information, and information identifying the lot and expiration of the drug.

EXAMPLES

[00240] **Example 1: Construction of CBD-XTEN-Cys, a cysteine-engineered XTEN**

[00241] A cysteine island (CysIsland) encoding the amino acid sequence GGSPAGSCTSP containing one cysteine was introduced by annealed oligos in the CBD-stuffer-GFP vector to obtain CBD-CysIsland-GFP, where CysIsland is flanked by the restriction sites BsaI and BbsI. The CBD-stuffer-GFP vector is a pET30 derivative from Novagen with TEV protease recognition site between CBD and the stuffer. Constructs were previously generated by replacing the stuffer in CBD-stuffer-GFP vector with genes encoding XTEN_AE288 and XTEN_AE576. The plasmid of CBD-XTEN_AE288-GFP was digested with BsaI/NcoI to generate the small fragment as the insert. The plasmid of CBD-CysIsland-GFP was digested with BbsI/NcoI to generate the large fragment as the vector. The insert and vector fragments were ligated and the ligation mixture was electroporated into BL21-Gold (DE3) cells to obtain transformants of CBD-CysIsland-XTEN_AE288-GFP. Similarly, the plasmid of CBD-CysIsland-XTEN_AE288-GFP was digested with BsaI/NcoI to generate the small fragment as the insert. The plasmid of CBD-XTEN_AE576-GFP was digested with BbsI/NcoI to generate the large fragment as the vector. The insert and vector fragments were ligated and the ligation mixture was electroporated into BL21-Gold (DE3) cells to obtain transformants of CBD-XTEN_AE576-CysIsland-XTEN_AE288-GFP. Finally, the plasmid of CBD-XTEN_AE576-CysIsland-XTEN_AE288-GFP was digested with BbsI/HindIII to remove GFP and ligate with annealed oligos for the stop codon, and the ligation mixture was electroporated into BL21-Gold (DE3) cells to obtain transformants of CBD-XTEN_AE576-CysIsland-XTEN_AE288, which has the DNA and encoded amino acid sequences that follow in

Table 10. Additional constructs can be created with cysteines inserted at different locations within the XTEN sequence by the selection of restriction sites appropriate for the given location, including multiple insertions. The method could also be utilized to create lysine-engineered XTEN by substitution of codons encoding lysine for those encoding cysteine in the oligonucleotides.

Table 10: DNA and amino acid sequence of Cys-engineered XTEN

Clone Name	DNA Sequence	Amino Acid Sequence
CBD-TEV-AE576-CysIsland-AE288	ATGGCAAATACACCGGTATCAGGCAATTTGAAGGTTGAATTCT	MANTPVSG
	ACAACAGCAATCCTTCAGATACTACTAACTCAATCAATCCTCA	NLKVEFYNS
	GTTCAAGGTTACTAATACCGGAAGCAGTGCAATTGATTTGTCC	NPSDTTNSIN
	AAACTCACATTGAGATATTATTATACAGTAGACGGACAGAAA	PQFKVTNTG
	GATCAGACCTTCTGGGCTGACCATGCTGCAATAATCGGCAGTA	SSAIDLSKLT
	ACGGCAGCTACAACGGAATTACTTCAAATGTAAAAGGAACAT	LRYYYTVD
	TTGTAAAAATGAGTTCCTCAACAAATAACGCAGACACCTACCT	GQKDQTFW
	TGAAATCAGCTTTACAGGCGGAACTCTTGAACCGGGTGCACAT	ADHAAIIGS
	GTTCAGATACAAGGTAGATTTGCAAAGAATGACTGGAGTAAC	NGSYNGITS
	TATACACAGTCAAATGACTACTCATTCAAGTCTGCTTCACAGT	NVKGTFVK
	TTGTTGAATGGGATCAGGTAACAGCATACTTGAACGGTGTTCT	MSSSTNNAD
	TGTATGGGGTAAAGAACCCGGTGGCAGTGTAGTAGGTTCAGG	TYLEISFTGG
	TTCAGGATCCGAAAATCTGTATTTTCAGGGTGGGTCTCCAGGT	TLEPGAHVQ
	AGCCCGGCTGGCTCTCCTACCTCTACTGAGGAAGGTACTTCTG	IQGRFAKND
	AAAGCGCTACTCCTGAGTCTGGTCCAGGTACCTCTACTGAACC	WSNYTQSN
	GTCCGAAGGTAGCGCTCCAGGTAGCCCAGCAGGCTCTCCGACT	DYSFKSASQ
	TCCACTGAGGAAGGTACTTCTACTGAACCTTCCGAAGGCAGCG	FVEWDQVT
	CACCAGGTACCTCTACTGAACCTTCTGAGGGCAGCGCTCCAGG	AYLNGVLV
	TACTTCTGAAAGCGCTACCCCGGAATCTGGCCCAGGTAGCGAA	WGKEPGGS
	CCGGCTACTTCTGGTTCTGAAACCCCAGGTAGCGAACCGGCTA	VVGSGSGSE
	CCTCCGGTTCTGAAACTCCAGGTAGCCCGGCAGGCTCTCCGAC	NLYFQGGSP
	CTCTACTGAGGAAGGTACTTCTGAAAGCGCAACCCCGGAGTCC	GSPAGSPTS
	GGCCCAGGTACCTCTACCGAACCGTCTGAGGGCAGCGCACCA	TEEGTSESA
	GGTACTTCTACCGAACCGTCCGAGGGTAGCGCACCAGGTAGCC	TPESGPGTST
	CAGCAGGTTCTCCTACCTCCACCGAGGAAGGTACTTCTACCGA	EPSEGSAPG
	ACCGTCCGAGGGTAGCGCACCAGGTACCTCTACTGAACCTTCT	SPAGSPTSTE
	GAGGGCAGCGCTCCAGGTACTTCTGAAAGCGCTACCCCGGAG	EGTSTEPSE
	TCCGGTCCAGGTACTTCTACTGAACCGTCCGAAGGTAGCGCAC	GSAPGTSTE
	CAGGTACTTCTGAAAGCGCAACCCCTGAATCCGGTCCAGGTAG	PSEGSAPGT
	CGAACCGGCTACTTCTGGCTCTGAGACTCCAGGTACTTCTACC	SESATPESGP
	GAACCGTCCGAAGGTAGCGCACCAGGTACTTCTACTGAACCGT	GSEPATSGS
	CTGAAGGTAGCGCACCAGGTACTTCTGAAAGCGCAACCCCGG	ETPGSEPATS
	AATCCGGCCCAGGTACCTCTGAAAGCGCAACCCCGGAGTCCG	GSETPGSPA
	GCCCAGGTAGCCCTGCTGGCTCTCCAACCTCCACCGAAGAAGG	GSPTSTEEG
	TACCTCTGAAAGCGCAACCCCTGAATCCGGCCCAGGTAGCGA	TSESATPESG
	ACCGGCAACCTCCGGTTCTGAAACCCCAGGTACCTCTGAAAGC	PGTSTEPSEG
	GCTACTCCGGAGTCTGGCCCAGGTACCTCTACTGAACCGTCTG	SAPGTSTEPS
	AGGGTAGCGCTCCAGGTACTTCTACTGAACCGTCCGAAGGTAG	EGSAPGSPA
	CGCACCAGGTACTTCTACCGAACCGTCCGAAGGCAGCGCTCCA	GSPTSTEEG
	GGTACCTCTACTGAACCTTCCGAGGGCAGCGCTCCAGGTACCT	TSTEPSEGSA
	CTACCGAACCTTCTGAAGGTAGCGCACCAGGTACTTCTACCGA	PGTSTEPSEG
	ACCGTCCGAGGGTAGCGCACCAGGTAGCCCAGCAGGTTCTCCT	SAPGTSESA
	ACCTCCACCGAGGAAGGTACTTCTACCGAACCGTCCGAGGGTA	TPESGPGTST
	GCGCACCAGGTACCTCTGAAAGCGCAACTCCTGAGTCTGGCCC	EPSEGSAPG
	AGGTAGCGAACCTGCTACCTCCGGCTCTGAGACTCCAGGTACC	TSESATPESG
	TCTGAAAGCGCAACCCCGGAATCTGGTCCAGGTAGCGAACCT	PGSEPATSG
	GCAACCTCTGGCTCTGAAACCCCAGGTACCTCTGAAAGCGCTA	SETPGTSTEP

Clone Name	DNA Sequence	Amino Acid Sequence
	CTCCTGAATCTGGCCCAGGTACTTCTACTGAACCGTCCGAGGG CAGCGCACCAGGTACTTCTGAAAGCGCTACTCCTGAGTCCGGC CCAGGTAGCCCGGCTGGCTCTCCGACTTCCACCGAGGAAGGTA GCCCGGCTGGCTCTCCAACTTCTACTGAAGAAGGTAGCCCGGC AGGCTCTCCGACCTCTACTGAGGAAGGTACTTCTGAAAGCGCA ACCCCGGAGTCCGGCCCAGGTACCTCTACCGAACCGTCTGAGG GCAGCGCACCAGGTGGTAGCCCGGCTGGCTCTTGTACCTCTCC AGGTACCTCTGAAAGCGCAACTCCTGAGTCTGGCCCAGGTAGC GAACCTGCTACCTCCGGCTCTGAGACTCCAGGTACCTCTGAAA GCGCAACCCCCGGAATCTGGTCCAGGTAGCGAACCTGCAACCTC TGGCTCTGAAACCCCAGGTACCTCTGAAAGCGCTACTCCTGAA TCTGGCCCAGGTACTTCTACTGAACCGTCCGAGGGCAGCGCAC CAGGTAGCCCTGCTGGCTCTCCAACCTCCACCGAAGAAGGTAC CTCTGAAAGCGCAACCCCTGAATCCGGCCCAGGTAGCGAACC GGCAACCTCCGGTTCTGAAACCCCAGGTACTTCTGAAAGCGCT ACTCCTGAGTCCGGCCCAGGTAGCCCGGCTGGCTCTCCGACTT CCACCGAGGAAGGTAGCCCGGCTGGCTCTCCAACTTCTACTGA AGAAGGTACTTCTACCGAACCTTCCGAGGGCAGCGCACCAGG TACTTCTGAAAGCGCTACCCCTGAGTCCGGCCCAGGTACTTCT GAAAGCGCTACTCCTGAATCCGGTCCAGGTACTTCTGAAAGCG CTACCCCGGAATCTGGCCCAGGTAGCGAACCGGCTACTTCTGG TTCTGAAACCCCAGGTAGCGAACCGGCTACCTCCGGTTCTGAA ACTCCAGGTAGCCCAGCAGGCTCTCCGACTTCCACTGAGGAAG GTACTTCTACTGAACCTTCCGAAGGCAGCGCACCAGGTACCTC TACTGAACCTTCTGAGGGCAGCGCTCCAGGTAGCGAACCTGCA ACCTCTGGCTCTGAAACCCCAGGTACCTCTGAAAGCGCTACTC CTGAATCTGGCCCAGGTACTTCTACTGAACCGTCCGAGGGCAG CGCACCAGGTTAA	SEGSAPGTS TEPSEGSAP GTSESATPES GPGTSESAT PESGPGSPA GSPTSTEEG TSESATPESG PGSEPATSG SETPGTSESA TPESGPGTST EPSEGSAPG TSTEPSEGSA PGTSTEPSEG SAPGTSTEPS EGSAPGTST EPSEGSAPG TSTEPSEGSA PGSPAGSPTS TEEGTSTEPS EGSAPGTSE SATPESGPG SEPATSGSET PGTSESATPE SGPGSEPAT SGSETPGTSE SATPESGPG TSTEPSEGSA PGTSESATPE SGPGSPAGS PTSTEEGSPA GSPTSTEEGS PAGSPTSTEE GTSESATPES GPGTSTEPSE GSAPGGSPA GSCTSPGTS ESATPESGP GSEPATSGS ETPGTSESA TPESGPGSEP ATSGSETPG TSESATPESG PGTSTEPSEG SAPGSPAGS PTSTEEGTSE SATPESGPG SEPATSGSET PGTSESATPE SGPGSPAGS PTSTEEGSPA GSPTSTEEG TSTEPSEGSA PGTSESATPE SGPGTSESA TPESGPGTSE SATPESGPG SEPATSGSET PGSEPATSG

Clone Name	DNA Sequence	Amino Acid Sequence
		SETPGSPAG SPTSTEEGTS TEPSEGSAP GTSTEPSEGS APGSEPATS GSETPGTSES ATPESGPGT STEPSEGSAP G

[00242] Example 2: Construction of Cys3-XTEN

[00243] A pair of primers was designed to introduce the restriction site BamHI and cysteine island 1 of sequence GRATAEAAGCGTAEAA at the N-terminus of XTEN_AE432-1, and a partial cysteine island 2 of sequence TAEAAG and restriction site BbsI at C-terminus of XTEN_AE432-1. A second pair of primers was designed to introduce the restriction site BsaI and a partial cysteine island 2 of sequence GCGTAEAA at the N-terminus of XTEN_AE432-2, and cysteine island 4 of sequence TAEAAGCGTAEAAR with an 8xHis-tag (H8) and restriction site HindIII at the C-terminus of XTEN_AE432-2. The XTEN_AE432-1 contains the 1-432 amino acid sequence and XTEN_AE432-2 contains the 433-864 amino acid sequence encoded by the XTEN_AE864 gene. These two pairs of primers were used to amplify the XTEN_AE432-1 and XTEN_AE432-2 gene, respectively, by polymerase chain reaction (PCR). The PCR products of correct sizes were gel-purified and digested with the restriction enzymes BamHI/BbsI and BsaI/HindIII, respectively, as the inserts for ligation. A destination vector, a derivative of pET30 (Novagen) includes a CBD (cellulose binding domain)-stuffer with the flanking restriction sites BamHI and HindIII. The destination vector was digested with the restriction enzymes BamHI/HindIII to remove the stuffer and prepared as the vector. The vector was ligated with the BamHI/BbsI digested PCR product of XTEN_AE432-1 and BsaI/HindIII digested PCR of XTEN_AE432-2 above. The ligation mixture was transformed into *E. coli* TOP10 competent cells. Transformants were screened by DNA miniprep and the desired constructs were confirmed by DNA sequencing. Thus, the final plasmid yields the CBD-cysteine island 1-XTEN_AE432-cysteine island 2-XTEN_AE432-cysteine island 3-H8 gene under the control of a T7 promoter. The DNA sequences and protein sequences are provided in Table 11.

Table 11: Cys3-XTEN DNA and amino acid sequences

Clone Name	DNA Sequence	Amino Acid Sequence
CBD-Cys1-AE432-Cys2-AE432-Cys3-H8 (AC673)	ATGGCTAATACCCCAGTGAGCGGCAACCTGAAAGTGGAAT TCTACAATAGCAACCCGAGCGACACCACCAACAGCATTAA TCCGCAGTTCAAAGTGACCAACACGGGTAGCTCCGCGATC GATCTGTCGAAGCTGACGCTGCGTTACTATTACACGGTTG ACGGTCAGAAAGATCAGACGTTCTGGGCTGACCATGCGGC CATTATTGGCAGCAACGGTTCCTACAACGGTATCACGAGC AATGTCAAAGGCACTTTTGTTAAGATGAGCTCTTCGACCA	MANTPVSGNL KVEFY NSNPSDTTNSIN PQF KVTNTGSSAID LSKL TLRYYYTVDG

Clone Name	DNA Sequence	Amino Acid Sequence
	ACAATGCCGATACCTATCTGGAGATTAGCTTCACCGGTGG	QKDQT
	TACTCTGGAGCCGGGTGCACACGTTCAAATCCAAGGTCGC	FWADHAAIIGS
	TTCGCAAAGAATGACTGGAGCAACTATACCCAGTCCAATG	NGSY
	ACTACAGCTTCAAAAGCGCTAGCCAATTTGTTGAATGGGA	NGITSNVKGTF
	TCAGGTCACCGCATACCTGAACGGCGTGCTGGTCTGGGGC	VKMS
	AAGGAACCGGGTGGTAGCGTTGTCGGTTCTGGCAGCGGAT	SSTNNADTYLE
	CCggtcgtGCGACGGCAGAAGCCGCTGGCtgcGGTACTGCTGA	ISFT
	AGCGGCAGGTAGCCCAGCTGGTAGCCCAACCTCTACCGAA	GGTLEPGAHVQ
	GAAGGTACCTCTGAATCCGCTACTCCAGAATCCGGTCCTG	IQGR
	GTACTAGCACTGAGCCAAGCGAAGGTTCTGCTCCAGGCTC	FAKNDWSNYT
	CCCGGCAGGTAGCCCTACCTCTACCGAAGAGGGCACTAGC	QSNDY
	ACCGAACCATCTGAGGGTTCCGCTCCTGGCACCTCCACTG	SFKSASQFVEW
	AACCGTCCGAAGGCAGTGCTCCGGGTACTTCCGAAAGCGC	DQVT
	AACTCCGGAATCCGGCCCTGGTTCTGAGCCTGCTACTTCCG	AYLNGVLVWG
	GCTCTGAAACTCCAGGTAGCGAGCCAGCGACTTCTGGTTC	KEPGG
	TGAAACTCCAGGTTCACCGGCGGGTAGCCCGACGAGCACG	SVVGSGSGSGR
	GAGGAAGGTACCTCTGAGTCGGCCACTCCTGAGTCCGGTC	ATAE
	CGGGCACGAGCACCGAGCCGAGCGAGGGTTCAGCCCCGG	AAGCGTAEAA
	GTACCAGCACGGAGCCGTCCGAGGGTAGCGCACCGGGTTC	GSPAG
	TCCGGCGGGCTCCCCTACGTCTACGGAAGAGGGTACGTCC	SPTSTEEGTSES
	ACTGAACCTAGCGAGGGCAGCGCGCCAGGCACCAGCACT	ATP
	GAACCGAGCGAAGGCAGCGCACCTGGCACTAGCGAGTCT	ESGPGTSTEPSE
	GCGACTCCGGAGAGCGGTCCGGGTACGAGCACGGAACCA	GSA
	AGCGAAGGCAGCGCCCCAGGTACCTCTGAATCTGCTACCC	PGSPAGSPTSTE
	CAGAATCTGGCCCGGGTTCCGAGCCAGCTACCTCTGGTTC	EGT
	TGAAACCCCAGGTACTTCCACTGAACCAAGCGAAGGTAGC	STEPSEGSAPGT
	GCTCCTGGCACTTCTACTGAACCATCCGAAGGTTCCGCTCC	STE
	TGGTACGTCTGAAAGCGCTACCCCTGAAAGCGGCCCAGGC	PSEGSAPGTSES
	ACCTCTGAAAGCGCTACTCCTGAGAGCGGTCCAGGCTCTC	ATP
	CAGCAGGTTCTCCAACCTCCACTGAAGAAGGCACCTCTGA	ESGPGSEPATS
	GTCTGCTACCCCTGAATCTGGTCCTGGCTCCGAACCTGCTA	GSET
	CCTCTGGTTCCGAAACTCCAGGTACCTCGGAATCTGCGAC	PGSEPATSGSET
	TCCGGAATCTGGCCCGGGCACGAGCACGGAGCCGTCTGAG	PGS
	GGTAGCGCACCAGGTACCAGCACTGAGCCTTCTGAGGGCT	PAGSPTSTEEGT
	CTGCACCGGGTACCTCCACGGAACCTTCGGAAGGTTCTGC	SES
	GCCGGGTACCTCCACTGAGCCATCCGAGGGTTCAGCACCA	ATPESGPGTSTE
	GGTACTAGCACGGAACCGTCCGAGGGCTCTGCACCAGGTA	PSE
	CGAGCACCGAACCGTCGGAGGGTAGCGCTCCAGGTAGCCC	GSAPGTSTEPSE
	AGCGGGCTCTCCGACAAGCACCGAAGAAGGCACCAGCAC	GSA
	CGAGCCGTCCGAAGGTTCCGCACCAACCGCTGAAGCCGCA	PGSPAGSPTSTE
	GGTtgtGGCACTGCGGAAGCTGCAGGTACAAGCGAGAGCGC	EGT
	GACTCCTGAATCTGGTCCGGGTAGCGAGCCTGCAACCAGC	STEPSEGSAPGT
	GGTTCTGAGACGCCGGGCACTTCCGAATCTGCGACCCCGG	STE
	AGTCCGGTCCAGGTTCAGAGCCGGCGACGAGCGGTTCGGA	PSEGSAPGTSES
	AACGCCGGGTACGTCTGAATCAGCCACGCCGGAGTCTGGT	ATP
	CCGGGTACCTCGACCGAACCAAGCGAAGGTTCGGCACCGG	ESGPGTSTEPSE
	GTACTAGCGAGAGCGCAACCCCTGAAAGCGGTCCGGGCA	GSA
	GCCCGGCAGGTTCTCCAACCAGCACCGAAGAAGGTTCCCC	PGTSESATPESG
	TGCTGGTAGCCCGACCTCTACGGAGGAAGGTAGCCCTGCA	PGS
	GGTTCCCCAACTTCTACTGAGGAAGGTACTTCTGAGTCCG	EPATSGSETPGT
	CTACCCCAGAAAGCGGTCCTGGTACCTCCACTGAACCGTC	STE
	TGAAGGCTCTGCACCAGGCACTTCTGAGTCTGCTACTCCA	PSEGSAPGTSTE
	GAAAGCGGCCCAGGTTCTGAACCAGCAACTTCTGGCTCTG	PSE
	AGACTCCAGGCACTTCTGAGTCCGCAACGCCTGAATCCGG	GSAPGTSESAT
	TCCTGGTTCTGAACCAGCTACTTCCGGCAGCGAAACCCCA	PESG
	GGTACCTCTGAGTCTGCGACTCCAGAGTCTGGTCCTGGTA	PGTSESATPESG
	CTTCCACTGAGCCTAGCGAGGGTTCCGCACCAGGTTCTCC	PGS
	GGCTGGTAGCCCGACCAGCACGGAGGAGGGTACGTCTGA	PAGSPTSTEEGT

Clone Name	DNA Sequence	Amino Acid Sequence
	ATCTGCAACGCCGGAATCGGGCCCAGGTTCGGAGCCTGCA ACGTCTGGCAGCGAAACCCCGGGTACCTCCGAATCTGCTA CACCGGAAAGCGGTCCTGGCAGCCCTGCTGGTTCTCCAAC CTCTACCGAGGAGGGTTCACCGGCAGGTAGCCCGACTAGC ACTGAAGAAGGTACTAGCACGGAGCCGAGCGAGGGTAGT GCTCCGGGTACGAGCGAGAGCGCAACGCCAGAGAGCGGT CCAGGCACCAGCGAATCGGCCACCCCTGAGAGCGGCCCA GGTACTTCTGAGAGCGCCACTCCTGAATCCGGCCCTGGTA GCGAGCCGGCAACCTCCGGCTCAGAAACTCCTGGTTCGGA ACCAGCGACCAGCGGTTCTGAAACTCCGGGTAGCCCGGCA GGCAGCCCAACGAGCACCGAAGAGGGTACCAGCACGGAA CCGAGCGAGGGTTCTGCCCCGGGTACTTCCACCGAACCAT CGGAGGGCTCTGCACCTGGTAGCGAACCTGCGACGTCTGG TTCTGAAACGCCGGGTACCAGCGAAAGCGCTACCCCAGAA TCCGGTCCGGGCACTAGCACCGAGCCATCGGAGGGCTCCG CACCAACTGCAGAAGCGGCTGGTtgtGGCACCGCCGAAGCA GCTcgtCATCACCATCACCACCATCATCACTAA	SES ATPESGPGSEP ATSG SETPGTSESATP ESG PGTSTEPSEGSA PGT STEPSEGSAPGT STE PSEGSAPGTSTE PSE GSAPGTSTEPSE GSA PGTSTEPSEGSA PGS PAGSPTSTEEGT STE PSEGSAPTAEA AGCG TAEAAGTSESA TPES GPGSEPATSGS ETPG TSESATPESGPG SEP ATSGSETPGTSE SAT PESGPGTSTEPS EGS APGTSESATPES GPG SPAGSPTSTEEG SPA GSPTSTEEGSPA GSP TSTEEGTSESAT PES GPGTSTEPSEGS APG TSESATPESGPG SEP ATSGSETPGTSE SAT PESGPGSEPATS GSE TPGTSESATPES GPG TSTEPSEGSAPG SPA GSPTSTEEGTSE SAT PESGPGSEPATS GSE TPGTSESATPES GPG SPAGSPTSTEEG SPA GSPTSTEEGTST

Clone Name	DNA Sequence	Amino Acid Sequence
		EPS EGSAPGTSESA TPES GPGTSESATPES GPG TSESATPESGPG SEP ATSGSETPGSEP ATS GSETPGSPAGSP TST EEGTSTEPSEGS APG TSTEPSEGSAPG SEP ATSGSETPGTSE SAT PESGPGTSTEPS EGS APTAEAAGCGT AEAA RHHHHHHHHH

[00244] Example 3: Construction of Lys2-XTEN

[00245] A pair of primers was designed to introduce the restriction site BamHI and the amino acid sequence GRGSP at the N-terminus of XTEN_AE432-1, and the amino acid sequence TAEAAG and restriction site BbsI at the C-terminus of XTEN_AE432-1. A second pair of primers was designed to introduce the restriction site BsaI and an amino acid sequence with an incorporated lysine of GKPGTAEAA at the N-terminus of XTEN_AE432-2, and an amino acid sequence with an incorporated lysine of GKAT with an 8xHis-tag (H8) and restriction site HindIII at C-terminus of XTEN_AE432-2. XTEN_AE432-1 contains the 1-432 amino acid sequence and XTEN_AE432-2 contains the 433-864 amino acid sequence encoded by the XTEN_AE864 gene. These two pairs of primers were used to amplify the XTEN_AE432-1 and XTEN_AE432-2 gene, respectively, by polymerase chain reaction (PCR). The PCR products of right sizes were gel-purified and digested with the restriction enzymes BamHI/BbsI and BsaI/HindIII, respectively, as the inserts for ligation. A destination vector, derivative of pET30 (Novagen), of CBD (cellulose binding domain)-stuffer with the flanking restriction sites BamHI and HindIII was digested with the restriction enzymes BamHI/HindIII to remove the stuffer and prepared as the vector. Ligate the vector with the BamHI/BbsI digested PCR product of XTEN_AE432-1 and BsaI/HindIII digested PCR of XTEN_AE432-2 above. The ligation mixture was transformed into *E. coli* TOP10 competent cells. Transformants were screened by DNA miniprep and the desired constructs were confirmed by DNA sequencing. Thus, the final plasmid yields the CBD-GRGSP-XTEN_AE432-TAEAAGKPGTAEAA-XTEN_AE432-GKAT-H8 gene under the control of a T7 promoter. The DNA sequences and protein sequences are provided in Table 12.

Table 12: Lys2-XTEN DNA and amino acid sequences

Clone Name	DNA Sequence	Amino Acid Sequence
CBD-R-AE432-K-AE432-K-H8 (AC698)	ATGGCTAATACCCCAGTGAGCGGCAACCTGAAAGTGGAAT	MANTPVSGNLK
	TCTACAATAGCAACCCGAGCGACACCACCAACAGCATTAA	VEFY
	TCCGCAGTTCAAAGTGACCAACACGGGTAGCTCCGCGATC	NSNPSDTTNSIN
	GATCTGTCGAAGCTGACGCTGCGTTACTATTACACGGTTG	PQF
	ACGGTCAGAAAGATCAGACGTTCTGGGCTGACCATGCGGC	KVTNTGSSAIDL
	CATTATTGGCAGCAACGGTTCCTACAACGGTATCACGAGC	SKL
	AATGTCAAAGGCACTTTTGTTAAGATGAGCTCTTCGACCA	TLRYYYTVDGQ
	ACAATGCCGATACCTATCTGGAGATTAGCTTCACCGGTGG	KDQT
	TACTCTGGAGCCGGGTGCACACGTTCAAATCCAAGGTCGC	FWADHAAIIGSN
	TTCGCAAAGAATGACTGGAGCAACTATACCCAGTCCAATG	GSY
	ACTACAGCTTCAAAAGCGCTAGCCAATTTGTTGAATGGGA	NGITSNVKGTFV
	TCAGGTCACCGCATACCTGAACGGCGTGCTGGTCTGGGGC	KMS
	AAGGAACCGGGTGGTAGCGTTGTCGGTTCTGGCAGCGGAT	SSTNNADTYLEI
	CCggtcgtGGGTCTCCAGGTAGCCCAGCTGGTAGCCCAACCT	SFT
	CTACCGAAGAAGGTACCTCTGAATCCGCTACTCCAGAATC	GGTLEPGAHVQI
	CGGTCCTGGTACTAGCACTGAGCCAAGCGAAGGTTCTGCT	QGR
	CCAGGCTCCCCGGCAGGTAGCCCTACCTCTACCGAAGAGG	FAKNDWSNYTQ
	GCACTAGCACCGAACCATCTGAGGGTTCCGCTCCTGGCAC	SNDY
	CTCCACTGAACCGTCCGAAGGCAGTGCTCCGGGTACTTCC	SFKSASQFVEW
	GAAAGCGCAACTCCGGAATCCGGCCCTGGTTCTGAGCCTG	DQVT
	CTACTTCCGGCTCTGAAACTCCAGGTAGCGAGCCAGCGAC	AYLNGVLVWG
	TTCTGGTTCTGAAACTCCAGGTTCACCGGCGGGTAGCCCG	KEPGG
	ACGAGCACGGAGGAAGGTACCTCTGAGTCGGCCACTCCTG	SVVGSGSGSGR
	AGTCCGGTCCGGGCACGAGCACCGAGCCGAGCGAGGGTT	GSPG
	CAGCCCCGGGTACCAGCACGGAGCCGTCCGAGGGTAGCG	SPAGSPTSTEEG
	CACCGGGTTCTCCGGCGGGCTCCCCTACGTCTACGGAAGA	TSE
	GGGTACGTCCACTGAACCTAGCGAGGGCAGCGCGCCAGG	SATPESGPGTST
	CACCAGCACTGAACCGAGCGAAGGCAGCGCACCTGGCAC	EPS
	TAGCGAGTCTGCGACTCCGGAGAGCGGTCCGGGTACGAGC	EGSAPGSPAGSP
	ACGGAACCAAGCGAAGGCAGCGCCCCAGGTACCTCTGAA	TST
	TCTGCTACCCCAGAATCTGGCCCGGGTTCCGAGCCAGCTA	EEGTSTEPSEGS
	CCTCTGGTTCTGAAACCCCAGGTACTTCCACTGAACCAAG	APG
	CGAAGGTAGCGCTCCTGGCACTTCTACTGAACCATCCGAA	TSTEPSEGSAPG
	GGTTCCGCTCCTGGTACGTCTGAAAGCGCTACCCCTGAAA	TSE
	GCGGCCCAGGCACCTCTGAAAGCGCTACTCCTGAGAGCGG	SATPESGPGSEP
	TCCAGGCTCTCCAGCAGGTTCTCCAACCTCCACTGAAGAA	ATS
	GGCACCTCTGAGTCTGCTACCCCTGAATCTGGTCCTGGCTC	GSETPGSEPATS
	CGAACCTGCTACCTCTGGTTCCGAAACTCCAGGTACCTCG	GSE
	GAATCTGCGACTCCGGAATCTGGCCCGGGCACGAGCACGG	TPGSPAGSPTST
	AGCCGTCTGAGGGTAGCGCACCAGGTACCAGCACTGAGCC	EEG
	TTCTGAGGGCTCTGCACCGGGTACCTCCACGGAACCTTCG	TSESATPESGPG
	GAAGGTTCTGCGCCGGGTACCTCCACTGAGCCATCCGAGG	TST
	GTTCAGCACCAGGTACTAGCACGGAACCGTCCGAGGGCTC	EPSEGSAPGTST
	TGCACCAGGTACGAGCACCGAACCGTCGGAGGGTAGCGCT	EPS
	CCAGGTAGCCCAGCGGGCTCTCCGACAAGCACCGAAGAA	EGSAPGSPAGSP
	GGCACCAGCACCGAGCCGTCCGAAGGTTCCGCACCAACCG	TST
	CTGAAGCCGCAGGTaaaccgGGCACTGCGGAAGCTGCAGGTA	EEGTSTEPSEGS
	CAAGCGAGAGCGCGACTCCTGAATCTGGTCCGGGTAGCGA	APG
	GCCTGCAACCAGCGGTTCTGAGACGCCGGGCACTTCCGAA	TSTEPSEGSAPG
	TCTGCGACCCCGGAGTCCGGTCCAGGTTCAGAGCCGGCGA	TSE
	CGAGCGGTTCGGAAACGCCGGGTACGTCTGAATCAGCCAC	SATPESGPGTST
	GCCGGAGTCTGGTCCGGGTACCTCGACCGAACCAAGCGAA	EPS
	GGTTCGGCACCGGGTACTAGCGAGAGCGCAACCCCTGAAA	EGSAPGTSESAT
	GCGGTCCGGGCAGCCCGGCAGGTTCTCCAACCAGCACCGA	PES
	AGAAGGTTCCCCTGCTGGTAGCCCGACCTCTACGGAGGAA	GPGSEPATSGSE
	GGTAGCCCTGCAGGTTCCCCAACTTCTACTGAGGAAGGTA	TPG
	CTTCTGAGTCCGCTACCCCAGAAAGCGGTCCTGGTACCTC	TSTEPSEGSAPG

Clone Name	DNA Sequence	Amino Acid Sequence
	CACTGAACCGTCTGAAGGCTCTGCACCAGGCACTTCTGAG TCTGCTACTCCAGAAAGCGGCCCAGGTTCTGAACCAGCAA CTTCTGGCTCTGAGACTCCAGGCACTTCTGAGTCCGCAAC GCCTGAATCCGGTCCTGGTTCTGAACCAGCTACTTCCGGC AGCGAAACCCCAGGTACCTCTGAGTCTGCGACTCCAGAGT CTGGTCCTGGTACTTCCACTGAGCCTAGCGAGGGTTCCGC ACCAGGTTCTCCGGCTGGTAGCCCGACCAGCACGGAGGAG GGTACGTCTGAATCTGCAACGCCGGAATCGGGCCCAGGTT CGGAGCCTGCAACGTCTGGCAGCGAAACCCCGGGTACCTC CGAATCTGCTACACCGGAAAGCGGTCCTGGCAGCCCTGCT GGTTCTCCAACCTCTACCGAGGAGGGTTCACCGGCAGGTA GCCCGACTAGCACTGAAGAAGGTACTAGCACGGAGCCGA GCGAGGGTAGTGCTCCGGGTACGAGCGAGAGCGCAACGC CAGAGAGCGGTCCAGGCACCAGCGAATCGGCCACCCCTG AGAGCGGCCCAGGTACTTCTGAGAGCGCCACTCCTGAATC CGGCCCTGGTAGCGAGCCGGCAACCTCCGGCTCAGAAACT CCTGGTTCGGAACCAGCGACCAGCGGTTCTGAAACTCCGG GTAGCCCGGCAGGCAGCCCAACGAGCACCGAAGAGGGTA CCAGCACGGAACCGAGCGAGGGTTCTGCCCCGGGTACTTC CACCGAACCATCGGAGGGCTCTGCACCTGGTAGCGAACCT GCGACGTCTGGTTCTGAAACGCCGGGTACCAGCGAAAGCG CTACCCCAGAATCCGGTCCGGGCACTAGCACCGAGCCATC GGAGGGCTCCGCACCAggtAAAgcgaccCATCACCATCACCAC CATCATCACTAA	TST EPSEGSAPGTSE SAT PESGPGTSESAT PES GPGSPAGSPTST EEG TSESATPESGPG SEP ATSGSETPGTSE SAT PESGPGTSTEPSE GS APGTSTEPSEGS APG TSTEPSEGSAPG TST EPSEGSAPGTST EPS EGSAPGTSTEPS EGS APGSPAGSPTST EEG TSTEPSEGSAPT AEA AGKPGTAEAAG TSES ATPESGPGSEPA TSG SETPGTSESATPE SG PGSEPATSGSET PGT SESATPESGPGT STE PSEGSAPGTSES ATP ESGPGSPAGSPT STE EGSPAGSPTSTE EGS PAGSPTSTEEGT SES ATPESGPGTSTE PSE GSAPGTSESATP ESG PGSEPATSGSET PGT SESATPESGPGS EPA TSGSETPGTSES ATP ESGPGTSTEPSE GSA PGSPAGSPTSTE EGT SESATPESGPGS

Clone Name	DNA Sequence	Amino Acid Sequence
		EPA TSGSETPGTSES ATP ESGPGSPAGSPT STE EGSPAGSPTSTE EGT STEPSEGSAPGT SES ATPESGPGTSES ATP ESGPGTSESATP ESG PGSEPATSGSET PGS EPATSGSETPGS PAG SPTSTEEGTSTEP SE GSAPGTSTEPSE GSA PGSEPATSGSET PGT SESATPESGPGT STE PSEGSAPGKATH HHH HHHH

[00246] Example 4: Host Strain and Promoter for Expression of XTEN for Conjugation

[00247] Plasmids pCW1054 expressing CBD-TEV site-XTEN_AE864 under the control of the T7/lac promoter, pLCW0968.003 expressing CBD-TEV site-XTEN_AE864-GFP under the control of the T7/lac promoter, and pLCW0970.030 expressing CBD-TEV site-XTEN_AE864-GFP under the control of the PhoA promoter, were transformed into both *E. coli* BL21 DE3 and *E. coli* BL21 DE3 rne-131 (Lopez, P. J., et al. (1999) Mol. Microbiol. 33, 188-199). The rne-131 mutation disrupts the 3' endoribonucleolytic activity of RNase E (Kido, M., et al. (1996) Journal of Bacteriology, 178: 3917-3925). Starter cultures of all six tranformants were prepared by picking single colonies to inoculate 2 mL of LB Broth media containing the appropriate selective antibiotic. The starter cultures were grown overnight and 0.5mL was used to inoculate, in quadruplicate, 25 mL of 2xYT broth, for cells containing pCW1054 and pLCW0968.003, or 25mL of PhoA induction broth (Amunix recipe 136-1) for cells containing pLCW0970.030. The cultures were shaken for 3 hours at 37°C. The temperature was then reduced to 26°C for all of the cultures; and for cells containing pCW1054 and pLCW0968.003 protein expression was induced with IPTG at 1.0 mM final concentration. Induction of the PhoA promoter in pLCW070.030 is auto-induced upon depletion of phosphate from the culture media. The cultures were then shaken overnight at 26°C. Samples of each culture were lysed and 20 µl of each lysate were subjected to

non-reducing SDS-PAGE using NuPAGE 4-12% Bis-Tris gel from Invitrogen, according to manufacturer's specifications, with Coomassie staining. The results (FIG. 43) showed that when either CBD-TEV site-XTEN_AE864-GFP or CBD-TEV site-XTEN_AE864 is expressed under the control of the T7/lac promoter the yield of the recombinant protein is significantly higher in the E. coli BL21 DE3 rne-131 compared to the E. coli BL21 DE3 cells. Cells were also measured for GFP fluorescence using a fluorescent plate reader. The results (FIG. 44) showed that expression in E. coli BL21 DE3 cells of CBD-TEV site-XTEN_AE864-GFP under the control of the PhoA promoter is nearly three times greater than the expression of CBD-TEV site-XTEN_AE864-GFP under the control of the T7/lac promoter. In the E. coli BL21 DE3 rne-131 both the T7/lac and PhoA promoters yielded similar levels of expression. Therefore, only upon inhibition of Rnase E 3' riboendocleolytic activity does T7/lac induction yield titers similar to that obtained using PhoA induction. Likely the faster rate of T7 RNAP transcription (Studier, W., et al. (1986) Journal of Molecular Biology 189: 113–130), relative to translation results in mRNA susceptible to endonucleolytic attack, while in the case of PhoA induction where transcription is mediated by the native E. coli RNA polymerase II, the rate of transcription is closely coupled to the rate of translation protecting the mRNA from endonucleolytic attack.

[00248] Example 5: Construction of 1xAmino-XTEN288

[00249] A pair of primers AE278BsaIfor-AACG and AE278-RH8HindIIIrev were used to PCR the plasmid containing XTEN_AE864_003 in order to obtain the PCR product XTEN_AE278. Gel-purification of the band of the right size was performed, followed by digestion with BsaI/HindIII as the insert of XTEN_AE278-R-H8. A digestion of plasmid pCW1161, which encodes the gene of N-term-RP11-R-AE432_3Cys-R-H8 (construct AC763), with BsaI/HindIII was performed to remove the fragment of AE432_3Cys-R-H8 and gel-purify the large fragment as the vector. Ligation of the vector with the insert was performed and was used to transform BL21 competent cells to obtain the construct of N-term-RP11-R-AE288-R-H8. The XTEN length between the two trypsin digestion sites (R, Arginine) was calculated by XTEN_AE278 plus some flanking amino acids and equaled exactly 288 amino acids. Thus, the construct is designed to produce the precursor N-term-RP11-R-AE288-R-H8 (sequence in Table 13, below) and generate 1xAmino-XTEN288 (Seg 178 of Table 1, with an N-terminal amino group for conjugation) after removal of the N-term-RP11 tag and C-term 8xHis-tag by trypsin digestion.

Table 13: DNA and amino acid sequence for 1xAmino-XTEN288

Clone Name	DNA Sequence	Amino Acid Sequence
N-term-RP11-R-AE288-R-H8	ATGAAAAACCCAGAGCAAGCAGAAGAACAAGCTGAAGAACA GCGCGAAGAAACACGTCCGCGTCCTCGCCCACGTCCACGTCCG CGTCCACGCCCTCGTCCTCGTCCGCGCCCTCGTCCGagcgcgtctcgtt ccgctGGGTCTCCAAcgGGCCCAGGTTCTGAACCAGCAACTTCTG GCTCTGAGACTCCAGGCACTTCTGAGTCCGCAACGCCTGAATC CGGTCCTGGTTCTGAACCAGCTACTTCCGGCAGCGAAACCCCA	MKNPEQAEE QAEEQREETR PRPRPRPRPR PRPRPRPRPR PSASRSAGSP TGPGSEPATS

Clone Name	DNA Sequence	Amino Acid Sequence
	GGTACCTCTGAGTCTGCGACTCCAGAGTCTGGTCCTGGTACTT CCACTGAGCCTAGCGAGGGTTCCGCACCAGGTTCTCCGGCTGG TAGCCCGACCAGCACGGAGGAGGGTACGTCTGAATCTGCAAC GCCGGAATCGGGCCCAGGTTCGGAGCCTGCAACGTCTGGCAG CGAAACCCCGGGTACCTCCGAATCTGCTACACCGGAAAGCGG TCCTGGCAGCCCTGCTGGTTCTCCAACCTCTACCGAGGAGGGT TCACCGGCAGGTAGCCCGACTAGCACTGAAGAAGGTACTAGC ACGGAGCCGAGCGAGGGTAGTGCTCCGGGTACGAGCGAGAGC GCAACGCCAGAGAGCGGTCCAGGCACCAGCGAATCGGCCACC CCTGAGAGCGGCCCAGGTACTTCTGAGAGCGCCACTCCTGAAT CCGGCCCTGGTAGCGAGCCGGCAACCTCCGGCTCAGAAACTCC TGGTTCGGAACCAGCGACCAGCGGTTCTGAAACTCCGGGTAGC CCGGCAGGCAGCCCAACGAGCACCGAAGAGGGTACCAGCACG GAACCGAGCGAGGGTTCTGCCCCGGGTACTTCCACCGAACCAT CGGAGGGCTCTGCACCTGGTAGCGAACCTGCGACGTCTGGTTC TGAAACGCCGGGTACCAGCGAAAGCGCTACCCCAGAATCCGG TCCGGGCACTAGCACCGAGCCATCGGAGGGCTCCGCACCAagcg cctctcgctccgcaCATCACCATCACCACCATCATCACTAA	GSETPGTSES ATPESGPGSE PATSGSETPG TSESATPESG PGTSTEPSEG SAPGSPAGSP TSTEEGTSES ATPESGPGSE PATSGSETPG TSESATPESG PGSPAGSPTS TEEGSPAGSP TSTEEGTSTE PSEGSAPGTS ESATPESGPG TSESATPESG PGTSESATPE SGPGSEPATS GSETPGSEPA TSGSETPGSP AGSPTSTEEG TSTEPSEGSA PGTSTEPSEG SAPGSEPATS GSETPGTSES ATPESGPGTS TEPSEGSAPS ASRSAHHHH HHHH

[00250] Example 6: Optimization for expression of XTEN constructs with RP11 affinity tag

[00251] 1. 1. Design and construction of constructs encoding N-term-RP11-AE864-GFP and MalEss-AE48-RP11-AE864

[00252] A group of highly expressed native E. coli proteins described by Ishihama Y. et al, BMC Genomics 2008, 9:102 were used to generate a list of the first 20 N-terminal amino acids (column 3 of Table 14), from which the hydrophobic amino acids F, I, W, L, V, M, C were converted to alanine or serine, or were deleted in order to generate candidates to create helper sequences containing at least 11 amino acids (column 4 of Table 14). For comparative purposes, the first 20 amino acids of a known CBD sequence from a well expressed construct built at Amunix (AC616) was also included as a control.

Table 14: List of N-terminal helpers designed in the study

Plasmid	Protein description	1st 20aa	N-term Helpers
pSD0107	30S ribosomal protein S9	MAENQYYGTGRRKSSAAR VF	MAENQYYGTGRRKSSAAR
pSD0108	50S ribosomal protein L32	MAVQQNKPTRSKRGMRRS HD	MAAQQNKPTRSKRGARRS HD
pSD0109	30S ribosomal protein S20	MANIKSAKKRAIQSEKARK H	MANAKSAKKRAAQSEKAR KH
pSD0110	GrpE protein HSP-70	MSSKEQKTPEGQAPEEIIMD	MSSKEQKTPEGQAPEE

	cofactor		
pSD0111	30S ribosomal protein S14	MAKQSMKAREVKRVALADKY	MAKQSAKAREAKRAASADKY
pSD0112	30S ribosomal protein S12	MATVNQLVRKPRARKVAKSN	MATANQAARKPRARKAAKSN
pSD0113	Elongation factor Tu (EF-Tu) (P-43)	MSKEKFERTKPHVNVGTIGH	MSKEKAERTKPHANAGT
pSD0114	30S ribosomal protein S15	MSLSTEATAKIVSEFGRDAN	MSASTEATAKAASEAGRDAN
pSD0115	Superoxide dismutase [Mn] (EC 1.15.1.1) (MnSOD)	MSYTLPSLPYAYDALEPHFD	MSYTAPSAPYAYDAAEPH
pSD0116	rraB ribonuclease E inhibitor protein B	MANPEQLEEQREETRLIIEE	MANPEQAEEQREETR
pSD0117	Cellulose Binding Protein (CBD)	MANTPVSGNLKVEFYNSNPS	MANTPASGNAKAEAYNSNPS
pSD0118	CBD (from previous construct AC616)	MANTPVSGNLKVEFYNSNPS	MANTPVSGNLKVEFYNSNPS

[00253] DNA oligonucleotides for the 107N-F&R to 119N-F&R series and RP11F&R sequences of Table 15 were synthesized at Elim Biopharm (Hayward, CA). Solutions of each DNA pair (107N-F and 107N-R, 108N-F and 108N-R, etc) was mixed at a 1:1 molar ratio, were denatured at 95oC for 3 min, followed by cooling to 25oC at 0.1oC/min to allow double strand DNA annealing. The base vector LCW0970.030 (encoding CBD-AE864-GFP) was digested with NdeI/BsaI and the larger fragment was gel-purified as the vector. The vector was ligated with the annealed oligos 107-118N-F&R and PNK treated annealed RP11F&R oligos, and the ligation products were transformed into *E. coli* BL21 competent cells (New England Biolabs) to obtain the colonies designated pSD0107 to pSD118. The clones pSD0107-109, pSD0111-112, and pSD0114-118 were obtained and verified by DNA sequencing; clone pSD0110.001 had one mutation of frame-shift and was used as the stuffer vector.

[00254] The plasmid construct pCW1110 (encoding RP11-AE864) was digested with BsaI/NotI and the smaller band of corresponding to the nucleotides encoding AE864 was gel-purified as the insert. pCW1139 (encoding MalEss-AE48-payload-AE864) was digested by XhoI/BstXI/NotI and the larger fragment was gel-purified as the vector. The annealed product of oligos of 119-AEN-F&R was ligated with the insert and vector, and then transformed into E. coli BL21 to obtain colonies with the plasmid, designated pSD0119. The clones were sequence verified.

[00255] 2. Construction of N-term helper libraries based on pSD0116

[00256] DNA oligonucleotide pairs of Stuffer-RP5for & Stuffer-RP5revP, RP6-SASRSAforP & RP6-SASRSArev, L2for & L2rev, L3for & L3rev, L4for & L4rev, L5for & L5rev, L6for & L6rev, L7for & L7rev, L8for & L8rev, L9for & L9rev, L10for & L10rev, L11for & L11rev, L12for & L12rev, and L13for & L13rev (Table 15) were synthesized at Elim Biopharm

(Hayward, CA) and each pair was annealed as described above (Section 1) to generate double strand DNA.

[00257] Plasmid pSD0110 was digested with NdeI/BsaI and the larger fragment was gel-purified as the vector. The vector was ligated with annealed oligos of Stuffer-RP5for&revP and RP6-SASRSAforP&rev, and then transformed into E. coli BL21 to obtain the colonies with the stuffer vector plasmid pCW1146 (Stuffer-RP11-XTEN_AE864_003-GFP). The clone was sequence verified.

[00258] The NdeI/BsaI digested pSD0110 vector was ligated with L5for&rev annealed oligos to obtain colonies of LCW1160 (L5).

[00259] The stuffer vector pCW1146 was digested with NdeI/BsaI and the larger fragment was gel-purified as the vector. The vector was ligated with annealed oligos of L2-4, and L6-13 for&rev as in Table 15, and then transformed into E. coli BL21 to obtain the colonies of constructs LCW1157 (L2), LCW1158 (L3), LCW1159 (L4), LCW1163 (L6), LCW1171 (L7), LCW1172 (L8), LCW1203 (L9), LCW1204 (L10), LCW1208 (L11), LCW1209 (L12), and LCW1210 (L13).

Table 15: List of DNA oligonucleotides

Name	DNA Sequence
RP11F	CGTCCGCGTCCTCGCCCACGTCCACGTCCGCGTCCACGCCCTCGTCCTCGTCCGCGC CCTCGTCCGagcgcgtctcgttccgctGGGTCTCC
RP11R	ACCTGGAGACCCAGCGGAACGAGACGCGCTCGGACGAGGGCGCGGACGAGGACG AGGGCGTGGACGCGGACGTGGACGTGGGCGAGGACGCG
107N-F	TATGGCTGAAAATCAATATTATGGTACGGGGCGCCGGAAGAGTTCGGCCGCC
107N-R	GACGGGCGGCCGAACTCTTCCGGCGCCCCGTACCATAATATTGATTTTCAGCCA
108N-F	TATGGCAGCTCAGCAGAATAAGCCTACACGAAGTAAAAGAGGCGCGCGCCGGTCG CACGAT
108N-R	GACGATCGTGCGACCGGCGCGCGCCTCTTTTACTTCGTGTAGGCTTATTCTGCTGAG CTGCCA
109N-F	TATGGCAAATGCTAAGAGTGCAAAGAAACGGGCGGCACAAAGCGAAAAAGCTCGG AAACAT
109N-R	GACGATGTTTCCGAGCTTTTTCGCTTTGTGCCGCCCGTTTCTTTGCACTCTTAGCATT TGCCA
110N-F	TATGTCCAGCAAAGAACAGAAGACTCCGGAAGGTCAAGCGCCAGAGGAG
110N-R	GACGCTCCTCTGGCGCTTGACCTTCCGGAGTCTTCTGTTCTTTGCTGGACA
111N-F	TATGGCCAAACAAAGCGCTAAAGCCCGCGAGGCGAAACGTGCAGCCTCTGCGGAC AAATAT
111N-R	GACGATATTTGTCCGCAGAGGCTGCACGTTTCGCCTCGCGGGCTTTAGCGCTTTGTT TGGCCA
112N-F	TATGGCTACTGCAAATCAGGCCGCCCGTAAACCTCGAGCACGAAAGGCTGCTAAAT CAAAT
112N-R	GACGATTTGATTTAGCAGCCTTTCGTGCTCGAGGTTTACGGGCGGCCTGATTTGCAG TAGCCA
113N-F	TATGTCCAAAGAAAAAGCCGAACGGACCAAACCTCATGCTAACGCTGGCACG
113N-R	GACGCGTGCCAGCGTTAGCATGAGGTTTGGTCCGTTCGGCTTTTTCTTTGGACA
114N-F	TATGTCAGCGTCTACGGAGGCAACCGCAAAAGCTGCTAGTGAAGCGGGCCGTGAT GCGAAT
114N-R	GACGATTCGCATCACGGCCCGCTTCACTAGCAGCTTTTGCGGTTGCCTCCGTAGAC

Name	DNA Sequence
	GCTGACA
115N-F	TATGAGCTATACTGCACCGAGCGCACCGTATGCTTATGATGCAGCCGAACCTCAC
115N-R	GACGGTGAGGTTCGGCTGCATCATAAGCATACGGTGCGCTCGGTGCAGTATAGCTC A
116N-F	TATGGCAAACCCCGAACAGGCTGAGGAACAGAGAGAAGAAACA
116N-R	GACGTGTTTCTTCTCTCTGTTCCTCAGCCTGTTCGGGGTTTGCCA
117N-F	TATGGCTAATACCCCTGCGAGCGGGAACGCCAAGGCGGAAGCTTACAACAGTAAT CCAAGC
117N-R	GACGGCTTGGATTACTGTTGTAAGCTTCCGCCTTGGCGTTCCCGCTCGCAGGGGTAT TAGCCA
118N-F	TATGGCAAATACACCGGTATCAGGCAATTTGAAGGTTGAATTCTACAACAGCAATC CTTCA
118N-R	GACGTGAAGGATTGCTGTTGTAGAATTCAACCTTCAAATTGCCTGATACCGGTGTA TTTGCCA
119-AEN-F	TCGAGCACGGGCAGCCCA
119-AEN-R	GACGTGGGCTGCCCGTGC
Stuffer-RP5for	TATGggctgaGGGTCTCaCGTCCGCGTCCTCGCCCACGTCCACGTCCGCGT
Stuffer-RP5revP	GGACGTGGACGTGGGCGAGGACGCGGACGtGAGACCCtcagccCA
RP6-SASRSAforP	CCACGCCCTCGTCCTCGTCCGCGCCCTCGTCCGagcgcgtctcgttccgc
RP6-SASRSArev	ACCTgcggaacgagacgcgctCGGACGAGGGCGCGGACGAGGACGAGGGCGTGGACGC
L2for	tATGaaaAAYCCNGARCARGCNGARGARCARMGYGARGARACa
L2rev	GACGTGTYTCYTCRCKYTGYTCYTCNGCYTGYTCNGGRTTTTTCA
L3for	tATGGCNAAYCCNGARCARGCNGARGARCARMGYGARGARACa
L3rev	GACGTGTYTCYTCRCKYTGYTCYTCNGCYTGYTCNGGRTTNGCCA
L4for	tATGaaaAAcCCVGARCARGCDGARGAaCARGCDGARGAaCAgMGYGAaGARACa
L4rev	GACGTGTYTCTTCRCKCTGTTCYTCHGCYTGTTCYTCHGCYTGYTCBGGGTTTTTCA
L5for	tATGaRaCCNCGNCCNCGNCCNCGNCCNCGNCCNCGNCCNCGNCCNCGNCCNCGNC CNCGNCCNCGNCCNGGGTCTCC
L5rev	ACCTGGAGACCCNGGNCGNGGNCGNGGNCGNGGNCGNGGNCGNGGNCGNGGNCG NGGNCGNGGNCGNGGNCGNGGTYTCA
L6for	tATGaaaAAHMMVGARCARGCWGARGAaCARGCDGARGAaCAgMGYGAaGARACa
L6rev	GACGTGTYTCTTCRCKCTGTTCYTCHGCYTGTTCYTCWGCYTGYTCBKKDTTTTTCA
L7for	tATGAAAAAWCAMRARMARRMWRAARAAMAARMDRAARAACAGMGCGAAGARA CA
L7rev	GACGTGTYTCTTCGCKCTGTTYTTYHKYTTKTTYTTYWKYYTKYTYKTGWTTTTTC A
L8for	tATGAAAAAWCAMRARMARRMWRAARAAMAAGCDGAAGAACAGMGCGAAGARA CA
L8rev	GACGTGTYTCTTCGCKCTGTTCTTCHGCTTKTTYTTYWKYYTKYTYKTGWTTTTTCA
L9for	tATGAAAAANCMMGAACAAGAARAARAAMAAGCNGAAGAACARCGYGARGARAC A
L9rev	GACGTGTYTCYTCRCGYTGTTCTTCNGCTTKTTYTTYTTCTTGTTCKKGNTTTTTCA
L10for	tATGAAAAANCMMGAACAAGAARAARAAMAAGCNGAAGAAMARMRHRARRARA

Name	DNA Sequence
	MA
L10rev	GACGTKTYTYYTYDYKYTKTTCTTCNGCTTKTTYTTYTTCTTGTTCKKGNTTTTTCA
L11for	tATGAAAAAACAAGAACAAGAAAAAGAACAAGCGGAAGAACAAKCNVARKCNVA RCGTGAGGAGACA
L11rev	GACGTGTCTCCTCACGYTBNGMYTBNGMTTGTTCTTCCGCTTGTTCTTTTTCTTGTT CTTGTTTTTTCA
L12for	tATGAAAAAACAAGAACAAGAAAAAGAACAAGCGGAAGAACAAKCVVAAKCVVA AKCVVAAKCVVAACGTGAGGAGACA
L12rev	GACGTGTCTCCTCACGTTBBGMTTBBGMTTBBGMTTBBGMTTGTTCTTCCGCTTGTT CTTTTTCTTGTTCTTGTTTTTTCA
L13for	tATGAAAAAACAAGAACAAGAAAAAGAACAAGCGGAAGAACAANNNNNNNNNNN NCGTGAGGAGACA
L13rev	GACGTGTCTCCTCACGNNNNNNNNNNNNTTGTTCTTCCGCTTGTTCTTTTTCTTGTT CTTGTTTTTTCA

[00260] 3. Screening and analysis of the N-terminal helper libraries LCW1157-1159

[00261] E. coli hosts transformed with the plasmids pSD0107 to pSD0118 were expressed in shake flasks and the expression levels in each were evaluated by measuring the fluorescence from the C-terminal GFP reporter. Briefly, an overnight culture was grown for each construct in SB media (with 12.5 µg/ml tetracycline), which was then used to inoculate a 200 ml culture of PhoA phosphate depletion autoinduction media with 12.5 µg/ml tetracycline (3 shake flasks were grown for each construct). After growing at 26°C with 225 rpm for 48 h, 100 µl aliquot was taken from each culture and the GFP expression level was measured with a fluorescence plate reader with excitation wavelength of 395 nm and emission wavelength of 510 nm. Two readings were taken for each shake flask. Among the constructs tested, pSD0116 had the highest fluorescence signal, followed by pSD0114 (FIG. 34), but the expression levels of these two constructs were significantly lower than the LCW0970.030 (pLCW970) positive control.

[00262] In order to further improve expression, three libraries (LCW1157, 1158, and 1159) were built based on pSD0106, and were screened in a high through-put format. Large numbers of colonies from these libraries (Table 16) were picked to grow individually in 500 µl SB media (with 12.5 µg/ml tetracycline) in 96 deep well plates overnight at 37oC shaking with 300 rpm. 20 µl of the saturated culture was used to inoculate 500 µl of PhoA phosphate depletion autoinduction media (with 12.5 µg/ml tetracycline) in 96 deep well plates that were incubated at 26oC shaking with 300 rpm for 22-24 h. Expression was then determined by placing 100 µl of the culture into a 96 well plate measuring the fluorescence from the C-terminal GFP reporter.

Table 16: Libraries LCW1157-1159

Library	Description	N-terminal helper sequence (DNA)	Diversity	Number Screened
LCW 1157	Codon optimize helper sequence, Change 2nd	ATGaaaAAY(C/T)CCN(A/G/C/T) GAR(A/G)CAR(A/G)GCN(A/C/T/	98304	406

	codon to Lys (AAA)	G)GAR(A/G)GAR(A/G)CAR(A/G)M(A/C)GY(C/T)GAR(A/G)GAR(A/G)ACa		
LCW 1158	Codon optimize helper sequence	ATGGCN(A/G/C/T)AAY(C/T)CCN(A/G/C/T)GAR(A/G)CAR(A/G)GCN(A/G/C/T)GAR(A/G)GAR(A/G)CAR(A/G)M(A/C)GY(C/T)GAR(A/G)GAR(A/G)ACa	393216	85
LCW 1159	Insert AEEQ into helper sequence, Use fewer codons than LCW1157 to keep diversity manageable	ATGaaaAAcCCV(A/G/C)GAR(A/G)CAR(A/G)GCD(A/G/T)GAR(A/G)GAaCAR(A/G)GCD(A/G/T)GAR(A/G)GAaCAgM(A/C)GY(C/T)GAaGAR(A/G)ACa	20736	146

[00263] After evaluation of the fluorescence signal, the six highest expression clones and two low expression clones were chosen from each 96 deep well plate tested and the expression of these clones were tested again with 4 replicates. For all three libraries, clones having higher expression than the pSD0116 construct were identified (FIG. 35A-C), and the correlation between the original tests and retests was good (FIG. 35D-F). Among the 3 libraries tested, library LCW1159 gave higher expression level in general, and the highest expression construct from this round of screening (LCW1159.004) came from this library.

[00264] The 8 constructs with the highest expression levels from these three libraries and controls were treated with Pop Culture (EMD Millipore), were heat treated, and the resulting lysates were analyzed with SDS-PAGE/Coomassie staining (FIG. 36). Expression levels of the 8 constructs were confirmed to be higher than controls (negative control, LSD0114 and LSD0116), and full-length proteins were produced from these constructs. Lysates of 4 top expression constructs (after Pop Culture and heat treatment) and a negative control (AP32, a purified XTEN-GFP protein without RP11 tag) were each loaded onto a MacroCap SP column at pH of 8 and conductivity of 6.5 mS/cm. The column was washed with 20mM sodium phosphate, pH 8, 100mM NaCl, and the protein was eluted with 20mM sodium phosphate, pH 8, 500mM NaCl. The samples of the load, flow through, wash and elutate were analyzed by SDS-PAGE gel (FIG. 37). The results demonstrated that the expression protein of the four, which came from all three libraries, can be captured by MacroCap SP; thus the binding was contributed to the presence of the RP11 tag in the protein since the negative control (not containing the RP11 tag) didn't bind to the MacroCap SP column.

[00265] The plasmids of the clones chosen for retests were minipreped and the DNA sequences of the N-terminal helpers were analyzed. Codon bias was observed at several locations (Table 17). For example, the 3rd amino acid of LCW1157, N, is encoded by AAC or AAT. Most of the high expression clones (77%) in LCW1157 are encoded by AAT at the 3rd amino acid, while most of the low expression clones (88%) are encoded by AAC, indicating AAT is preferred over AAC at this position for high expression. Similarly, CCG is preferred at the 4th amino acid, while GCG at

the 7th amino acid is accumulated in low expression clones. These trends were also observed in libraries LCW1158 and LCW1159, as well.

Table 17: Analysis of sequence results of high and low expression clones in libraries LCW1157-1159*

LCW1157															
amino acid#		2	3	4	5	6	7	8	9	10	11		12	13	14
		K(AAA)	N(AA+)	P(CC+)	E(GA+)	Q(CA+)	A(GC+)	E(GA+)	E(GA+)	Q(CA+)	R or S(+G+)		E(GA+)	E(GA+)	T(AC+)
codon			3rd	3rd	3rd	3rd	3rd	3rd	3rd	3rd	1st	3rd	3rd	3rd	
high	A		0%	10%	57%	67%	33%	80%	70%	50%	30%	0%	63%	57%	
total:30	G		0%	43%	43%	33%	7%	20%	30%	50%	0%	0%	37%	43%	
	C		23%	23%	0%	0%	13%	0%	0%	0%	70%	67%	0%	0%	
	T	K(AAA)	77%	23%	0%	0%	47%	0%	0%	0%	0%	33%	0%	0%	T(ACA)
low	A		0%	13%	88%	50%	25%	38%	63%	25%	63%	0%	75%	63%	
total:8	G		0%	13%	13%	50%	63%	63%	38%	75%	0%	0%	25%	38%	
	C		88%	0%	0%	0%	0%	0%	0%	0%	38%	38%	0%	0%	
	T		13%	75%	0%	0%	13%	0%	0%	0%	0%	63%	0%	0%	

LCW1158															
amino acid#		2	3	4	5	6	7	8	9	10	11		12	13	14
		A(GC+)	N(AA+)	P(CC+)	E(GA+)	Q(CA+)	A(GC+)	E(GA+)	E(GA+)	Q(CA+)	R or S(+G+)		E(GA+)	E(GA+)	T(AC+)
codon		3rd	3rd	3rd	3rd	3rd	3rd	3rd	3rd	3rd	1st	3rd	3rd	3rd	
high	A	83%	0%	33%	17%	67%	50%	83%	50%	83%	83%	0%	67%	50%	
total:6	G	0%	0%	0%	83%	33%	0%	17%	50%	17%	0%	0%	33%	50%	
	C	0%	33%	33%	0%	0%	33%	0%	0%	0%	17%	100%	0%	0%	
	T	17%	67%	33%	0%	0%	17%	0%	0%	0%	0%	0%	0%	0%	T(ACA)
low	A	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	100%	
total:2	G	100%	0%	100%	0%	100%	100%	100%	100%	100%	0%	0%	100%	0%	
	C	0%	0%	0%	0%	0%	0%	0%	0%	0%	100%	100%	0%	0%	
	T	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	

LCW1159																				
amino acid#		2	3	4	5	6	7	8	9	10	11	12	13	14	15		16	17	18
		K(AAA)	N(AA+)	P(CC+)	E(GA+)	Q(CA+)	A(GC+)	E(GA+)	E(GA+)	Q(CA+)	A(GC+)	E(GA+)	E(GA+)	Q(CA+)	R or S(+G+)		E(GA+)	E(GA+)	T(AC+)
codon			3rd	3rd	3rd	3rd	3rd	3rd	3rd	3rd	3rd	3rd	3rd	3rd	1st	3rd	3rd	3rd	
high	A			42%	67%	67%	50%	75%		67%	50%	58%			25%	0%		58%	
total:12	G			17%	33%	33%	0%	25%		33%	17%	42%			0%	0%		42%	
	C			42%	0%	0%	0%	0%		0%	0%	0%			75%	75%		0%	
	T	K(AAA)	N(AAC)	0%	0%	0%	50%	0%	E(GAA)	0%	33%	0%	E(GAA)	Q(CAG)	0%	25%	E(GAA)	0%	T(ACA)
low	A			33%	100%	67%	0%	67%		67%	100%	100%			33%	0%		33%	
total:3	G			33%	0%	33%	100%	33%		33%	0%	0%			0%	0%		67%	
	C			33%	0%	0%	0%	0%		0%	0%	0%			67%	67%		0%	
	T			0%	0%	0%	0%	0%		0%	0%	0%			0%	33%		0%	

* The total numbers of high and low expression clones analyzed are specified for each library, and the percentages of A, G, C, or T in each of the location where the codon was varied were analyzed.

[00266] 4. Screening and analysis of the N-terminal helper library LCW1163

[00267] Since library LCW1159 had generally higher expression than LCW1157 and LCW1158, the next library design was based on LCW1159, with the introduction of more variants in the N-terminal helper domain coding region. A total of 672 clones from this library (theoretical diversity of 55296) was screened and retested in the same way as libraries LCW1157-1159. Clones with higher expression than LCW1159.004 (the highest from the previous round of screening) were observed (FIG. 38). The top expression clone, LCW1163.029, achieved a 40% improvement from LCW1159.004, while its expression level was 27% lower than the CBD control. The sequence of the high and low expression clones were analyzed, and the preferred nucleotides for expression

were identified and summarized (see Table 18). Most of the locations have shown a preference of codon A for high expression, while others have a preference for C.

Table 18: Analysis of sequence results of high and low expression clones in the library LCW1163*

		Expression (Percentage)													
nucleotide number		#9	#10	#11	#12	#15	#18	#21	#24	#30	#33	#36	#43	#45	#51
high expression	A	74	17	74	40	37	40	66	77	60	43	66	40	0	34
	C	17	83	26	49	0	0	0	0	0	0	0	60	83	0
	T	9	0	0	0	0	0	34	0	0	29	0	0	17	0
	G	0	0	0	11	63	60	0	23	40	29	34	0	0	66
low expression	A	20	70	3	20	40	20	60	50	30	10	20	50	0	60
	C	60	30	70	30	0	0	0	0	0	0	0	50	50	0
	T	20	0	0	0	0	0	40	0	0	40	0	0	50	0
	G	0	0	0	50	60	80	0	50	70	50	80	0	0	40
preferred		A	C	A	A/C				A	A		A		C	

[00268] * The percentages of A, G, C, or T in each of the location where the codon was varied were analyzed, and the identified preferred nucleotides were summarized in the bottom row

[00269] 5. Screening and analysis of the RP11 library LCW1160

[00270] At the same time, 168 colonies from library LCW1160 (the library varying the coding region of RP11 tag without any N-terminal helper domain (total theoretical diversity of $8.8X10^{12}$) were screened and analyzed. However, this library had very low expression level in general (FIG. 39), and the one outlier with high expression was found to encode a truncated RP11 tag after performing plasmid miniprep and DNA sequencing analysis. The screening results of the library, under these experimental conditions, strongly suggest that an N-terminal helper is required to achieve high expression levels.

[00271] 6. Screening of the N-terminal helper libraries LCW1171, LCW1172, LCW1203, and LCW1204.

[00272] More N-terminal libraries (LCW1171, LCW1172, LCW1203, and LCW1204) were screened and analyzed in the same way as those described above. LCW1171 and 1172 were designed similarly, while LCW1171 allowed more amino acid changes in the helper sequence than LCW1172. The screening results showed that LCW1171 in general had much lower expression level than LCW1172 (FIGS. 94A and B). LCW1203 and 1204 were designed similarly, focusing on randomizing a different region of the helper sequence compared with LCW1171 and 1172. LCW1204 allowed more amino acid changes than LCW1203, which resulted in lower expression than LCW1203 in general (FIGS 94C and D). These results suggest that the expression level is sensitive to amino acid changes in the helper domain sequences.

[00273] 7. Screening of the N-terminal helper libraries LCW1208, LCW1209, and LCW1210.

[00274] Three new N-terminal libraries (LCW1208, LCW1209, and LCW1210) were designed to investigate the effect of further elongation the N-terminal helper sequence (Table 19). LCW1208 and LCW1210 introduced 4 more residues to the helper domain, while LCW1209 introduced 8 more residues. The screening results showed a general trend that LCW1209 had highest expression, followed by LCW1208, and then LCW1210 (FIG.95), which confirmed the beneficial effect of adding more amino acid in the helper sequence.

Table 19: Libraries LCW1208-1210

Library	N-terminal Helper Sequence (amino acid)	Diversity	Number Screened
LCW1208	MKKQEQEKEQAEEQ*BXBX*REET (B=A/S; X=E/K/Q)	2304	336
LCW1209	MKKQEQEKEQAEEQ*BXBXBXBX*REET (B=A/S; X=E/K/Q)	104976	336
LCW1210	MKKQEQEKEQAEEQ*ZZZZ*REET (Z=any amino acid)	262144	336

[00275] Additional residues in the helper sequences were underlined.

[00276] Three new N-terminal libraries (LCW1208, LCW1209, and LCW1210) were designed to investigate the effect of further elongation the N-terminal helper sequence (Table 20). LCW1208 and LCW1210 introduced 4 more residues to the helper domain, while LCW1209 introduced 8 more residues. The screening results showed a general trend that LCW1209 had highest expression, followed by LCW1208, and then LCW1210 (FIG. 43), which confirmed the beneficial effect of adding more amino acids in the helper sequence for improving expression.

[00277] The 4 top constructs with the highest expression levels from the three libraries were chosen from the retest and the sequence of their N-terminal helper sequences were analyzed (Table 20). The current highest expressed construct (LCW 1209.029) achieved 90% of the expression level of the CBD control by comparing the average florescence after subtracting the negative control.

Table 20: Constructs with the highest expression levels from Libraries LCW1208-1210 and controls.

Sample name	Avg. Fluorescence	Helper sequence*
LCW1208.009	9070	MKKQEQEKEQAEEQ*AESE*REET
LCW 1208.007	8870	MKKQEQEKEQAEEQ*AESE*REET
LCW 1208.008	8800	MKKQEQEKEQAEEQ*SQSQ*REET
LCW 1208.010	8740	MKKQEQEKEQAEEQ*SESE*REET
LCW 1209.029	10010	MKKQEQEKEQAEEQ*AKAESEAE*REET
LCW 1209.015	9440	MKKQEQEKEQAEEQ*SKSQAEAE*REET
LCW 1209.023	8980	MKKQEQEKEQAEEQ*AQAQAEDE*REET
LCW 1209.010	8580	MKKQEQEKEQAEEQ*SKSKAEDE*REET

LCW 1210.030	8650	MKKQEQEKEQAEEQ*PEVQ*REET
LCW 1210.032	8310	MKKQEQEKEQAEEQ*VENP*REET
LCW 1210.025	8100	MKKQEQEKEQAEEQ*ELCE*REET
LCW 1210.009	8030	MKKQEQEKEQAEEQ*GIDT*REET
CBD control	10940	n/a
Negative control	1630	n/a

*Additional residues in the helper sequences were underlined.

[00278] In summary, the screening results, under these experimental conditions, strongly suggest that an N-terminal helper contributes in achieving high expression levels.

[00279] Example 7: Fermentation of XTEN using PhoA induction- Evaluation of Expression Yields

[00280] *E. coli* BL21 carrying the plasmids encoding Helper_LCW1159.004-RP11-AE288-His8 (AC767), Helper_LCW1172.033-RP11-AE576-His8 (AC780), and Helper_LCW1172.033-RP11-AE864-His8 (AC786) were transformed into the *E. coli* BL21 strain. Three 10L fermentations were run for each of the 3 strains. Glycerol stocks were used to inoculate 125 mL of LB broth media containing 10 µg/mL tetracycline. The starter cultures were then shaken overnight at 37°C. The starter culture was used to inoculate 4L of fermentation batch media containing -20 g ammonium sulfate, 10.4 g potassium phosphate dibasic anhydrous, 5 g sodium citrate dihydrate; 4.6 g sodium phosphate monobasic monohydrate; 106 g NZ BL4 soy peptone (Kerry Bioscience #5X00043); 54 g yeast extract (Teknova #Y9020); 3.6 L water; 0.05 mL polypropylene glycol; 5.2 mL trace elements solution (Amunix recipe 144-1); 35mL 1M magnesium sulfate; and 4 mL Kanamycin (50mg/mL)- in 10 L glass jacketed vessel with a B. Braun Biostat B controller. The fermentation control settings were: pH = 6.9 +/- 0.1; dO_2 = 10%; dissolved oxygen cascade in stirrer only mode with a range of 125 – 1180 rpm; air flow of 5 liters per minute of 90% oxygen; initial temperature 37°C; base control 13% ammonium hydroxide; and no acid control. After 6 hours of culture a 70% glycerol feed was initiated at a rate of 40 g/hr. Upon the cultures reaching an OD600 of 50 +/- 10 OD the culture temperature was lowered to 26°C, 54 mL of 1M magnesium sulfate was added, and a salt feed consisting of: 10 g/l ammonium sulfate, 26 g/l potassium phosphate dibasic anhydrous, 2 g/l sodium citrate dihydrate; 13 g/l sodium phosphate monobasic monohydrate; 15 g/l potassium phosphate monobasic anhydrous; 0.08% trace elements solution, was started at a rate of 33 g/L and continued for 6 hours. After a total fermentation run time of 64-70 hours the culture was harvested by centrifugation yielding cell pellets between 1.6-2.3 kilograms in wet weight. The pellets were stored frozen at -80°C until further use. For titer analysis, end of run fermentation whole broth samples were frozen in a 0.2mL volume, then later thawed, then 0.2mL of water was added, then to lyse and flocculate host proteins the samples were incubated at 85oC for 15 minutes, then transferred to 4°C for 15 minutes, followed centrifugation for 10 minutes at 20,000g. The resulting flocculated soluble lysates were assayed by C18 reversed phased HPLC, and the A214 absorbance area corresponding

of the peaks representing the Helper-RP11-XTEN-His8 was compared to that of purified reference standard. Next, to determine the dry cell weight (DCW), aliquots of cells were pelleted and the supernatant discarded. The cell pellet was washed once with water, and was then dried in an oven at 80° C for 3 days. The tubes containing cell pellets were weighed, the weight of the tube was subtracted from the measured weights, and the remaining weight was divided by the initial volume of each sample (0.0002 L) to obtain the DCW. The results of the fermentation growth, titer analysis, and dry cell weight are summarized in Table 21 below. In 96-well plate screening assays, when a library of RP11-XTEN-His8 constructs without an N-terminal helper, LCW1160 were screened (Example 6, FIG. 39) the expression level of the LCW1160.006, the highest expression construct, was only 50% of the expression of LCW1159.004 or LCW1172.033; constructs with helper sequences. Therefore, it is expected that XTEN constructs with N-terminal helper sequences will result in significantly higher expression titers compared to XTEN constructs not having helper sequences.

Table 21: Measured expression parameters

XTEN Series	Fermentation #	Final O.D.	Total Run Time (hours)	Dry cell weight per Liter of Ferm (g/L)	Titer by culture volume (g/L)	Average titer (micromoles/L)	Titer / Dry weight ferm (mg of XTEN / g of *E. coli*)	MW (g/mol)	Titer (micromoles /L)	Average titer (micromoles/L)	Std error of the mean (micromoles/L)
Helper_ LCW1159.004-RP11-AE288-His8 (AC767)	EC846	140	67.5	95	0.5	0.7	5.6	26399	212	271	76
	EC888	140	64	85	0.8		9.4	26399	357		
	EC903	136	69	103	0.7		6.4	26399	243		
Helper LCW1172.033-RP11-AE576-His8 (AC780)	EC850	125	66	107	1.1	2.0	10.0	52929	189	388	182
	EC869	140	70	90	2.1		22.8	52929	430		
	EC883	170	67	102	2.9		28.8	52929	545		
Helper LCW1172.033-RP11-AE864-His8 (AC786)	EC873	135	68	91	2.0	2.1	22.2	79284	280	272	41
	EC889	145	65.5	95	2.3		24.4	79284	308		
	EC913	108	67	108	2.0		18.1	79284	228		

[00281] Example 8: Purification of XTEN with RP11 and His8 tags

[00282] 1. Expression, lysis and clarification

[00283] The fusion protein MKNPEQAEEQAEEQREET-RP11-SASRSA-XTEN_AE432(C12,C217,C422)-SASRSA-His(8)], with the N-terminal helper sequence from the library member LCW1159.004 described in Example 4, with two affinity tags linked to XTEN at the N- and C-terminus, respectively, was expressed in *E.coli* using a 4 L fermenter using conditions described herein. After growth, the cells were harvested by centrifugation and frozen at -80°C until use. The cell pellet was resuspended in lysis buffer (20 mM sodium phosphate, 50 mM NaCl, 2 mM EDTA pH 8.0, 3 ml buffer per gram cell paste). The cells were lysed by passing through an APV homogenizer three times at a pressure of 830-900 bar. Lysis buffer (1 ml buffer per gram cell paste) was used as a chase to retrieve hold up volume from the homogenizer. The homogenized lysate was incubated in a water bath at 85°C for 20 minutes, followed by quick cooling in ice water bath for 20 minutes. After the heating and cooling treatment, the lysate was centrifuged at 11000 RPM for 90 minutes in a SORVALL centrifuge. After centrifugation, the supernatant was filtered through two CUNO Bio cap 25 (BC0025L90SP08A) filters. The clarified supernatant was stored at 4°C overnight.

[00284] 2. Capture Step: Toyopearl IMAC Chromatography

[00285] IMAC affinity chromatography was used as a capture step for binding the XTEN with an intact C-terminal His-tag. Briefly, the chromatography column BPG140/12 (GE Life Sciences) was packed with 2000 ml Toyopearl IMAC 650 M resin (TOSOH Biosciences). The column was equilibrated with 2 column volumes (CVs) of equilibrium buffer (20 mM sodium phosphate, 500 mM NaCl, pH 8.0). Clarified cell lysate was adjusted to a final NaCl concentration of 500 mM using 5 M NaCl stock solution, and then was loaded onto the IMAC resin. The column was washed with 2 column volumes of equilibrium buffer, and then 2 column volumes of 20 mM sodium phosphate, 500 mM NaCl, 5 mM Imidazole pH 8.0, followed by 2 column volumes of 20 mM sodium phosphate, 5 mM imidazole pH 8.0 to remove salt. Elution was performed with 2 column volumes of 20 mM sodium phosphate, 100 mM imidazole, pH 8.0. The flow through, wash and elution fractions were analyzed by non-reducing 4 -12% Bis-Tris SDS-PAGE/Coomassie staining and the fractions with the desired product were pooled.

[00286] 3. Polishing/Capture Step: MacroCap SP Chromatography

[00287] Cation exchange chromatography was used as a polishing step to ensure the N-terminal integrity of the product. MacroCap SP resin (GE Life Sciences) was selected among several cation exchange media due to its superior capacity and selectivity for the product. 1000 ml of MacroCap SP resin was packed in a BPG100/13 (GE Life Sciences) chromatography column and equilibrated with 20 mM sodium phosphate pH 8.0, 20 mM NaCl. The IMAC pool was loaded onto the column and the resin was washed with 2 column volumes of 20 mM sodium phosphate,

50 mM NaCl, pH 8.0 and 2 column volumes of 20 mM sodium phosphate pH 8.0, 150 mM NaCl. The protein was eluted with 5 column volumes of linear gradient from 150 to 500 mM NaCl in 20 mM sodium phosphate pH 8.0. Fractions were collected and analyzed by 4-12% Bis-Tris SDS/PAGE. Fractions the with desired product were combined for the next step.

[00288] 4. Trypsin digestion of Macrocap SP elution pool

[00289] Trypsin (Sigma, Trypsin from Bovine Pancreas) digestion of the SP elution pool was performed at 1:200 m/m enzyme/protein ratio overnight at 37°C.

[00290] 5. Polishing Step: Macrocap Q Chromatography

[00291] After trypsin digestion, the cleaved tags were separated from the final product using Macrocap Q chromatography. The BPG100/19 column (GE Life Sciences) was packed with 1500 ml column volume of Macrocap Q resin (GE Life Sciences). The trypsin digested Macrocap SP elution pool was incubated for 15 min at 80o C with 20 mM DTT and 2 mM EDTA to reduce disulfide bonds and to inactivate trypsin. The cooled protein solution was diluted to a conductivity below 5 mS/cm with Milli-Q water and loaded onto the Macrocap Q column equilibrated with 20 mM HEPES, 50 mM NaCl, pH 7.0. The column was washed with 2 column volumes of 20 mM HEPES, 50 mM NaCl, pH 7.0, then 2 column volumes of 20 mM HEPES, 2 mM TCEP, 150mM NaCl pH7.0. The protein was eluted with a linear gradient from 150mM NaCl to 500 mM NaCl in 20 mM HEPES, pH 7.0 in 20 column volumes. Fractions were analyzed by SDS-PAGE/silver staining.

[00292] 6. Concentration and Diafiltration (Final Formulation)

[00293] Selected MacroCap Q fractions were combined and concentrated and using 10 KD Pellicon mini (Millipore) at a feed pressure < 20 psi and retentate < 8 psi, followed by 10X diafiltration with 20 mM HEPES, 50 mM NaCl, pH 7.0 to achieve a final protein concentration of > 5 mg/ml.

[00294] 9. Purity analysis of proteins purified with different methods

[00295] One batch (Batch 1) was purified through three purification steps as described above. Another batch (Batch 2) was purified from the same fermented material but the MacroCap SP polishing step was omitted. Truncated species of XTEN were detected by SDS-PAGE/silver staining in MacroCap Q elution fractions for Batch 2 (FIG. 40A), while the MacroCap Q elution fractions for Batch 1 were essentially free from truncations (FIG. 40B). These results support that, under the conditions employed, the MacroCap SP step based on the RP11 tag is essential to ensure N-terminal integrity and overall product quality.

[00296] Example 9: Construction of 1xAmino, 9xThiol-XTEN432

[00297] The following sets of primers 5Afor&CI1BbsIrev-TGGC, CI1BsaIfor-TGGC&CI2-AE38BbsIrev, and CI2-AE38BsaIfor&AatIICI3-2P were used to PCR plasmid pCW1164 containing XTEN_AE432 (C422) in order to obtain the PCR products of AE-CI1, CI1-2 and CI2-

3, respectively. CI1, 2&3 were designated Cysteine Island1, 2&3, having the same amino acid sequence TAEAAGCGTAEAA but with different codon usages. Gel-purification of the PCR products was performed to obtain bands of the right sizes, which were digested with restriction enzymes SbfI/BbsI, BsaI/BbsI and BsaI/AatII, respectively, as the inserts. Digestion of the plasmid pCW1164, which encodes the gene of N-term-RP11-R-XTEN_AE432 (C422)-R-H8, was performed with SbfI/AatII to remove the fragment of about 290 amino acids within XTEN_AE432 and gel-purification was performed on the remaining large fragment as the vector. Ligation of the vector with the three inserts of PCR products, above, was performed and used to transform BL21 competent cells in order to obtain the construct N-term-RP11-R-XTEN_AE432 (C319, C370, C422)-R-H8. PCR was performed on this construct with primers CI1BsaIfor-TGGC&AatIICI3-2P to obtain a PCR product of around 360 bp in length. Gel-purification of the band of the right size was performed, followed by digestion with BsaI/AatII as the insert XTEN_AE120-3Cys, which contains three Cysteine Islands.

[00298] Simultaneously, the codon-optimized DNA fragment of XTEN_AE313-6Cys, containing six Cysteine Islands, was designed and synthesized (Genscript). The fragment was digested by the flanking restriction enzymes BsaI/BbsI and gel-purified as the insert containing the first six Cysteine Islands of XTEN_AE432. Digestion of the plasmid pCW1161, which encodes the gene of N-term-RP11-R-XTEN_AE432_3Cys-R-H8, with BsaI/AatII was performed to remove the fragment of XTEN_AE432_3Cys and gel-purification was performed to obtain the large fragment as the vector. Ligation of the vector with the BsaI/BbsI digested insert of XTEN_AE313-6Cys and BsaI/AatII digested insert of XTEN_AE120-3Cys, above, was performed. The ligated product was used to transform BL21 competent cells in order to obtain the construct N-term-RP11-R-XTEN_AE432 (C12, C63, C114, C165, C217, C268, C319, C370, C422)-R-H8. The construct was designed to produce the precursor N-term-RP11-R-AE432_9Cys-R-H8 (sequence in Table 22, below), the product of which was used to generate 1xAmino, 9-Thio-XTEN432 after removal of the N-term-RP11 tag and C-term 8xHis-tag by trypsin digestion. The final product contains nine cysteines in the XTEN432 sequence (Seg 177).

Table 22: DNA and amino acid sequence for 1xAmino, 9-Thio-XTEN432

Clone Name	DNA Sequence	Amino Acid Sequence
N-term-RP11-R-AE432_9Cys-R-H8	ATGAAAAACCCAGAGCAAGCAGAAGAACAAGCTGAAGAACA	MKNPEQAEE
	GCGCGAAGAAACACGTCCGCGTCCTCGCCCACGTCCACGTCCG	QAEE
	CGTCCACGCCCTCGTCCTCGTCCGCGCCCTCGTCCGagcgcgtctcgtt	QREETRPRPR
	ccgctGGGTCTCCAACGGCAGAGGCAGCAGGTTGTGGTACAGCA	PRP
	GAAGCAGCTCCGGGTAGCGAGCCTGCAACCAGCGGTTCTGAG	RPRPRPRPRP
	ACGCCGGGCACTTCCGAATCTGCGACCCCGGAGTCCGGTCCAG	RPR
	GTTCAGAGCCGGCGACGAGCGGTTCGGAAACGCCGGGTACTG	PSASRSAGSP
	CTGAAGCGGCTGGTTGTGGTACTGCTGAAGCTGCATCGACCGA	TAE
	ACCAAGCGAAGGTTCGGCACCGGGTACTAGCGAGAGCGCAAC	AAGCGTAEA
	CCCTGAAAGCGGTCCGGGCAGCCCGGCAGGTTCTCCAACCAG	APGS
	CACCGAAGAAGGTTCCCCTGCTACTGCCGAAGCTGCAGGCTGC	EPATSGSETP

Clone Name	DNA Sequence	Amino Acid Sequence
	GGTACTGCGGAGGCGGCGTCCCCAACTTCTACTGAGGAAGGT ACTTCTGAGTCCGCTACCCCAGAAAGCGGTCCTGGTACCTCCA CTGAACCGTCTGAAGGCTCTGCACCAGGCACTTCTGAGTCTGC TACTACCGCCGAAGCCGCTGGTTGTGGTACCGCAGAAGCTGCA TCTGAGACTCCAGGCACTTCTGAGTCCGCAACGCCTGAATCCG GTCCTGGTTCTGAACCAGCTACTTCCGGCAGCGAAACCCCAGG TACCTCTGAGTCTGCGACTCCAGAGTCTGGTACCGCGGAAGCG GCTGGTTGTGGTACTGCAGAGGCAGCTGGTTCTCCGGCTGGTA GCCCGACCAGCACGGAGGAGGGTACGTCTGAATCTGCAACGC CGGAATCGGGCCCAGGTTCGGAGCCTGCAACGTCTGGCAGCG AAACCCCGGGTACCACGGCGGAAGCGGCAGGTTGTGGCACCG CGGAGGCAGCAGCTGGTTCTCCAACCTCTACCGAGGAGGGTTC ACCGGCAGGTAGCCCGACTAGCACTGAAGAAGGTACTAGCAC GGAGCCGAGCGAGGGTAGTGCTCCGGGTACGAGCGAGAGCAC GGCAGAAGCCGCTGGCtgcGGTACTGCTGAAGCGGCAACCCCT GAGAGCGGCCCAGGTACTTCTGAGAGCGCCACTCCTGAATCCG GCCCTGGTAGCGAGCCGGCAACCTCCGGCTCAGAAACTCCTGG TTCGGAACCAGCGACCAGCGGTACCGCTGAAGCCGCAGGTtgtG GCACTGCGGAAGCTGCAACCGAAGAGGGTACCAGCACGGAAC CGAGCGAGGGTTCTGCCCCGGGTACTTCCACCGAACCATCGGA GGGCTCTGCACCTGGTAGCGAACCTGCGACGTCTGGTTCTGAA ACGCCGACTGCAGAAGCGGCTGGTtgtGGCACCGCCGAAGCAG CTagcgcctctcgctccgcaCATCACCATCACCACCATCATCACTAA	GTS ESATPESGPG SEP ATSGSETPGT AEA AGCGTAEAA STEP SEGSAPGTSE SAT PESGPGSPAG SPT STEEGSPATA EAA GCGTAEAAS PTST EEGTSESATP ESG PGTSTEPSEG SAP GTSESATTAE AAG CGTAEAASET PGT SESATPESGP GSE PATSGSETPG TSE SATPESGTAE AAG CGTAEAAGS PAGS PTSTEEGTSE SAT PESGPGSEPA TSG SETPGTTAEA AGC GTAEAAAGS PTST EEGSPAGSPT STE EGTSTEPSEG SAP GTSESTAEAA GCG TAEAATPESG PGT SESATPESGP GSE PATSGSETPG SEP ATSGTAEAA GCGT AEAATEEGTS TEP SEGSAPGTST EPS EGSAPGSEPA

Clone Name	DNA Sequence	Amino Acid Sequence
		TSG SETPTAEAAG CGT AEAASASRSA HHH HHHHH

[00299] Example 10: Construction of 1xAmino, 9xThiol-XTEN864

[00300] The primers PhoAfor&RP11-SASRSABsaIrevAGGT were used to PCR the plasmid containing N-term-RP11 tag to obtain the PCR product of N-term-RP11-R. Gel-purification of the band of the right size was performed and was digested with NdeI/BsaI as the first insert. Another PCR was performed with primers AE432BsaIforAGGT&AE432_001BbsIrev-AACG on the plasmid containing XTEN_AE864_003 in order to obtain the PCR product of XTEN_AE432. Gel purification was performed on the band of the right size, which was digested with BsaI/BbsI as the second insert. Digestion of the construct N-term-RP11-R-AE432_9Cys-R-H8 from Example 10 was performed with NdeI/BsaI to remove the N-term-RP11-R fragment and gel purification was performed to obtain the large fragment as the vector. Ligation of the vector with the first and second inserts, above, was performed and the product was used to transform BL21 competent cells in order to obtain the construct N-term-RP11-R-XTEN_AE864 (C444, C495, C546, C597, C649, C700, C751, C802, C854)-R-H8. The resulting construct was designed to produce the precursor N-term-RP11-R-AE864_9Cys-R-H8 (sequence in Table 23, below) the product of which would generate 1xAmino, 9-Thio-XTEN864 after removal of the N-term-RP11 tag and C-term 8xHis-tag by trypsin digestion. The resulting product contains an N-terminal amino group and nine cysteines in the XTEN864 sequence for conjugation (Seg 175).

Table 23: DNA and amino acid sequence for 1xAmino, 9-Thio-XTEN864

Clone Name	DNA Sequence	Amino Acid Sequence
N-term-RP11-R-AE864_9Cys-R-H8	ATGAAAAACCCAGAGCAAGCAGAAGAACAAGCTGAAGAACA GCGCGAAGAAACACGTCCGCGTCCTCGCCCACGTCCACGTCCG CGTCCACGCCCTCGTCCTCGTCCGCGCCCTCGTCCGagcgcgtctcgtt ccgctGGGTCTCCAGGTAGCCCAGCTGGTAGCCCAACCTCTACCG AAGAAGGTACCTCTGAATCCGCTACTCCAGAATCCGGTCCTGG TACTAGCACTGAGCCAAGCGAAGGTTCTGCTCCAGGCTCCCCG GCAGGTAGCCCTACCTCTACCGAAGAGGGCACTAGCACCGAA CCATCTGAGGGTTCCGCTCCTGGCACCTCCACTGAACCGTCCG AAGGCAGTGCTCCGGGTACTTCCGAAAGCGCAACTCCGGAAT CCGGCCCTGGTTCTGAGCCTGCTACTTCCGGCTCTGAAACTCC AGGTAGCGAGCCAGCGACTTCTGGTTCTGAAACTCCAGGTTCA CCGGCGGGTAGCCCGACGAGCACGGAGGAAGGTACCTCTGAG TCGGCCACTCCTGAGTCCGGTCCGGGCACGAGCACCGAGCCG AGCGAGGGTTCAGCCCCGGGTACCAGCACGGAGCCGTCCGAG GGTAGCGCACCGGGTTCTCCGGCGGGCTCCCCTACGTCTACGG AAGAGGGTACGTCCACTGAACCTAGCGAGGGCAGCGCGCCAG	MKNPEQAE EQAEE QREETRPRP RPRP RPRPRPRPRP RPR PSASRSAGS PGSP AGSPTSTEE GTSE SATPESGPG TSTE PSEGSAPGSP AGS PTSTEEGTST EPS

Clone Name	DNA Sequence	Amino Acid Sequence
	GCACCAGCACTGAACCGAGCGAAGGCAGCGCACCTGGCACTA GCGAGTCTGCGACTCCGGAGAGCGGTCCGGGTACGAGCACGG AACCAAGCGAAGGCAGCGCCCCAGGTACCTCTGAATCTGCTA CCCCAGAATCTGGCCCGGGTTCCGAGCCAGCTACCTCTGGTTC TGAAACCCCAGGTACTTCCACTGAACCAAGCGAAGGTAGCGC TCCTGGCACTTCTACTGAACCATCCGAAGGTTCCGCTCCTGGT ACGTCTGAAAGCGCTACCCCTGAAAGCGGCCCAGGCACCTCTG AAAGCGCTACTCCTGAGAGCGGTCCAGGCTCTCCAGCAGGTTC TCCAACCTCCACTGAAGAAGGCACCTCTGAGTCTGCTACCCCT GAATCTGGTCCTGGCTCCGAACCTGCTACCTCTGGTTCCGAAA CTCCAGGTACCTCGGAATCTGCGACTCCGGAATCTGGCCCGGG CACGAGCACGGAGCCGTCTGAGGGTAGCGCACCAGGTACCAG CACTGAGCCTTCTGAGGGCTCTGCACCGGGTACCTCCACGGAA CCTTCGGAAGGTTCTGCGCCGGGTACCTCCACTGAGCCATCCG AGGGTTCAGCACCAGGTACTAGCACGGAACCGTCCGAGGGCT CTGCACCAGGTACGAGCACCGAACCGTCGGAGGGTAGCGCTC CAGGTAGCCCAGCGGGCTCTCCGACAAGCACCGAAGAAGGCA CCAGCACCGAGCCGTCCGAAGGTTCCGCACCAACGGCAGAGG CAGCAGGTTGTGGTACAGCAGAAGCAGCTCCGGGTAGCGAGC CTGCAACCAGCGGTTCTGAGACGCCGGGCACTTCCGAATCTGC GACCCCGGAGTCCGGTCCAGGTTCAGAGCCGGCGACGAGCGG TTCGGAAACGCCGGGTACTGCTGAAGCGGCTGGTTGTGGTACT GCTGAAGCTGCATCGACCGAACCAAGCGAAGGTTCGGCACCG GGTACTAGCGAGAGCGCAACCCCTGAAAGCGGTCCGGGCAGC CCGGCAGGTTCTCCAACCAGCACCGAAGAAGGTTCCCCTGCTA CTGCCGAAGCTGCAGGCTGCGGTACTGCGGAGGCGGCGTCCC CAACTTCTACTGAGGAAGGTACTTCTGAGTCCGCTACCCCAGA AAGCGGTCCTGGTACCTCCACTGAACCGTCTGAAGGCTCTGCA CCAGGCACTTCTGAGTCTGCTACTACCGCCGAAGCCGCTGGTT GTGGTACCGCAGAAGCTGCATCTGAGACTCCAGGCACTTCTGA GTCCGCAACGCCTGAATCCGGTCCTGGTTCTGAACCAGCTACT TCCGGCAGCGAAACCCCAGGTACCTCTGAGTCTGCGACTCCAG AGTCTGGTACCGCGGAAGCGGCTGGTTGTGGTACTGCAGAGG CAGCTGGTTCTCCGGCTGGTAGCCCGACCAGCACGGAGGAGG GTACGTCTGAATCTGCAACGCCGGAATCGGGCCCAGGTTCGGA GCCTGCAACGTCTGGCAGCGAAACCCCGGGTACCACGGCGGA AGCGGCAGGTTGTGGCACCGCGGAGGCAGCAGCTGGTTCTCC AACCTCTACCGAGGAGGGTTCACCGGCAGGTAGCCCGACTAG CACTGAAGAAGGTACTAGCACGGAGCCGAGCGAGGGTAGTGC TCCGGGTACGAGCGAGAGCACGGCAGAAGCCGCTGGCtgcGGT ACTGCTGAAGCGGCAACCCCTGAGAGCGGCCCAGGTACTTCTG AGAGCGCCACTCCTGAATCCGGCCCTGGTAGCGAGCCGGCAA CCTCCGGCTCAGAAACTCCTGGTTCGGAACCAGCGACCAGCGG TACCGCTGAAGCCGCAGGTtgtGGCACTGCGGAAGCTGCAACCG AAGAGGGTACCAGCACGGAACCGAGCGAGGGTTCTGCCCCGG GTACTTCCACCGAACCATCGGAGGGCTCTGCACCTGGTAGCGA ACCTGCGACGTCTGGTTCTGAAACGCCGACTGCAGAAGCGGCT GGTtgtGGCACCGCCGAAGCAGCTagcgcctctcgctccgcaCATCACCAT CACCACCATCATCACTAA	EGSAPGTST EPSE GSAPGTSES ATPE SGPGSEPAT SGSE TPGSEPATS GSET PGSPAGSPTS TEE GTSESATPES GPG TSTEPSEGSA PGT STEPSEGSAP GSP AGSPTSTEE GTST EPSEGSAPG TSTE PSEGSAPGT SESA TPESGPGTST EPS EGSAPGTSE SATP ESGPGSEPA TSGS ETPGTSTEPS EGS APGTSTEPSE GSA PGTSESATPE SGP GTSESATPES GPG SPAGSPTSTE EGT SESATPESGP GSE PATSGSETP GTSE SATPESGPG TSTE PSEGSAPGT STEP SEGSAPGTS TEPS EGSAPGTST EPSE GSAPGTSTE PSEG SAPGTSTEPS EGS APGSPAGSP TSTE EGTSTEPSE GSAP

Clone Name	DNA Sequence	Amino Acid Sequence
		TAEAAGCGT AEAA PGSEPATSG SETP GTSESATPES GPG SEPATSGSET PGT AEAAGCGT AEAAS TEPSEGSAP GTSE SATPESGPG SPAG SPTSTEEGSP ATA EAAGCGTAE AASP TSTEEGTSES ATP ESGPGTSTEP SEG SAPGTSESA TTAE AAGCGTAE AASET PGTSESATPE SGP GSEPATSGS ETPG TSESATPESG TAE AAGCGTAE AAGSP AGSPTSTEE GTSE SATPESGPG SEPA TSGSETPGT TAEA AGCGTAEA AAGSP TSTEEGSPA GSPT STEEGTSTEP SEG SAPGTSEST AEAA GCGTAEAAT PESG PGTSESATPE SGP GSEPATSGS ETPG SEPATSGTA EAAG CGTAEAATE EGTS

Clone Name	DNA Sequence	Amino Acid Sequence
		TEPSEGSAP GTST EPSEGSAPG SEPA TSGSETPTA EAAG CGTAEAASA SRSA HHHHHHHH

[00301] Example 11: Fermentation of XTEN for Conjugation

[00302] Starter cultures were prepared by inoculating glycerol stocks of *E. coli* carrying the plasmid containing the appropriate XTEN for conjugation protein sequences into 125 mL of LB Broth media containing 50 µg/mL kanamycin. The cultures were then shaken overnight at 37°C. The starter culture was used to inoculate 2L of fermentation batch media containing -12.5g ammonium sulfate, 15g potassium phosphate dibasic anhydrous, 2.5g sodium citrate dihydrate; 8.5g sodium phosphate monobasic monohydrate; 50g NZ BL4 soy peptone (Kerry Bioscience #5X00043); 25g yeast extract (Teknova #Y9020); 1.8L water; 0.5mL polypropylene glycol; 2.5mL trace elements solution (Amunix recipe 144-1); 17.5mL 1M magnesium sulfate; and 2mL Kanamycin (50mg/mL)- in 5L glass jacketed vessel with a B. Braun Biostat B controller. The fermentation control settings were: pH = 6.9 +/- 0.1; dO2 = 10%; dissolved oxygen cascade in stirrer only mode with a range of 125 – 1180 rpm; air flow of 5 liters per minute of 90% oxygen; initial temperature 37°C; base control 13% ammonium hydroxide; and no acid control. After 6 hours of culture a 50% glucose feed was initiated at a rate of 30g/hr. After 20 hours of culture, 25mL of 1M magnesium sulfate and 3mL of 1M IPTG were added. After a total fermentation run time of 45 hours the culture was harvested by centrifugation yielding cell pellets between 0.45-1.1 kilograms in wet weight for all constructs. The pellets were stored frozen at -80°C until further use. Culture samples at multiple time points in the fermentation were taken, the cells were lysed, then cell debris was flocculated with heat and rapid cooling, clarified soluble lysates were prepared by centrifugation and analyzed by a regular non-reducing SDS-PAGE using NuPAGE 4-12% Bis-Tris gel from Invitrogen according to manufacturer's specifications with Coomassie staining. An example of the accumulation of XTEN fusion protein as a function of fermentation run time is shown in FIG. 45. The results showed that the XTEN fusion protein constructs were expressed at fermentation scale with titers > 1g/L, with an apparent MW of about 160 kDa (note: the actual molecular weight are 100 kDa. The observed migration in SDS-PAGE was comparable to that observed for other XTEN-containing fusion proteins).

[00303] Example 12: Purification and Assessment of RP11/His8-XTEN Two-tag System

[00304] Using expression vectors as described above, two-tagged XTEN proteins were constructed to encode fusion proteins with the following amino acid sequences or components: MKIKTGARILALSALTTMMFSASALAAPTTAGAG-Tag-XTEN_AE869(Am1)-RHHHHHHHH, where: MKIKTGARILALSALTTMMFSASALA is a MalE recognition sequence that is cleaved from the polypeptide expressed and transported to the host cell periplasm, APTTAGAG is a spacer, and Tag is from the following (tag name followed by sequence in parentheses): RP5 (RPRPRPRPRPGR); RP7 (RPRPRPRPRPRPRPGR); KP5 (KPKPKPKPKPGR); RP9 (RPRPRPRPRPRPRPRPRPGR); RP11 (RPRPRPRPRPRPRPRPRPRPRPGR); P5K4P9 (RPRPKPRPKPRPKPRPKPGR); and R6K5P11 (RPRPKPRPKPRPKPRPKPRPKPGR). All variations of the constructs with tags were made and the proteins were expressed in *E.coli* using the methods as described in Example 3. Soluble extracts were prepared from the host cell for SDS-PAGE/Coomassie staining analysis. By analysis, the N-terminal tag length and amino acid composition did not affect protein expression noticeably (see FIG. 23) for expression of constructs with RP7, RP9, RP11 and R6K5P11 tags. Expressed proteins with the RP5, RP7, RP9 and RP11 tags proteins were further tested for binding to MacroCap SP resin. . Proteins with longer tags bound more efficiently to cation exchange resin, and remained bound under more stringent wash conditions. Thus RP11 tag was selected as N-terminal tag for purification of XTEN using cation exchange chromatography. An additional experiment showed that RP11-XTEN-His8 polypeptide was efficiently expressed in fermenter and most of the expressed protein was found in the cell pellet fraction (FIG. 24).

[00305] Example 13: Purification of 1xAmino-XTEN reagent with RP11/H8 two-tag system

[00306] 1. Expression

[00307] The RP11-XTEN-His8 precursor of 1xAmino-XTEN was produced by expression in transformed *E. coli* using a 4 L fermentation reaction as described. Cells were harvested by centrifugation and frozen at -80°C until use.

[00308] 2. Lysis and clarification

[00309] 25 g of cell paste was resuspended in 75 mL of 20 mM sodium phosphate pH 8.0, 50 mM NaCl, 2 mM EDTA. Lysis was performed by passing the resuspended paste through a homogenizer at 800-900 bar three times. Homogenate was held at 85°C in a water bath for 15 min before it was quickly cooled down using an ice/water bath until the temperature dropped to below 10°C. The treated homogenate was then centrifuged at 10, 000 rpm in SLA-3000 rotor for 60 minutes. Supernatant was collected and filtered using a 0.22 μm bottle top filter.

[00310] 3. Cation exchange capture step

[00311] Cation exchange chromatography was used as a capture step to ensure N-terminal integrity of the product. MacroCap SP resin (GE Healthcare) was selected among several cation exchange media due to its superior capacity and selectivity for the product. A 20 mL MacroCap SP column was packed in a Redi-Sep housing and equilibrated with 20 mM sodium phosphate (pH 8.0), 20 mM NaCl buffer. The lysate was loaded onto the column by gravity. Three column volumes (CV) of 20 mM sodium phosphate pH 8.0, 100 mM NaCl was applied as the wash step before protein was step-eluted with 3 CVs of 20 mM sodium phosphate pH 8.0, 500 mM NaCl. Half CV fractions (10 mL) were collected for elutions and analyzed by 4-12% Bis-Tris SDS/PAGE (FIG. 25). Elution fractions 2-4 were combined for the subsequent chromatography step.

[00312] 4. IMAC polishing step

[00313] A 20-mL ToyoPearl AF-Chelate column was packed in a Redi-Sep column housing and charged with 100 mM nickel sulfate. The column was equilibrated with 20 mM sodium phosphate pH 8.0, 500 mM NaCl before the MacroCap SP pool was loaded onto the column by gravity. Two wash steps were applied using 20 mM sodium phosphate (pH 8.0), 500 mM NaCl, 5 mM imidazole buffer followed by 20 mM sodium phosphate (pH 8.0), 5 mM imidazole. Product was then eluted from the column using 20 mM sodium phosphate, 100 mM imidazole and half CV (10 mL) fractions were collected for 4 CVs of elution. Samples from each step were examined by 4-12% Bis-Tris SDS/PAGE (FIG. 26). Based on the gel, elutions 2 and 3 were pooled for further processing. The overall yield was 30%.

[00314] 5. Trypsin digestion of IMAC elution pool

[00315] Trypsin digestion of the IMAC pool was performed at 1:200 and 1:500 m/m ratio by adding 1 mg/mL bovine trypsin (Sigma, Cat # T1426, Trypsin from Bovine Pancreas) to the IMAC pool. The reaction mixtures were held at 37°C overnight and the completion of digestion was confirmed by MALDI-TOF mass spectrometry. Pre- and post-digest samples were analyzed by 4-12% Bis-Tris SDS/PAGE stained by both Coomassie and silver stain (FIG. 27). Fainter staining on Coomassie-stained gel as well as the shifting of the molecule weight for the post-digestion samples, when compared to pre-digestion sample, indicates the successful removal of both N- and C-terminal tags. A silver stained gel showed homogeneous bands after digestion, suggesting the absence of truncated species in the sample, supporting the conclusion that the RP11/H8 two-tag system and purification methods provides a homogeneous final XTEN product.

[00316] **Example 14: Purification of XTEN with RP11 and His8 affinity tags**

[00317] 1. Expression

[00318] The fusion protein RP11-XTEN-His8, with two affinity tags linked to XTEN at the N- and C-terminus, respectively, was expressed in E.coli using a 4 L fermentation reaction using conditions described.

[00319] 2. Lysis and clarification

[00320] After growth, the cells were harvested by centrifugation and frozen at -80°C until use. The cell pellet was resuspended in lysis buffer (20 mM sodium phosphate, 50 mM NaCl, 2 mM EDTA pH 8.0, 3 ml buffer per gram cell paste). The cells were lysed by passing through an APV homogenizer three times at a pressure of 830-900 bar. Lysis buffer (1 ml buffer per gram cell paste) was used as a chase to retrieve hold up volume from the homogenizer. The homogenized lysate was incubated in a water bath at 85°C for 20 minutes, followed by quick cooling in ice water bath for 20 minutes. After the heating and cooling treatment, the lysate was centrifuged at 11000 RPM for 90 minutes in a SORVALL centrifuge. After centrifugation, the supernatant was filtered through two CUNO Bio cap 25 (BC0025L90SP08A) filters. The clarified supernatant was stored at 4°C overnight.

[00321] 3. Capture Step: Toyopearl IMAC Chromatography

[00322] IMAC affinity chromatography was used as a capture step for binding the XTEN with an intact C-terminal His-tag. Briefly, the chromatography column BPG140/12 (GE Life Sciences) was packed with 2000 ml Toyopearl IMAC 650 M resin (TOSOH Biosciences). The column was equilibrated with 2 column volumes (CVs) of equilibrium buffer (20 mM sodium phosphate, 500 mM NaCl, pH 8.0). Clarified cell lysate was adjusted to a final NaCl concentration of 500 mM using 5 M NaCl stock solution, and then was loaded onto the IMAC resin. The column was washed with 2 column volumes of equilibrium buffer, and then 2 column volumes of 20 mM sodium phosphate, 500 mM NaCl, 5 mM Imidazole pH 8.0, followed by 2 column volumes of 20 mM sodium phosphate, 5 mM Imidazole pH 8.0 to remove salt. Elution was performed with 2 column volumes of 20 mM sodium phosphate, 100 mM Imidazole, pH 8.0. The flow through, wash and elution fractions were analyzed by non-reducing 4 -12% Bis-Tris SDS-PAGE/Coomassie staining and the fractions with the desired product were pooled.

[00323] 4. Polishing/Capture Step: MacroCap SP Chromatography

[00324] Cation exchange chromatography was used as a polishing step to ensure the N-terminal integrity of the product. MacroCap SP resin (GE Life Sciences) was selected among several cation exchange media due to its superior capacity and selectivity for the product. 1000 ml of MacroCap SP resin was packed in a BPG100/13 (GE Life Sciences) chromatography column and equilibrated with 20 mM sodium phosphate pH 8.0, 20 mM NaCl. The IMAC pool was loaded onto the column and the resin was washed with 2 column volumes of 20 mM sodium phosphate, 50 mM NaCl, pH 8.0 and 2 column volumes of 20 mM sodium phosphate pH 8.0, 150 mM NaCl. The protein was eluted with 5 column volumes of linear gradient from 150 to 500 mM NaCl in 20 mM sodium phosphate pH 8.0. Fractions were collected and analyzed by 4-12% Bis-Tris SDS/PAGE. Fractions the with desired product were combined for the next step.

[00325] 5. Trypsin digestion of Macrocap SP elution pool

[00326] Trypsin (Sigma, Trypsin from Bovine Pancreas) digestion of the SP elution pool was performed at 1:200 m/m enzyme/protein ratio overnight at 37°C.

[00327] 6. Polishing Step: Macrocap Q Chromatography

[00328] After trypsin digestion, the cleaved tags were separated from the final product using Macrocap Q chromatography. The BPG100/19 column (GE Life Sciences) was packed with 1500 ml column volume of Macrocap Q resin (GE Life Sciences). The trypsin digested Macrocap SP elution pool was incubated for 15 min at 80o C with 20 mM DTT and 2 mM EDTA to reduce disulfide bonds and to inactivate trypsin. The cooled protein solution was diluted to a conductivity below 5 mS/cm with Milli-Q water and loaded onto the Macrocap Q column equilibrated with 20 mM HEPES, 50 mM NaCl, pH 7.0. The column was washed with 2 column volumes of 20 mM HEPES, 50 mM NaCl, pH 7.0, then 2 column volumes of 20 mM HEPES, 2 mM TCEP, 150mM NaCl pH7.0. The protein was eluted with a linear gradient from 150mM NaCl to 500 mM NaCl in 20 mM HEPES, pH 7.0 in 20 column volumes. Fractions were analyzed by SDS-PAGE/silver staining.

[00329] 7. Concentration and Diafiltration (Final Formulation)

[00330] Selected MacroCap Q fractions were combined and concentrated and using 10 KD Pellicon mini (Millipore) at a feed pressure < 20 psi and retentate < 8 psi, followed by 10X diafiltratiion with 20 mM HEPES, 50 mM NaCl, pH 7.0 to achieve a final protein concentration of > 5 mg/ml.

[00331] 8. Purity analysis of proteins purified with different methods

[00332] One batch (Batch 1) was purified through three purification steps as described above. Another batch (Batch 2) was purified from the same fermented material but the MacroCap SP polishing step was omitted. Truncated species of XTEN were detected by SDS-PAGE/silver staining in MacroCap Q elution fractions for Batch 2 (FIG. 40A), while the MacroCap Q elution fractions for Batch 1 were essentially free from truncations (FIG. 40B). These results support that, under the conditions employed, the MacroCap SP step based on the RP11 tag is essential to ensure N-terminal integrity and overall product quality and that the RP11/H8 two-tag system and purification methods provides a homogeneous final XTEN product.

[00333] Example 15: Fermentation and Purification of Cysteine-engineered XTEN for Conjugation

[00334] *E. coli* containing AC292 on a plasmid was grown to saturation overnight in 2xYT and then 200 ml of this culture was used to inoculate a 25L culture of 2xYT media in a wavebag. Both cultures were in the presence of 50 µg/ml kanamycin. The second culture was grown to an OD600 of ~1.0 at 37°C, chilled to 26°C, and induced with 12 ml of 1M IPTG overnight. The cell pellet was harvested at 4000 rpm in a SLA-3000 rotor spinning for 20 minutes. The cell pellet (184g) was resuspended in 736 ml of 20 mM Tris pH 6.8, 50 mM NaCl. The resuspended cells were

lysed with a microfluidizer at 20,000 psi and then heated to 75°C for 15 minutes, followed by rapid cooling on ice for 30 minutes. The lysate was then clarified by centrifugation. The clarified lysate was then loaded on to a DE52 column, previously sanitized with NaOH and equilibrated with 20 mM Tris pH 6.8, 50 mM NaCl. The column was washed with 5 column volumes of 20 mM Tris pH 6.8, 50 mM NaCl, 5 column volumes of 20 mM Tris pH 6.8, 150 mM NaCl and eluted with 5 column volumes of 20 mM Tris pH 6.8, 250 mM NaCl. The pooled elution fractions. were then loaded on to a macrocapQ column, previously sanitized with NaOH and equilibrated with 20 mM Tris pH 6.8, 50 mM NaCl. The column was washed with 9 column volumes of 20 mM Tris pH 6.8, 50 mM NaCl, 9 column volumes of 20 mM Tris pH 6.8, 100 mM NaCl and eluted with 9 column volumes of 20 mM Tris pH 6.8, 250 mM NaCl. The pooled elution fractions were adjusted to a 15% w/v sodium sulfate and then loaded on to a octyl sepharose FF column, previously sanitized with NaOH and equilibrated with Tris pH 7.5. The column was washed with 4 column volumes of 20 mM Tris pH 7.5 15% w/v sodium sulfate, and eluted with 4 column volumes of 20 mM Tris pH 7.5, 5% w/v sodium sulfate. The sample was stored at 4°C and given the lot # AP197. The purified cysteine-engineered XTEN could then serve as a suitable reactant for conjugation with a payload, such as a drug, resulting in an XTEN-drug conjugate.

[00335] Example 16: Preparation of A Bispecific Conjugate from Monospecific XTEN Precursors Linked by the N-termini.

[00336] The example describes the creation of an XTEN-payload composition by linking two different XTEN-payload precursors in an N- to N-terminus configuration; one with a payload A and one with a payload B, resulting in a bispecific conjugate.

[00337] As a first step, XTEN molecules containing multiple cysteines (cysteine-engineered XTEN) are prepared using a RP11-His8 two-tag purification system, described above, and are formulated in 20 mM HEPES, pH 7.0, 50 mM NaCl. A Payload A-maleimide is dissolved in aqueous solution 20 mM HEPES, pH 7.0, DMF or DMCO or any other appropriate solvent depending on reagent solubility. The Payload A-maleimide is added to the cysteine-engineered XTEN in a 2-6 molar excess over XTEN and incubated for 1 hr at 25°C. Completion of modification is monitored by C18 RP-HPLC. The resulting Payload A-XTEN conjugate is purified from contaminants and unreacted components using preparative C4-C18 RP-HPLC. The Payload A-XTEN conjugate is formulated in 20 mM HEPES, pH 7.0, 50 mM NaCl. Next, the Payload A-XTEN conjugate is further modified by adding dibenzylcyclooctyne (DBCO)-NHS ester or DBCO-sulfo-NHS ester in a 10-50 molar excess to the XTEN and incubating 1-2 hrs at 25°C. Completion of the modification is monitored by analytical C18 RP-HPLC. If the conjugation efficiency is low (for example, <90%) or multiple unspecific products are formed, the DBCO-Payload A-XTEN conjugate is purified using preparative C4-C18 RP-HPLC. If the

efficiency of DBCO-NHS ester conjugation is high (>90%) with no significant side products, the DBCO-Payload A-XTEN conjugate is purified from excess reagent by buffer exchange using a 10-30 kDa MWCO centrifugal device, acetonitrile precipitation or anion exchange chromatography.

[00338] To create the second XTEN-payload precursor, a Payload B-maleimide is dissolved in aqueous solution 20 mM HEPES, pH 7.0, DMF or DMCO or any other appropriate solvent depending on reagent solubility. Payload B-maleimide is added to the second cysteine-containing XTEN in 2-6 molar excess over XTEN concentration and incubated for 1 hr at 25°C. Completion of modification is monitored by analytical C18 RP-HPLC. The resulting Payload B-XTEN conjugate is purified from contaminants and reactants using preparative C4-C18 RP-HPLC. The Payload B-XTEN conjugate is formulated in 20 mM HEPES, pH 7.0, 50 mM NaCl. Azide-PEG4-NHS ester is added in 10-50 molar excess to the Payload B-XTEN and incubated 1-2 hrs at 25°C. Completion of modification is monitored by C18 RP-HPLC. If the conjugation efficiency is low (for example <90%) or multiple unspecific products are formed, the azide-Payload B-XTEN conjugate is purified using preparative C4-C18 RP-HPLC. If the efficiency of DBCO-NHS ester conjugation is high (>90%) with no significant side products, the azide-Payload B-XTEN conjugate is purified from excess reagent by buffer exchange using a 10-30 kDa MWCO centrifugal device, acetonitrile precipitation or anion exchange chromatography. The final product is created by mixing purified and concentrated DBCO-Payload A-XTEN and azide-Payload B-XTEN proteins in an equilmolar ratio in 20 mM HEPES pH 7.0 buffer, 50 mM NaCl and incubated at 25°C for 1 hr or longer until the reaction is complete. Completion of modification is monitored by C4 or C18 RP-HPLC. If necessary, the bispecific conjugate Payload A-XTEN-Payload B is purified by preparative RP-HPLC, hydrophobic interaction chromatography or anion exchange chromatography.

[00339] Example 17: Preparation of 1xAzide,3xMMAE-XTEN by conjugation

[00340] An aliquot of the fusion protein 1xAmino,3xThiol-XTEN432 (XTEN_AE432(Am1,C12,C217,C422)) was prepared as 196 µM (7.7 mg/ml) solution in 20 mM HEPES, pH 7.0, 50 mM NaCl. FIG. 41A shows a C18 RP-HPLC and ESI-MS analysis of the protein. MA6-Val-Cit-PAB-MMAE (MMAE-Mal, custom synthesis by Concortis Biosystems, Inc.) was dissolved in dimethylsulfoxide (DMSO) to a final concentration 1 mg/ml. A 2.2 ml volume of 1xAmino,3xThiol-XTEN432 solution was mixed with 2.5 ml of MMAE-Mal (3.5x molar excess over protein). The reaction mixture was incubated for 1 hr at 25°C and the products of the reaction were analyzed by C18 RP-HPLC (FIG. 41B). Azide-PEG4-NHS ester (Click Chemistry Tools, Inc., cat. # A103) was dissolved in anhydrous DMF to a final concentration of 500 mM. The reaction mixture (4.7 ml) was mixed with 0.235 ml 1 M HEPES, pH 8.0 and with 9.75ul 500 mM azide-PEG4-NHS (10x molar excess over protein). The reaction mixture was

incubated for 2 hr at 25°C and the products of reaction were analyzed by C18 RP-HPLC (FIG. 41C). The conjugation mixture was diluted to 15 ml with 0.01% TFA and the pH was adjusted to ~3 using 10% TFA. The protein solution was loaded onto a preparative C4 RP-HPLC column Vydac C4 250 x 22 mm (Grace Davison Discovery Sciences, cat. # 214TP1022). The protein was eluted with 1200 ml linear 5-50% gradient of acetonitrile in 0.01% TFA at 15 ml/min flow rate. Fractions containing 1xAzide,3xMMAE-XTEN432 were adjusted to ~pH7 with 1 M HEPES pH 8 and concentrated by vacuum evaporation to yield the final product.

[00341] Example 18: Preparation of 1xDBCO,3xFolate(γ)-XTEN by conjugation

[00342] An aliquot of the fusion protein 1xAmino,3xThiol-XTEN432 (XTEN_AE432(Am1,C12,C217,C422)) was prepared as a 203 μM (8.0 mg/ml) solution in 20 mM HEPES, pH 7.0, 50 mM NaCl. FIG. 45A shows a C18 RP-HPLC and ESI-MS analysis of the protein. Folate-γ-aminopentyl-maleimide (FA(γ)-Mal, custom synthesis by CPC Scientific) was dissolved in dimethylformamide (DMF) to a final concentration 21 mg/ml (9.8 mM). A 1.1 ml volume of 1xAmino,3xThiol-XTEN432 solution was mixed with 0.08 ml of FA(γ)-Mal (3.5 molar excess over protein). The reaction mixture was incubated for 2 hrs at 25°C and the products of the reaction were analyzed by C18 RP-HPLC (FIG. 45B). pH of the reaction mixture was adjusted with 0.06 ml of 1 M HEPES pH 8.0 buffer. DBCO-sulfo-NHS (Click Chemistry Tools, Inc., cat. # A124) was dissolved in anhydrous DMF to a final concentration of 50 mM. 0. 53 ml of DBCO-sulfo-NHS solution was added to protein solution (11x molar excess over protein). The reaction mixture was incubated for 2 hours at 25°C and products of reaction were analyzed by C18 RP-HPLC (FIG. 45C). A solution of 1M ethanolamine pH 8.0 was added to a final concentration of 50 mM to quench the unreacted DBCO-sulfo-NHS. The reaction mixture was incubated for 2 hours at 25°C and then overnight at 4°C. The reaction mixture was diluted to 15mL with 0.01% TFA and pH adjusted to ~3 using 10% TFA solution. The protein solution was loaded on a preparative C4 RP-HPLC column Vydac C4 250 x 10 mm (Grace Davison Discovery Sciences, cat. # 214TP510). The protein was eluted with 180 ml linear 5-50% gradient of acetonitrile in 0.01% TFA at 2 ml/min flow rate. Fractions containing 1xDBCO,3xFA(γ)-XTEN432 were adjusted to pH ~ 7 with 1 M HEPES pH 8 and were concentrated by vacuum evaporation to yield the final product.

[00343] The 1xDBCO,3xFolate(α)-XTEN conjugate was prepared essentially as described above using folate-alpha-Maleimide (FA(α)-Mal, custom synthesis by CPC Scientific).

[00344] Example 19: Preparation of 1xDBCO,3xFolate(γ)-XTEN by conjugation

[00345] An aliquot of the fusion protein 1xAmino,3xThiol-XTEN432 (XTEN_AE432(Am1,C12,C217,C422)) was prepared as a 203 μM (8.0 mg/ml) solution in 20 mM HEPES, pH 7.0, 50 mM NaCl. FIG. 45A shows a C18 RP-HPLC and ESI-MS analysis of the protein. Folate-γ-aminopentyl-maleimide (FA(γ)-Mal, custom synthesis by CPC Scientific) was

dissolved in dimethylformamide (DMF) to a final concentration 21 mg/ml (9.8 mM). A 1.1 ml volume of 1xAmino,3xThiol-XTEN432 solution was mixed with 0.08 ml of FA(γ)-Mal (3.5 molar excess over protein). The reaction mixture was incubated for 2 hrs at 25°C and the products of the reaction were analyzed by C18 RP-HPLC (FIG. 45B). pH of the reaction mixture was adjusted with 0.06 ml of 1 M HEPES pH 8.0 buffer. DBCO-sulfo-NHS (Click Chemistry Tools, Inc., cat. # A124) was dissolved in anhydrous DMF to a final concentration of 50 mM. 0. 53 ml of DBCO-sulfo-NHS solution was added to protein solution (11x molar excess over protein). The reaction mixture was incubated for 2 hours at 25°C and products of reaction were analyzed by C18 RP-HPLC (FIG. 45C). A solution of 1M ethanolamine pH 8.0 was added to a final concentration of 50 mM to quench the unreacted DBCO-sulfo-NHS. The reaction mixture was incubated for 2 hours at 25°C and then overnight at 4°C. The reaction mixture was diluted to 15mL with 0.01% TFA and pH adjusted to ~3 using 10% TFA solution. The protein solution was loaded on a preparative C4 RP-HPLC column Vydac C4 250 x 10 mm (Grace Davison Discovery Sciences, cat. # 214TP510). The protein was eluted with 180 ml linear 5-50% gradient of acetonitrile in 0.01% TFA at 2 ml/min flow rate. Fractions containing 1xDBCO,3xFA(γ)-XTEN432 were adjusted to pH ~ 7 with 1 M HEPES pH 8 and were concentrated by vacuum evaporation to yield the final product.

[00346] The 1xDBCO,3xFolate(α)-XTEN conjugate was prepared essentially as described above using folate-alpha-Maleimide (FA(α)-Mal, custom synthesis by CPC Scientific).

[00347] Example 20: Preparation of 3xFA(γ),3xMMAE-XTEN by conjugation

[00348] An aliquot of the fusion protein 1xDBCO,3xFA(γ)-XTEN432 was prepared as 191 μM (8.8 mg/ml) solution in 20 mM HEPES, pH 7.0. 1xAzide,3xMMAE-XTEN432 (prepared as described in Example 17) was prepared as a 242 μM (11.1 mg/ml) solution in 20 mM HEPES, pH 7.0. The two protein reactants were mixed in solution to yield a 1.1 molar excess of 1xDBCO,3xFA(γ)-XTEN432. The reaction mixture was incubated overnight at 25°C. Completion of the click chemistry reaction was analyzed by RP-HPLC (FIG. 108). The conjugation mixture was diluted to 15 ml with 0.01% TFA and pH adjusted to ~3 using 10% TFA. The protein solution was loaded onto a preparative C4 RP-HPLC column Vydac C4 250 x 10 mm (Grace Davison Discovery Sciences, cat. # 214TP510). The protein was eluted with 180 ml linear 5-50% gradient of acetonitrile in 0.01% TFA at 2 ml/min flow rate. Fractions containing 3xFA(γ),3xMMAE-XTEN were adjusted to ~pH7 with 1 M HEPES pH 8 and concentrated by vacuum evaporation to yield the final product. The final product was analyzed by size exclusion chromatography (SEC-HPLC) (FIG. 109A), RP-HPLC (FIG. 109B) and ESI-MS (FIG. 109C).

[00349] The 3xFA(α),3xMMAE-XTEN conjugate was prepared essentially as described above using click reaction between 1xDBCO,3xFA(α)-XTEN432 and 1xAzide,3xMMAE-XTEN432. The structure is shown in FIG. 49.

[00350] Example 21: *In vitro* cell-based screening of folate-XTEN-drug conjugates for activity and specificity

[00351] Folate-XTEN-drug conjugates are first subjected to an *in vitro* activity and selectivity screen. Each folate-XTEN-drug conjugate, its corresponding non-targeting XTEN-drug molecule and respective free drug control are tested in a CellTiter-Glo anti-proliferation assay against a panel of folate receptor positive and negative cell lines listed in Table 24. As culture media contain high folic acid content, cells will be grown and the assay performed in folic acid free-media containing 5-10% heat-inactivated fetal calf serum (FCS) at 37°C, in an atmosphere of 5% CO_2 (heat-inactivated FCS contains endogenous level of folic acid sufficient for folate receptor expressing cells to survive and proliferate). Appropriate assay conditions are established, including optimal cell density and incubation times, using folate-free media containing 5-10% FCS using the respective free drug as control. Folate-XTEN-drug conjugates are then tested as follows: cells in log-phase are collected, counted and plated at pre-determined cell density in 96-well microtiter assay plates. Adherent cells are allowed to attach to the plate by an overnight incubation at 37°C, 5% CO_2. Folate-XTEN-drug conjugates and corresponding controls are introduced using appropriate dose range dilutions, in duplicate, and the plates are incubated for an additional 2-5 days. After the appropriate incubation period, CellTiter-Glo reagent is added to each well and is mixed for 2 minutes on an orbital shaker. The plate is then centrifuged at 90 x g and incubated at room temperature for an additional 10 minutes to stabilize the luminescent signal. Luminescence signals are then read on a luminometer and the IC_{50} (half maximal inhibitory concentration) values are calculated with GraphPad Prism or equivalent software. Quantitative comparisons of the IC_{50} values will enable ranking of the constructs' activity for inhibition of cell growth and selectivity against folate receptor positive versus negative cell lines. It is expected that the results would support the finding that the folate-XTEN-drug conjugates will show highly selective potent killing on folate receptor positive cells but not on folate receptor negative cells. This will be in contrast to the free drug moiety where no discrimination in cytotoxicity is expected between folate receptor positive and negative cell lines. The XTEN-drug control is expected to yield poor cytotoxic activity. Folate-XTEN-drug conjugates with favorable activity and cell line selectivity relative to controls will be further verified for folate receptor association by the addition of free competitive folic acid in the assay, demonstrating impaired folate-XTEN-drug cytotoxicity, further verifying the selective activity of the constructs.

Table 24: Folate receptor positive and negative cell lines

Cell line	Tissue	Folate receptor status
KB	Nasopharyngeal	Positive
IGROV	Ovarian	Positive
SK-OV-3	Ovarian	Positive

Cell line	Tissue	Folate receptor status
HeLa	Cervical	Positive
LoVo	Colorectal	Positive
SW620	Colorectal	Positive
A549	Lung	Negative
A375	Multiple melanoma	Negative
LS-174T	Colorectal	Negative
SK-BR-3	Breast	Negative

[00352] Example 22: *In vitro* serum stability of folate-XTEN-drug conjugates

[00353] As a measure of stability, folate-XTEN-drug conjugates are incubated independently in normal human, cynomolgus monkey and mouse plasma at 37°C for up to 2 weeks with aliquots removed at periodic intervals and stored at -80°C until analysis. The stability of folate-XTEN-drug conjugate is assessed either by the amount of free drug or the integrity of the folate-XTEN-drug conjugate over time. Free drug is quantitated with HPLC and/or LC-MS/MS whereas the amount of intact folate-XTEN-drug conjugate is determined using an XTEN/drug and/or folate/drug ELISA.

[00354] For RP- HPLC analysis, plasma samples are treated with organic solvents such as acetonitrile or acetone to precipitate proteins. Soluble fractions are evaporated under vacuum, redissolved in loading solutions and analyzed by RP-HPLC. Analytes are detected by UV absorption at wavelength specific for a particular drug, compared to known drug standards. For example, doxorubicin is detected at 480 nm. For LC-MS/MS analysis, plasma samples will be treated with organic solvents such as acetonitrile or acetone to precipitate proteins. Soluble fractions will be evaporated in vacuum, redissolved in loading solutions and analyzed by RP-HPLC. Analytes will be in-line detected and quantitated by triple quadrupole tandem mass spectrometry. Parental ion–daughter ion pairs will be determined experimentally for each drug. Calibration standards will be prepared by adding known amounts of free drug to corresponding plasma type and will be treated in parallel with experimental samples. For quantitative ELISA, optimal concentrations of antibodies for folate-XTEN-drug conjugate in the ELISAs is determined using criss-cross serial dilution analysis. An appropriate capture antibody recognizing one component of the conjugate is coated onto a 96-well microtiter plate by an overnight incubation at 4°C. The wells are blocked, washed and serum stability samples added to the wells, each at varying dilutions to allow optimal capture of the folate-XTEN-drug conjugate by the coated antibody. After washing, detection antibody recognizing another component of the conjugate is added and allowed to bind to the conjugate captured on the plate. Wells are then washed again and either streptavidin-horseradish peroxidase (complementary to biotinylated version of detection antibody) or an appropriate secondary antibody-horseradish peroxidase (complementary to non-biotifnylated version of detection antibody) is then added. After appropriate incubation

and a final wash step, tetramethylbenzidine (TMB) substrate is added and the plate is read at 450nM. Concentrations of intact conjugate are then calculated for each time point by comparing the colorimetric response to a calibration curve prepared with folate-XTEN-drug in the relevant plasma type. The *t*1/2 of the decay of the conjugate in human, cyno and mouse serum is then defined using linear regression analysis of the log concentrations vs. time.

[00355] Example 23: *In Vivo* and *ex vivo* imaging of folate-XTEN-Cy5.5 conjugates

[00356] A Cy5.5 fluorescent tagged folate-XTEN molecule is used as a surrogate to investigate the targeting and biodistribution efficiency of folate-XTEN-drug conjugates. Experiments will be carried out in nude mice bearing subcutaneous grown xenografts of folate receptor positive tumor cells using *in vivo*, followed by *ex vivo*, fluorescence imaging with IVIS 50 optical imaging system (Caliper Life Sciences, Hopkinton, MA). As culture media contain high folate content, folate receptor positive tumor cells to be transplanted onto these mice will be grown in folate-free cell culture media containing 5-10% heat-inactivated FCS with no antibiotics. Similarly, normal rodent chow contains a high concentration of folic acid; nude mice used in this study will be maintained on folate-free diet 2 weeks prior to tumor implantation and for the duration of the imaging analysis to reduce serum folate concentration.

[00357] In brief, female nu/nu mice bearing folate receptor positive tumor cells are given a single intravenous injection of high or low dose folate-XTEN-Cy5.5 and corresponding doses of non-targeting Cy5.5 tagged XTEN control. Whole body scans are acquired pre-injection and then at approximately 8, 24, 48 and 72 hours post-injection on live anesthetized animals using the IVIS 50 optical imaging system. After measuring the distribution of fluorescence in the entire animal at the last time point of 72h, tumor and healthy organs including liver, lung, heart, spleen and kidneys are excised and their fluorescence registered and processed by the imaging system. Cy5.5 excitation (615-665nm) and emission (695-770nm) filters are selected to match the fluorescence agents' wavelengths. Small and medium binning of the CCD chip is used and the exposure time optimized to obtain at least several thousand counts from the signals that were observable in each mouse in the image and to avoid saturation of the CCD chip. To normalize images for quantification, a background fluorescence image is acquired using background excitation and emission filters for the Cy5.5 spectral region. The intensity of fluorescence is expressed as different colors with blue color reflecting the lowest intensity and red being indicative of the highest intensity, and the resulting images are used to assess the uptake of the conjugates and controls.

[00358] Example 24: Pharmacokinetic analysis of folate-XTEN-drug conjugates

[00359] The *in vivo* pharmacokinetics of folate-XTEN-drug constructs are assessed using standard methods for protein compositions using mice, rats, cynomolgus monkeys, and dogs. As normal feed contains a high concentration of folic acid (approx. 6mg/kg mouse chow), animals to

be used in pharmacokinetic studies of folate conjugates will be maintained on folate-free diet for 2 weeks prior to study initiation and for the duration of the study. The compositions are administered at appropriate doses and via multiple routes: most preferably via intravenous or subcutaneous routes. Blood samples are collected at appropriate time points ranging from 0.08 to 504 hours, and processed into plasma. Plasma samples are analyzed for concentration of folate-XTEN-drug conjugates by a variety of methods including ELISA, HPLC and/or LC-MS/MS.

[00360] ELISA analysis are performed using a sandwich ELISA format that can recognize 2 components of the folate-XTEN-drug conjugate, for instance, XTEN/folate, XTEN/drug moiety, folate/drug moiety and/or XTEN/XTEN combinations. Typically antibody recognizing one component of the folate-XTEN-drug conjugate is coated onto wells of a 96-well microtiter plate. The wells are blocked, washed and plasma samples are then added to the wells at varying dilutions to allow capture of the conjugate by the coated antibody. Wells are then washed extensively, and bound protein detected using either a biotinylated antibody or an appropriate secondary antibody against the second folate-XTEN-drug conjugate component. Wells are then washed again and streptavidin-horseradish peroxidase (complementary to the biotinylated detection antibody) or a secondary antibody-horseradish peroxidase (complementary to a non biotinylated detection antibody) is then added. After appropriate incubation and a final wash step, tetramethylbenzidine (TMB) substrate is added and the plate is read at 450 nM. Concentrations of conjugate are then calculated for each time point by comparing the colorimetric response to a folate-XTEN-drug calibration curve. Pharmacokinetic parameters are calculated using the WinNonLin software package.

[00361] For RP-HPLC analysis, plasma samples are treated with organic solvents such as acetonitrile or acetone to precipitate proteins. Soluble fractions are evaporated in vacuum, redissolved in loading solutions and analyzed by RP-HPLC. Analytes are detected by UV absorption at wavelength specific for a particular drug. For example, doxorubicin is detected at 480 nm. Calibration standards are prepared by adding known amounts of free drug to corresponding plasma type and are assayed in parallel with experimental samples.

[00362] For LC-MS/MS analysis, plasma samples are treated with organic solvents such as acetonitrile or acetone to precipitate proteins. Soluble fractions are evaporated under vacuum, redissolved in loading solutions and analyzed by RP-HPLC. Analytes are in-line detected and quantitated by triple quadrupole tandem mass spectrometry. Parental ion–daughter ion pairs are determined experimentally for each drug. Calibration standards are prepared by adding known amounts of free drug to corresponding plasma type and are assayed in parallel with experimental samples.

[00363] It is expected that the results would support the finding that addition of an XTEN to folate and drug moiety will greatly increase the terminal half-life and enhance the pharmacokinetic properties of targeting and drug moiety not linked to XTEN.

[00364] Example 25: In vivo efficacy and toxicity analysis of folate-XTEN-drug conjugates

[00365] Folate-XTEN-drug conjugates are intended for targeted delivery of the toxin component to folate receptor positive tumor cells. The *in vivo* pharmacologic activity of folate-XTEN-drug constructs is assessed using human tumor cell expressing folate receptor xenograft onto nude mice. Prior to beginning the efficacy study, an initial assessment in nude mice is carried out to establish the maximum tolerated dose (MTD) of the folate-XTEN-drug candidates. The MTD, the highest dose that will be tolerated by the animal for the study duration, is then used to calculate the dose range for the efficacy and toxicity study in the standard xenograft model. As normal rodent chow contains a high concentration of folic acid (6mg/kg chow), mice to be used in these studies are maintained on a folate-free diet for 2 weeks prior to study initiation and for the duration of the study. The MTD experiment is carried out with 5 mice per group evaluating the intravenous administration of folate-XTEN-drug conjugates at various dose level, interval and duration. The starting MTD dose and number of dose groups required is based on scientific literature, knowledge of the targeting folate moiety, the nature of the drug moiety conjugated, toxicological properties of closely related compounds, and data from the initial pharmacokinetic studies (see PK Example above). Standard MTD parameters such as reduction in body weight, food and water consumption and signs of piloerection, hunched, behavior patterns, respiratory pattern, tremors, convulsions, prostration and self-mutilation are carefully monitored on a daily basis. The highest dose of folate-XTEN-drug that does not cause unacceptable toxicity is assigned as the MTD.

[00366] The tumor xenograft study includes 3 to 4 dosing levels of folate-XTEN-drug, depending on the results from the MTD study, with other parameters depending on the tumor cell line chosen. Table 24 describes examples of tumor lines that can be used in the xenograft study. To reduce folate content, folate receptor positive tumor cells to be transplanted onto nude mice are grown in folate-free cell culture media containing 5-10% heat-inactivated fetal calf serum with no antibiotics. Similarly, to reduce serum folate concentration, mice used in the xenograft studies are maintained on folate-free diet 2 weeks prior to tumor implantation and for the duration of the study. An appropriate number of folate receptor positive cells from the relevant line are injected subcutaneously and allowed to form tumors, the size of which are measured with calipers and the volume calculated as $0.5 \times L \times W^2$, where L = measurement of longest axis in millimeters and W = measurement of axis perpendicular to L in millimeters. Following randomization of mice containing tumor volume in the desired size range into groups of 8-10 animals, vehicle control,

free drug control and folate-XTEN-drug is administered intravenously at the chosen doses and intervals. Cessation or regression of tumor growth is determined through measuring the tumor size and volume at selected time points with calipers. Body weight and food consumption is measured every 1 to 2 days to assess gross toxicity. Survival of animals is monitored daily. At the end of study, all animals are sacrificed and major organs will be removed for clinical pathology and histopathology examination.

[00367] It is anticipated that the targeted chemotherapeutic folate-XTEN-drug conjugate will be more effective and less toxic than free cytotoxic drug alone on folate receptor positive tumors in the mouse model.

[00368] Example 26: Selective cytotoxicity of 3xFA(γ),3xMMAE-XTEN on KB cells

[00369] The ability to selectively target and kill cells bearing folate receptors was evaluated. Test articles of free MMAE, a non-targeting 3xMMAE-XTEN conjugate (XTEN linked to toxin) and the folate receptor-targeted 3xFA(γ),3xMMAE-XTEN conjugate were evaluated in a CellTiter-Glo anti-proliferation assay using the folate receptor-positive KB cell line. As culture media contain high folic acid content, KB cells were grown in folic acid-free media containing 10% heat-inactivated fetal calf serum at 37°C, 5% CO2 for at least 7 days prior to the commencement of the cell viability experiment, This medium was also utilized for the execution of the experiment. In brief, KB cells were plated at 10,000 cells per well onto a 96-well microtiter assay plate. KB cells were allowed to adhere to the plate by an overnight incubation at 37°C, 5% CO2. The spent media was then removed and wells designated to contain folic acid competitor received assay medium containing folic acid, while wells not designated to have folic acid competitor received assay medium only. The plate was incubated for 30 min at 37°C, 5% CO2 before the assay media was aspirated and plate washed with assay media. Free MMAE, 3xMMAE-XTEN and 3xFA(γ),3xMMAE-XTEN in the presence or absence of folic acid competitor was then added at an appropriate range of doses. The plate was then further incubated for 2- 4 h at 37°C, 5% CO2. Media was then removed, the plate washed and fresh media introduced and the plate was allowed to incubate for an additional 48-72h. After the appropriate incubation period, CellTiter-Glo reagent was added and the plate was read on a luminometer. The IC50 of each test article was determined using a 4 parameter logistic curve fit using GraphPad Prism.

[00370] Results: As shown in FIG. 48, free MMAE drug moiety shows highly potent killing of KB cells, with an IC50 of 0.8nM, while 3 copies of MMAE conjugated to non-targeting XTEN resulted in at least a 3 log reduction in cell killing (IC50 >1,000nM). Significantly, the addition of 3 copies of folate targeting domains to the 3xMMAE-XTEN conjugate restored the cell killing, with an IC 50 of 4.2 nM; a level of activity close to that observed for free MMAE. Of equal importance, the introduction of folic acid as a competitor to the targeted conjugate impaired the

observed cell killing activity of 3xFA(γ),3xMMAE-XTEN on the KB cell line. This reduction from potent cell killing of the folate-XTEN drug conjugate (from 4.2 nM to >1,000nM) supports the conclusion that the detected cell toxicity was, under the experimental conditions, facilitated by use of the folate as the targeting mechanism for the drug conjugate against the KB cell line.

[00371] Example 27: Clinical applications of folate-XTEN-drug conjugates

[00372] Targeted chemotherapy is a modern approach aimed at increasing the efficacy of systemic chemotherapy and reducing its side effects. Brentuximab vedotin (Adcentris), approved for Hodgkin lymphoma and systemic anaplastic large cell lymphoma is a leading example of effective toxin targeted therapy. Folate, also known as folic acid, vitamin B_9, is a vital nutrient required by all living cells for nucleotide biosynthesis and function as cofactor in certain biological pathways. It is especially important in aiding rapid cell division and growth. As such, the folate receptor is a focus for the development of therapies to treat fast dividing malignancy in particular ovarian cancer and non-small cell lung carcinoma. Some ovarian tumor type is likely to recur after initial successes with surgery and platinum-based chemotherapy, to which the regrowth can become resistant to available therapies. While folate receptor expression is negligible in normal ovary, ~90% of epithelial ovarian cancers overexpress the folate receptor, as do many lung adenocarinomas, thereby opening the possibility of directed therapies. In support of this, clinical studies with EC-145, a targeted vinca alkaloid analog of folate, indicated that platinum-resistant ovarian cancer and non-small cell lung carcinoma are susceptible to folate-based therapies. Fusion of a XTEN carrying ≥1 copy of folate to a XTEN bearing ≥ 3 drug molecules to create a targeted peptide-drug conjugate is expected to have vastly improved therapeutic index and half-life that will enable dosing at levels way below maximum tolerated dose (MTD), reduce dosing frequency and cost (reduced drug required per dose).

[00373] Clinical evaluation of folate-XTEN-drug composition is conducted in patients with relapsed or refractory advanced tumors or specifically in patients suffering from platinum-resistant ovarian cancer and non-small cell lung carcinoma who have failed numerous chemotherapies. Clinical trials are designed such that the efficacy and advantages of the folate-XTEN-drug conjugate can be verified in humans. Such studies in patients would comprise three phases. First, a Phase I safety and pharmacokinetics study is conducted to determine the MTD and to characterize the dose-limiting toxicity, pharmacokinetics and preliminary pharmacodynamics in humans. These initial studies are performed in patients with relapsed or refractory advanced tumors and for which standard curative or palliative measures could not be used or were no longer effective or tolerated. To enhance treatment efficacy, folate receptor positive status is an enrollment criteria, determined by immunohistochemistry of primary tumors or metastatic specimens and/or by folate-targeted molecular imaging agent. The scheme of the phase I study is to use single escalating doses of folate-XTEN-drug conjugate and measure the

biochemical, PK, and clinical parameters. This would permit the determination of the MTD and establish the threshold and maximum concentrations in dosage and in circulating drug that constitute the therapeutic window to be used in subsequent Phase II and Phase III trials. It also defines potential toxicities and adverse events to be tracked in future studies.

[00374] Phase II clinical studies of human patients are independently conducted in folate receptor positive platinum-resistant ovarian cancer patient population; non-small cell lung carcinoma patients having failed numerous chemotherapies; and patients suffering from relapsed or refractory advanced tumors. The trial evaluates the efficacy and safety of folate-XTEN-drug conjugate alone and in combination with a current chemotherapy employed in the specific indication. Patients receive intravenously administered folate-XTEN-drug conjugate at a dose level and regimen pre-determined in Phase I with or without the standard chemo-agent. A control arm comprising of the chemo-agent plus placebo is included. The primary endpoint is response rate as defined by the Response Evaluation Criteria in Solid Tumors (RECIST). Secondary endpoints would include safety and tolerability, time-to-progression and overall survival.

[00375] A phase III efficacy and safety study is structured to replicate or modify the phase II trial design in folate receptor positive platinum-resistant ovarian cancer patients; non-small cell lung carcinoma patients; and advanced tumor relapsed or refractory patients depending on the phase II clinical observations. Refinement of patient enrollment criteria, further patient stratification (example folate receptor expression level), dosage, regimen, status of standard chemo-agent etc., is further adjusted. The primary endpoint is progression-free-survival, as measured by RECIST, in patients defined as folate receptor positive. The trial will also be statistically powered for overall survival as a secondary endpoint with projected enrollment in excess of 400 patients. Incidence of adverse events, serious adverse event and deaths is also assessed.

[00376] It is anticipated that folate-XTEN-drug candidate will demonstrate anticancer activity without severe toxicity even in these highly pretreated patient populations.

[00377] Example 28: Serum stability of XTEN

[00378] A fusion protein containing XTEN_AE864 fused to the N-terminus of GFP was incubated in monkey plasma and rat kidney lysate for up to 7 days at 37°C. Samples were withdrawn at time 0, Day 1 and Day 7 and analyzed by SDS PAGE followed by detection using Western analysis and detection with antibodies against GFP as shown in FIG. 29. The sequence of XTEN_AE864 showed negligible signs of degradation over 7 days in plasma. However, XTEN_AE864 was rapidly degraded in rat kidney lysate over 3 days. The in vivo stability of the fusion protein was tested in plasma samples wherein the GFP_AE864 was immunoprecipitated and analyzed by SDS PAGE as described above. Samples that were withdrawn up to 7 days after injection showed very few signs of degradation. The results demonstrate the resistance of aaT-

XTEN to degradation due to serum proteases; a factor in the enhancement of pharmacokinetic properties of the aaT-XTEN fusion proteins.

[00379] Example 29: Characterization of secondary structure of XTEN linked to exendin-4

[00380] The XTEN_AE864-Ex4 was evaluated for degree of secondary structure by circular dichroism spectroscopy. CD spectroscopy was performed on a Jasco J-715 (Jasco Corporation, Tokyo, Japan) spectropolarimeter equipped with Jasco Peltier temperature controller (TPC-348WI). The concentration of protein was adjusted to 0.2 mg/mL in 20 mM sodium phosphate pH 7.0, 50 mM NaCl. The experiments were carried out using HELLMA quartz cells with an optical path-length of 0.1 cm. The CD spectra were acquired at 5°, 25°, 45°, and 65°C and processed using the J-700 version 1.08.01 (Build 1) Jasco software for Windows. The samples were equilibrated at each temperature for 5 min before performing CD measurements. All spectra were recorded in duplicate from 300 nm to 185 nm using a bandwidth of 1 nm and a time constant of 2 sec, at a scan speed of 100 nm/min. The CD spectrum shown in FIG. 30 shows no evidence of stable secondary structure and is consistent with an unstructured polypeptide.

[00381] Example 30: Increasing solubility and stability of a peptide payload by linking to XTEN

[00382] In order to evaluate the ability of XTEN to enhance the physicochemical properties of solubility and stability, fusion proteins of glucagon plus shorter-length XTEN were prepared and evaluated. The test articles were prepared in Tris-buffered saline at neutral pH and characterization of the Gcg-XTEN solution was by reverse-phase HPLC and size exclusion chromatography to affirm that the protein was homogeneous and non-aggregated in solution. The data are presented in Table 25. For comparative purposes, the solubility limit of unmodified glucagon in the same buffer was measured at 60 μM (0.2 mg/mL), and the result demonstrate that for all lengths of XTEN added, a substantial increase in solubility was attained. Importantly, in most cases the glucagon-XTEN fusion proteins were prepared to achieve target concentrations and were not evaluated to determine the maximum solubility limits for the given construct. However, in the case of glucagon linked to the AF-144 XTEN, the limit of solubility was determined, with the result that a 60-fold increase in solubility was achieved, compared to glucagon not linked to XTEN. In addition, the glucagon-AF144 was evaluated for stability, and was found to be stable in liquid formulation for at least 6 months under refrigerated conditions and for approximately one month at 37°C (data not shown).

[00383] The data support the conclusion that the linking of short-length XTEN polypeptides to a biologically active protein such as glucagon can markedly enhance the solubility properties of the protein by the resulting fusion protein, as well as confer stability at the higher protein concentrations.

Table 25: Solubility of Glucagon-XTEN constructs

Test Article	Solubility
Glucagon	60 μM
Glucagon-Y36	>370 μM
Glucagon-Y72	>293 μM
Glucagon-AF108	>145 μM
Glucagon-AF120	>160 μM
Glucagon-Y144	>497 μM
Glucagon-AE144	>467 μM
Glucagon-AF144	>3600 μM
Glucagon-Y288	>163 μM

[00384] Example 31: Analytical size exclusion chromatography of XTEN linked with diverse payloads

[00385] Size exclusion chromatography analyses were performed on fusion proteins containing various therapeutic proteins and unstructured recombinant proteins of increasing length. An exemplary assay used a TSKGel-G4000 SWXL (7.8mm x 30cm) column in which 40 μg of purified glucagon fusion protein at a concentration of 1 mg/ml was separated at a flow rate of 0.6 ml/min in 20 mM phosphate pH 6.8, 114 mM NaCl. Chromatogram profiles were monitored using OD214nm and OD280nm. Column calibration for all assays were performed using a size exclusion calibration standard from BioRad; the markers include thyroglobulin (670 kDa), bovine gamma-globulin (158 kDa), chicken ovalbumin (44 kDa), equine myoglobuin (17 kDa) and vitamin B12 (1.35 kDa). Representative chromatographic profiles of Glucagon-Y288, Glucagon-Y144, Glucagon-Y72, Glucagon-Y36 are shown as an overlay in FIG. 31. The data show that the molecular weight of each compound is proportional to the length of the attached XTEN sequence. However, the data also show that the apparent molecular weight of each construct is significantly larger than that expected for a globular protein (as shown by comparison to the standard proteins run in the same assay). Based on the SEC analyses for all constructs evaluated, the apparent molecular weights, the apparent molecular weight factor (expressed as the ratio of apparent molecular weight to the calculated molecular weight) and the hydrodynamic radius (R_H in nm) are shown in Table 26. The results indicate that incorporation of different XTENs of 576 amino acids or greater confers an apparent molecular weight for the fusion protein of approximately 339 kDa to 760, and that XTEN of 864 amino acids or greater confers an apparent molecular weight greater than approximately 800 kDA. The results of proportional increases in apparent molecular weight to actual molecular weight were consistent for fusion proteins created with XTEN from several different motif families; i.e., AD, AE, AF, AG, and AM, with increases of at least four-fold and

ratios as high as about 17-fold. Additionally, the incorporation of XTEN fusion partners with 576 amino acids or more into fusion proteins with the various payloads (and 288 residues in the case of glucagon fused to Y288) resulted with a hydrodynamic radius of 7 nm or greater; well beyond the glomerular pore size of approximately 3-5 nm. Accordingly, it is expected that fusion proteins comprising growth and XTEN have reduced renal clearance, contributing to increased terminal half-life and improving the therapeutic or biologic effect relative to a corresponding un-fused biologic payload protein.

Table 26: SEC analysis of various polypeptides

Construct Name	XTEN or fusion partner	Therapeutic Protein	Actual MW (kDa)	Apparent MW (kDa)	Apparent Molecular Weight Factor	R_H (nm)
AC14	Y288	Glucagon	28.7	370	12.9	7.0
AC28	Y144	Glucagon	16.1	117	7.3	5.0
AC34	Y72	Glucagon	9.9	58.6	5.9	3.8
AC33	Y36	Glucagon	6.8	29.4	4.3	2.6
AC89	AF120	Glucagon	14.1	76.4	5.4	4.3
AC88	AF108	Glucagon	13.1	61.2	4.7	3.9
AC73	AF144	Glucagon	16.3	95.2	5.8	4.7
AC53	AG576	GFP	74.9	339	4.5	7.0
AC39	AD576	GFP	76.4	546	7.1	7.7
AC41	AE576	GFP	80.4	760	9.5	8.3
AC52	AF576	GFP	78.3	526	6.7	7.6
AC398	AE288	FVII	76.3	650	8.5	8.2
AC404	AE864	FVII	129	1900	14.7	10.1
AC85	AE864	Exendin-4	83.6	938	11.2	8.9
AC114	AM875	Exendin-4	82.4	1344	16.3	9.4
AC143	AM875	hGH	100.6	846	8.4	8.7
AC302	AE912 + AE144	hGH	119.1	2,287	19.2	11.0
AC227	AM875	IL-1ra	95.4	1103	11.6	9.2
AC228	AM1318	IL-1ra	134.8	2286	17.0	10.5
AC493	AE864	FIX	127.7*	3967	31.1	12.2
AC616	AE864	GLP2-2G	83.1	1427	17.2	10
AC647	AE864	Ghrelin	82.7	996	12	9.2
AC659	AE864	C-peptide	82.7	822	10	8.8
AC663	AE1296	C-peptide	122.2	2348	19.2	11.1
AC434	AE288	aaT	71.1	500	7.0	7.7
AC435	AE576	aaT	97.5	1,127	11.6	9.5
AC345	AM875	aaT	122.6	1,390	11.3	9.9
AC450	AE288	aHer2_scFv	56.2	312	5.5	6.7

Construct Name	XTEN or fusion partner	Therapeutic Protein	Actual MW (kDa)	Apparent MW (kDa)	Apparent Molecular Weight Factor	R_H (nm)
AC451	AE576	aHer2_scFv	82.6	760	9.2	8.6
AC452	AE864	aHer2_scFv	109.1	1,390	12.7	9.9

* excluding glycosylation

[00386] Example 32: Analysis of sequences for secondary structure by prediction algorithms

[00387] Amino acid sequences can be assessed for secondary structure via certain computer programs or algorithms, such as the well-known Chou-Fasman algorithm (Chou, P. Y., *et al.* (1974) *Biochemistry*, **13**: 222-45) and the Garnier-Osguthorpe-Robson, or "GOR" method (Garnier J, Gibrat JF, Robson B. (1996). GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol 266:540-553). For a given sequence, the algorithms can predict whether there exists some or no secondary structure at all, expressed as 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.

[00388] Several representative sequences from XTEN "families" have been assessed using two algorithm tools for the Chou-Fasman and GOR methods to assess the degree of secondary structure in these sequences. The Chou-Fasman tool was provided by William R. Pearson and the University of Virginia, at the "Biosupport" internet site, URL located on the World Wide Web at .fasta.bioch.virginia.edu/fasta_www2/fasta_www.cgi?rm=misc1 as it existed on June 19, 2009. The GOR tool was provided by Pole Informatique Lyonnais at the Network Protein Sequence Analysis internet site, URL located on the World Wide Web at .npsa-pbil.ibcp.fr/cgi-bin/secpred_gor4.pl as it existed on June 19, 2008.

[00389] As a first step in the analyses, a single XTEN sequence was analyzed by the two algorithms. The AE864 composition is a XTEN with 864 amino acid residues created from multiple copies of four 12 amino acid sequence motifs consisting of the amino acids G, S, T, E, P, and A. The sequence motifs are characterized by the fact that there is limited repetitiveness within the motifs and within the overall sequence in that the sequence of any two consecutive amino acids is not repeated more than twice in any one 12 amino acid motif, and that no three contiguous amino acids of full-length the XTEN are identical. Successively longer portions of the AF 864 sequence from the N-terminus were analyzed by the Chou-Fasman and GOR algorithms (the latter requires a minimum length of 17 amino acids). The sequences were analyzed by entering the FASTA format sequences into the prediction tools and running the analysis. The results from the analyses are presented in Table 27.

[00390] The results indicate that, by the Chou-Fasman calculations, short XTEN of the AE and AG families, up to at least 288 amino acid residues, have no alpha-helices or beta sheets, but

amounts of predicted percentage of random coil by the GOR algorithm vary from 78-99%. With increasing XTEN lengths of 504 residues to greater than 1300, the XTEN analyzed by the Chou-Fasman algorithm had predicted percentages of alpha-helices or beta sheets of 0 to about 2%, while the calculated percentages of random coil increased to from 94-99%. Those XTEN with alpha-helices or beta sheets were those sequences with one or more instances of three contiguous serine residues, which resulted in predicted beta-sheet formation. However, even these sequences still had approximately 99% random coil formation.

[00391] The analysis supports the conclusion that: 1) XTEN created from multiple sequence motifs of G, S, T, E, P, and A that have limited repetitiveness as to contiguous amino acids are predicted to have very low amounts of alpha-helices and beta-sheets; 2) that increasing the length of the XTEN does not appreciably increase the probability of alpha-helix or beta-sheet formation; and 3) that progressively increasing the length of the XTEN sequence by addition of non-repetitive 12-mers consisting of the amino acids G, S, T, E, P, and A results in increased percentage of random coil formation. Based on the numerous sequences evaluated by these methods, it is concluded that XTEN created from sequence motifs of G, S, T, E, P, and A that have limited repetitiveness (defined as no more than two identical contiguous amino acids in any one motif) are expected to have very limited secondary structure. With the exception of motifs containing three contiguous serines, generally any order or combination of sequence motifs from Table 1 can be used to create an XTEN polypeptide that will result in an XTEN sequence that is substantially devoid of secondary structure, and that the effects of three contiguous serines is ameliorated by increasing the length of the XTEN. Such sequences are expected to have the characteristics described in the XTEN-containing composition embodiments of the invention disclosed herein.

Table 27: CHOU-FASMAN and GOR prediction calculations of polypeptide sequences

SEQ NAME	Sequence	No. Residues	Chou-Fasman Calculation	GOR Calculation
AE36: LCW0402_002	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP	36	Residue totals: H: 0 E: 0 percent: H: 0.0 E: 0.0	94.44%
AE36: LCW0402_003	GTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAP	36	Residue totals: H: 0 E: 0 percent: H: 0.0 E: 0.0	94.44%
AG36: LCW0404_001	GASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSP	36	Residue totals: H: 0 E: 0 percent: H: 0.0 E: 0.0	77.78%
AG36: LCW0404_003	GSSTPSGATGSPGSSPSASTGTGPGSSTPSGATGSP	36	Residue totals: H: 0 E: 0 percent: H: 0.0 E:	83.33 %

SEQ NAME	Sequence	No. Residues	Chou-Fasman Calculation	GOR Calculation
			0.0	
AE42_1	TEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGS	42	Residue totals: H: 0 E: 0 percent: H: 0.0 E: 0.0	90.48%
AE42_1	TEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGS	42	Residue totals: H: 0 E: 0 percent: H: 0.0 E: 0.0	90.48%
AG42_1	GAPSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGPSGP	42	Residue totals: H: 0 E: 0 percent: H: 0.0 E: 0.0	88.10%
AG42_2	GPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASP	42	Residue totals: H: 0 E: 0 percent: H: 0.0 E: 0.0	88.10%
AE144	GSEPATSGSETPGTSESATPESGPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGSEPATSGSETPGSEPATSGSETPGSEPATSGSETPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAP	144	Residue totals: H: 0 E: 0 percent: H: 0.0 E: 0.0	98.61%
AG144_1	PGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSS	144	Residue totals: H: 0 E: 0 percent: H: 0.0 E: 0.0	91.67%
AE288	GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP	288	Residue totals: H: 0 E: 0 percent: H: 0.0 E: 0.0	99.31%
AG288_2	GSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSP	288	Residue totals: H: 0 E: 0 percent: H: 0.0 E: 0.0	92.71
AF504	GASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSNPSASTGTGPGASPGTSSTG	504	Residue totals: H: 0 E: 0	94.44%

SEQ NAME	Sequence	No. Residues	Chou-Fasman Calculation	GOR Calculation
	SPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGASPGTSSTGSPGASPG TSSTGSPGTPGSGTASSSPGSSTPSGAT GSPGASPGTSSTGSPGTPGSGTASSSP GSSTPSGATGSPGSNPSASTGTGPGSS PSASTGTGPGSSTPSGATGSPGSSTPSG ATGSPGASPGTSSTGSPGASPGTSSTG SPGASPGTSSTGSPGTPGSGTASSSPG ASPGTSSTGSPGASPGTSSTGSPGASP GTSSTGSPGSSPSASTGTGPGTPGSGT ASSSPGASPGTSSTGSPGASPGTSSTGS PGASPGTSSTGSPGSSTPSGATGSPGSS TPSGATGSPGASPGTSSTGSPGTPGSG TASSSPGSSTPSGATGSPGSSTPSGATG SPGSSTPSGATGSPGSSPSASTGTGPG ASPGTSSTGSP		percent: H: 0.0 E: 0.0	
AD 576	GSSESGSSEGGPGSGGEPSESGSSGSSE SGSSEGGPGSSESGSSEGGPGSSESGSS EGGPGSSESGSSEGGPGSSESGSSEGG PGESPGGSSGSESGSEGSSGPGESSGSS ESGSSEGGPGSSESGSSEGGPGSSESGS SEGGPGSGGEPSESGSSGESPGGSSGS ESGESPGGSSGSESGSGGEPSESGSSGS SESGSSEGGPGSGGEPSESGSSGSGGE PSESGSSGSEGSSGPGESSGESPGGSSG SESGSGGEPSESGSSGSGGEPSESGSSG SGGEPSESGSSGSSESGSSEGGPGESPG GSSGSESGESPGGSSGSESGESPGGSS GSESGESPGGSSGSESGESPGGSSGSES GSSESGSSEGGPGSGGEPSESGSSGSE GSSGPGESSGSSESGSSEGGPGSGGEP SESGSSGSSESGSSEGGPGSGGEPSESG SSGESPGGSSGSESGESPGGSSGSESGS SESGSSEGGPGSGGEPSESGSSGSSESG SSEGGPGSGGEPSESGSSGSGGEPSES GSSGESPGGSSGSESGSEGSSGPGESS GSSESGSSEGGPGSEGSSGPGESS	576	Residue totals: H: 7 E: 0 percent: H: 1.2 E: 0.0	99.65%
AE576	GSPAGSPTSTEEGTSESATPESGPGTST EPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGTSESATPESGP GSEPATSGSETPGSEPATSGSETPGSPA GSPTSTEEGTSESATPESGPGTSTEPSE GSAPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSE SATPESGPGTSTEPSEGSAPGTSESATP ESGPGSEPATSGSETPGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGTSE SATPESGPGSPAGSPTSTEEGTSESATP ESGPGSEPATSGSETPGTSESATPESGP GTSTEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEPSE GSAPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSESATPESGPGSEP ATSGSETPGTSESATPESGPGSEPATS GSETPGTSESATPESGPGTSTEPSEGSA PGTSESATPESGPGSPAGSPTSTEEGSP	576	Residue totals: H: 2 E: 0 percent: H: 0.4 E: 0.0	99.65%

SEQ NAME	Sequence	No. Residues	Chou-Fasman Calculation	GOR Calculation
	AGSPTSTEEGSPAGSPTSTEEGTSESAT PESGPGTSTEPSEGSAP			
AG576	PGTPGSGTASSSPGSSTPSGATGSPGSS PSASTGTGPGSSPSASTGTGPGSSTPSG ATGSPGSSTPSGATGSPGASPGTSSTG SPGASPGTSSTGSPGASPGTSSTGSPGT PGSGTASSSPGASPGTSSTGSPGASPG TSSTGSPGASPGTSSTGSPGSSPSASTG TGPGTPGSGTASSSPGASPGTSSTGSP GASPGTSSTGSPGASPGTSSTGSPGSST PSGATGSPGSSTPSGATGSPGASPGTS STGSPGTPGSGTASSSPGSSTPSGATGS PGSSTPSGATGSPGSSTPSGATGSPGSS PSASTGTGPGASPGTSSTGSPGASPGT SSTGSPGTPGSGTASSSPGASPGTSSTG SPGASPGTSSTGSPGASPGTSSTGSPG ASPGTSSTGSPGTPGSGTASSSPGSSTP SGATGSPGTPGSGTASSSPGSSTPSGA TGSPGTPGSGTASSSPGSSTPSGATGSP GSSTPSGATGSPGSSPSASTGTGPGSSP SASTGTGPGASPGTSSTGSPGTPGSGT ASSSPGSSTPSGATGSPGSSPSASTGTG PGSSPSASTGTGPGASPGTSSTGS	576	Residue totals: H: 0 E: 3 percent: H: 0.4 E: 0.5	99.31%
AF540	GSTSSTAESPGPGSTSSTAESPGPGSTS ESPSGTAPGSTSSTAESPGPGSTSSTAE SPGPGTSTPESGSASPGSTSESPSGTAP GTSPSGESSTAPGSTSESPSGTAPGSTS ESPSGTAPGTSPSGESSTAPGSTSESPS GTAPGSTSESPSGTAPGTSPSGESSTAP GSTSESPSGTAPGSTSESPSGTAPGSTS ESPSGTAPGTSTPESGSASPGSTSESPS GTAPGTSTPESGSASPGSTSSTAESPGP GSTSSTAESPGPGTSTPESGSASPGTST PESGSASPGSTSESPSGTAPGTSTPESG SASPGTSTPESGSASPGSTSESPSGTAP GSTSESPSGTAPGSTSESPSGTAPGSTS STAESPGPGTSTPESGSASPGTSTPESG SASPGSTSESPSGTAPGSTSESPSGTAP GTSTPESGSASPGSTSESPSGTAPGSTS ESPSGTAPGTSTPESGSASPGTSPSGES STAPGSTSSTAESPGPGTSPSGESSTAP GSTSSTAESPGPGTSTPESGSASPGSTS ESPSGTAP	540	Residue totals: H: 2 E: 0 percent: H: 0.4 E: 0.0	99.65
AD836	GSSESGSSEGGPGSSESGSSEGGPGESP GGSSGSESGSGGEPSESGSSGESPGGS SGSESGESPGGSSGSESGSSESGSSEGG PGSSESGSSEGGPGSSESGSSEGGPGES PGGSSGSESGESPGGSSGSESGESPGG SSGSESGSSESGSSEGGPGSSESGSSEG GPGSSESGSSEGGPGSSESGSSEGGPG SSESGSSEGGPGSSESGSSEGGPGSGG EPSESGSSGESPGGSSGSESGESPGGSS GSESGSGGEPSESGSSGSEGSSGPGESS GSSESGSSEGGPGSGGEPSESGSSGSE GSSGPGESSGSSESGSSEGGPGSGGEP SESGSSGESPGGSSGSESGSGGEPSESG	836	Residue totals: H: 0 E: 0 percent: H: 0.0 E: 0.0	98.44%

SEQ NAME	Sequence	No. Residues	Chou-Fasman Calculation	GOR Calculation
	SSGSGGEPSESGSSGSSESGSSEGGPGS GGEPSESGSSGSGGEPSESGSSGSEGSS GPGESSGESPGGSSGSESGSEGSSGPG ESSGSEGSSGPGESSGSGGEPSESGSSG SSESGSSEGGPGSSESGSSEGGPGESPG GSSGSESGSGGEPSESGSSGSEGSSGP GESSGESPGGSSGSESGSEGSSGPGSSE SGSSEGGPGSGGEPSESGSSGSEGSSG PGESSGSEGSSGPGESSGSEGSSGPGES SGSGGEPSESGSSGSGGEPSESGSSGES PGGSSGSESGESPGGSSGSESGSGGEP SESGSSGSEGSSGPGESSGESPGGSSGS ESGSSESGSSEGGPGSSESGSSEGGPGS SESGSSEGGPGSGGEPSESGSSGSSESG SSEGGPGESPGGSSGSESGSGGEPSES GSSGSSESGSSEGGPGESPGGSSGSES GSGGEPSESGSSGESPGGSSGSESGSG GEPSESGSS			
AE864	GSPAGSPTSTEEGTSESATPESGPGTST EPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGTSESATPESGP GSEPATSGSETPGSEPATSGSETPGSPA GSPTSTEEGTSESATPESGPGTSTEPSE GSAPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSE SATPESGPGTSTEPSEGSAPGTSESATP ESGPGSEPATSGSETPGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGTSE SATPESGPGSPAGSPTSTEEGTSESATP ESGPGSEPATSGSETPGTSESATPESGP GTSTEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEPSE GSAPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSESATPESGPGSEP ATSGSETPGTSESATPESGPGSEPATS GSETPGTSESATPESGPGTSTEPSEGSA PGTSESATPESGPGSPAGSPTSTEEGSP AGSPTSTEEGSPAGSPTSTEEGTSESAT PESGPGTSTEPSEGSAPGTSESATPESG PGSEPATSGSETPGTSESATPESGPGSE PATSGSETPGTSESATPESGPGTSTEPS EGSAPGSPAGSPTSTEEGTSESATPES GPGSEPATSGSETPGTSESATPESGPGS PAGSPTSTEEGSPAGSPTSTEEGTSTEP SEGSAPGTSESATPESGPGTSESATPES GPGTSESATPESGPGSEPATSGSETPGS EPATSGSETPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGSEPATSGSE TPGTSESATPESGPGTSTEPSEGSAP	864	Residue totals: H: 2 E: 3 percent: H: 0.2 E: 0.4	99.77%
AF864	GSTSESPSGTAPGTSPSGESSTAPGSTS ESPSGTAPGSTSESPSGTAPGTSTPESG SASPGTSTPESGSASPGSTSESPSGTAP GSTSESPSGTAPGTSPSGESSTAPGSTS ESPSGTAPGTSPSGESSTAPGTSPSGES STAPGSTSSTAESPGPGTSPSGESSTAP GTSPSGESSTAPGSTSSTAESPGPGTST	875	Residue totals: H: 2 E: 0 percent: H: 0.2 E: 0.0	95.20%

SEQ NAME	Sequence	No. Residues	Chou-Fasman Calculation	GOR Calculation
	PESGSASPGTSTPESGSASPGSTSESPS GTAPGSTSESPSGTAPGTSTPESGSASP GSTSSTAESPGPGTSTPESGSASPGSTS ESPSGTAPGTSPSGESSTAPGSTSSTAE SPGPGTSPSGESSTAPGTSTPESGSASP GSTSSTAESPGPGSTSSTAESPGPGSTS STAESPGPGSTSSTAESPGPGTSPSGES STAPGSTSESPSGTAPGSTSESPSGTAP GTSTPESGPXXXGASASGAPSTXXXX SESPSGTAPGSTSESPSGTAPGSTSESP SGTAPGSTSESPSGTAPGSTSESPSGTA PGSTSESPSGTAPGTSTPESGSASPGTS PSGESSTAPGTSPSGESSTAPGSTSSTA ESPGPGTSPSGESSTAPGTSTPESGSAS PGSTSESPSGTAPGSTSESPSGTAPGTS PSGESSTAPGSTSESPSGTAPGTSTPES GSASPGTSTPESGSASPGSTSESPSGTA PGTSTPESGSASPGSTSSTAESPGPGST SESPSGTAPGSTSESPSGTAPGTSPSGE SSTAPGSTSSTAESPGPGTSPSGESSTA PGTSTPESGSASPGTSPSGESSTAPGTS PSGESSTAPGTSPSGESSTAPGSTSSTA ESPGPGSTSSTAESPGPGTSPSGESSTA PGSSPSASTGTGPGSSTPSGATGSPGSS TPSGATGSP			
AG864	GASPGTSSTGSPGSSPSASTGTGPGSSP SASTGTGPGTPGSGTASSSPGSSTPSG ATGSPGSSPSASTGTGPGASPGTSSTG SPGTPGSGTASSSPGSSTPSGATGSPGT PGSGTASSSPGASPGTSSTGSPGASPG TSSTGSPGTPGSGTASSSPGSSTPSGAT GSPGASPGTSSTGSPGTPGSGTASSSP GSSTPSGATGSPGSSPSASTGTGPGSSP SASTGTGPGSSTPSGATGSPGSSTPSG ATGSPGASPGTSSTGSPGASPGTSSTG SPGASPGTSSTGSPGTPGSGTASSSPG ASPGTSSTGSPGASPGTSSTGSPGASP GTSSTGSPGSSPSASTGTGPGTPGSGT ASSSPGASPGTSSTGSPGASPGTSSTGS PGASPGTSSTGSPGSSTPSGATGSPGSS TPSGATGSPGASPGTSSTGSPGTPGSG TASSSPGSSTPSGATGSPGSSTPSGATG SPGSSTPSGATGSPGSSPSASTGTGPG ASPGTSSTGSPGASPGTSSTGSPGTPGS GTASSSPGASPGTSSTGSPGASPGTSST GSPGASPGTSSTGSPGASPGTSSTGSP GTPGSGTASSSPGSSTPSGATGSPGTP GSGTASSSPGSSTPSGATGSPGTPGSG TASSSPGSSTPSGATGSPGSSTPSGATG SPGSSPSASTGTGPGSSPSASTGTGPG ASPGTSSTGSPGTPGSGTASSSPGSSTP SGATGSPGSSPSASTGTGPGSSPSAST GTGPGASPGTSSTGSPGASPGTSSTGS PGSSTPSGATGSPGSSPSASTGTGPGA SPGTSSTGSPGSSPSASTGTGPGTPGSG TASSSPGSSTPSGATGSPGSSTPSGATG	864	Residue totals: H: 0 E: 0 percent: H: 0.0 E: 0.0	94.91%

SEQ NAME	Sequence	No. Residues	Chou-Fasman Calculation	GOR Calculation
	SPGASPGTSSTGSP			
AM875	GTSTEPSEGSAPGSEPATSGSETPGSPA GSPTSTEEGSTSSTAESPGPGTSTPESG SASPGSTSESPSGTAPGSTSESPSGTAP GTSTPESGSASPGTSTPESGSASPGSEP ATSGSETPGTSESATPESGPGSPAGSPT STEEGTSTEPSEGSAPGTSESATPESGP GTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSE GSAPGTSESATPESGPGTSESATPESGP GTSTEPSEGSAPGTSTEPSEGSAPGTSE SATPESGPGTSTEPSEGSAPGSEPATSG SETPGSPAGSPTSTEEGSSTPSGATGSP GTPGSGTASSSPGSSTPSGATGSPGTS TEPSEGSAPGTSTEPSEGSAPGSEPATS GSETPGSPAGSPTSTEEGSPAGSPTSTE EGTSTEPSEGSAPGASASGAPSTGGTS ESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGSTSSTAESPGPGSTSESPSGTA PGTSPSGESSTAPGTPGSGTASSSPGSS TPSGATGSPGSSPSASTGTGPGSEPAT SGSETPGTSESATPESGPGSEPATSGSE TPGSTSSTAESPGPGSTSSTAESPGPGT SPSGESSTAPGSEPATSGSETPGSEPAT SGSETPGTSTEPSEGSAPGSTSSTAESP GPGTSTPESGSASPGSTSESPSGTAPGT STEPSEGSAPGTSTEPSEGSAPGTSTEP SEGSAPGSSTPSGATGSPGSSPSASTGT GPGASPGTSSTGSPGSEPATSGSETPG TSESATPESGPGSPAGSPTSTEEGSSTP SGATGSPGSSPSASTGTGPGASPGTSS TGSPGTSESATPESGPGTSTEPSEGSAP GTSTEPSEGSAP	875	Residue totals: H: 7 E: 3 percent: H: 0.8 E: 0.3	98.63%
AM1318	GTSTEPSEGSAPGSEPATSGSETPGSPA GSPTSTEEGSTSSTAESPGPGTSTPESG SASPGSTSESPSGTAPGSTSESPSGTAP GTSTPESGSASPGTSTPESGSASPGSEP ATSGSETPGTSESATPESGPGSPAGSPT STEEGTSTEPSEGSAPGTSESATPESGP GTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSE GSAPGTSESATPESGPGTSESATPESGP GTSTEPSEGSAPGTSTEPSEGSAPGTSE SATPESGPGTSTEPSEGSAPGSEPATSG SETPGSPAGSPTSTEEGSSTPSGATGSP GTPGSGTASSSPGSSTPSGATGSPGTS TEPSEGSAPGTSTEPSEGSAPGSEPATS GSETPGSPAGSPTSTEEGSPAGSPTSTE EGTSTEPSEGSAPGPEPTGPAPSGGSEP ATSGSETPGTSESATPESGPGSPAGSPT STEEGTSESATPESGPGSPAGSPTSTEE GSPAGSPTSTEEGTSESATPESGPGSPA GSPTSTEEGSPAGSPTSTEEGSTSSTAE SPGPGSTSESPSGTAPGTSPSGESSTAP GSTSESPSGTAPGSTSESPSGTAPGTSP SGESSTAPGTSTEPSEGSAPGTSESATP	1318	Residue totals: H: 7 E: 0 percent: H: 0.7 E: 0.0	99.17%

SEQ NAME	Sequence	No. Residues	Chou-Fasman Calculation	GOR Calculation
	ESGPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGTSESATPESGPGTST EPSEGSAPGTSESATPESGPGTSTEPSE GSAPGTSPSGESSTAPGTSPSGESSTAP GTSPSGESSTAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGSSPSAST GTGPGSSTPSGATGSPGSSTPSGATGS PGSSTPSGATGSPGSSTPSGATGSPGA SPGTSSTGSPGASASGAPSTGGTSPSG ESSTAPGSTSSTAESPGPGTSPSGESST APGTSESATPESGPGTSTEPSEGSAPG TSTEPSEGSAPGSSPSASTGTGPGSSTP SGATGSPGASPGTSSTGSPGTSTPESG SASPGTSPSGESSTAPGTSPSGESSTAP GTSESATPESGPGSEPATSGSETPGTST EPSEGSAPGSTSESPSGTAPGSTSESPS GTAPGTSTPESGSASPGSPAGSPTSTEE GTSESATPESGPGTSTEPSEGSAPGSPA GSPTSTEEGTSESATPESGPGSEPATSG SETPGSSTPSGATGSPGASPGTSSTGSP GSSTPSGATGSPGSTSESPSGTAPGTSP SGESSTAPGSTSSTAESPGPGSSTPSGA TGSPGASPGTSSTGSPGTPGSGTASSSP GSPAGSPTSTEEGSPAGSPTSTEEGTST EPSEGSAP			
AM923	MAEPAGSPTSTEEGASPGTSSTGSPGS STPSGATGSPGSSTPSGATGSPGTSTEP SEGSAPGSEPATSGSETPGSPAGSPTST EEGSTSSTAESPGPGTSTPESGSASPGS TSESPSGTAPGSTSESPSGTAPGTSTPE SGSASPGTSTPESGSASPGSEPATSGSE TPGTSESATPESGPGSPAGSPTSTEEGT STEPSEGSAPGTSESATPESGPGTSTEP SEGSAPGTSTEPSEGSAPGSPAGSPTST EEGTSTEPSEGSAPGTSTEPSEGSAPGT SESATPESGPGTSESATPESGPGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPES GPGTSTEPSEGSAPGSEPATSGSETPGS PAGSPTSTEEGSSTPSGATGSPGTPGS GTASSSPGSSTPSGATGSPGTSTEPSEG SAPGTSTEPSEGSAPGSEPATSGSETPG SPAGSPTSTEEGSPAGSPTSTEEGTSTE PSEGSAPGASASGAPSTGGTSESATPE SGPGSPAGSPTSTEEGSPAGSPTSTEEG STSSTAESPGPGSTSESPSGTAPGTSPS GESSTAPGTPGSGTASSSPGSSTPSGA TGSPGSSPSASTGTGPGSEPATSGSETP GTSESATPESGPGSEPATSGSETPGSTS STAESPGPGSTSSTAESPGPGTSPSGES STAPGSEPATSGSETPGSEPATSGSETP GTSTEPSEGSAPGSTSSTAESPGPGTST PESGSASPGSTSESPSGTAPGTSTEPSE GSAPGTSTEPSEGSAPGTSTEPSEGSAP GSSTPSGATGSPGSSPSASTGTGPGAS PGTSSTGSPGSEPATSGSETPGTSESAT PESGPGSPAGSPTSTEEGSSTPSGATGS	924	Residue totals: H: 4 E: 3 percent: H: 0.4 E: 0.3	98.70%

SEQ NAME	Sequence	No. Residues	Chou-Fasman Calculation	GOR Calculation
	PGSSPSASTGTGPGASPGTSSTGSPGTS ESATPESGPGTSTEPSEGSAPGTSTEPS EGSAP			
AE912	MAEPAGSPTSTEEGTPGSGTASSSPGS STPSGATGSPGASPGTSSTGSPGSPAG SPTSTEEGTSESATPESGPGTSTEPSEG SAPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGTSESATPESGPGSEPA TSGSETPGSEPATSGSETPGSPAGSPTS TEEGTSESATPESGPGTSTEPSEGSAPG TSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPGTSTEPSEGSAPGTSESATPE SGPGTSTEPSEGSAPGTSESATPESGPG SEPATSGSETPGTSTEPSEGSAPGTSTE PSEGSAPGTSESATPESGPGTSESATPE SGPGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTE PSEGSAPGTSTEPSEGSAPGTSTEPSEG SAPGTSTEPSEGSAPGTSTEPSEGSAPG TSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPGTSES ATPESGPGSPAGSPTSTEEGSPAGSPTS TEEGSPAGSPTSTEEGTSESATPESGPG TSTEPSEGSAPGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGTSTEPSEGSAPG SPAGSPTSTEEGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGSPAGSPTS TEEGSPAGSPTSTEEGTSTEPSEGSAPG TSESATPESGPGTSESATPESGPGTSES ATPESGPGSEPATSGSETPGSEPATSGS ETPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGSEPATSGSETPGTSES ATPESGPGTSTEPSEGSAP	913	Residue totals: H: 8 E: 3 percent: H: 0.9 E: 0.3	99.45%
BC 864	GTSTEPSEPGSAGTSTEPSEPGSAGSEP ATSGTEPSGSGASEPTSTEPGSEPATS GTEPSGSEPATSGTEPSGSEPATSGTEP SGSGASEPTSTEPGTSTEPSEPGSAGSE PATSGTEPSGTSTEPSEPGSAGSEPATS GTEPSGSEPATSGTEPSGTSTEPSEPGS AGTSTEPSEPGSAGSEPATSGTEPSGS EPATSGTEPSGTSEPSTSEPGAGSGAS EPTSTEPGTSEPSTSEPGAGSEPATSGT EPSGSEPATSGTEPSGTSTEPSEPGSAG TSTEPSEPGSAGSGASEPTSTEPGSEPA TSGTEPSGSEPATSGTEPSGSEPATSGT EPSGSEPATSGTEPSGTSTEPSEPGSAG SEPATSGTEPSGSGASEPTSTEPGTSTE PSEPGSAGSEPATSGTEPSGSGASEPTS TEPGTSTEPSEPGSAGSGASEPTSTEPG SEPATSGTEPSGSGASEPTSTEPGSEPA TSGTEPSGSGASEPTSTEPGTSTEPSEP GSAGSEPATSGTEPSGSGASEPTSTEP GTSTEPSEPGSAGSEPATSGTEPSGTST		Residue totals: H: 0 E: 0 percent: H: 0 E: 0	99.77%

SEQ NAME	Sequence	No. Residues	Chou-Fasman Calculation	GOR Calculation
	EPSEPGSAGSEPATSGTEPSGTSTEPSE PGSAGTSTEPSEPGSAGTSTEPSEPGSA GTSTEPSEPGSAGTSTEPSEPGSAGTST EPSEPGSAGTSEPSTSEPGAGSGASEPT STEPGTSTEPSEPGSAGTSTEPSEPGSA GTSTEPSEPGSAGSEPATSGTEPSGSG ASEPTSTEPGSEPATSGTEPSGSEPATS GTEPSGSEPATSGTEPSGSEPATSGTEP SGTSEPSTSEPGAGSEPATSGTEPSGSG ASEPTSTEPGTSTEPSEPGSAGSEPATS GTEPSGSGASEPTSTEPGTSTEPSEPGS A			
Seg 176	SAGSPTAEAAGCGTAEAAGTSESATP ESGPGSEPATSGSETPGTSESATPESGP GTSTEPSEGSAPGTSESATPESGPGSPA GSPTSTEEGSPAGSPTSTEEGSPAGSPT STEEGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGSEPATSGSETPGTSE SATPESGPGSEPATSGSETPGTSESATP ESGPGTSTEPSEGSAPTAEAAGCGTAE AAGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGSPAG SPTSTEEGSPAGSPTSTEEGTSTEPSEG SAPGTSESATPESGPGTSESATPESGPG TSESATPESGPGSEPATSGSETPGSEPA TSGSETPGSPAGSPTSTEEGTSTEPSEG SAPGTSTEPSEGSAPGSEPATSGSETPT AEAAGCGTAEAASASR		Residue totals H: 16 E: 0 Percent H: 3.7 E: 0.0	94.9%
Seg 177	SAGSPTAEAAGCGTAEAAPGSEPATS GSETPGTSESATPESGPGSEPATSGSET PGTAEAAGCGTAEAASTEPSEGSAPG TSESATPESGPGSPAGSPTSTEEGSPAT AEAAGCGTAEAASPTSTEEGTSESATP ESGPGTSTEPSEGSAPGTSESATTAEA AGCGTAEAASETPGTSESATPESGPGS EPATSGSETPGTSESATPESGTAEAAG CGTAEAAGSPAGSPTSTEEGTSESATP ESGPGSEPATSGSETPGTTAEAAGCGT AEAAAGSPTSTEEGSPAGSPTSTEEGT STEPSEGSAPGTSESTAEAAGCGTAEA ATPESGPGTSESATPESGPGSEPATSGS ETPGSEPATSGTAEAAGCGTAEAATE EGTSTEPSEGSAPGTSTEPSEGSAPGSE PATSGSETPTAEAAGCGTAEAASASR		Residue totals H: 64 E: 0 Percent: H: 14.8 E: 0.0	86.8%
Seg 194	SAGSPGSPAGSPTSTEEGTSESATPESG PGTSTEPSEGSAPGSPAGSPTSTEEGTS TEPSEGSAPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGSEPATSGSET PGSPAGSPTSTEEGTSESATPESGPGTS TEPSEGSAPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEGSA PGTSESATPESGPGTSTEPSEGSAPGTS ESATPESGPGSEPATSGSETPGTSTEPS EGSAPGTSTEPSEGSAPGTSESATPES GPGTSESATPESGPGSPAGSPTSTEEG TSESATPESGPGSEPATSGSETPGTSES		Residue Totals H: 0 E: 3 percent: H: 0.0 E: 0.3	99.4%

SEQ NAME	Sequence	No. Residues	Chou-Fasman Calculation	GOR Calculation
	ATPESGPGTSTEPSEGSAPGTSTEPSEG SAPGTSTEPSEGSAPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGSPAG SPTSTEEGTSTEPSEGSAPGTSESATPE SGPGSEPATSGSETPGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTE PSEGSAPGTSESATPESGPGSPAGSPTS TEEGSPAGSPTSTEEGSPAGSPTSTEEG TSESATPESGPGTSTEPSEGSAPGTSES ATPESGPGSEPATSGSETPGTSESATPE SGPGSEPATSGSETPGTSESATPESGPG TSTEPSEGSAPGSPAGSPTSTEEGTSES ATPESGPGSEPATSGSETPGTSESATPE SGPGSPAGSPTSTEEGSPAGSPTSTEEG TSTEPSEGSAPGTSESATPESGPGTSES ATPESGPGTSESATPESGPGSEPATSGS ETPGSEPATSGSETPGSPAGSPTSTEEG TSTEPSEGSAPGTSTEPSEGSAPGSEPA TSGSETPGTSESATPESGPGTESASK			

* H: alpha-helix E: beta-sheet

[00392] Example 33: Analysis of polypeptide sequences for repetitiveness

[00393] Polypeptide amino acid sequences can be assessed for repetitiveness by quantifying the number of times a shorter subsequence appears within the overall polypeptide. For example, a polypeptide of 200 amino acid residues has 192 overlapping 9-amino acid subsequences (or 9-mer "frames"), but the number of unique 9-mer subsequences will depend on the amount of repetitiveness within the sequence. In the present analysis, different sequences were assessed for repetitiveness by summing the occurrence of all unique 3-mer subsequences for each 3-amino acid frame across the first 200 amino acids of the polymer portion divided by the absolute number of unique 3-mer subsequences within the 200 amino acid sequence. The resulting subsequence score is a reflection of the degree of repetitiveness within the polypeptide.

[00394] The results, shown in Table 28, indicate that the unstructured polypeptides consisting of 2 or 3 amino acid types have high subsequence scores, while those of consisting of 12 amino acids motifs of the six amino acids G, S, T, E, P, and A with a low degree of internal repetitiveness, have subsequence scores of less than 10, and in some cases, less than 5. For example, the L288 sequence has two amino acid types and has short, highly repetitive sequences, resulting in a subsequence score of 50.0. The polypeptide J288 has three amino acid types but also has short, repetitive sequences, resulting in a subsequence score of 33.3. Y576 also has three amino acid types, but is not made of internal repeats, reflected in the subsequence score of 15.7 over the first 200 amino acids. W576 consists of four types of amino acids, but has a higher degree of internal repetitiveness, e.g., "GGSG", resulting in a subsequence score of 23.4. The AD576 consists of four types of 12 amino acid motifs, each consisting of four types of amino

acids. Because of the low degree of internal repetitiveness of the individual motifs, the overall subsequence score over the first 200 amino acids is 13.6. In contrast, XTEN's consisting of four motifs contains six types of amino acids, each with a low degree of internal repetitiveness have lower subsequence scores; i.e., AE864 (6.1), AF864 (7.5), and AM875 (4.5).

[00395] Conclusions: The results indicate that the combination of 12 amino acid subsequence motifs, each consisting of four to six amino acid types that are essentially non-repetitive, into a longer XTEN polypeptide results in an overall sequence that is non-repetitive. This is despite the fact that each subsequence motif may be used multiple times across the sequence. In contrast, polymers created from smaller numbers of amino acid types resulted in higher subsequence scores, although the actual sequence can be tailored to reduce the degree of repetitiveness to result in lower subsequence scores.

Table 28: Subsequence score calculations of polypeptide sequences

Seq Name	Amino Acid Sequence	Score
J288	GSGGEGGSGGEGGSGGEGGSGGEGGSGGEGGSGGEGGSGGEGGSGGEGG SGGEGGSGGEGGSGGEGGSGGEGGSGGEGGSGGEGGSGGEGGSGGEGGS GGEGGSGGEGGSGGEGGSGGEGGSGGEGGSGGEGGSGGEGGSGGEGGSG GEGGSGGEGGSGGEGGSGGEGGSGGEGGSGGEGGSGGEGGSGGEGGSGG EGGSGGEGGSGGEGGSGGEGGSGGEGGSGGEGGSGGEGGSGGEGGSGGE GGSGGEGGSGGEGGSGGEGGSGGEGGSGGEGGSGGEGGSGGEGG	33.3
K288	GEGEGGGEGGEGEGGGEGGEGEGGGEGGEGEGGGEGGEGEGGGEGGEGE GGGEGGEGEGGGEGGEGEGGGEGGEGEGGGEGGEGEGGGEGGEGEGGG EGGEGEGGGEGGEGEGGGEGGEGEGGGEGGEGEGGGEGGEGEGGGEGGE GEGGGEGGEGEGGGEGGEGEGGGEGGEGEGGGEGGEGEGGGEGGEGEG GGEGGEGEGGGEGGEGEGGGEGGEGEGGGEGGEGEGGGEGGEGEGGGE GGEGEGGGEGGEGEGGGEGGEGEGGGEGGEGEGGGEGGEGEGGGEG	46.9
L288	SSESSESSSSESSSESSESSSSESSSESSESSSSESSSESSESSSSESSSESSESSSS ESSSESSESSSSESSSESSESSSSESSSESSESSSSESSSESSESSSSESSSESSESS SSESSSESSESSSSESSSESSESSSSESSSESSESSSSESSSESSESSSSESSSESSE SSSSESSSESSESSSSESSSESSESSSSESSSESSESSSSESSSESSESSSSESSSES SESSSSESSSESSESSSSESSSESSESSSSESSSESSESSSSESSSESSESSSSES	50.0
Y288	GEGSGEGSEGEGSEGSGEGEGSEGSGEGEGGSEGSEGEGGSEGSEGEGGSE GSEGEGSGEGSEGEGGSEGSEGEGSGEGSEGEGSEGGSEGEGGSEGSEGEG SGEGSEGEGGEGGSEGEGSEGSGEGEGSGEGSEGEGSEGSGEGEGSGEGSE GEGSEGSGEGEGSEGSGEGEGGSEGSEGEGSEGSGEGEGGEGSGEGEGSGE GSEGEGGGEGSEGEGSGEGGEGEGSEGGSEGEGGSEGGEGEGSEGSGEGE GSEGGSEGEGSEGGSEGEGSEGSGEGEGSEGSGE	26.8
Q576	GGKPGEGGKPEGGGGKPGGKPEGEGEGKPGGKPEGGGKPGGGEGGKPEG GKPEGEGKPGGGEGKPGGKPEGGGGKPEGEGKPGGGGGKPGGKPEGEGK PGGGEGGKPEGKPGEGGEGKPGGKPEGGGEGKPGGGKPGEGGKPGEGKP GGGEGGKPEGGKPEGEGKPGGGEGKPGGKPGEGGKPEGGGEGKPGGKPG EGGEGKPGGGKPEGEGKPGGGKPGGGEGGKPEGEGKPGGKPEGGGEGKP GGKPEGGGKPEGGGEGKPGGGKPGEGGKPGEGEGKPGGKPEGEGKPGGE GGGKPEGKPGGGEGGKPEGGKPGEGGKPEGGKPGEGGEGKPGGGKPGEG GKPEGGGKPEGEGKPGGGGKPGEGGKPEGGKPEGGGEGKPGGGKPEGEG KPGGGEGKPGGKPEGGGGKPGEGGKPEGGKPGGEGGGKPEGEGKPGGKP GEGGGGKPGGKPEGEGKPGEGGEGKPGGKPEGGGEGKPGGKPEGGGEGK PGGGKPGEGGKPEGGGKPGEGGKPGEGGKPEGEGKPGGGEGKPGGKPGE GGKPEGGGEGKPGGKPGGEGGGKPEGGKPGEGGKPEG	18.5
U576	GEGKPGGKPGSGGGKPGEGGKPGSGEGKPGGKPGSGGSGKPGGKPGEGG KPEGGSGGKPGGGGKPGGKPGGEGSGKPGGKPEGGGKPEGGSGGKPGGK	18.1

Seq Name	Amino Acid Sequence	Score
	PEGGSGGKPGGKPGSGEGGKPGGGKPGGEGKPGSGKPGGEGSGKPGGKPE GGSGGKPGGKPEGGSGGKPGGSGKPGGKPGEGGKPEGGSGGKPGGSGKP GGKPEGGGSGKPGGKPGEGGKPGSGEGGKPGGGKPGGEGKPGSGKPGGE GSGKPGGKPGSGGEGKPGGKPEGGSGGKPGGGKPGGEGKPGSGGKPGEG GKPGSGGGKPGGKPGGEGEGKPGGKPGEGGKPGGEGSGKPGGGGKPGGK PGGEGGKPEGSGKPGGGSGKPGGKPEGGGGKPEGSGKPGGGGKPEGSGKP GGGKPEGGSGGKPGGSGKPGGKPGEGGGKPEGSGKPGGGSGKPGGKPEG GGKPEGGSGGKPGGKPEGGSGGKPGGKPGGEGSGKPGGKPGSGEGGKPG GKPGEGSGGKPGGKPEGGSGGKPGGSGKPGGKPEGGGSGKPGGKPGEGG KPGGEGSGKPGGSGKPG	
W576	GGSGKPGKPGGSGSGKPGSGKPGGGSGKPGSGKPGGGSGKPGSGKPGGGS GKPGSGKPGGGGKPGSGSGKPGGGKPGGSGGKPGGGSGKPGKPGSGGSG KPGSGKPGGGSGGKPGKPGSGGSGGKPGKPGSGGGSGKPGKPGSGGSGGK PGKPGSGGSGGKPGKPGSGGSGKPGSGKPGGGSGKPGSGKPGSGGSGKPG KPGSGGSGKPGSGKPGSGSGKPGSGKPGGGSGKPGSGKPGSGGSGKPGKP GSGGGKPGSGSGKPGGGKPGSGSGKPGGGKPGGSGGKPGGSGGKPGKPGS GGGSGKPGKPGSGGGSGKPGKPGGSGSGKPGSGKPGGGSGKPGSGKPGSG GSGKPGKPGSGGSGGKPGKPGSGGGKPGSGSGKPGGGKPGSGSGKPGGGK PGSGSGKPGGGKPGSGSGKPGGSGKPGSGKPGGGSGGKPGKPGSGGSGKP GSGKPGSGGSGKPGKPGGSGSGKPGSGKPGGGSGKPGSGKPGGGSGKPGS GKPGGGSGKPGSGKPGGGGKPGSGSGKPGGSGGKPGKPGSGGSGGKPGKP GSGGSGKPGSGKPGGGSGGKPGKPGSGG	23.4
Y576	GEGSGEGSEGEGSEGSGEGEGSEGSGEGEGGSEGSEGEGSEGSGEGEGGEG SGEGEGSGEGSEGEGGGEGSEGEGSGEGGEGEGSEGGSEGEGGSEGGEGE GSEGSGEGEGSEGGSEGEGSEGGSEGEGSEGSGEGEGSEGSGEGEGSEGSG EGEGSEGSGEGEGSEGGSEGEGGSEGSEGEGSGEGSEGEGGSEGSEGEGGG EGSEGEGSGEGSEGEGGSEGSEGEGGSEGSEGEGGEGSGEGEGSEGSGEGE GSGEGSEGEGSEGSGEGEGSEGSGEGEGGSEGSEGEGSGEGSEGEGSEGSG EGEGSEGSGEGEGGSEGSEGEGGSEGSEGEGGSEGSEGEGGEGSGEGEGSE GSGEGEGSGEGSEGEGSEGSGEGEGSEGSGEGEGGSEGSEGEGSEGSGEGE GGEGSGEGEGSGEGSEGEGGGEGSEGEGSEGSGEGEGSEGSGEGEGSEGGS EGEGGSEGSEGEGSEGGSEGEGSEGGSEGEGSEGSGEGEGSEGSGEGEGSG EGSEGEGGSEGGEGEGSEGGSEGEGSEGGSEGEGGEGSGEGEGGGEGSEGE GSEGSGEGEGSGEGSE	15.7
AD576	GSSESGSSEGGPGSGGEPSESGSSGSSESGSSEGGPGSSESGSSEGGPGSSESG SSEGGPGSSESGSSEGGPGSSESGSSEGGPGESPGGSSGSESGSEGSSGPGESS GSSESGSSEGGPGSSESGSSEGGPGSSESGSSEGGPGSGGEPSESGSSGESPG GSSGSESGESPGGSSGSESGSGGEPSESGSSGSSESGSSEGGPGSGGEPSESGS SGSGGEPSESGSSGSEGSSGPGESSGESPGGSSGSESGSGGEPSESGSSGSGG EPSESGSSGSGGEPSESGSSGSSESGSSEGGPGESPGGSSGSESGESPGGSSGS ESGESPGGSSGSESGESPGGSSGSESGESPGGSSGSESGSSESGSSEGGPGSG GEPSESGSSGSEGSSGPGESSGSSESGSSEGGPGSGGEPSESGSSGSSESGSSE GGPGSGGEPSESGSSGESPGGSSGSESGESPGGSSGSESGSSESGSSEGGPGS GGEPSESGSSGSSESGSSEGGPGSGGEPSESGSSGSGGEPSESGSSGESPGGSS GSESGSEGSSGPGESSGSSESGSSEGGPGSEGSSGPGESS	13.6
AE576	AGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTE PSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSET PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAG SPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSA PGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSES ATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSET PGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTE PSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSA PGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTE EGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP	6.1
AF540	GSTSSTAESPGPGSTSSTAESPGPGSTSESPSGTAPGSTSSTAESPGPGSTSST	8.8

Seq Name	Amino Acid Sequence	Score
	AESPGPGTSTPESGSASPGSTSESPSGTAPGTSPSGESSTAPGSTSESPSGTAP GSTSESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGSTSESPSGTAPGTSPSG ESSTAPGSTSESPSGTAPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSASP GSTSESPSGTAPGTSTPESGSASPGSTSSTAESPGPGSTSSTAESPGPGTSTPES GSASPGTSTPESGSASPGSTSESPSGTAPGTSTPESGSASPGTSTPESGSASPG STSESPSGTAPGSTSESPSGTAPGSTSESPSGTAPGSTSSTAESPGPGTSTPESG SASPGTSTPESGSASPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGS TSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSPSGESSTAPGSTSSTAES PGPGTSPSGESSTAPGSTSSTAESPGPGTSTPESGSASPGSTSESPSGTAP	
AF504	GASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPS GATGSPGSNPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGS PGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTP SGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSNPSASTGT GPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTGSPGAS PGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSST GSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTGSPGA SPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPGTSS TGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGS SPSASTGTGPGASPGTSSTGSP	7.0
AE864	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETP GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGS PTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESA TPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETP GTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAP GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEE GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPAT SGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP GSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGS PTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGP GTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEP SEGSAP	6.1
AF864	GSTSESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGSTSESPSGTAPGTSTPES GSASPGTSTPESGSASPGSTSESPSGTAPGSTSESPSGTAPGTSPSGESSTAPG STSESPSGTAPGTSPSGESSTAPGTSPSGESSTAPGSTSSTAESPGPGTSPSGES STAPGTSPSGESSTAPGSTSSTAESPGPGTSTPESGSASPGTSTPESGSASPGS TSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGSTSSTAESPGPGTSTPESGS ASPGSTSESPSGTAPGTSPSGESSTAPGSTSSTAESPGPGTSPSGESSTAPGTS TPESGSASPGSTSSTAESPGPGSTSSTAESPGPGSTSSTAESPGPGSTSSTAESP GPGTSPSGESSTAPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGPXXXGAS ASGAPSTXXXXSESPSGTAPGSTSESPSGTAPGSTSESPSGTAPGSTSESPSGT APGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSPSGESSTAPGTSP SGESSTAPGSTSSTAESPGPGTSPSGESSTAPGTSTPESGSASPGSTSESPSGT APGSTSESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGTSTPESGSASPGTST PESGSASPGSTSESPSGTAPGTSTPESGSASPGSTSSTAESPGPGSTSESPSGT APGSTSESPSGTAPGTSPSGESSTAPGSTSSTAESPGPGTSPSGESSTAPGTST PESGSASPGTSPSGESSTAPGTSPSGESSTAPGTSPSGESSTAPGSTSSTAESP GPGSTSSTAESPGPGTSPSGESSTAPGSSPSASTGTGPGSSTPSGATGSPGSST PSGATGSP	7.5
AG868	GGSPGASPGTSSTGSPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPG SSTPSGATGSPGSNPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSG ATGSPGTPGSGTASSSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSP GSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSNPS ASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGATGSPGASPGTSSTG	7.5

Seq Name	Amino Acid Sequence	Score
	SPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASP GTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGTSSTG SPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASP GTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATG SPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASP GTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASS SPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSST PSGATGSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSST GSPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGA SPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSPSASTGTGPGASPGTSS TGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPG ASPGTSSTGSP	
AM875	GTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGSTSSTAESPGPGTSTPE SGSASPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSTPESGSASP GSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGTSESA TPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAP GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSTEP SEGSAPGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEE GSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTSTEPSEGSAPGTSTEP SEGSAPGSEPATSGSETPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAP GASASGAPSTGGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSTSST AESPGPGSTSESPSGTAPGTSPSGESSTAPGTPGSGTASSSPGSSTPSGATGSP GSSPSASTGTGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGSTSST AESPGPGSTSSTAESPGPGTSPSGESSTAPGSEPATSGSETPGSEPATSGSETP GTSTEPSEGSAPGSTSSTAESPGPGTSTPESGSASPGSTSESPSGTAPGTSTEP SEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSSTPSGATGSPGSSPSASTGTGP GASPGTSSTGSPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSSTPS GATGSPGSSPSASTGTGPGASPGTSSTGSPGTSESATPESGPGTSTEPSEGSAP GTSTEPSEGSAP	4.5
AM1318	GTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEEGSTSSTAESPGPGTSTPE SGSASPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSTPESGSASP GSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGTSTEPSEGSAPGTSESA TPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAP GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSTEP SEGSAPGTSESATPESGPGTSTEPSEGSAPGSEPATSGSETPGSPAGSPTSTEE GSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGTSTEPSEGSAPGTSTEP SEGSAPGSEPATSGSETPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAP GPEPTGPAPSGGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGTSESAT PESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGSPAGSPTSTEEG SPAGSPTSTEEGSTSSTAESPGPGSTSESPSGTAPGTSPSGESSTAPGSTSESPS GTAPGSTSESPSGTAPGTSPSGESSTAPGTSTEPSEGSAPGTSESATPESGPGT SESATPESGPGSEPATSGSETPGTSESATPESGPGTSESATPESGPGTSTEPSE GSAPGTSESATPESGPGTSTEPSEGSAPGTSPSGESSTAPGTSPSGESSTAPGT SPSGESSTAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGSSPSAST GTGPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPG ASPGTSSTGSPGASASGAPSTGGTSPSGESSTAPGSTSSTAESPGPGTSPSGES STAPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSSPSASTGTGPGS STPSGATGSPGASPGTSSTGSPGTSTPESGSASPGTSPSGESSTAPGTSPSGES STAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGSTSESPSGTAPGS TSESPSGTAPGTSTPESGSASPGSPAGSPTSTEEGTSESATPESGPGTSTEPSE GSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGSSTPSGATGSPG ASPGTSSTGSPGSSTPSGATGSPGSTSESPSGTAPGTSPSGESSTAPGSTSSTA ESPGPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTASSSPGSPAGSPTSTEEG SPAGSPTSTEEGTSTEPSEGSAP	4.5

[00396] Example 34: Calculation of TEPITOPE scores

[00397] TEPITOPE scores of 9mer peptide sequence can be calculated by adding pocket potentials as described by Sturniolo [Sturniolo, T., et al. (1999) Nat Biotechnol, 17: 555]. In the present Example, separate Tepitope scores were calculated for individual HLA alleles. Table 29 shows as an example the pocket potentials for HLA*0101B, which occurs in high frequency in the Caucasian population. To calculate the TEPITOPE score of a peptide with sequence P1-P2-P3-P4-P5-P6-P7-P8-P9, the corresponding individual pocket potentials in Table 29 were added. The HLA*0101B score of a 9mer peptide with the sequence FDKLPRTSG is the sum of 0, -1.3, 0, 0.9, 0, -1.8, 0.09, 0, 0.

[00398] To evaluate the TEPITOPE scores for long peptides one can repeat the process for all 9mer subsequences of the sequences. This process can be repeated for the proteins encoded by other HLA alleles. Tables 30-33 give pocket potentials for the protein products of HLA alleles that occur with high frequency in the Caucasian population.

[00399] TEPITOPE scores calculated by this method range from approximately -10 to +10. However, 9mer peptides that lack a hydrophobic amino acid (FKLMVWY) in P1 position hafve calculated TEPITOPE scores in the range of -1009 to -989. This value is biologically meaningless and reflects the fact that a hydrophobic amino acid serves as an anchor residue for HLA binding and peptides lacking a hydrophobic residue in P1 are considered non binders to HLA. Because most XTEN sequences lack hydrophobic residues, all combinations of 9mer subsequences will have TEPITOPEs in the range in the range of -1009 to -989. This method confirms that XTEN polypeptides may have few or no predicted T-cell epitopes.

Table 29: Pocket potential for HLA*0101B allele.

Amino Acid	P1	P2	P3	P4	P5	P6	P7	P8	P9
A	-999	0	0	0	-	0	0	-	0
C	-999	0	0	0	-	0	0	-	0
D	-999	-1.3	-1.3	-2.4	-	-2.7	-2	-	-1.9
E	-999	0.1	-1.2	-0.4	-	-2.4	-0.6	-	-1.9
F	0	0.8	0.8	0.08	-	-2.1	0.3	-	-0.4
G	-999	0.5	0.2	-0.7	-	-0.3	-1.1	-	-0.8
H	-999	0.8	0.2	-0.7	-	-2.2	0.1	-	-1.1
I	-1	1.1	1.5	0.5	-	-1.9	0.6	-	0.7
K	-999	1.1	0	-2.1	-	-2	-0.2	-	-1.7
L	-1	1	1	0.9	-	-2	0.3	-	0.5
M	-1	1.1	1.4	0.8	-	-1.8	0.09	-	0.08
N	-999	0.8	0.5	0.04	-	-1.1	0.1	-	-1.2
P	-999	-0.5	0.3	-1.9	-	-0.2	0.07	-	-1.1
Q	-999	1.2	0	0.1	-	-1.8	0.2	-	-1.6
R	-999	2.2	0.7	-2.1	-	-1.8	0.09	-	-1

Amino Acid	P1	P2	P3	P4	P5	P6	P7	P8	P9
S	-999	-0.3	0.2	-0.7	-	-0.6	-0.2	-	-0.3
T	-999	0	0	-1	-	-1.2	0.09	-	-0.2
V	-1	2.1	0.5	-0.1	-	-1.1	0.7	-	0.3
W	0	-0.1	0	-1.8	-	-2.4	-0.1	-	-1.4
Y	0	0.9	0.8	-1.1	-	-2	0.5	-	-0.9

Table 30: Pocket potential for HLA*0301B allele.

Amino acid	P1	P2	P3	P4	P5	P6	P7	P8	P9
A	-999	0	0	0	-	0	0	-	0
C	-999	0	0	0	-	0	0	-	0
D	-999	-1.3	-1.3	2.3	-	-2.4	-0.6	-	-0.6
E	-999	0.1	-1.2	-1	-	-1.4	-0.2	-	-0.3
F	-1	0.8	0.8	-1	-	-1.4	0.5	-	0.9
G	-999	0.5	0.2	0.5	-	-0.7	0.1	-	0.4
H	-999	0.8	0.2	0	-	-0.1	-0.8	-	-0.5
I	0	1.1	1.5	0.5	-	0.7	0.4	-	0.6
K	-999	1.1	0	-1	-	1.3	-0.9	-	-0.2
L	0	1	1	0	-	0.2	0.2	-	-0
M	0	1.1	1.4	0	-	-0.9	1.1	-	1.1
N	-999	0.8	0.5	0.2	-	-0.6	-0.1	-	-0.6
P	-999	-0.5	0.3	-1	-	0.5	0.7	-	-0.3
Q	-999	1.2	0	0	-	-0.3	-0.1	-	-0.2
R	-999	2.2	0.7	-1	-	1	-0.9	-	0.5
S	-999	-0.3	0.2	0.7	-	-0.1	0.07	-	1.1
T	-999	0	0	-1	-	0.8	-0.1	-	-0.5
V	0	2.1	0.5	0	-	1.2	0.2	-	0.3
W	-1	-0.1	0	-1	-	-1.4	-0.6	-	-1
Y	-1	0.9	0.8	-1	-	-1.4	-0.1	-	0.3

Table 31: Pocket potential for HLA*0401B allele.

Amino acid	P1	P2	P3	P4	P5	P6	P7	P8	P9
A	-999	0	0	0	-	0	0	-	0
C	-999	0	0	0	-	0	0	-	0
D	-999	-1.3	-1.3	1.4	-	-1.1	-0.3	-	-1.7
E	-999	0.1	-1.2	1.5	-	-2.4	0.2	-	-1.7
F	0	0.8	0.8	-0.9	-	-1.1	-1	-	-1
G	-999	0.5	0.2	-1.6	-	-1.5	-1.3	-	-1
H	-999	0.8	0.2	1.1	-	-1.4	0	-	0.08
I	-1	1.1	1.5	0.8	-	-0.1	0.08	-	-0.3
K	-999	1.1	0	-1.7	-	-2.4	-0.3	-	-0.3
L	-1	1	1	0.8	-	-1.1	0.7	-	-1

Amino acid	P1	P2	P3	P4	P5	P6	P7	P8	P9
M	-1	1.1	1.4	0.9	-	-1.1	0.8	-	-0.4
N	-999	0.8	0.5	0.9	-	1.3	0.6	-	-1.4
P	-999	-0.5	0.3	-1.6	-	0	-0.7	-	-1.3
Q	-999	1.2	0	0.8	-	-1.5	0	-	0.5
R	-999	2.2	0.7	-1.9	-	-2.4	-1.2	-	-1
S	-999	-0.3	0.2	0.8	-	1	-0.2	-	0.7
T	-999	0	0	0.7	-	1.9	-0.1	-	-1.2
V	-1	2.1	0.5	-0.9	-	0.9	0.08	-	-0.7
W	0	-0.1	0	-1.2	-	-1	-1.4	-	-1
Y	0	0.9	0.8	-1.6	-	-1.5	-1.2	-	-1

Table 32: Pocket potential for HLA*0701B allele.

Amino acid	P1	P2	P3	P4	P5	P6	P7	P8	P9
A	-999	0	0	0	-	0	0	-	0
C	-999	0	0	0	-	0	0	-	0
D	-999	-1.3	-1.3	-1.6	-	-2.5	-1.3	-	-1.2
E	-999	0.1	-1.2	-1.4	-	-2.5	0.9	-	-0.3
F	0	0.8	0.8	0.2	-	-0.8	2.1	-	2.1
G	-999	0.5	0.2	-1.1	-	-0.6	0	-	-0.6
H	-999	0.8	0.2	0.1	-	-0.8	0.9	-	-0.2
I	-1	1.1	1.5	1.1	-	-0.5	2.4	-	3.4
K	-999	1.1	0	-1.3	-	-1.1	0.5	-	-1.1
L	-1	1	1	-0.8	-	-0.9	2.2	-	3.4
M	-1	1.1	1.4	-0.4	-	-0.8	1.8	-	2
N	-999	0.8	0.5	-1.1	-	-0.6	1.4	-	-0.5
P	-999	-0.5	0.3	-1.2	-	-0.5	-0.2	-	-0.6
Q	-999	1.2	0	-1.5	-	-1.1	1.1	-	-0.9
R	-999	2.2	0.7	-1.1	-	-1.1	0.7	-	-0.8
S	-999	-0.3	0.2	1.5	-	0.6	0.4	-	-0.3
T	-999	0	0	1.4	-	-0.1	0.9	-	0.4
V	-1	2.1	0.5	0.9	-	0.1	1.6	-	2
W	0	-0.1	0	-1.1	-	-0.9	1.4	-	0.8
Y	0	0.9	0.8	-0.9	-	-1	1.7	-	1.1

Table 33: Pocket potential for HLA*1501B allele.

Amino acid	P1	P2	P3	P4	P5	P6	P7	P8	P9
A	-999	0	0	0	-	0	0	-	0
C	-999	0	0	0	-	0	0	-	0
D	-999	-1.3	-1.3	-0.4	-	-0.4	-0.7	-	-1.9
E	-999	0.1	-1.2	-0.6	-	-1	-0.7	-	-1.9
F	-1	0.8	0.8	2.4	-	-0.3	1.4	-	-0.4

Amino acid	P1	P2	P3	P4	P5	P6	P7	P8	P9
G	-999	0.5	0.2	0	-	0.5	0	-	-0.8
H	-999	0.8	0.2	1.1	-	-0.5	0.6	-	-1.1
I	0	1.1	1.5	0.6	-	0.05	1.5	-	0.7
K	-999	1.1	0	-0.7	-	-0.3	-0.3	-	-1.7
L	0	1	1	0.5	-	0.2	1.9	-	0.5
M	0	1.1	1.4	1	-	0.1	1.7	-	0.08
N	-999	0.8	0.5	-0.2	-	0.7	0.7	-	-1.2
P	-999	-0.5	0.3	-0.3	-	-0.2	0.3	-	-1.1
Q	-999	1.2	0	-0.8	-	-0.8	-0.3	-	-1.6
R	-999	2.2	0.7	0.2	-	1	-0.5	-	-1
S	-999	-0.3	0.2	-0.3	-	0.6	0.3	-	-0.3
T	-999	0	0	-0.3	-	-0	0.2	-	-0.2
V	0	2.1	0.5	0.2	-	-0.3	0.3	-	0.3
W	-1	-0.1	0	0.4	-	-0.4	0.6	-	-1.4
Y	-1	0.9	0.8	2.5	-	0.4	0.7	-	-0.9

[00400] Example 35: Pharmacokinetics of GFP-XTEN and FITC-XTEN Conjugates

[00401] The pharmacokinetics of the GFP-XTEN and FITC-XTEN cross-linked conjugates prepared as described in the Examples above were tested in cynomolgus monkeys. GFP-XTEN and FITC-XTEN were administered to male cynos IV at 2 mg/kg and dose volumes of 0.77 and 0.68 mL respectively. Blood samples (1.0 mL) were collected into prechilled heparinized tubes at predose, 2, 4, 8, 24, 48, 72, 96, 120, 168, 216, 264, 336, 388, 432, 504 hour time points, and processed into plasma. Quantitation was performed by ELISA assay using the anti-XTEN antibody for both capture and detection in the case of GFP-XTEN and anti-XTEN capture and anti-FITC detection in the case of FITC-XTEN. A non-compartmental analysis was performed in WinNonLin with all time points included in the fit to determine the PK parameters. The pharmacokinetic results are summarized in Table 34 and FIG. 27. The data show XTEN can extend the half-life of molecules to which it is chemically conjugated in a manner comparable to genetic fusions to payloads of similar size.

Table 34: PK parameters of conjugated XTEN-payload compositions.

Construct	Cmax (ng/mL)	AUC (hr*ng/mL)	T ½ (hrs)
GFP-X-XTEN (AP197d)	52800	8220000	107
FITC-X-XTEN AP197e	18900	3930000	84.2

[00402] The pharmacokinetics of GFP-L288, GFP-L576, GFP-XTEN_AF576, GFP-XTEN_Y576 and XTEN_AD836-GFP were tested in cynomolgus monkeys to determine the effect of composition and length of the unstructured polypeptides on PK parameters. Blood samples were analyzed at various times after injection and the concentration of GFP in plasma

was measured by ELISA using a polyclonal antibody against GFP for capture and a biotinylated preparation of the same polyclonal antibody for detection. Results are summarized in FIG. 28. They show a surprising increase of half-life with increasing length of the XTEN sequence. For example, a half-life of 10 h was determined for GFP-XTEN_L288 (with 288 amino acid residues in the XTEN). Doubling the length of the unstructured polypeptide fusion partner to 576 amino acids increased the half-life to 20-22 h for multiple fusion protein constructs; i.e., GFP-XTEN_L576, GFP-XTEN_AF576, GFP-XTEN_Y576. A further increase of the unstructured polypeptide fusion partner length to 836 residues resulted in a half-life of 72-75 h for XTEN_AD836-GFP. Thus, increasing the polymer length by 288 residues from 288 to 576 residues increased in vivo half-life by about 10 h. However, increasing the polypeptide length by 260 residues from 576 residues to 836 residues increased half-life by more than 50 h. These results show that there is a surprising threshold of unstructured polypeptide length that results in a greater than proportional gain in in vivo half-life. Thus, fusion proteins comprising extended, unstructured polypeptides are expected to have the property of enhanced pharmacokinetics compared to polypeptides of shorter lengths.

Table 35: XTEN-cleavage sequence-affinity tag Polypeptide Sequences

Construct	Amino Acid Sequence
AC710	APTTAGAGRPRPRPRPRPRPRPRPRPRPGRGSPGSPAGSPTSTEEGTSESATPESGP GTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGP GSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP GTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGP GTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAP GTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETP GTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAP GTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGP GSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGP GTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETP GTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGP GTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP GRHHHHHHHHH
AC698	ANTPVSGNLKVEFYNSNPSDTTNSINPQFKVTNTGSSAIDLSKLTLRYYYTVDGQKD QTFWADHAAIIGSNGSYNGITSNVKGTFVKMSSSTNNADTYLEISFTGGTLEPGAHV QIQGRFAKNDWSNYTQSNDYSFKSASQFVEWDQVTAYLNGVLVWGKEPGGSVVG SGSGSGRGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEG TSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPG SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEG TSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPG SEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPG SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPG SPAGSPTSTEEGTSTEPSEGSAPTAEAAGKPGTAEAAGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPE SGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEG SAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPE SGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPE SGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPE

Construct	Amino Acid Sequence
	SGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEG SAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGKATHHHH HHHH
AC815	KNPEQAEEQAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPTAEAAGCGTA EAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSP TSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPS EGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATS GSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSP TSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPS EGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGS PTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPAT SGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGS PTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPAT SGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGS PTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGS PTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPAT SGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPAT SGSETPGTSESATPESGPGTSTEPSEGSAPSASRSAHHHHHHHHH
AC816	KNPEQAEEQAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPTAEAAGCGTA EAAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSP TSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPS EGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATS GSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSP TSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPS EGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGS PTSTEEGTSTEPSEGSAPTAEAAGCGTAEAAGTSESATPESGPGSEPATSGSETPGTS ESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSP AGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTS ESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTS TEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSP AGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTS ESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTS TEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPTAEAAGCGTAEAA SASRSAHHHHHHHHH
AC763	KNPEQAEEQAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPTAEAAGCGTA EAAGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATP ESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSE GSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATP ESGPGTSTEPSEGSAPTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSES ATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPTAEAAGCGTAEAAS ASRSAHHHHHHHHH
AC767	KNPEQAEEQAEEQREETRPRPRPRPRPRPRPRPRPRPRPSASRSAGSPTGPGSEPATS GSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSP TSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATS GSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATS GSETPGTSESATPESGPGTSTEPSEGSAPSASRSAHHHHHHHHH
AC769	KNPEQAEEQAEEQREETRPRPRPRPRPRPRPRPRPRPRPSASRSAGSPTAEAAGCGTA EAAPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTAEAAGCGTAEAASTEP SEGSAPGTSESATPESGPGSPAGSPTSTEEGSPATAEAAGCGTAEAASPTSTEEGTSE SATPESGPGTSTEPSEGSAPGTSESATTAEAAGCGTAEAASETPGTSESATPESGPGS EPATSGSETPGTSESATPESGTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGP GSEPATSGSETPGTTAEAAGCGTAEAAAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGS APGTSESTAEAAGCGTAEAATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSG TAEAAGCGTAEAATEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPTAEAAG

Construct	Amino Acid Sequence
	CGTAEAASASRSAHHHHHHHH
AC771	KNPEQAEEQAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPGSPAGSPTSTE EGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSA PGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPESG PGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSA PGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSA PGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESG PGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSA PGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSA PTAEAAGCGTAEAAPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTAEAAG CGTAEAASTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPATAEAAGCGTAEAA SPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATTAEAAGCGTAEAASETPGTSE SATPESGPGSEPATSGSETPGTSESATPESGTAEAAGCGTAEAAGSPAGSPTSTEEGT SESATPESGPGSEPATSGSETPGTTAEAAGCGTAEAAAGSPTSTEEGSPAGSPTSTEE GTSTEPSEGSAPGTSESTAEAAGCGTAEAATPESGPGTSESATPESGPGSEPATSGSE TPGSEPATSGTAEAAGCGTAEAATEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSG SETPTAEAAGCGTAEAASASRSAHHHHHHHH
AC817	KKQEQEEKKAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPGSPAGSPTST EEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGS APGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPES GPGTESASRSAHHHHHHHH
AC818	KKQEQEEKKAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPGSPAGSPTST EEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGS APGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPES GPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGS APGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGS APGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPES GPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGS APGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTESASRSA HHHHHHHHH
AC780	KKQEQEEKKAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPTEEGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPS EGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPS EGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGS PTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESA TPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEP SEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPAT SGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESA TPESGPGTSTEPSEGSAPSASRSAHHHHHHHH
AC785	KKQEQEEKKAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPTEGTSTEPSE GSAPGTSESTAEAAGCGTAEAATPESGPGTSESATPESGPGSEPATSGSETPGSEPAT SGTAEAAGCGTAEAATEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPTAEA AGCGTAEAASASRSAHHHHHHHH
AC765	KNPEQAEEQAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPTPGTSESATPE SGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEG SAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPE SGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPE SGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTS TEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPTAEAAGCGTAEAASASRSA HHHHHHHH
AC766	KNPEQAEEQAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPTGSEPATSGSE TPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPES GPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS APGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPES

Construct	Amino Acid Sequence
	GPGTSTEPSEGSAPTAEAAGCGTAEAAGSPAGSPTSTEEGTSESATPESGPGSEPATS GSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESTA EAAGCGTAEAATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAG SPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPTAEAAGCGTAEAASAS RSAHHHHHHHHH
AC768	KNPEQAEEQAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPTPGTSESATPE SGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEG SAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPE SGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESTAEA AGCGTAEAATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGTAEAAGCGTA EAATEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPTAEAAGCGTAEAASAS RSAHHHHHHHHH
AC786	KKQEQEEKKAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPGSPAGSPTST EEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGS APGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPES GPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGS APGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGS APGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPES GPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGS APGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGS APGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTST EEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPES GPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPES GPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGS APGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSE TPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPES GPGTESASRSAHHHHHHHHH
AC792	KKQEQEEKKAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPGSTSSTAESP GPGSTSSTAESPGPGCTSESPSGTAPGSTSSTAESPGPGSTSSTAESPGPGTSTPESGSA SPGSTSCSPSGEAPGTSPSGESSTAPGSTSESPSGTAPGSTSESPSGTAPETSPSGESCT APGSTSASRSAHHHHHHHHH
AC793	KKQEQEEKKAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPGTPGSGTASS SPGSSTPSGATGSPGCAGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGAT GSPGSSTCSGATGSPGSSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTA CSSPGSSSASRSAHHHHHHHHH
AC798	KKQEQEEKKAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPGSPAGSPTST EEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGS APGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPES GPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGS APGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGS APGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPES GPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGS APGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGS APGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTST EEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPES GPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPES GPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGS APGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSE TPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPES GPGTESASKSAHHHHHHHHH
AC809	KNPEQAEEQAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPTGPGSEPATS GSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSP TSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATS GSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATS

Construct	Amino Acid Sequence
	GSETPGTSESATPESGPGTSTACSEGSAPSASRSAHHHHHHHHH
AC810	KNPEQAEEQAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPTGPGSEPATS GSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSP TSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATS GSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATS GSETPGTSESATPESGPGTSTEPSEGSCASASRSAHHHHHHHHH
AC831	KKQEQEEKKAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPGSTSSTAESP GPGSTSSTAESPGPGSTSESPSGTAPGSTSSTAESPGPGSTSSTAESPGPGTSTPESGSA SPGSTSESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGSTSESPSGTAPGTSPSGESST APGSTSESPSGTAPGSTSESPSGTAPGTSPSGESSTAPGSTSESPSGTAPGSTSESPSGT APGSTSESPSGTAPGTSTPESGSASPGSTSESPSGTAPGTSTPESGSASPGSTSSTAESP GPGSTSSTAESPGPGTSTPESGSASPGTSTPESGSASPGSTSESPSGTAPGTSTPESGSA SPGTSTPESGSASPGSTSESPSGTAPGSTSESPSGTAPGSTSESPSGTAPGSTSSTAESP GPGTSTPESGSASPGTSTPESGSASPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSA SPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSPSGESSTAPGSTSSTAESP GPGTSPSGESSTAPGSTSSTAESPGPGTSTPESGSASPGSTSESPSGTAPGSTSSTAESP GPGTSTPESGSASPGTSTPESGSASPGSTSSTAESPGPGTSPSGESSTAPGTSTPESGSA SPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSPSGESSTAPGSTSSTAESP GPGTSPSGESSTAPGSTSSTAESPGPGSTSSTAESPGPGTSPSGESSTAPGTSESPSGT APGSTSESPSGTAPGTSTPESGSASPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSA SPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSPSGESSTAPGSTSSTAESP GPGTSPSGESSTAPGSTSSTAESPGPGSTSSTAESPGPGTSPSGESSTAPGSTSESPSGT APGSTSESPSGTAPGTSTPESGSASPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSA SPGSTSESPSGTAPGSTSESPSGTAPGTSTPESGSASPGTSPSGESSTAPGTSSASRSAHHHHHHHHH
AC832	KKQEQEEKKAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPGASPGTSSTG SPGSSPSASTGTGPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGSPGSSPSASTGT GPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGASPGTSST GSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGASPGTSSTGSPGTPGSGTA SSSPGSSTPSGATGSPGSSPSASTGTGPGSSPSASTGTGPGSSTPSGATGSPGSSTPSGA TGSPGASPGTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTS STGSPGASPGTSSTGSPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGASPGT SSTGSPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSPGSSTPSGATGSPGASPG TSSTGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGSSTPSGATGSPGSSPS ASTGTGPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGASPGTSSTGSPGASP GTSSTGSPGASPGTSSTGSPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPGTP GSGTASSSPGSSTPSGATGSPGTPGSGTASSSPGSSTPSGATGSPGSSTPSGATGSPGS SPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGTPGSGTASSSPGSSTPSGATGSPG SSPSASTGTGPGSSPSASTGTGPGASPGTSSTGSPGASPGTSSTGSPGSSTPSGATGSP GSSPSASTGTGPGASPGTSSTGSPGSSPSASTGTGPGTPGSGTASSSPGSSTPSGATGS PGSSTPSGATGSPGASSASRSAHHHHHHHHH
AC819	KKQEQEEKKAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPGSCAGSPTST EEGTSESACPESGPGTSTEPSEGSCPGSPAGSPTSTEEGTCTEPSEGSAPGTSTEPCSGS APGTSESATPESCPGSEPATSGSETPGSCPATSGSETPGSPAGSCTSTEEGTSESATPES CPGTESASRSAHHHHHHHHH
AC820	KNPEQAEEQAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPTGCGSEPATS GSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSP TSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGTSCTPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATS GSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATS GSETPGTSESATPESGPGTSTEPSCGSAPSASRSAHHHHHHHHH
AC821	KNPEQAEEQAEEQREETRPRPRPRPRPRPRPRPRPRPSASRSAGSPTGCGSEPATS GSETPGTSESATPESGPGSEPATSGSCTPGTSESATPESGPGTSTEPSEGSAPGSPAGSP CSTEEGTSESATPESGPGSEPATSGSETPGTSESCTPESGPGSPAGSPTSTEEGSPAGSP TSTEEGTSCTPSEGSAPGTSESATPESGPGTSESATPESGPGCSESATPESGPGSEPATS GSETPGSEPATSGSETCGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGCAPGSEPAT SGSETPGTSESATPESGPGTSTEPSCGSAPSASRSAHHHHHHHHH

CLAIMS

WHAT IS CLAIMED IS:

1. A conjugate comprising a first and a second extended recombinant polypeptide (XTEN), wherein each of said first and second XTEN is independently selected from the group consisting of the XTEN sequences set forth in Table 1 and wherein:

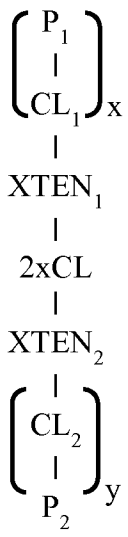
   a. the first and the second XTEN are conjugated to each other by a reaction product of an alkyne and an azide reactant, one of said alkyne and azide reactant being located at the N-terminus of the first XTEN and the other of said alkyne and azide reactant being linked to the N-terminus of the the second XTEN, wherein the alkyne and azide reactants are selected from the group consisting of the reactants set forth in Table 9;

   b. each of the first and the second XTEN comprises one or more cysteine residues and further comprises a first cross-linker conjugated to each cysteine residue of the first XTEN and a second cross-linker conjugated to each cysteine residue of the second XTEN, wherein the first and the second cross-linkers are independently selected from the group consisting of the cross-linkers set forth in Table 8;

   c. the conjugate comprises a first payload conjugated to each first cross-linker of the first XTEN, wherein the first payload is folic acid; and

   d. the conjugate comprises a second payload different from the first payload wherein the second payload is conjugated to each second cross-linker of the second XTEN, wherein the second payload is selected from the group consisting of auristatin E, auristatin F, monomethyl auristatin E, monomethyl auristatin F, valine-citrulline-*p*-aminobenzyloxycarbonyl-monomethyl auristatin E, and valine-citrulline-*p*-aminobenzyloxycarbonyl-monomethyl auristatin F.

2. The conjugate of claim 1 having the configuration of formula I



I

wherein independently for each occurrence :

a. P_1 is the first payload,
b. P_2 is the second payload;
c. CL_1 is the first cross-linker;
d. x is an integer from 1 to about 20;
e. CL_2 is the cross-linker;
f. y is an integer from 1 to about 20 with the proviso that x + y is ≥ 2;
g. 2xCL is the reaction product of the alkyne and azide reactants;
h. $XTEN_1$ is the first XTEN; and
i. $XTEN_2$ is the second XTEN.

3. The conjugate of claim 1 or claim 2, wherein the first XTEN and the second XTEN each independently has 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% sequence identity or is identical to a sequence selected from the group of sequences set forth in Table 1, when optimally aligned.

4. The conjugate of claim 1 or claim 2, wherein the first XTEN has 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% sequence identity or is identical, when optimally aligned, to a sequence selected from the group of sequences consisting of Seg 174, Seg 175, Seg 176, Seg 177, Seg 186, Seg 187, Seg 188, Seg 189, Seg 190, Seg 191, Seg 192, Seg 193, Seg 194, Seg 195, Seg 196, Seg 197, Seg 198, and Seg 199 set forth in Table 1 and the second

XTEN has 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% sequence identity or is identical, when optimally aligned, to a different sequence selected from the group consisting of Seg 174, Seg 175, Seg 176, Seg 177, Seg 186, Seg 187, Seg 188, Seg 189, Seg 190, Seg 191, Seg 192, Seg 193, Seg 194, Seg 195, Seg 196, Seg 197, Seg 198, and Seg 199 set forth in Table 1.

5. The conjugate of claim 1 or claim 2, wherein the first XTEN and the second XTEN are the identical are each has 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% sequence identity or is identical, when optimally aligned, to a sequence selected from the group consisting of Seg 174, Seg 175, Seg 176, Seg 177, Seg 186, Seg 187, Seg 188, Seg 189, Seg 190, Seg 191, Seg 192, Seg 193, Seg 194, Seg 195, Seg 196, Seg 197, Seg 198, and Seg 199 set forth in Table 1.

6. The conjugate of any one of claims 1-5, wherein the first cross-linker and the second cross linker are independently selected from the group consisting of N-maleimide, N-(5-aminopentyl)maleimide, 6-maleimidocaproic acid, iodoacetyl, pyridyl disulfide and vinyl sulfone, 3-propargyloxypropanoic acid, $(oxyethyl)_n$-acetylene where n is 1-10, dibenzylcyclooctyne (DBCO), cyclooctyne (COT), 3-azide-propionic acid, 6-azide-hexanoic acid, and $(oxyethyl)_n$-azide where n is 1-10.

7. The conjugate of claim 6, wherein the first cross-linker is N-(5-aminopentyl)maleimide and the second cross linker is 6-maleimidocaproic acid.

8. The conjugate of any one of claims 1-7, wherein the alkyne reactant is selected from the group consisting of 3-propargyloxypropanoic acid NHS ester, acetylene-$(oxyethyl)_n$ -NHS ester where n is 1-10, dibenzylcyclooctyne (DBCO)-NHS ester, DBCO-$(oxyethyl)_n$ -NHS ester where n is 1-10, cyclooctyne (COT)-NHS ester, COT-$(oxyethyl)_n$ -NHS ester where n is 1-10, COT-$(oxyethyl)_n$ -pentafluorophenyl (PFP) ester where n is 1-10, BCOT-NHS ester, BCOT-$(oxyethyl)_n$ -NHS ester where n is 1-10, BCOT-$(oxyethyl)_n$ -pentafluorophenyl (PFP) ester where n is 1-10, acetylene-$(oxyethyl)_n$-maleimide where n is 1-10, DBCO-maleimide, COT-maleimide, BCOT-maleimide and the azide reactant is selected from the group consisting of 3-azide-propionic acid NHS ester, 6-azide-hexanoic acid, NHS ester, 3-azide-propionic acid PFP ester, 6-azide-

hexanoic acid PFP ester, azide-(oxyethyl)$_n$ NHS ester where n is 1-10, azide-(oxyethyl)$_n$ - PFP ester where n is 1-10, azide-(oxyethyl)$_n$-maleimide where n is 1-10.

9. The conjugate of claim 8, wherein the alkyne reactant is 6-(11,12-didehydrodibenzo[b,f]azocin-5(6H)-yl)-6-oxohexanoic acid N-hydroxysulfosuccinimide ester and wherein the azide reactant 1-azido-3,6,9,12-tetraoxapentadecan-15-oic acid.

10. The conjugate of any one of claims 1-9, wherein at least 90%, 91%, 92%, 93%, 94%, or 95% of individual first XTEN molecules in said conjugate have an identical sequence length and wherein at least 90%, 91%, 92%, 93%, 94%, or 95% of individual second XTEN molecules in said conjugate have an identical sequence length.

11. The conjugate of any one of claims 1-3, or 5-10, wherein the first XTEN and the second XTEN are identical and the sequence is Seg 176 set forth in Table 1.

12. The conjugate of any one of claims 1-3, or 5-11, wherein x is 3 and y is 3.

13. The conjugate of any one of claims 1-3, or 5-10, wherein the first XTEN and the second XTEN are identical and the XTEN sequences are Seg 177 set forth in Table 1.

14. The conjugate of claim 13, wherein x is 9 and y is 9.

15. The conjugate of claim 4, wherein the first XTEN sequence is Seg 176 set forth in Table 1 and the second XTEN sequence is Seg 177 set forth in Table 1.

16. The conjugate of claim 15, wherein x is 3 and y is 9.

17. The conjugate of any one of claims 12, 14, or 16, wherein the second payload is valine-citrulline-*p*-aminobenzyloxycarbonyl-monomethylauristatin E.

18. The conjugate of any one of claims 12, 14, or 16, wherein the second payload is valine-citrulline-*p*-aminobenzyloxycarbonyl-monomethylauristatin F.

19. The conjugate having the structure set forth in FIG. 49.

20. The conjugate of any one of claims 1-3, or 5-14, wherein treatment of a folate receptor-bearing cell with the conjugate results in greater cytotoxicity compared to treatment with a comparable conjugate lacking folate.

21. The conjugate of claim 20, wherein the cytotoxicity is measured in an *in vitro* assay using a folate receptor-bearing cell selected from the group consisting of the cells set forth in Table 24 and the greater cytotoxicity is a reduction in IC50 by at least 100 nM, 200 nM, 300 nM, 400 nM, 500 nM, 600 nM, 700 nM, 800 nM, 900 nM or 1000 nM.

22. The conjugate of any one of claims 1-18, wherein the terminal half-life of the conjugate administered as a single dose to a subject is at least 2-fold, 4-fold, 8-fold, or 10-

fold greater compared to the second payload administered singly to the subject using a comparable molar dose.

23. The conjugate of any one of claims 1-3, or 5-14, wherein the first XTEN and the second XTEN utilized in the conjugate have the same XTEN sequence and are obtained from a substantially homogenous population of polypeptides prepared by the process comprising:

a. culturing a host cell that comprises a vector encoding the polypeptide in a fermentation reaction under conditions effective to express the polypeptide as a component of a crude expression product of the host cell, wherein the encoded polypeptide comprises the XTEN, a first cleavage sequence, a first affinity tag with binding affinity to a first chromatography substrate, a second cleavage sequence, and a second affinity tag wherein the first and the second cleavage sequences are capable of being cleaved by the same protease and wherein the second affinity tag has binding affinity to a second, different chromatography substrate than the first affinity tag;
b. adsorbing the polypeptide of the crude expression product onto a first chromatography substrate under conditions effective to capture the first affinity tag onto the first chromatography substrate;
c. eluting the polypeptide;
d. adsorbing the polypeptide onto a second chromatography substrate under conditions effective to capture the second affinity tag onto the second chromatography substrate;
e. eluting the polypeptide; and
f. recovering the polypeptide wherein at least 90%, 91%, 92%, 93%, 94%, or 95% of the polypeptides of the population have an identical sequence length.

24. The conjugate of claim 23, wherein the polypeptide has a sequence selected from the group consisting of the sequences set forth in Table 35.

25. The conjugate of claim 23, wherein the first chromatography substrate is selected from the group consisting of HIC, cation exchange, anion exchange, and IMAC.

26. The conjugate of any one of claims 23-25, wherein the affinity tag is selected from the group consisting of the affinity tags of Table 5.

27. The conjugate of claim 24, wherein the first chromatography substrate is cation exchange and the first affinity tag comprises a sequence selected from RPRPRPRPRPRPR and RPRPRPRPRPRPRPRPRPRPR.

28. The conjugate of claim 24, wherein the second chromatography substrate is IMAC and the second affinity tag comprises the sequence selected from HHHHHH and HHHHHHHH.

29. The conjugate of any one of claims 23-28, the process further comprising:

a. treating the polypeptide with a protease under conditions effective to cleave the cleavage sequences, thereby releasing the XTEN from the polypeptide;
b. adsorbing the XTEN onto an anion chromatography substrate under conditions effective to capture the XTEN;
c. eluting the XTEN; and
d. recovering the XTEN wherein at least 90%, or at least 91%, or at least 92%, or at least 93%, or at least 94%, or at least 95% of the individual XTEN molecules have an identical sequence length.

30. The conjugate of claim 29, wherein the anion chromatography substrate is selected from the group consisting of macrocap Q, capto Q, superQ-650M, and poros D.

31. The conjugate of claim 29 of claim 30, wherein the cleavage sequences are capable of being cleaved by trypsin wherein the cleavage sequence is selected from the group consisting of the sequences set forth in Table 8, and the protease is trypsin.

32. The conjugate of any one of claims 23-28, the process further comprising:

a. treating the polypeptide with a protease under conditions effective to cleave the cleavage sequence(s), thereby releasing the XTEN from the polypeptide;
b. adsorbing the protease onto a chromatography substrate under conditions effective to capture the protease but not the XTEN; and
c. recovering the XTEN in the eluate wherein at least 90%, 91%, 92%, 93%, 94%, or 95% of the XTEN have an identical sequence length..

33. The conjugate of claim 32, wherein the cleavage sequence is capable of being cleaved by trypsin wherein the cleavage sequence is selected from the group consisting of the sequences set forth in Table 8, and wherein the protease is trypsin.

34. The conjugate of claims 32 or claim 33, wherein the chromatography substrate is one or more of HIC, cation exchange, and IMAC.

35. A pharmaceutical composition, comprising the conjugate of any one of the preceding claims and a pharmaceutically acceptable carrier.

36. The pharmaceutical composition of claim 35 for treatment of a cancer.

37. The pharmaceutical composition of claim 35 for use in a pharmaceutical regimen for treatment of a subject with a cancer, said regimen comprising the pharmaceutical composition.

38. The pharmaceutical composition of claim 37, wherein the pharmaceutical regimen further comprises the step of determining the amount of pharmaceutical composition needed to achieve an improvement in a parameter associated with the cancer.

39. The pharmaceutical composition of claim 37, wherein the pharmaceutical regimen for treating the subject further comprises administering the pharmaceutical composition in an effective amount in two or more successive doses to the subject at an effective amount, wherein the administration results in at least a 10%, or 20%, or 30%, or 40%, or 50%, or 60%, or 70%, or 80%, or 90% greater improvement of at least one, two, or three parameters associated with the cancer compared to an untreated subject.

40. The pharmaceutical composition of claim 39, wherein the parameters are selected from the group consisting of response rate as defined by the Response Evaluation Criteria in Solid Tumors (RECIST), time-to-progression of the cancer (relapse), discovery of local recurrence, discovery of regional metastasis, discovery of distant metastasis, onset of symptoms, hospitalization, increase in pain medication requirement, requirement of salvage chemotherapy, requirement of salvage surgery, requirement of salvage radiotherapy, time-to-treatment failure, and increased time of survival.

41. A conjugate of any one of claims 1-34 for use in the preparation of a medicament for treatment of mammalian cells or treatment of a cancer.

42. An article of manufacture comprising a conjugate of any one of claims 1-34, a container, and a package insert or label indicating that the compound can be used to treat cancer.

43. A kit comprising a pharmaceutical composition of claim 38, a container, and a package insert or label indicating that the compound can be used to treat cancer.

44. A method of treating a cancer in a subject, comprising administering to the subject a pharmaceutical composition comprising an effective amount of a composition of any one of claims 1-34.

45. The method of claim 44, wherein the pharmaceutical composition comprises the conjugate having the structure set forth in FIG. 49.

46. The method of claim 44 or claim 45, wherein the cancer is non-small cell lung cancer or mesothelioma. platinum-resistant ovarian cancer, endometrial cancer, adenocarcinoma of the lung, and refractory advanced tumors.

47. The method of any one of claims 44-46, wherein the administration results in at least a 10%, or 20%, or 30%, or 40%, or 50%, or 60%, or 70%, or 80%, or 90% greater improvement of at least one, two, or three parameters associated with the cancer compared to an untreated subject.

48. The method of claim 47, wherein the parameters are selected from the group consisting of response rate as defined by the Response Evaluation Criteria in Solid Tumors (RECIST), time-to-progression of the cancer (relapse), discovery of local recurrence, discovery of regional metastasis, discovery of distant metastasis, onset of symptoms, hospitalization, increase in pain medication requirement, requirement of salvage chemotherapy, requirement of salvage surgery, requirement of salvage radiotherapy, time-to-treatment failure, and increased time of survival.

FIG. 1A

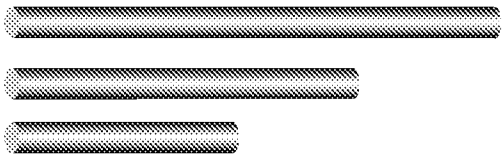


FIG. 1B

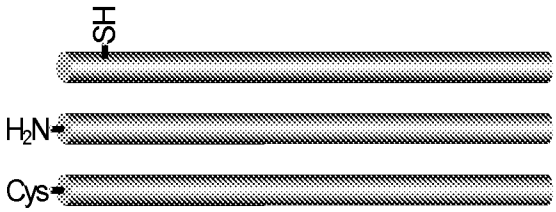


FIG. 1C

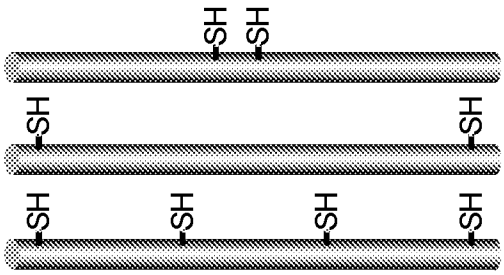


FIG. 1E

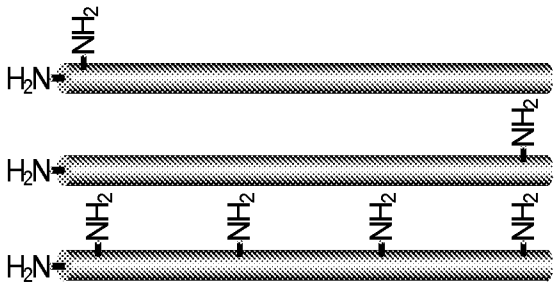


FIG. 1D

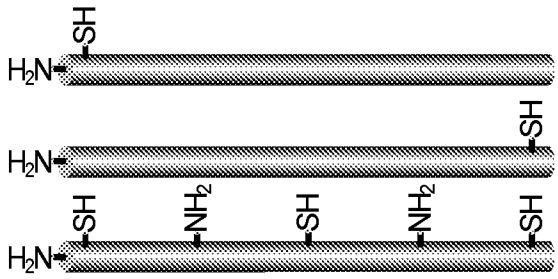


FIG. 1

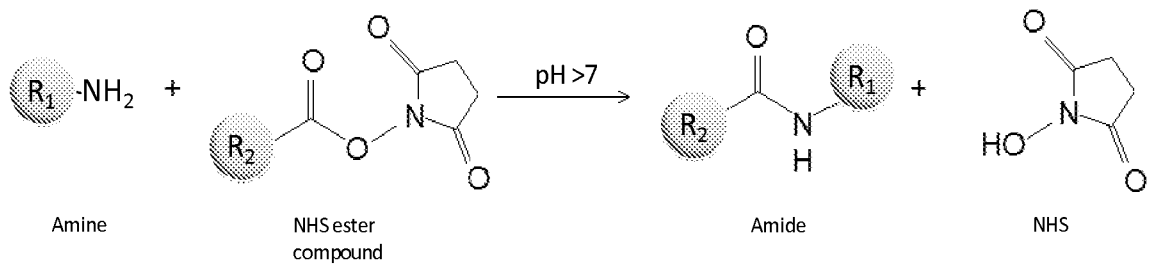

R1-NH2
+
R2
O
N
O
O
pH >7
R2
N
H
R1
+
HO
N
O
O
Amine
NHS ester compound
Amide
NHS

FIG. 2

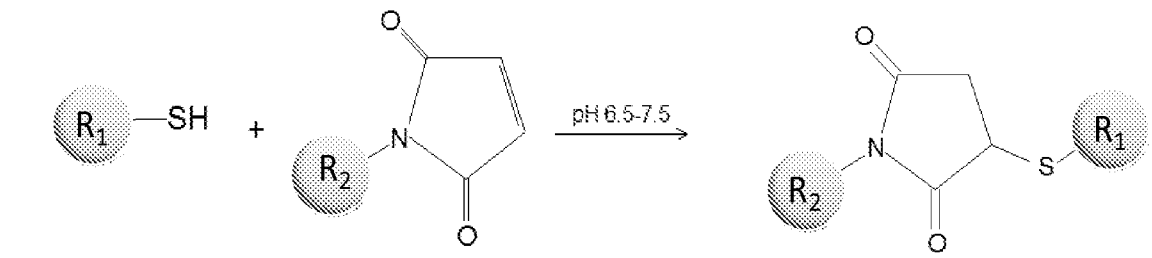

R1
SH
+
O
N
R2
O
pH 6.5-7.5
O
N
S
R1
R2
O
Sulfhydryl-containing molecule
Maleimide compound
Thioether bond

FIG. 3

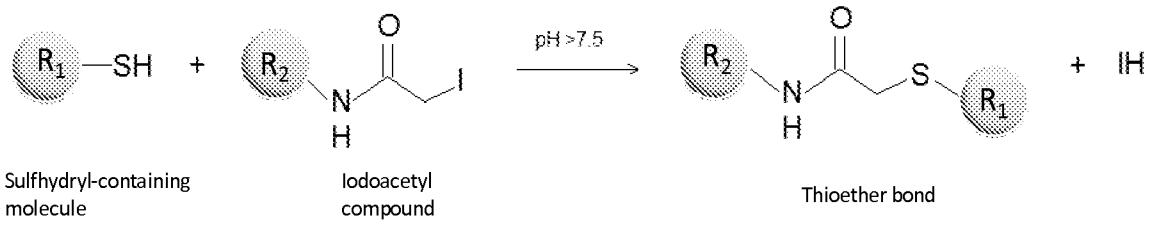

R1—SH
+
R2
O
N
H
I
pH >7.5
R2
O
N
H
S
R1
+ IH
Sulfhydryl-containing molecule
Iodoacetyl compound
Thioether bond

FIG. 4

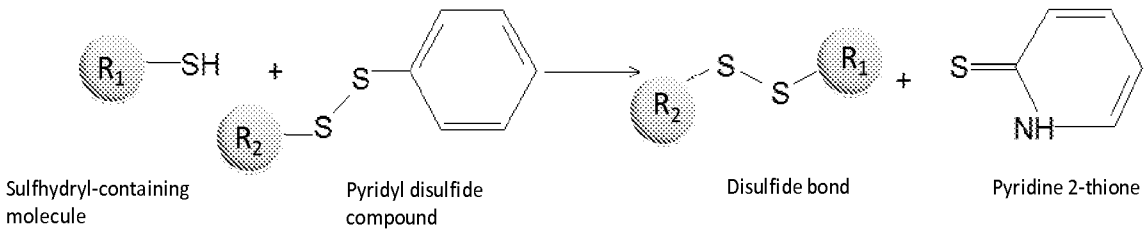

R1
SH
+
S
R2
S
R2
S
S
R1
+
S
NH
Sulfhydryl-containing molecule
Pyridyl disulfide compound
Disulfide bond
Pyridine 2-thione

FIG. 5

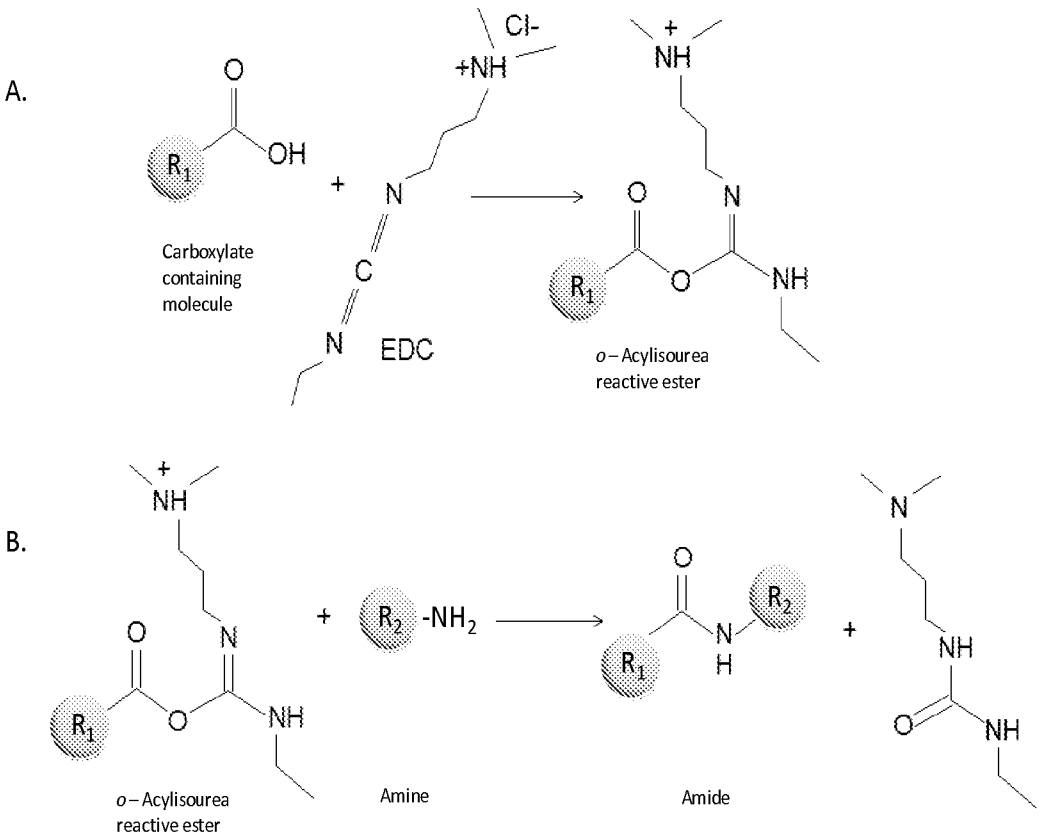

A.
Carboxylate containing molecule
EDC
o – Acylisourea reactive ester
B.
o – Acylisourea reactive ester
Amine
Amide

FIG. 6

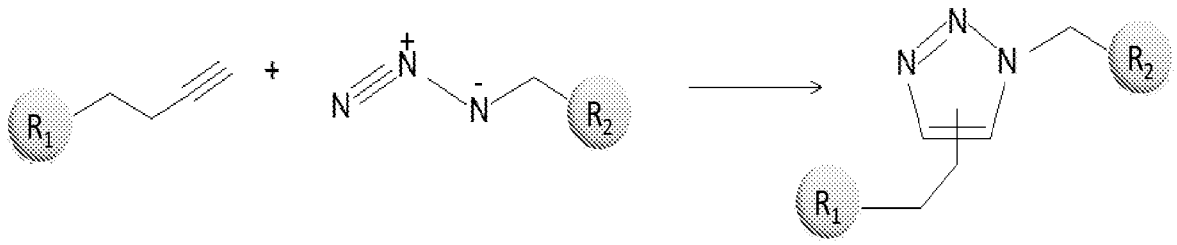


FIG. 7

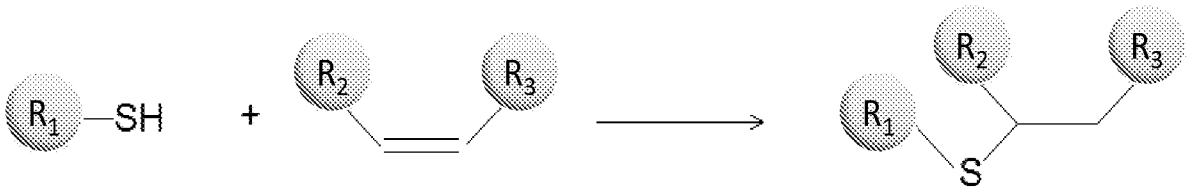


FIG. 8

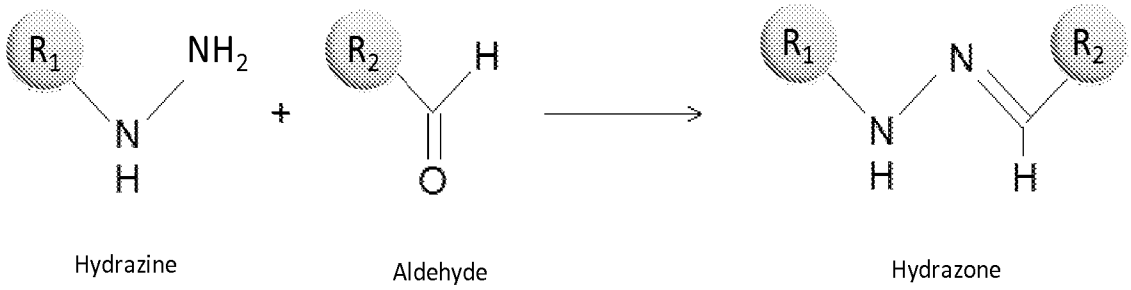


FIG. 9

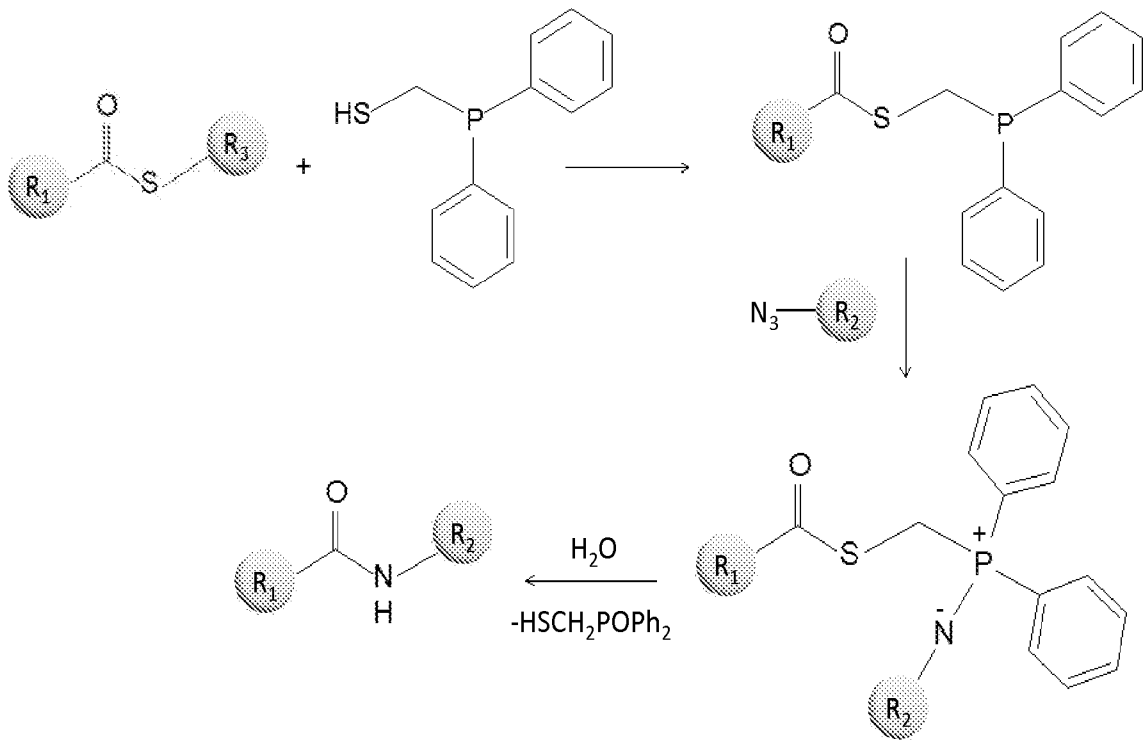


FIG. 10

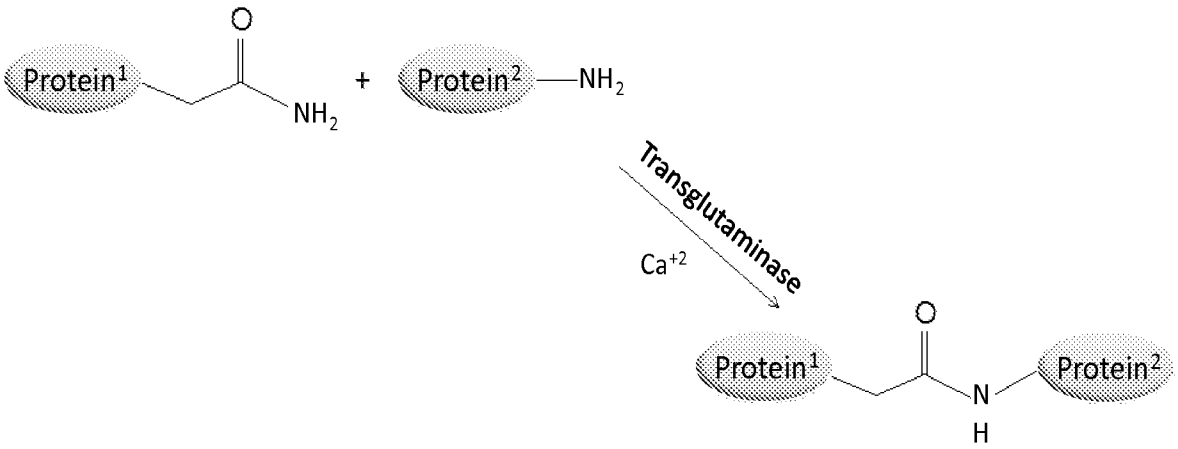
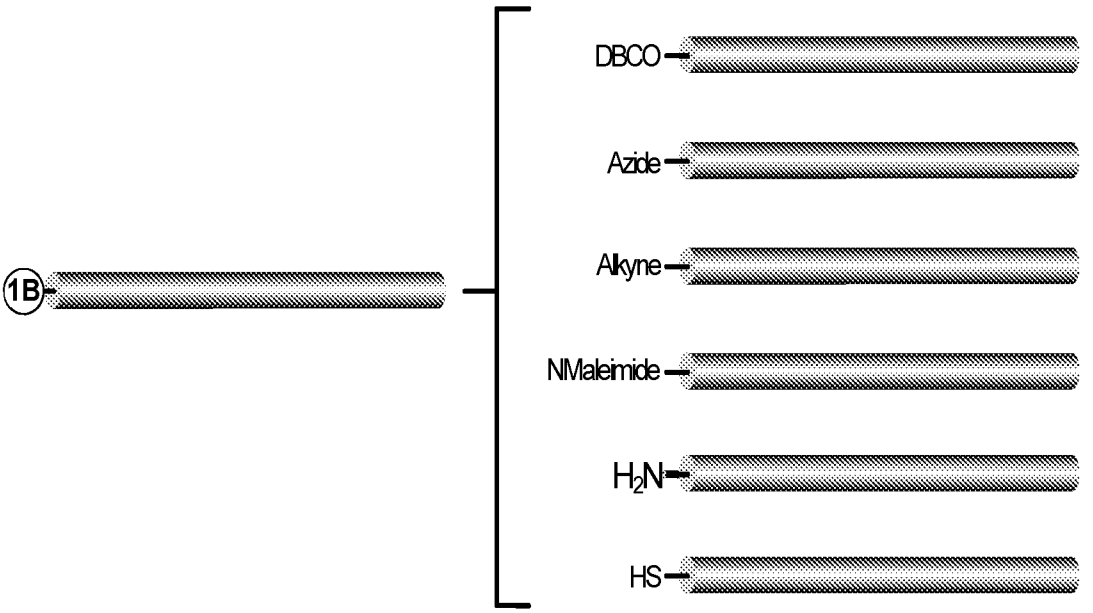

Protein1
O
NH2
+
Protein2
NH2
Transglutaminase
Ca+2
Protein1
O
N
H
Protein2

FIG. 11

A


1B
DBCO
Azide
Alkyne
NMaleimide
H2N
HS

B

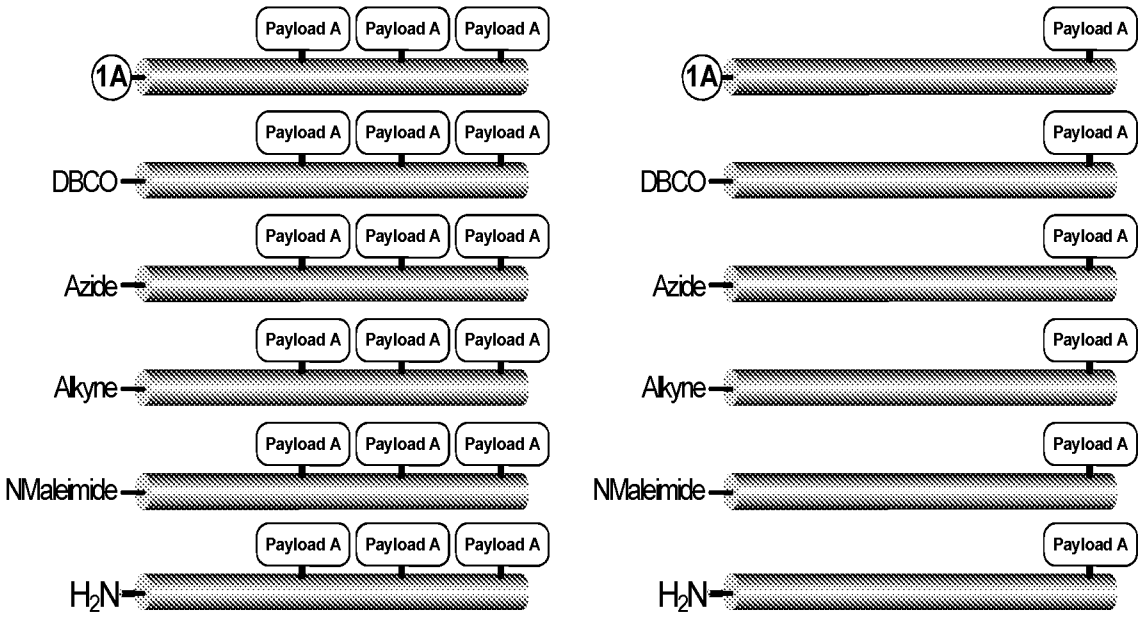

1A
Payload A
DBCO
Azide
Alkyne
NMaleimide
H2N

FIG. 12

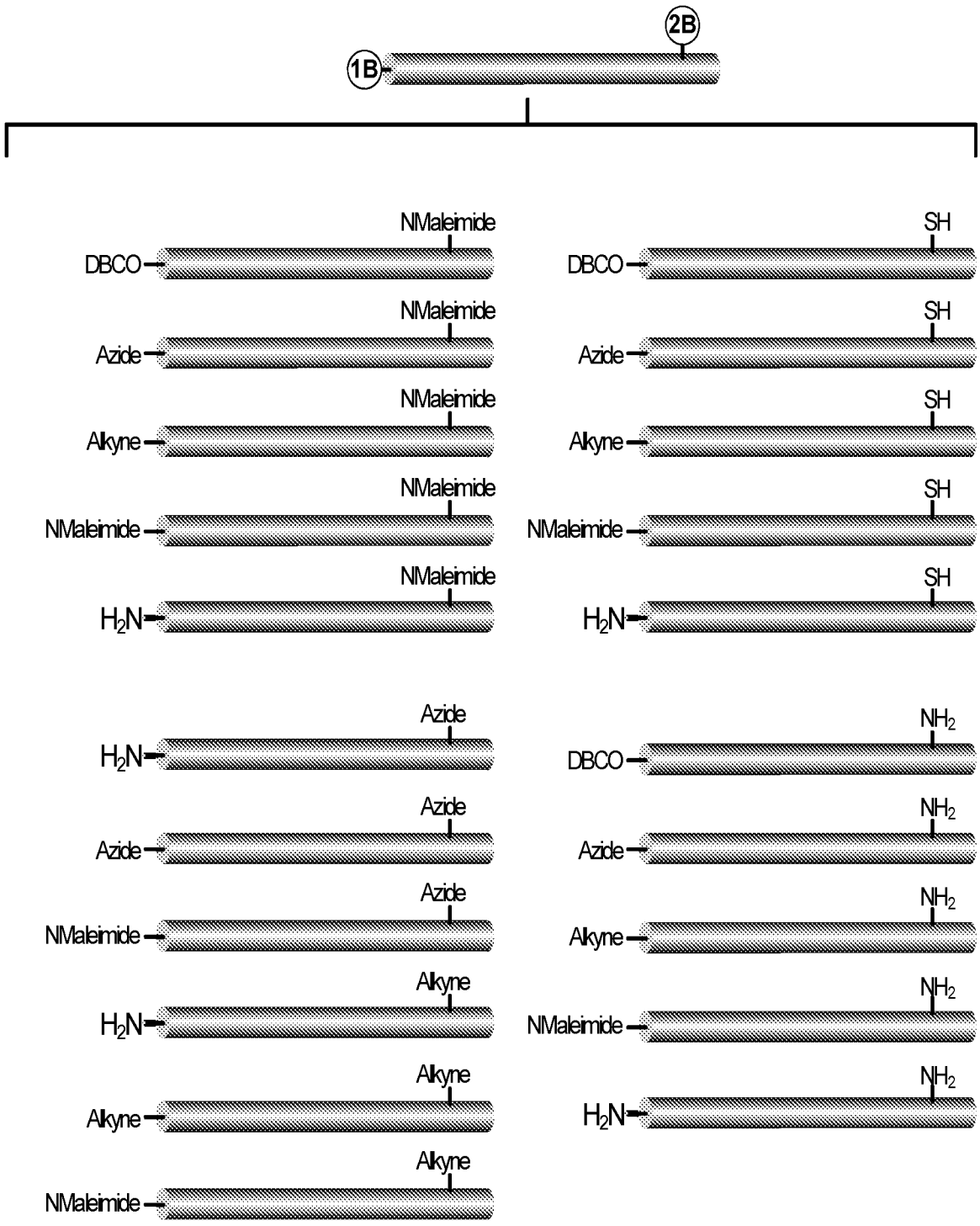

2B
1B
NMaleimide
DBCO
NMaleimide
Azide
NMaleimide
Alkyne
NMaleimide
NMaleimide
NMaleimide
H2N
SH
DBCO
SH
Azide
SH
Alkyne
SH
NMaleimide
SH
H2N
Azide
H2N
Azide
Azide
Azide
NMaleimide
Alkyne
H2N
Alkyne
Alkyne
Alkyne
NMaleimide
NH2
DBCO
NH2
Azide
NH2
Alkyne
NH2
NMaleimide
NH2
H2N

FIG. 13

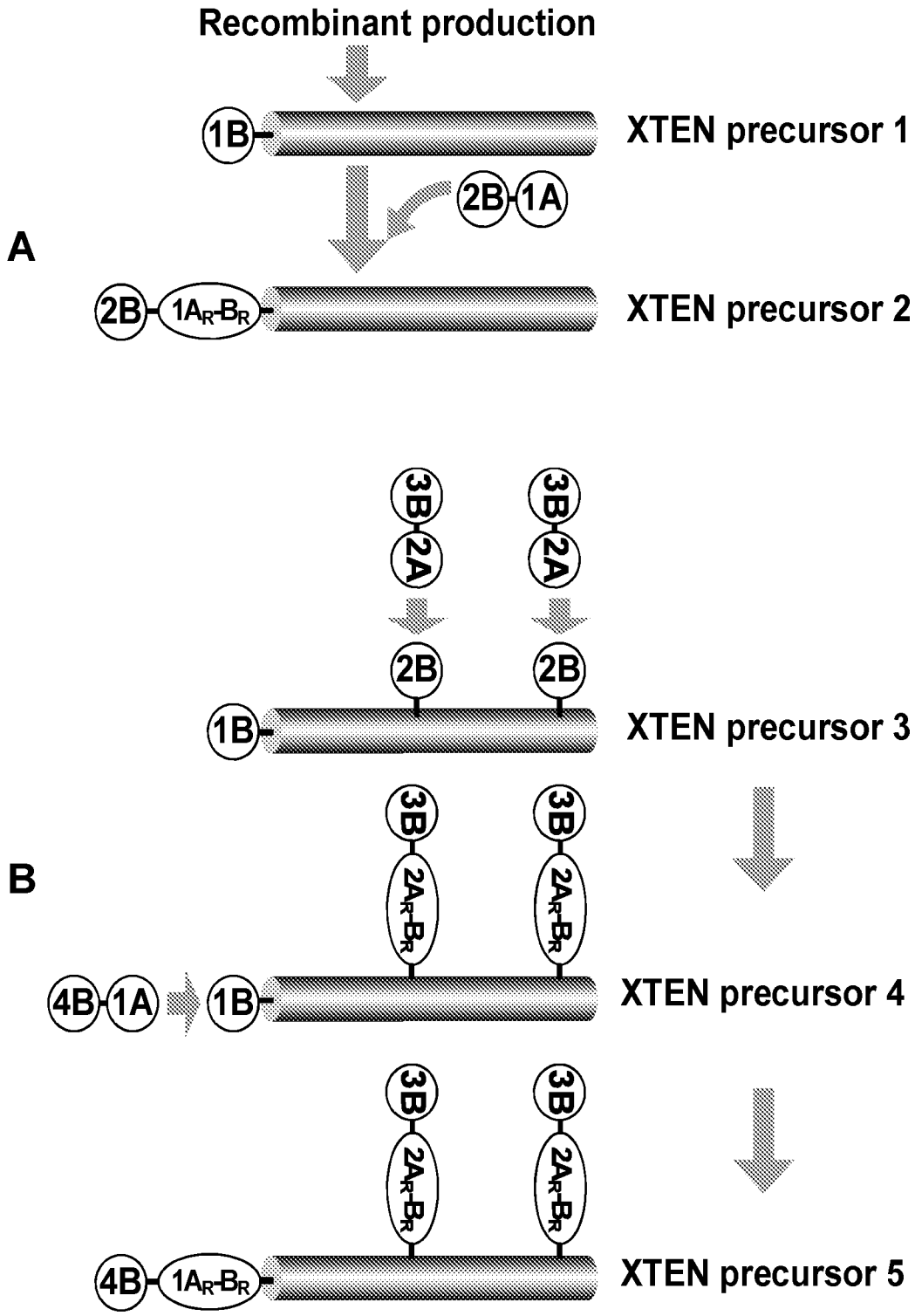

Recombinant production
A
1B
XTEN precursor 1
2B
1A
2B
1A$_R$-B$_R$
XTEN precursor 2
B
3B
2A
2B
1B
XTEN precursor 3
2A$_R$-B$_R$
4B
1A
XTEN precursor 4
XTEN precursor 5

FIG. 14

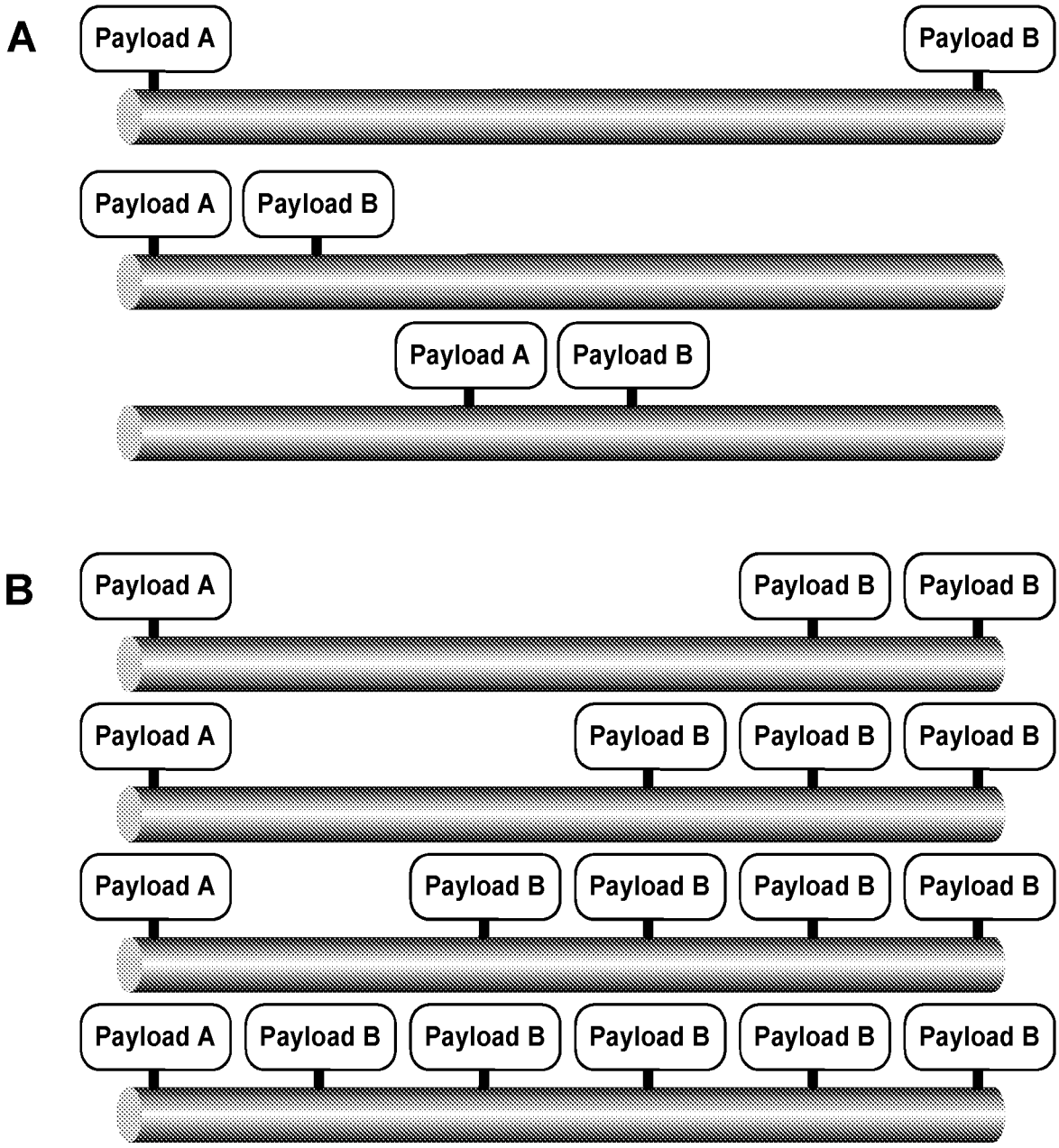

A
Payload A
Payload B
Payload A
Payload B
Payload A
Payload B
B
Payload A
Payload B
Payload B
Payload A
Payload B
Payload B
Payload B
Payload A
Payload B
Payload B
Payload B
Payload B
Payload A
Payload B
Payload B
Payload B
Payload B
Payload B

FIG. 15

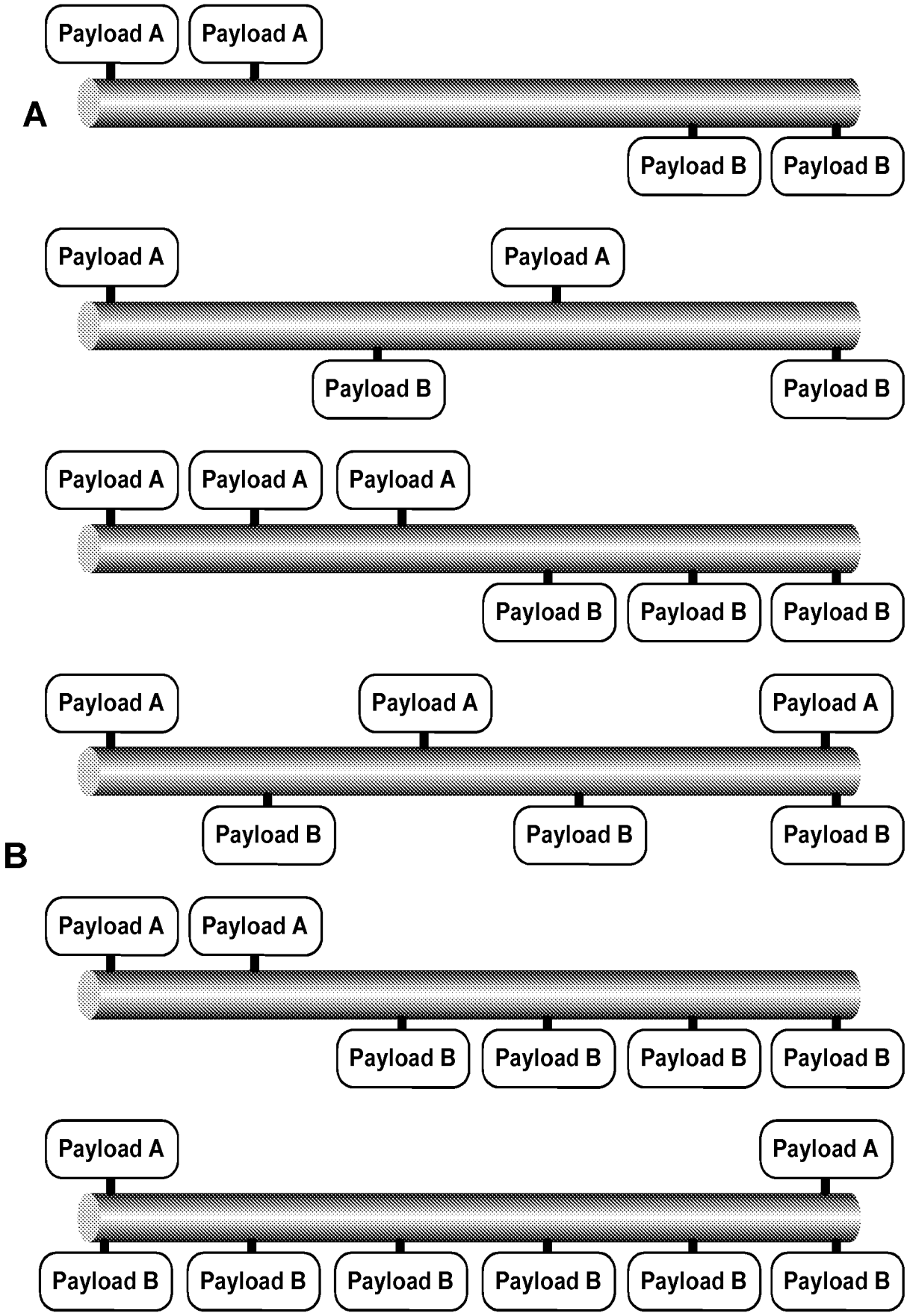

A
Payload A
Payload A
Payload B
Payload B
Payload A
Payload A
Payload B
Payload B
Payload A
Payload A
Payload A
Payload B
Payload B
Payload B
B
Payload A
Payload A
Payload A
Payload B
Payload B
Payload B
Payload A
Payload A
Payload B
Payload B
Payload B
Payload B
Payload A
Payload A
Payload B
Payload B
Payload B
Payload B
Payload B
Payload B

FIG. 16

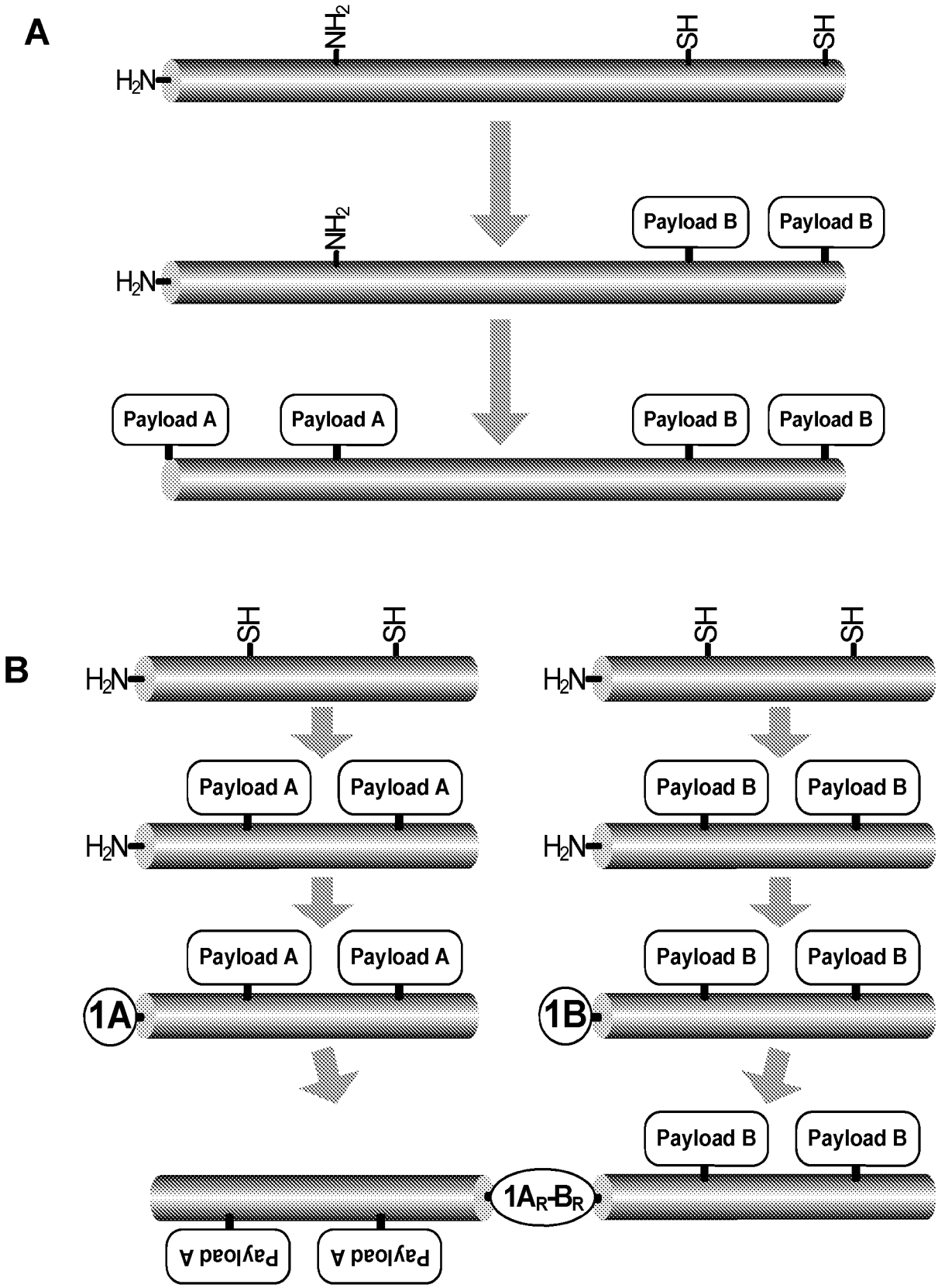

A
H_2N
NH_2
SH
SH
Payload B
Payload B
Payload A
Payload A
B
1A
1B
$1A_R$-B_R

FIG. 17

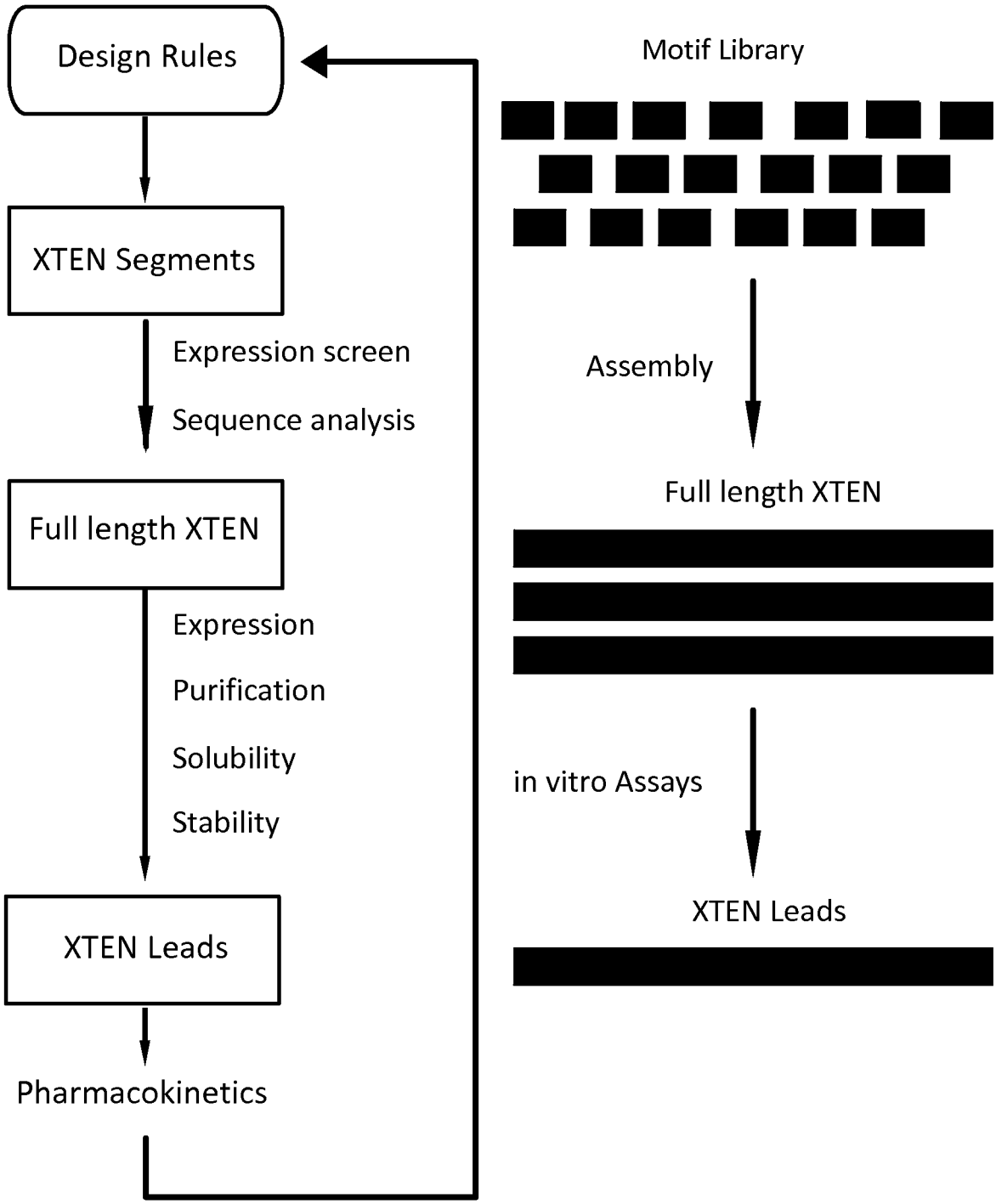

Design Rules
XTEN Segments
Expression screen
Sequence analysis
Full length XTEN
Expression
Purification
Solubility
Stability
XTEN Leads
Pharmacokinetics
Motif Library
Assembly
Full length XTEN
in vitro Assays
XTEN Leads

FIG. 18

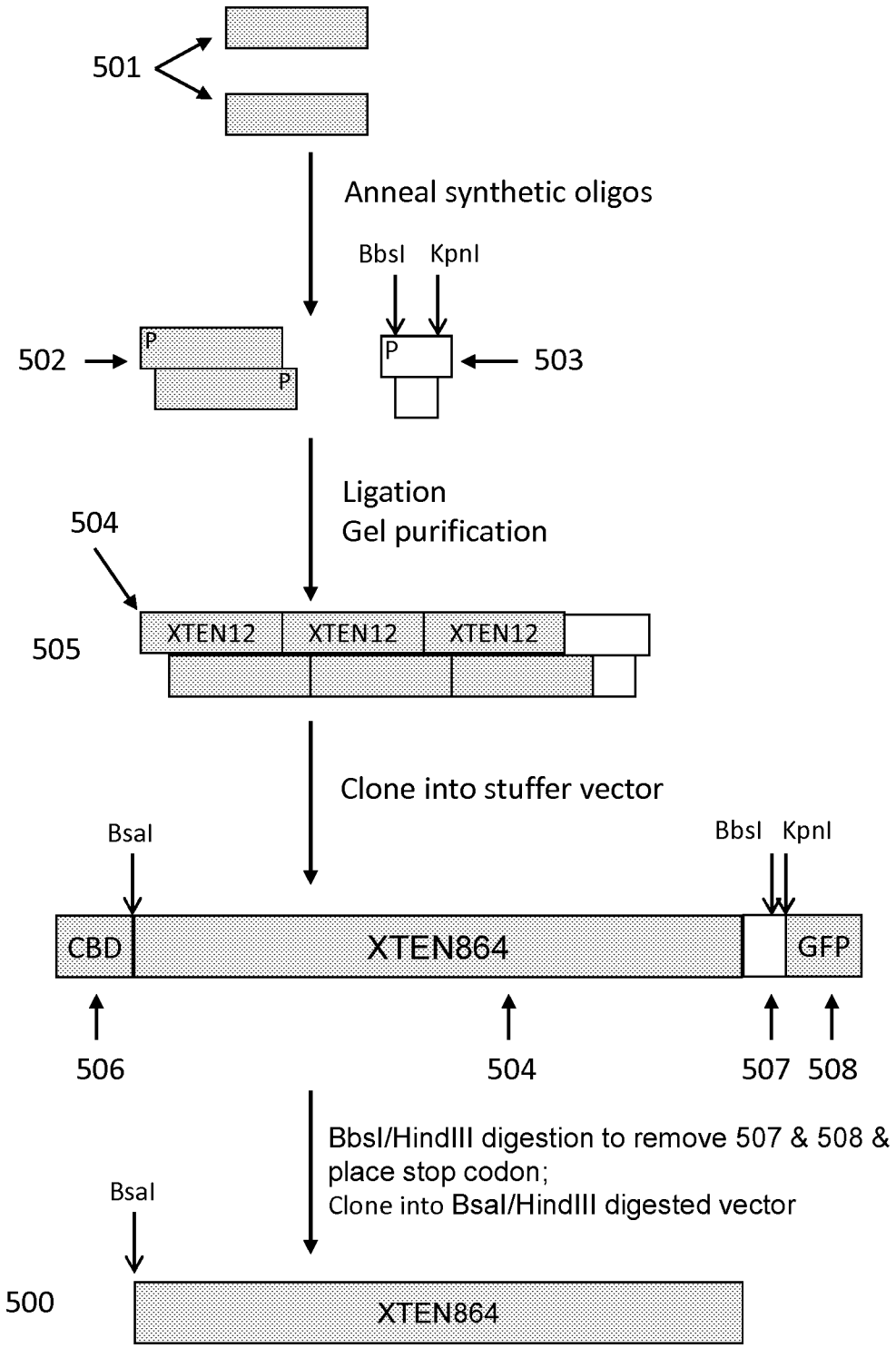

501
Anneal synthetic oligos
BbsI KpnI
502
P
P
P
503
Ligation
Gel purification
504
505
XTEN12
XTEN12
XTEN12
Clone into stuffer vector
BsaI
BbsI KpnI
CBD
XTEN864
GFP
506
504
507
508
BbsI/HindIII digestion to remove 507 & 508 &
place stop codon;
Clone into BsaI/HindIII digested vector
BsaI
500
XTEN864

FIG. 19

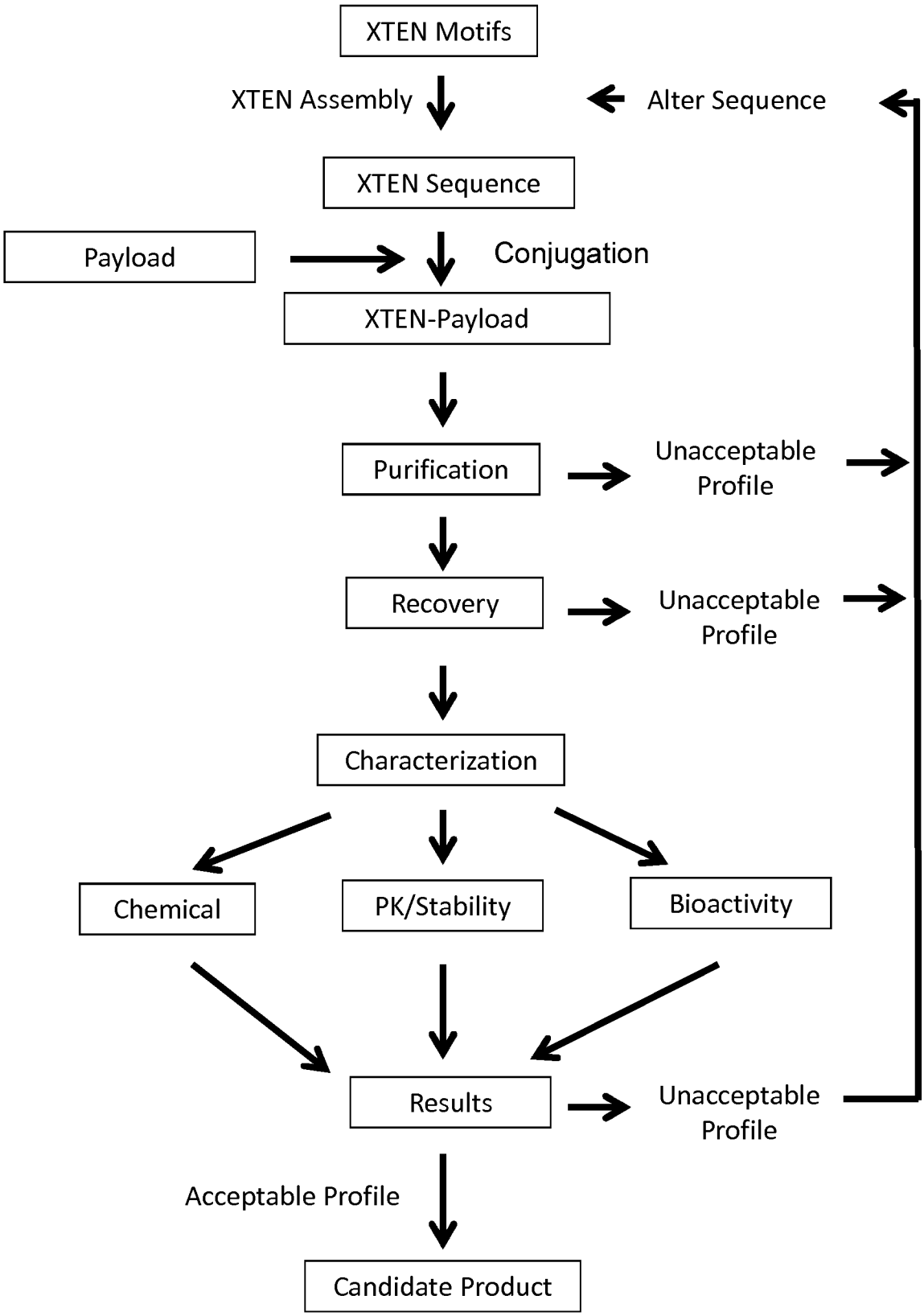

XTEN Motifs
XTEN Assembly
Alter Sequence
XTEN Sequence
Payload
Conjugation
XTEN-Payload
Purification
Unacceptable Profile
Recovery
Unacceptable Profile
Characterization
Chemical
PK/Stability
Bioactivity
Results
Unacceptable Profile
Acceptable Profile
Candidate Product

FIG. 20

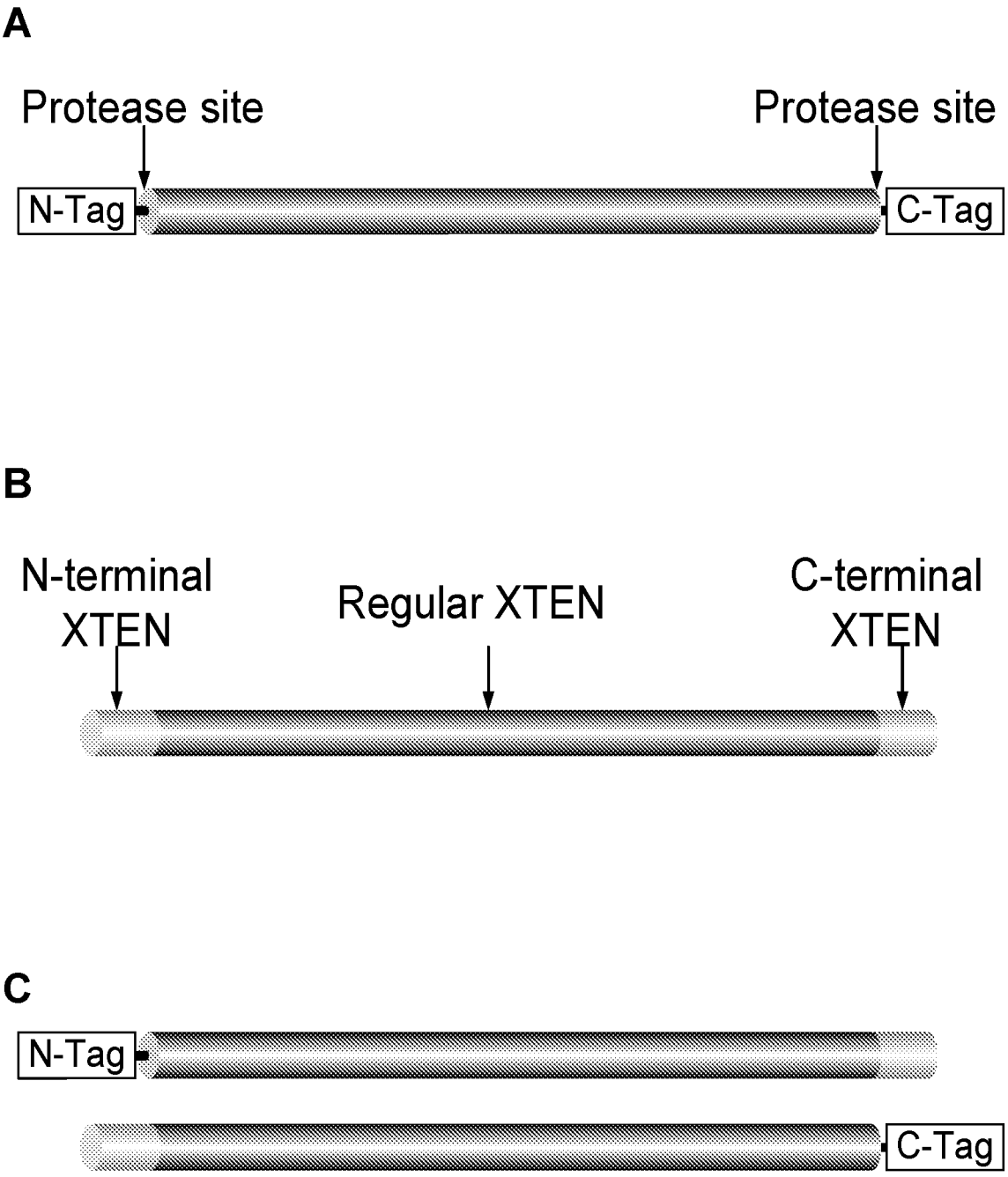

A
Protease site
Protease site
N-Tag
C-Tag
B
N-terminal XTEN
Regular XTEN
C-terminal XTEN
C
N-Tag
C-Tag

FIG. 21

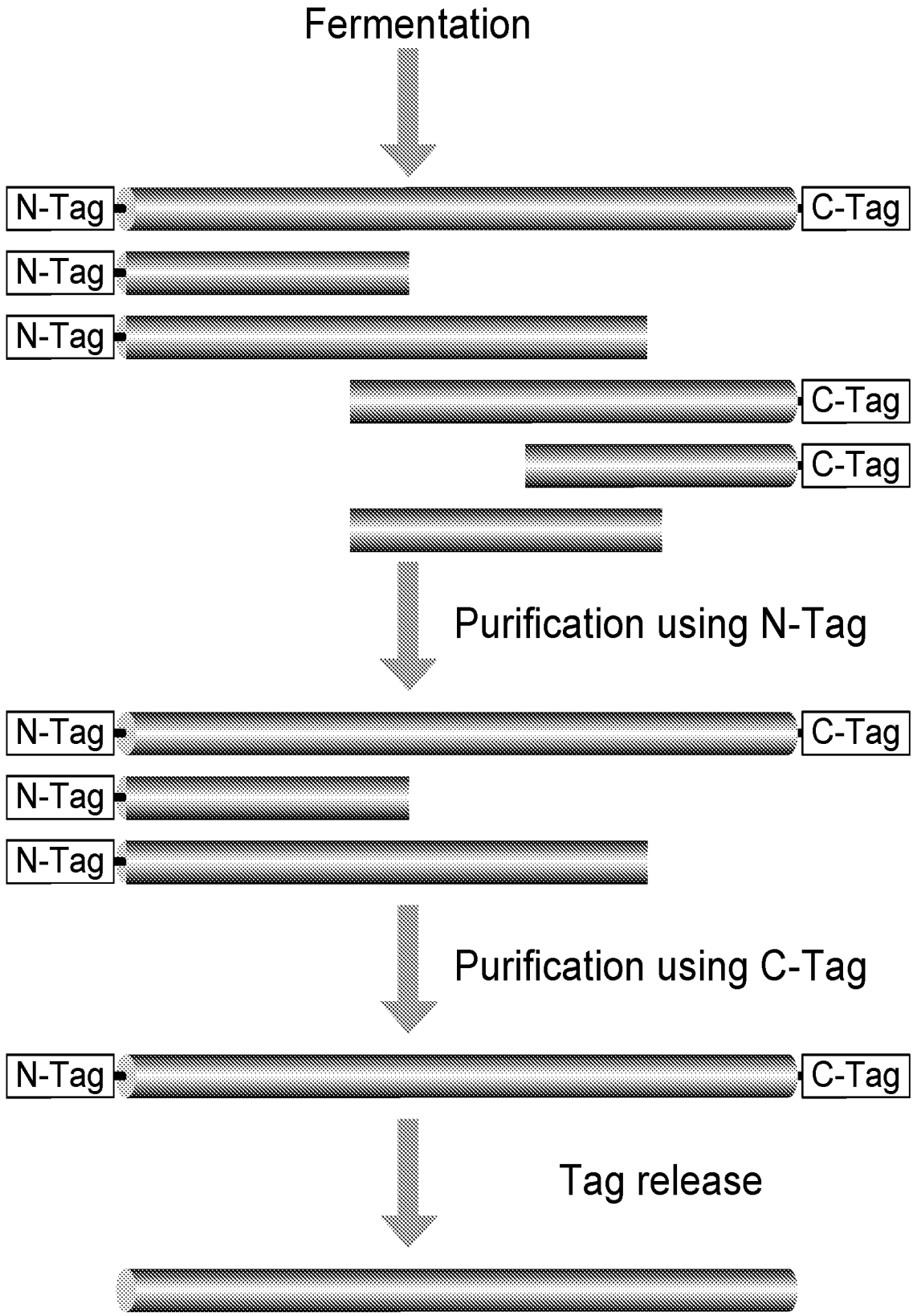

Fermentation
N-Tag
C-Tag
N-Tag
N-Tag
C-Tag
C-Tag
Purification using N-Tag
N-Tag
C-Tag
N-Tag
N-Tag
Purification using C-Tag
N-Tag
C-Tag
Tag release

FIG. 22

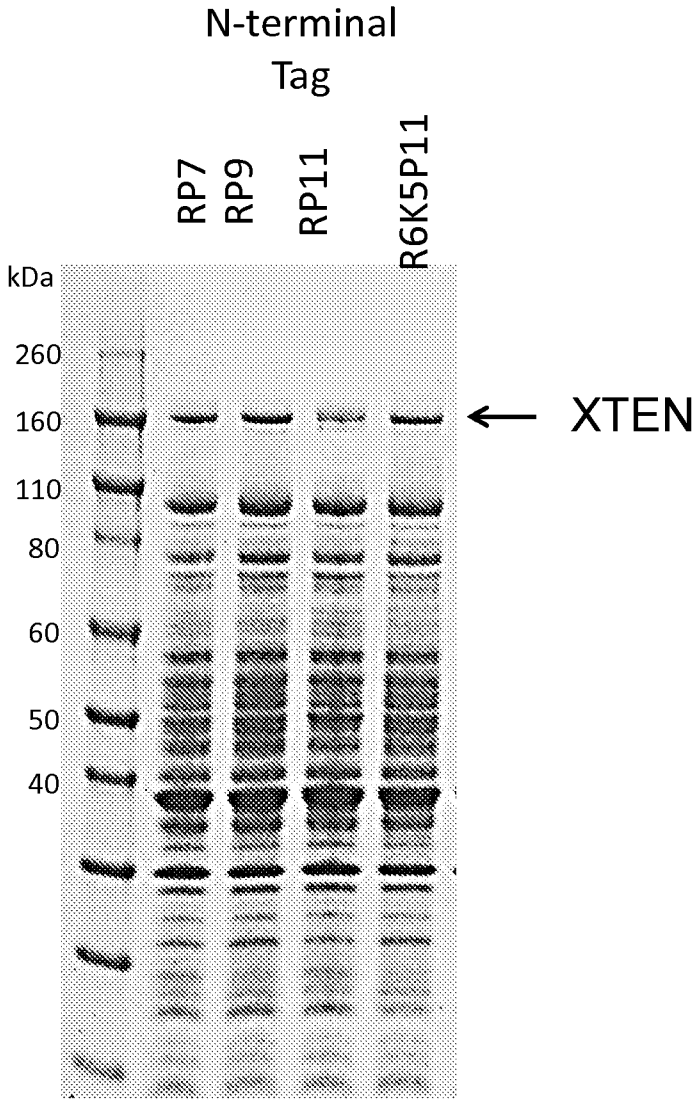

N-terminal
Tag
RP7
RP9
RP11
R6K5P11
kDa
260
160
110
80
60
50
40
XTEN

FIG. 23

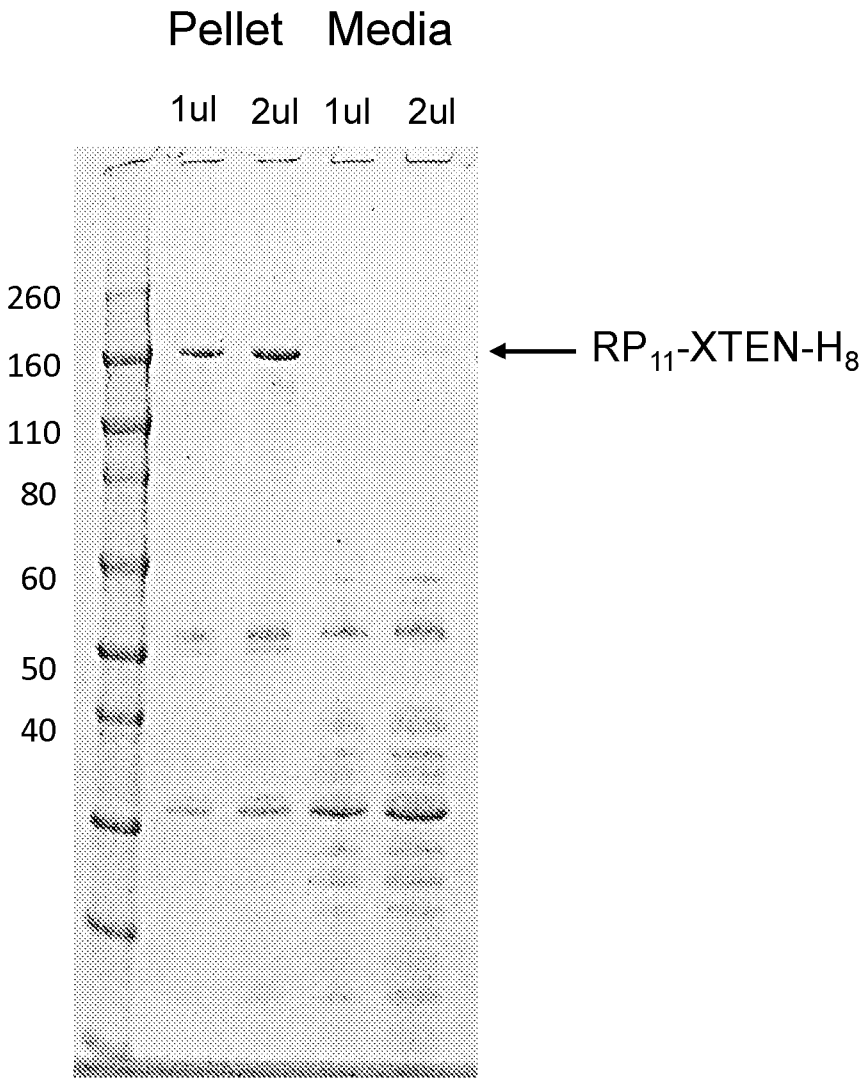

Pellet
Media
1ul
2ul
1ul
2ul
260
160
110
80
60
50
40
RP11-XTEN-H8

FIG. 24

Lane 1	Molecular weight standard
Lane 2	MacroCap SP load
Lane 3	MacroCap SP flow through
Lane 4	MacroCap SP chase
Lane 5	MacroCap SP wash
Lane 6	MacroCap SP elution 1
Lane 7	MacroCap SP elution 2
Lane 8	MacroCap SP elution 3
Lane 9	MacroCap SP elution 4
Lane 10	MacroCap SP elution 5
Lane 11	MacroCap SP elution 6

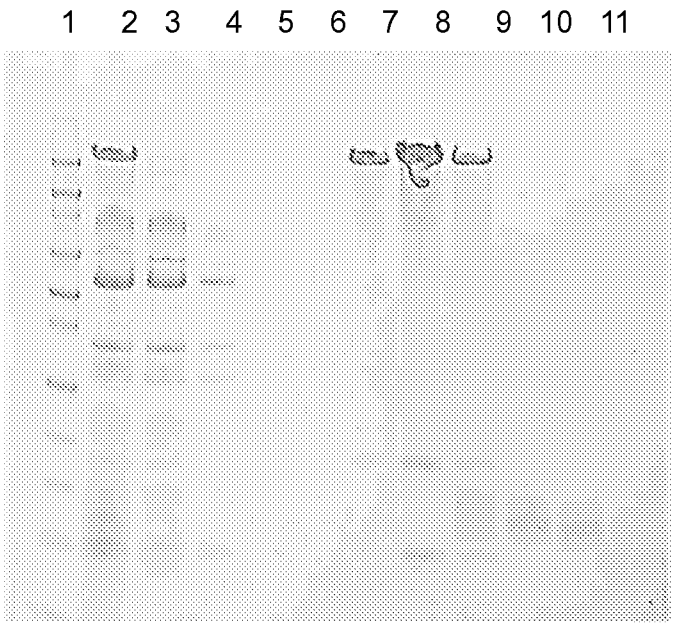


FIG. 25

Lane 1	Molecular weight standard
Lane 2	IMAC load
Lane 3	IMAC flow through
Lane 4	IMAC chase
Lane 5	IMAC wash
Lane 6	IMAC elution 1
Lane 7	IMAC elution 2
Lane 8	IMAC elution 3
Lane 9	IMAC elution 4
Lane 10	IMAC elution 5
Lane 11	IMAC elution 6
Lane 12	IMAC elution 7

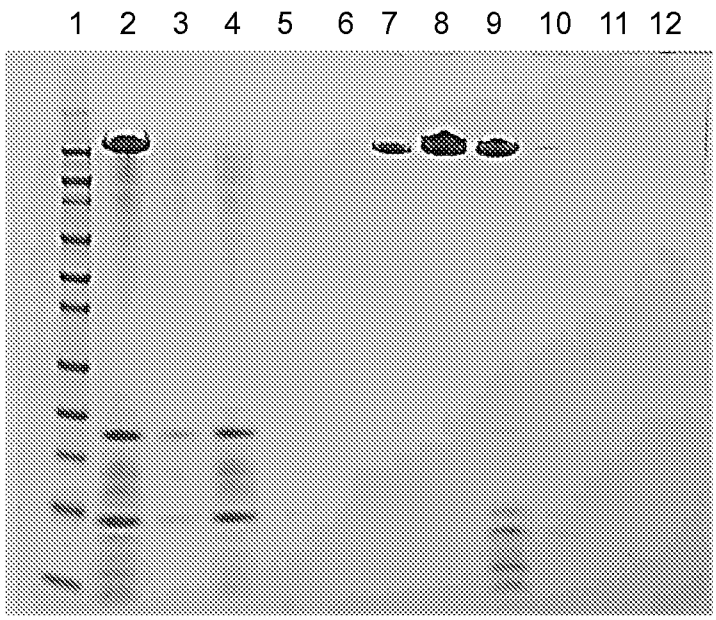
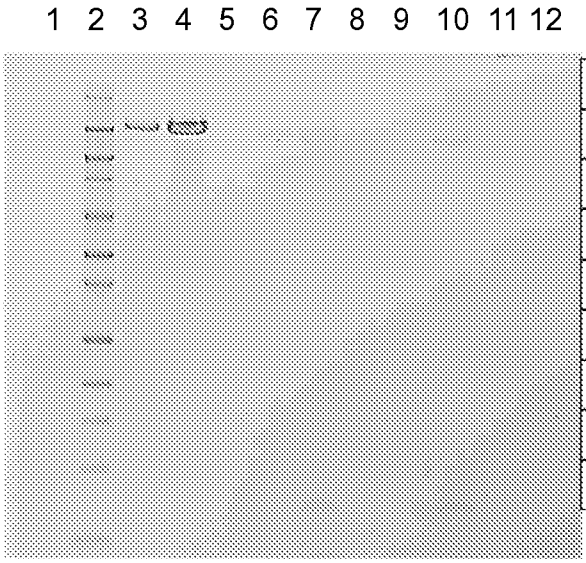


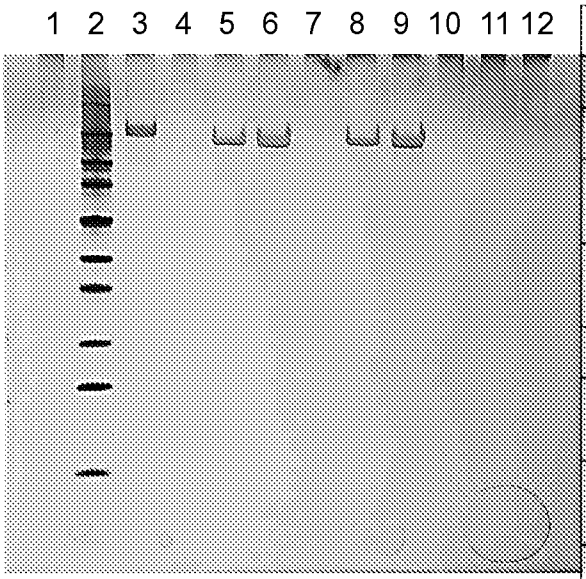
FIG. 26

A



Lane 2	Molecular weight standard
Lane 3	IMAC pool 2 μl
Lane 4	IMAC pool 10 μl
Lane 5	blank
Lane 6	IMAC pool 1:200 trypsin digest 2 μl
Lane 7	IMAC pool 1:200 trypsin digest 10 μl
Lane 8	blank
Lane 9	IMAC pool 1:500 trypsin digest 2 μl
Lane 10	IMAC pool 1:500 trypsin digest 10 μl

B



Lane 2	Molecular weight standard
Lane 3	IMAC pool 1:20 diluted 10 μl
Lane 4	blank
Lane 5	IMAC pool 1:200 trypsin digest 1:5 diluted 5 μl
Lane 6	IMAC pool 1:200 trypsin digest 1:5 diluted 10 μl
Lane 7	blank
Lane 8	IMAC pool 1:500 trypsin digest 1:5 diluted 5 μl
Lane 9	IMAC pool 1:500 trypsin digest 1:5 diluted 10 μl
Lane 10	blank

FIG. 27

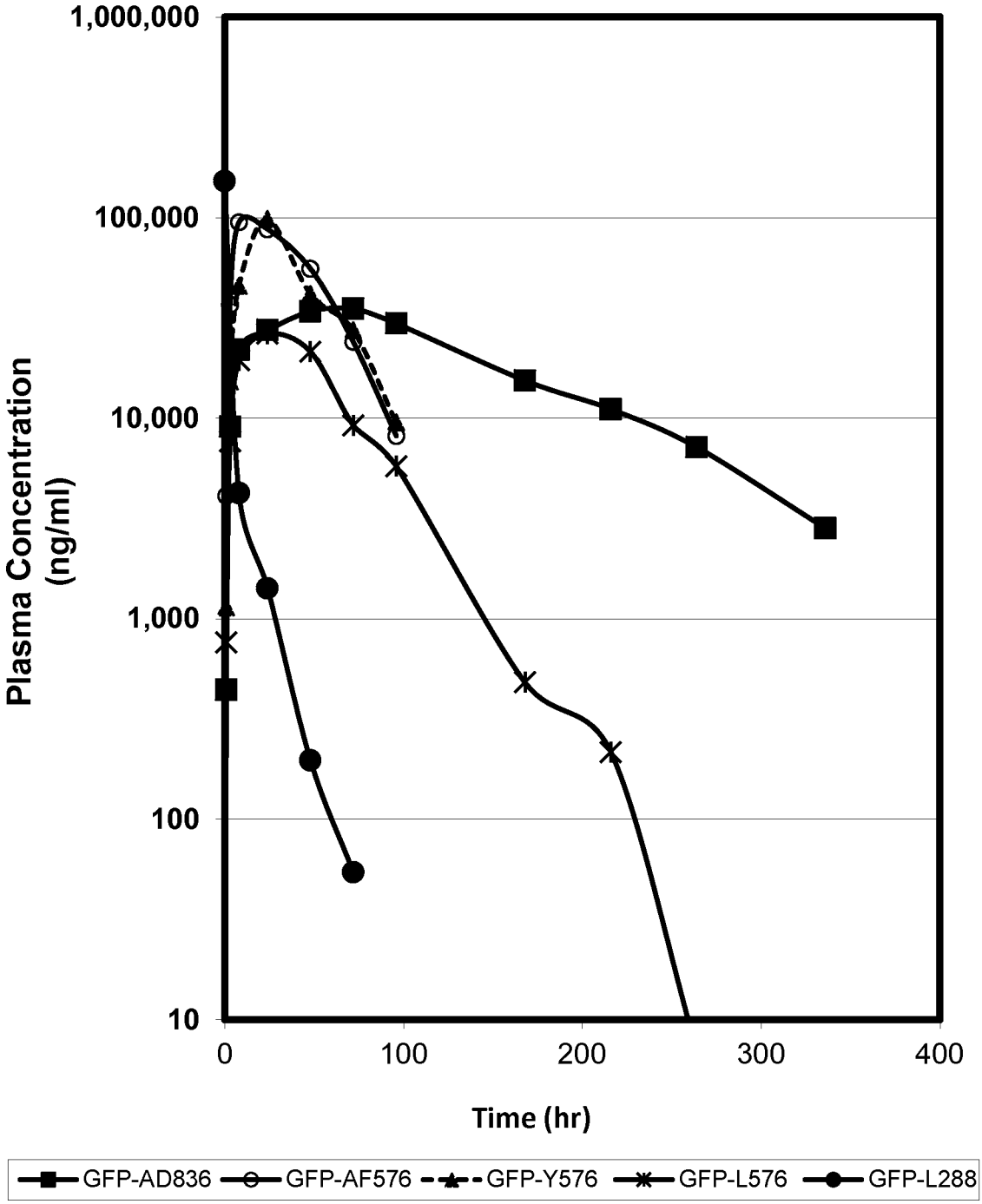

1,000,000
100,000
10,000
1,000
100
10
Plasma Concentration (ng/ml)
0
100
200
300
400
Time (hr)
GFP-AD836
GFP-AF576
GFP-Y576
GFP-L576
GFP-L288

FIG. 28

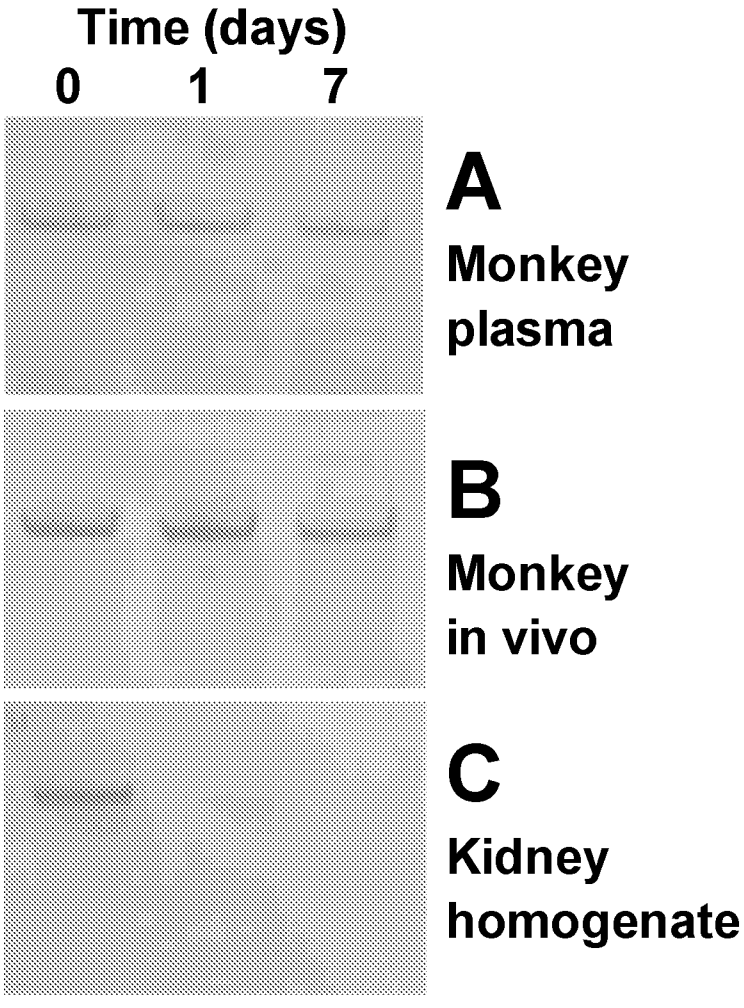

Time (days)
0
1
7
A
Monkey
plasma
B
Monkey
in vivo
C
Kidney
homogenate

FIG. 29

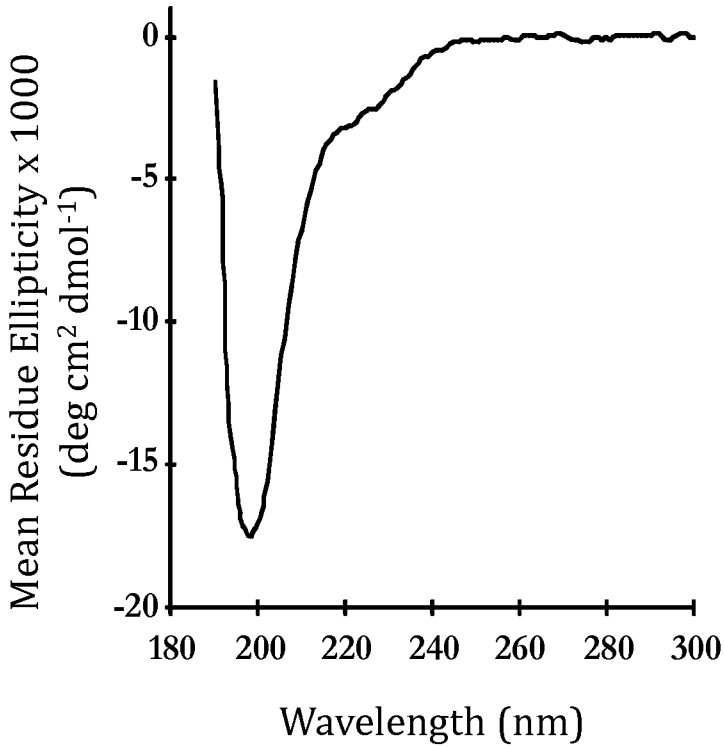
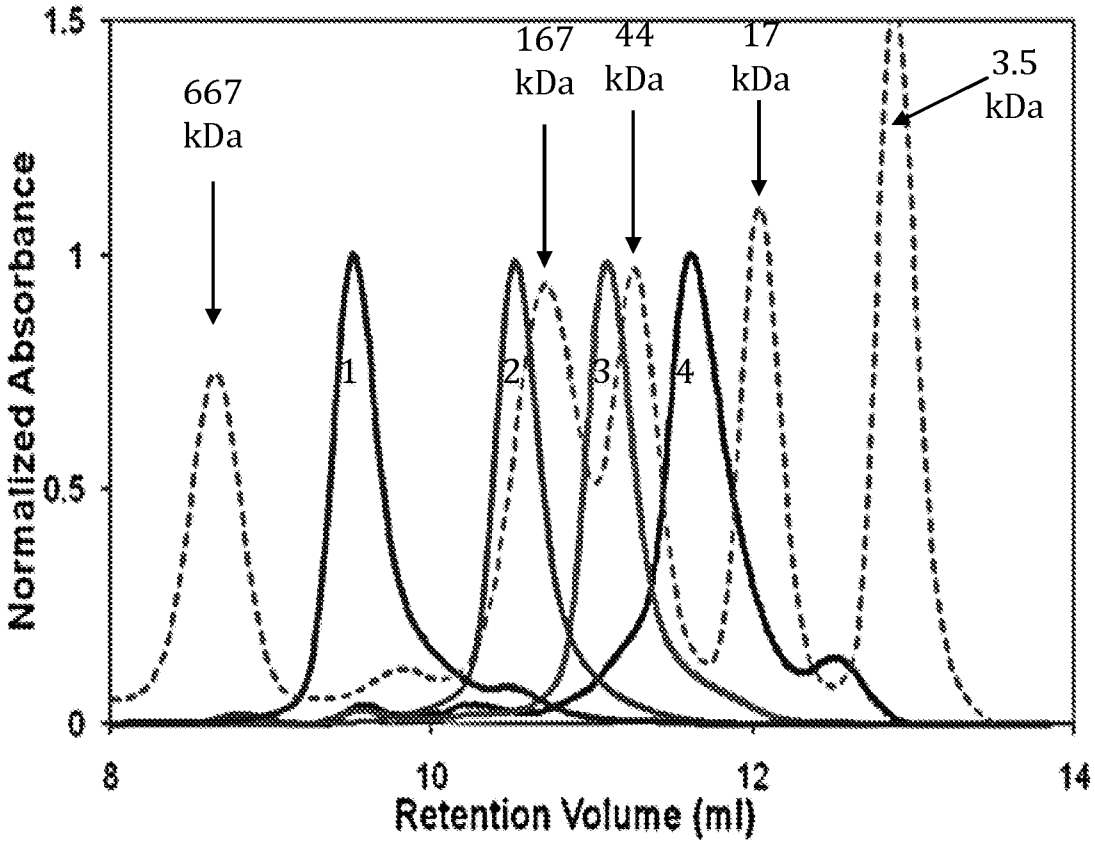

Mean Residue Ellipticity x 1000
(deg cm2 dmol-1)
0
-5
-10
-15
-20
180
200
220
240
260
280
300
Wavelength (nm)

FIG. 30



1. Glucagon-Y288
2. Glucagon-Y144
3. Glucagon-Y72
4. Glucagon-Y36

- - - - = Standards

FIG. 31

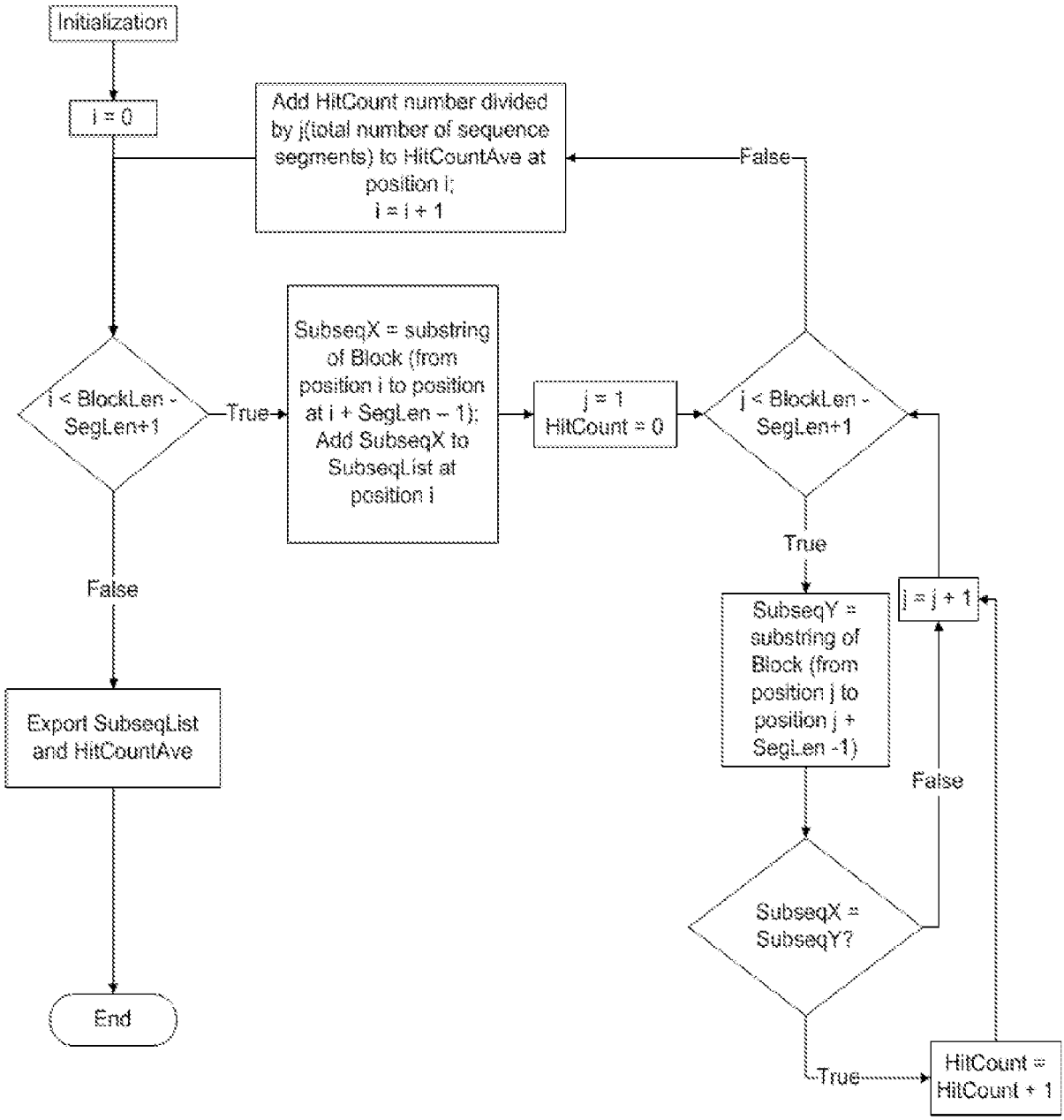

Initialization
i = 0
Add HitCount number divided by j(total number of sequence segments) to HitCountAve at position i;
i = i + 1
False
i < BlockLen - SegLen+1
True
SubseqX = substring of Block (from position i to position at i + SegLen – 1); Add SubseqX to SubseqList at position i
j = 1
HitCount = 0
j < BlockLen - SegLen+1
True
False
SubseqY = substring of Block (from position j to position j + SegLen -1)
j = j + 1
False
SubseqX = SubseqY?
True
HitCount = HitCount + 1
Export SubseqList and HitCountAve
End

FIG. 32

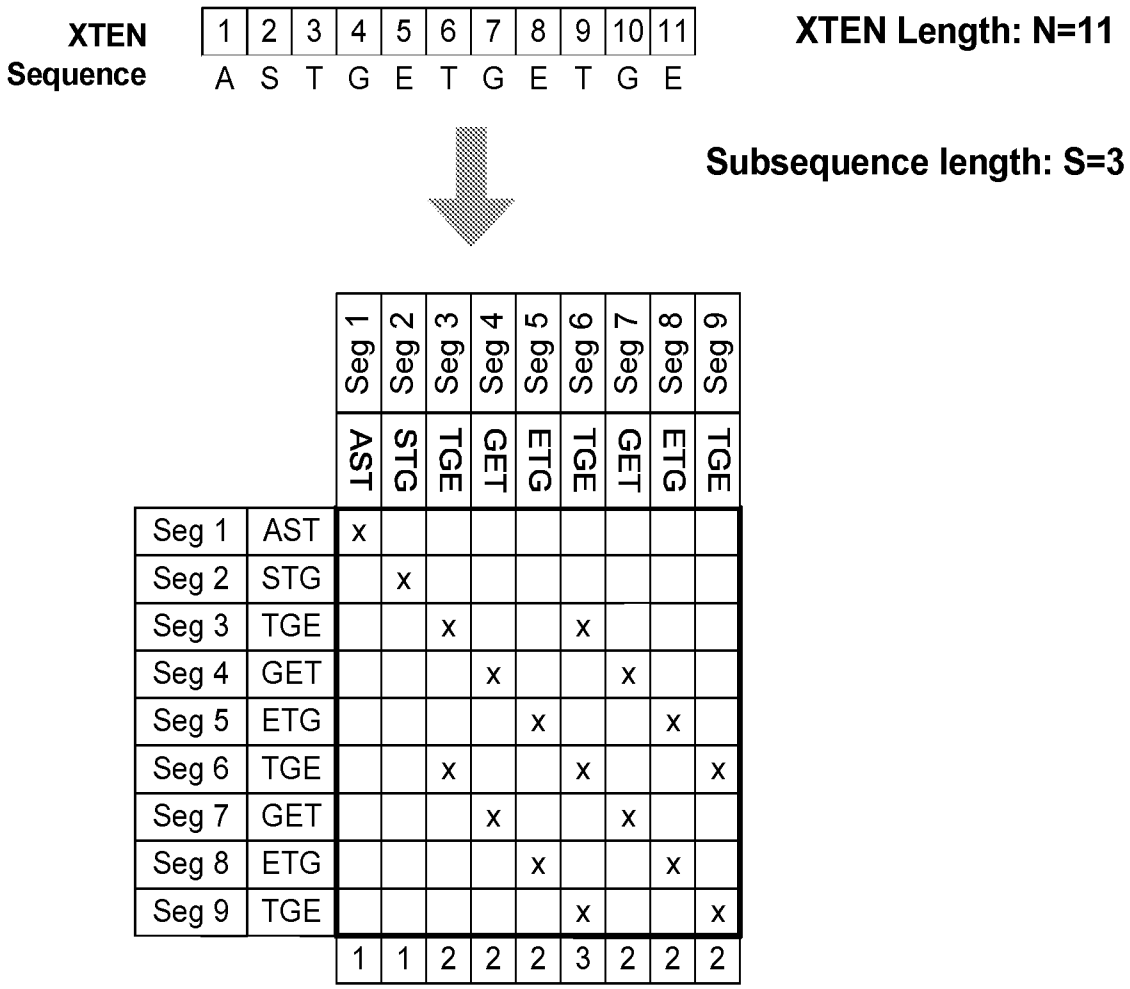

XTEN Sequence
1 2 3 4 5 6 7 8 9 10 11
A S T G E T G E T G E
XTEN Length: N=11
Subsequence length: S=3
Seg 1 Seg 2 Seg 3 Seg 4 Seg 5 Seg 6 Seg 7 Seg 8 Seg 9
AST STG TGE GET ETG TGE GET ETG TGE
Seg 1 AST x
Seg 2 STG x
Seg 3 TGE x x
Seg 4 GET x x
Seg 5 ETG x x
Seg 6 TGE x x x
Seg 7 GET x x
Seg 8 ETG x x
Seg 9 TGE x x
1 1 2 2 2 3 2 2 2
Subsequence Score = 1.89

FIG. 33

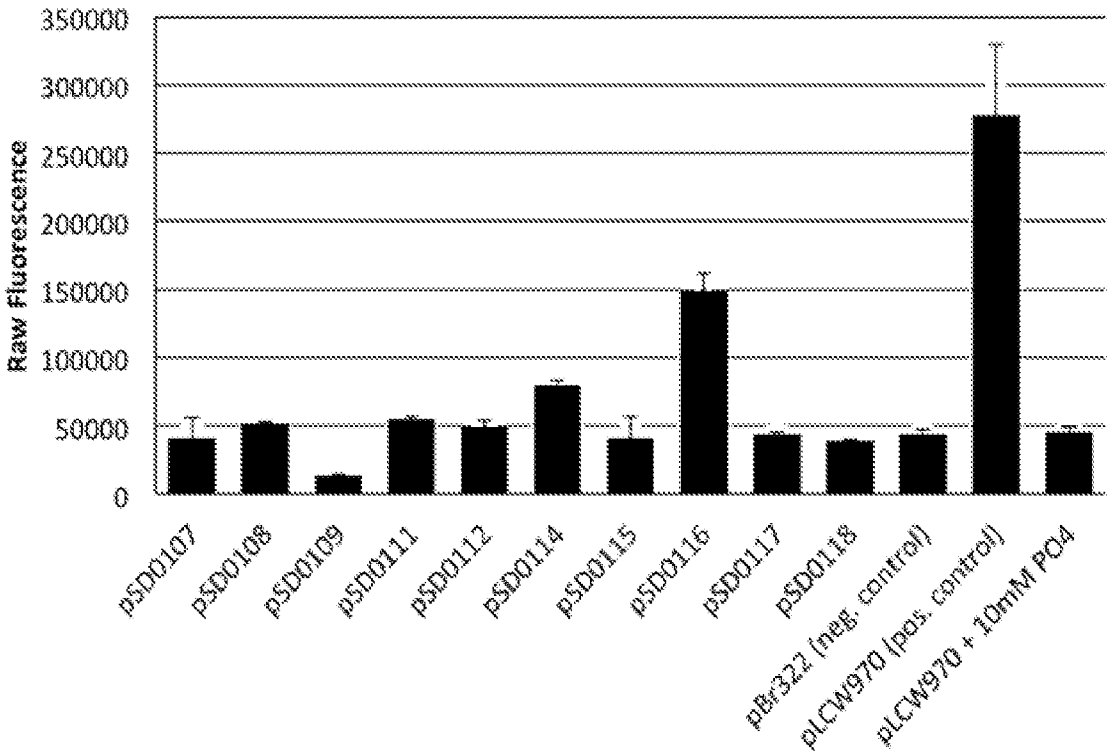

Raw Fluorescence
350000
300000
250000
200000
150000
100000
50000
0
pSD0107
pSD0108
pSD0109
pSD0111
pSD0112
pSD0114
pSD0115
pSD0116
pSD0117
pSD0118
pBr322 (neg. control)
pLCW970 (pos. control)
pLCW970 + 10mM PO4

FIG. 34

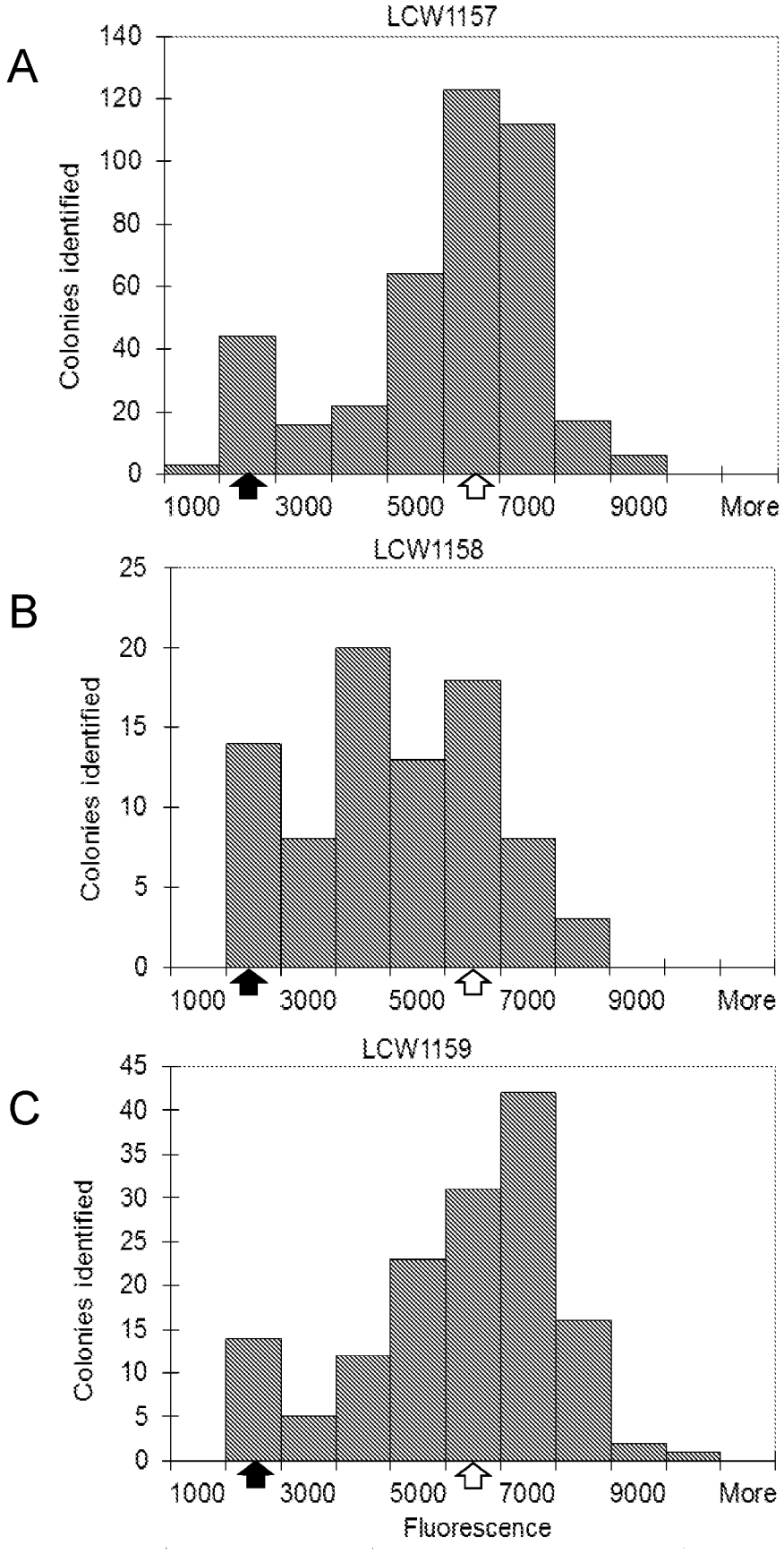

A
LCW1157
Colonies identified
140
120
100
80
60
40
20
0
1000
3000
5000
7000
9000
More
B
LCW1158
Colonies identified
25
20
15
10
5
0
1000
3000
5000
7000
9000
More
C
LCW1159
Colonies identified
45
40
35
30
25
20
15
10
5
0
1000
3000
5000
7000
9000
More
Fluorescence

FIG. 35, A, B, C

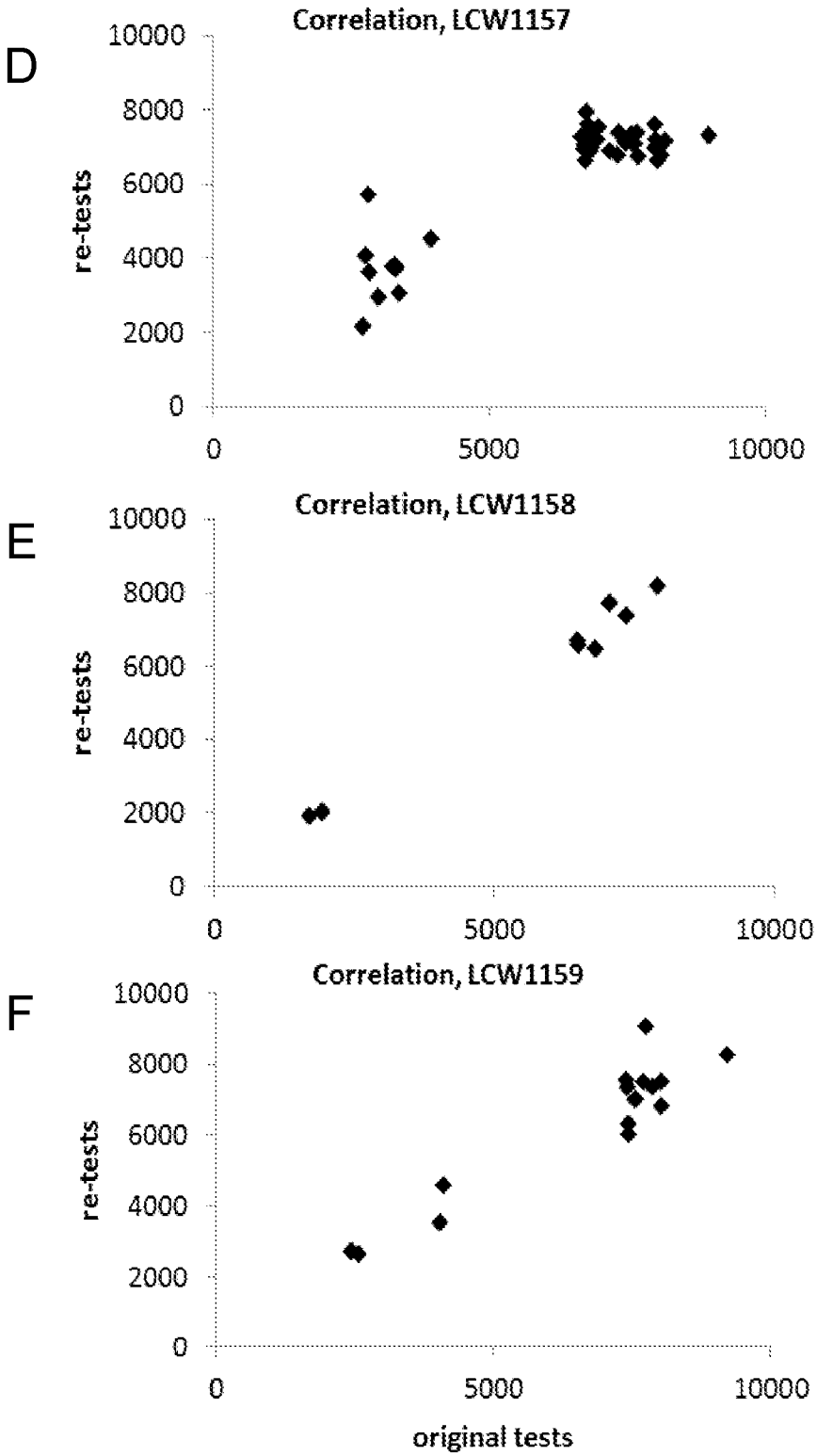

D
Correlation, LCW1157
re-tests
10000
8000
6000
4000
2000
0
0
5000
10000
E
Correlation, LCW1158
re-tests
10000
8000
6000
4000
2000
0
0
5000
10000
F
Correlation, LCW1159
re-tests
10000
8000
6000
4000
2000
0
0
5000
10000
original tests

FIG. 35, D, E, F

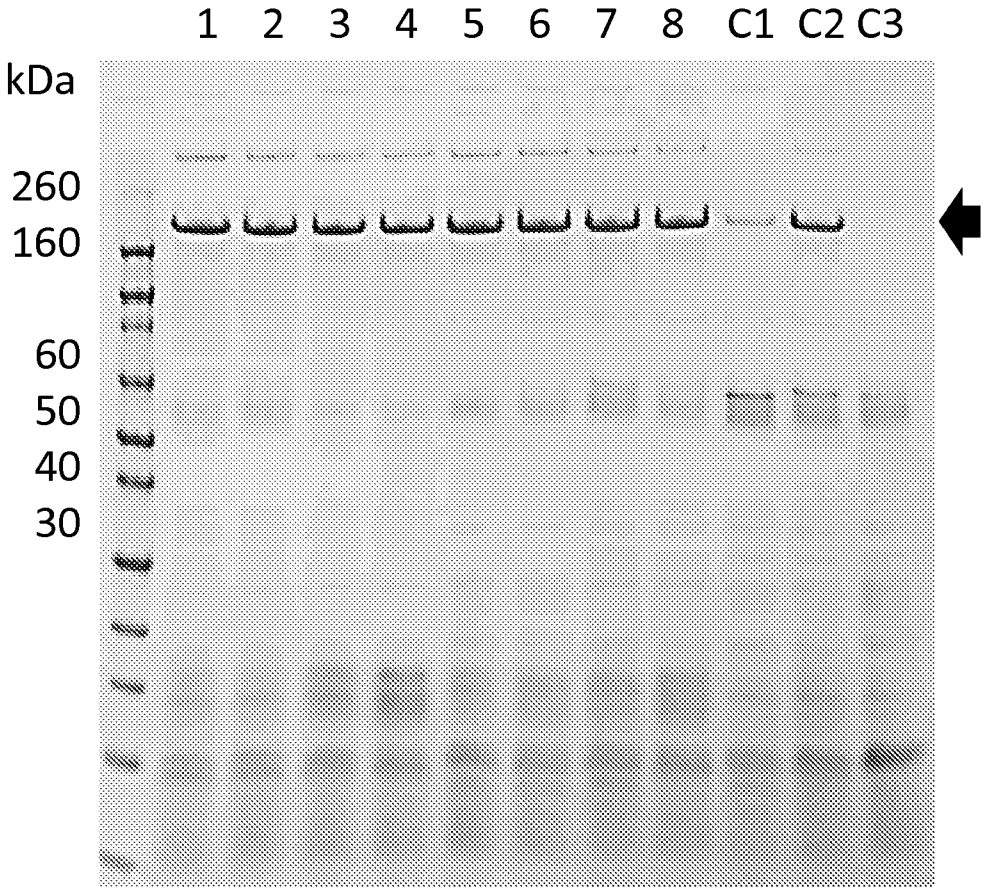


FIG. 36

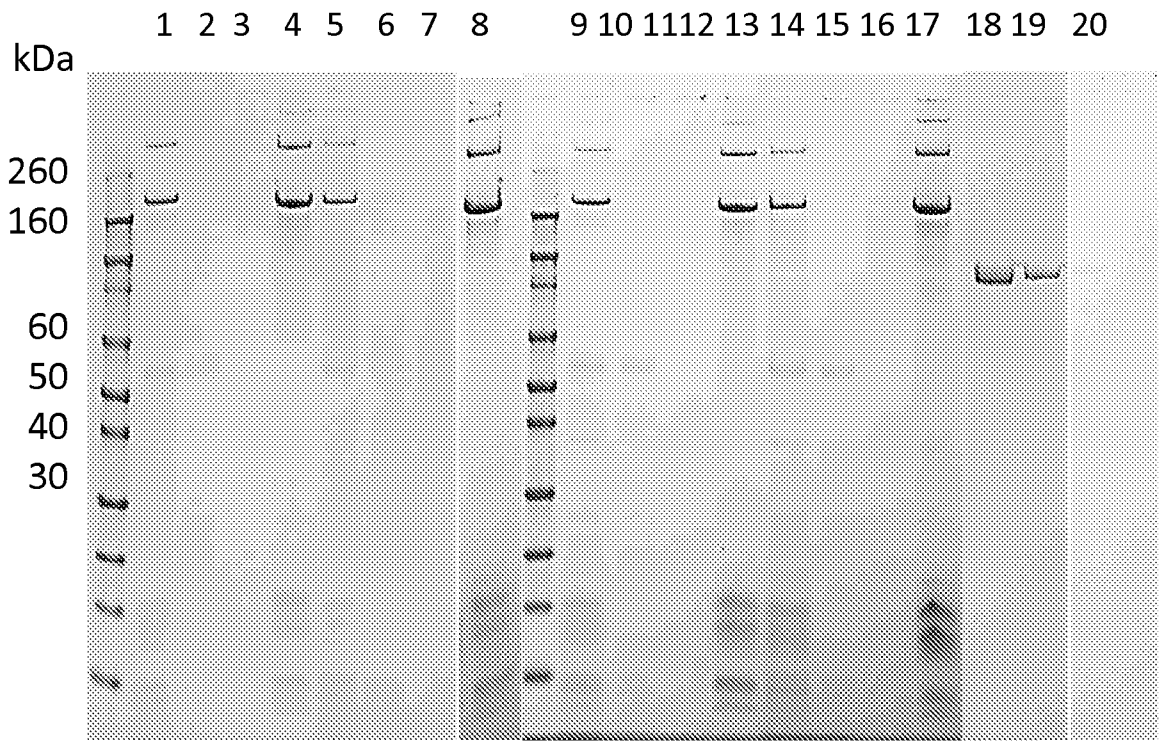


FIG. 37

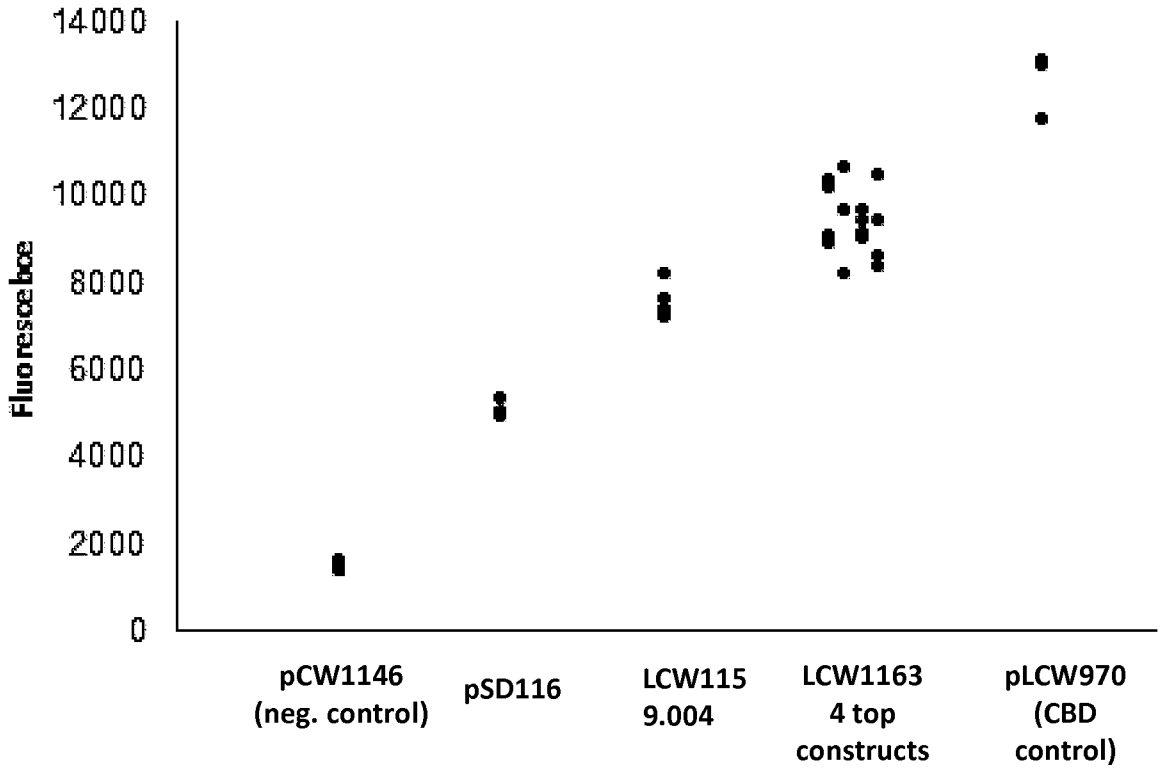

14000
12000
10000
8000
6000
4000
2000
0
Fluorescence
pCW1146 (neg. control)
pSD116
LCW1159.004
LCW1634 top constructs
pLCW970 (CBD control)

FIG. 38

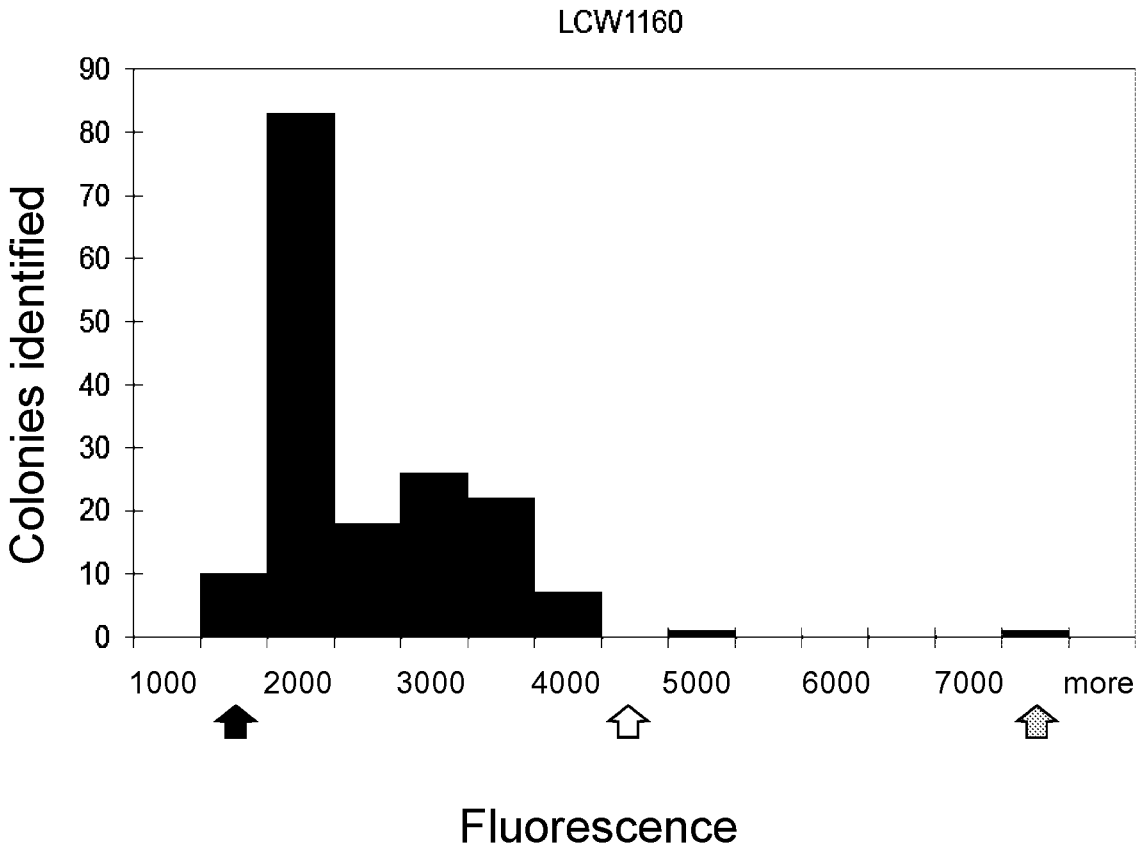

LCW1160
Colonies identified
90
80
70
60
50
40
30
20
10
0
1000
2000
3000
4000
5000
6000
7000
more
Fluorescence

FIG. 39

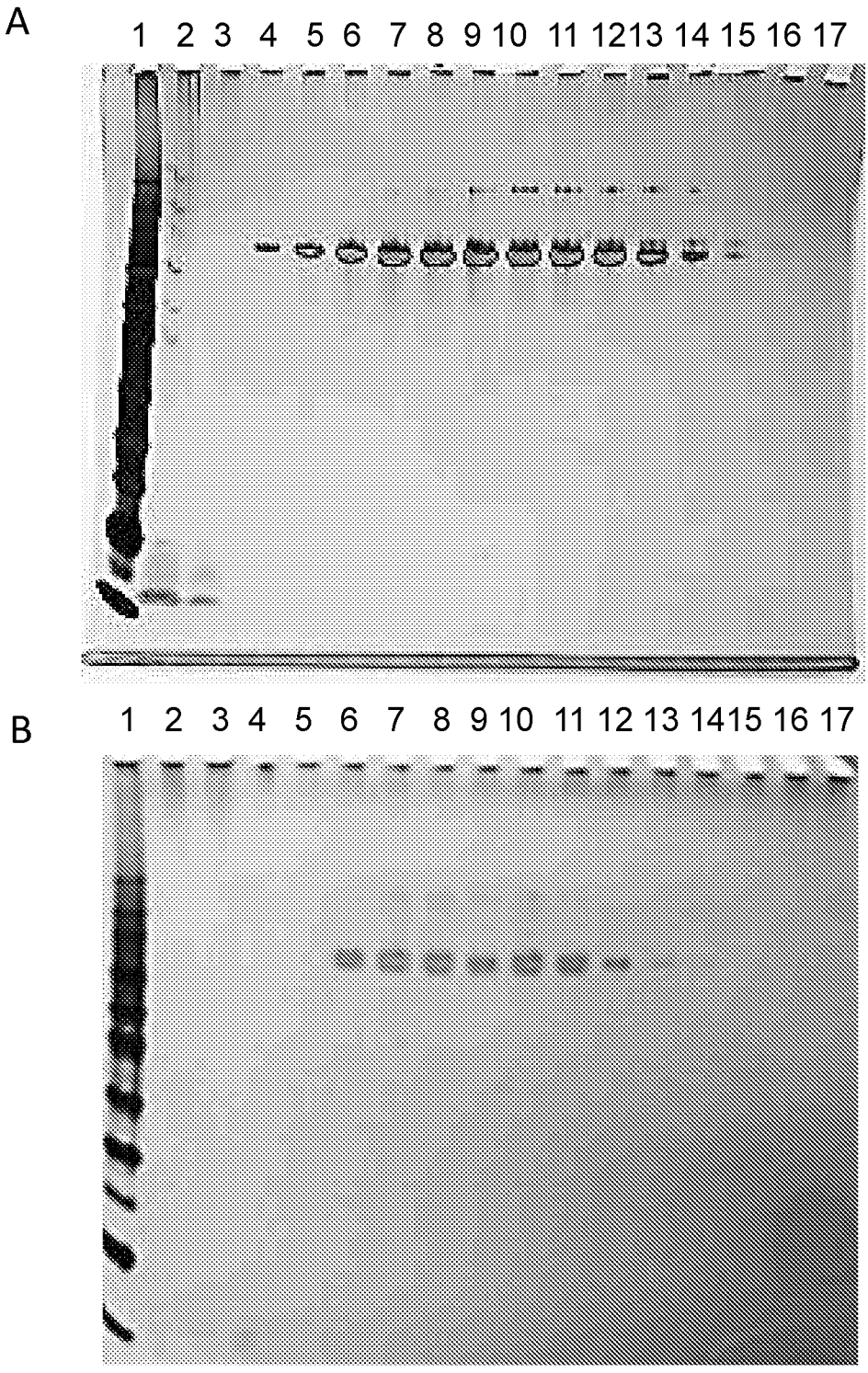

A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
B
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

FIG. 40

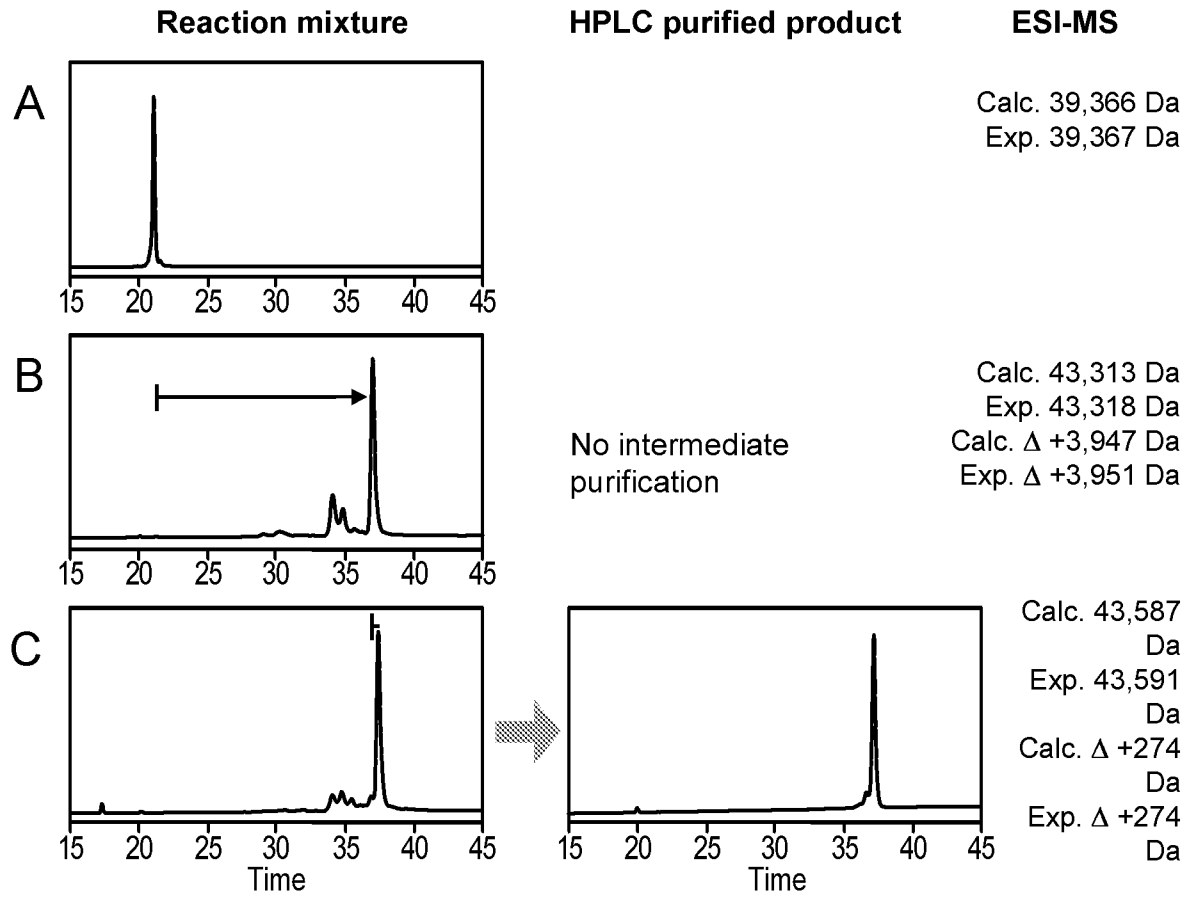

Reaction mixture
HPLC purified product
ESI-MS
A
Calc. 39,366 Da
Exp. 39,367 Da
B
No intermediate purification
Calc. 43,313 Da
Exp. 43,318 Da
Calc. Δ +3,947 Da
Exp. Δ +3,951 Da
C
Calc. 43,587 Da
Exp. 43,591 Da
Calc. Δ +274 Da
Exp. Δ +274 Da
15 20 25 30 35 40 45
Time

FIG. 41

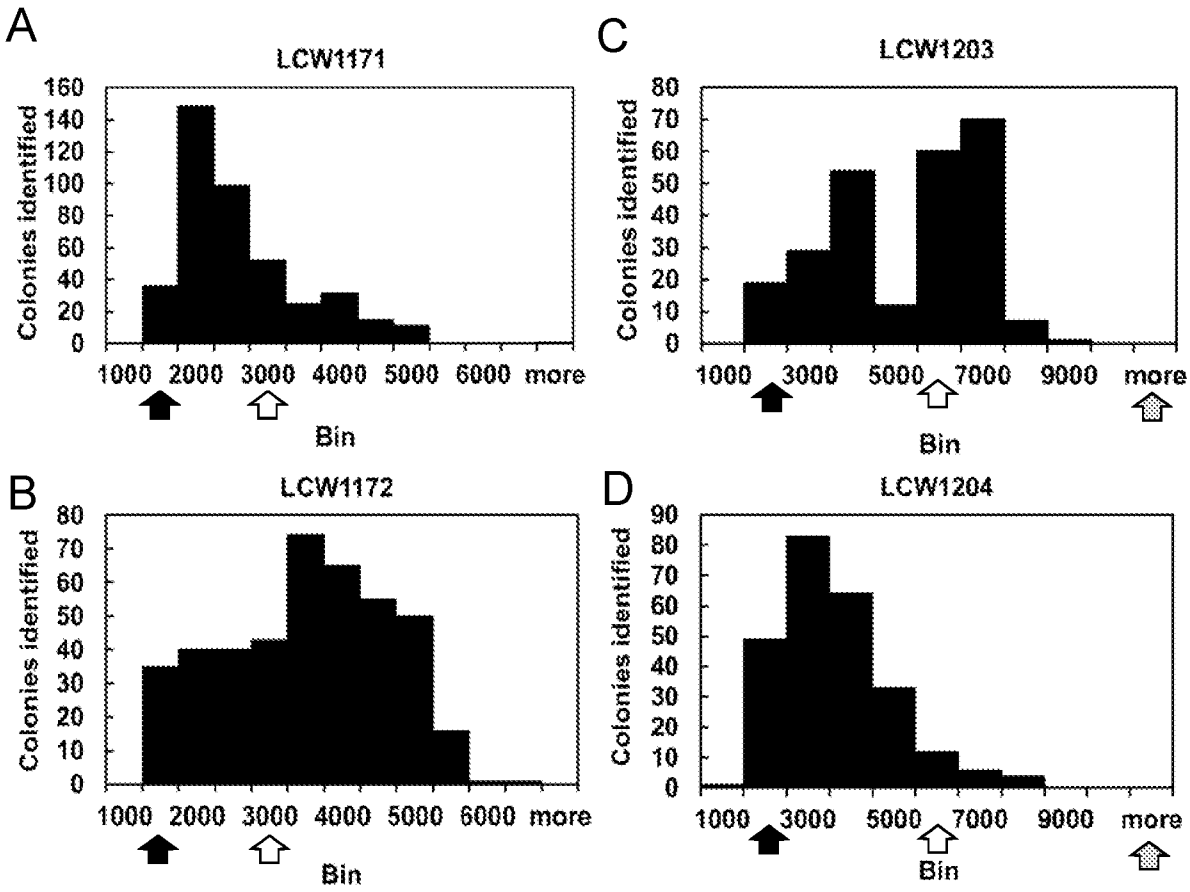

A
LCW1171
Colonies identified
160
140
120
100
80
60
40
20
0
1000 2000 3000 4000 5000 6000 more
Bin
C
LCW1203
Colonies identified
80
70
60
50
40
30
20
10
0
1000 3000 5000 7000 9000 more
Bin
B
LCW1172
Colonies identified
80
70
60
50
40
30
20
10
0
1000 2000 3000 4000 5000 6000 more
Bin
D
LCW1204
Colonies identified
90
80
70
60
50
40
30
20
10
0
1000 3000 5000 7000 9000 more
Bin

FIG. 42

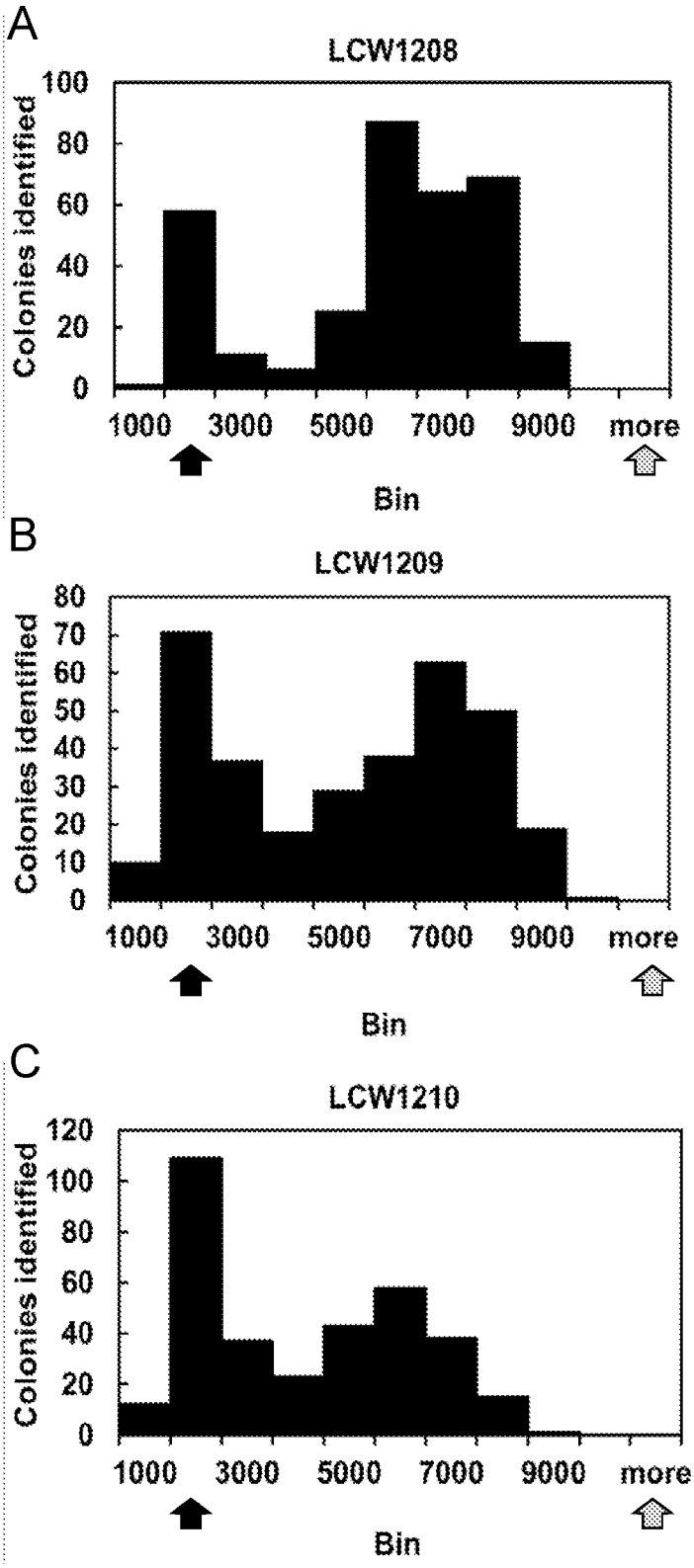

A
LCW1208
Colonies identified
100
80
60
40
20
0
1000
3000
5000
7000
9000
more
Bin
B
LCW1209
Colonies identified
80
70
60
50
40
30
20
10
0
1000
3000
5000
7000
9000
more
Bin
C
LCW1210
Colonies identified
120
100
80
60
40
20
0
1000
3000
5000
7000
9000
more
Bin

FIG. 43

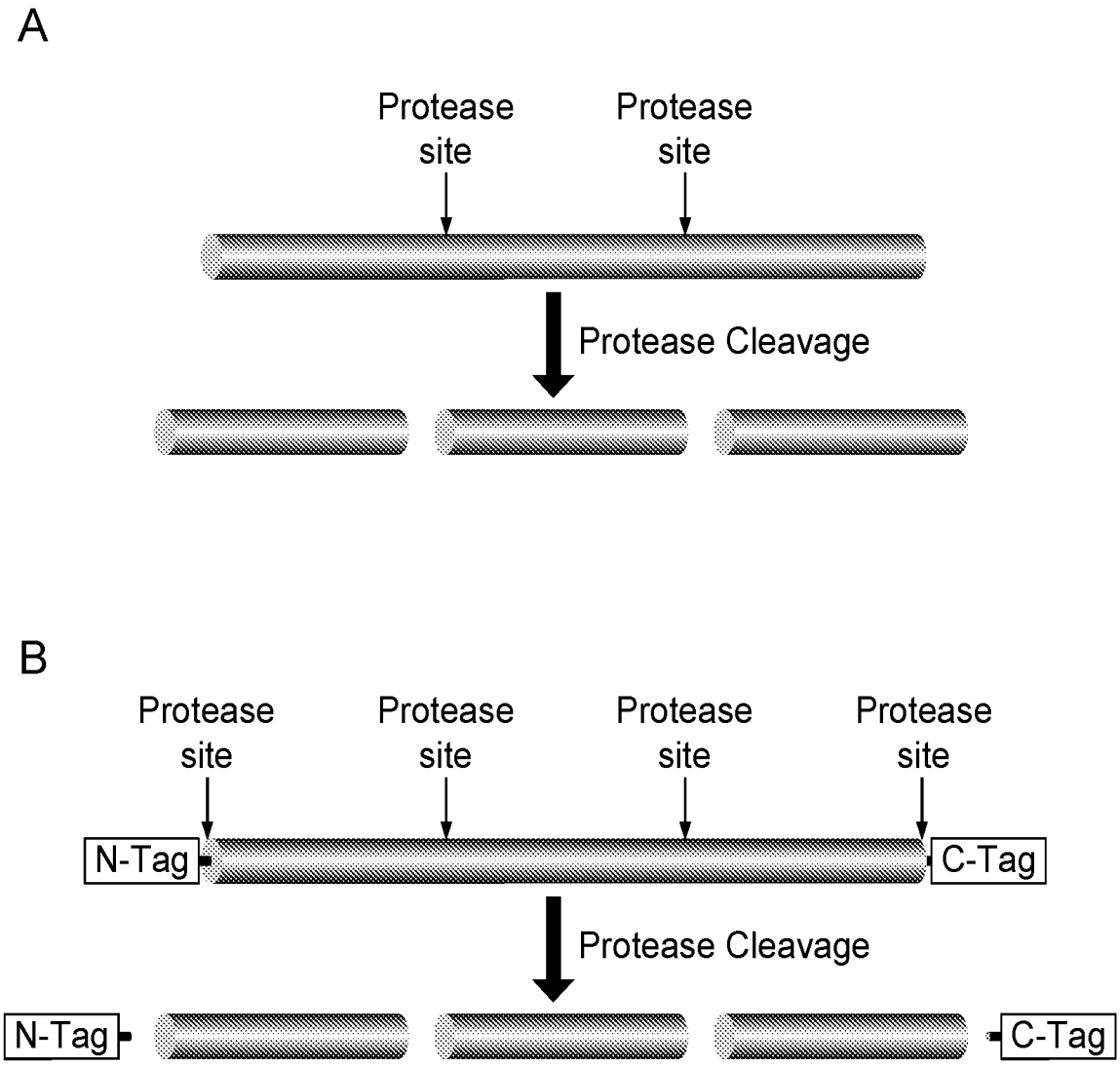

A
Protease site
Protease site
Protease Cleavage
B
Protease site
Protease site
Protease site
Protease site
N-Tag
C-Tag
Protease Cleavage
N-Tag
C-Tag

FIG. 44

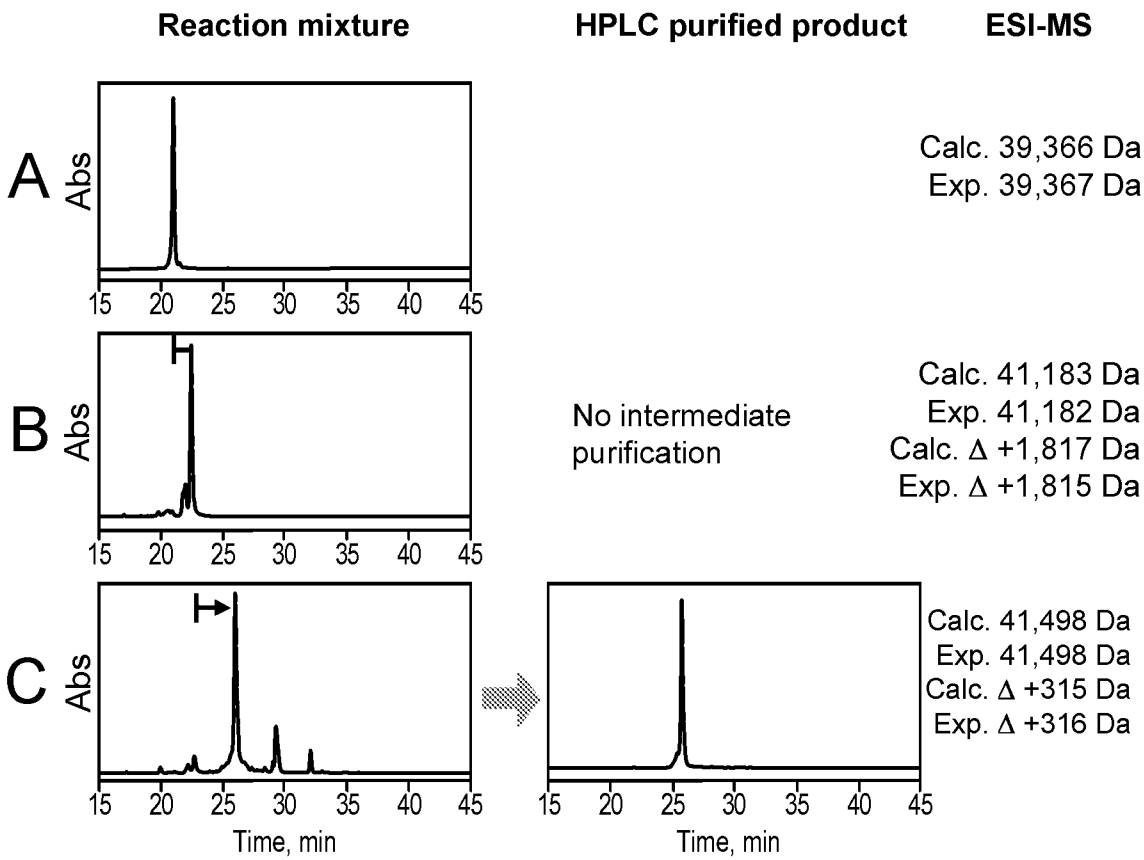

Reaction mixture
HPLC purified product
ESI-MS
A
Abs
Calc. 39,366 Da
Exp. 39,367 Da
B
Abs
No intermediate purification
Calc. 41,183 Da
Exp. 41,182 Da
Calc. Δ +1,817 Da
Exp. Δ +1,815 Da
C
Abs
Calc. 41,498 Da
Exp. 41,498 Da
Calc. Δ +315 Da
Exp. Δ +316 Da
15 20 25 30 35 40 45
Time, min

FIG. 45

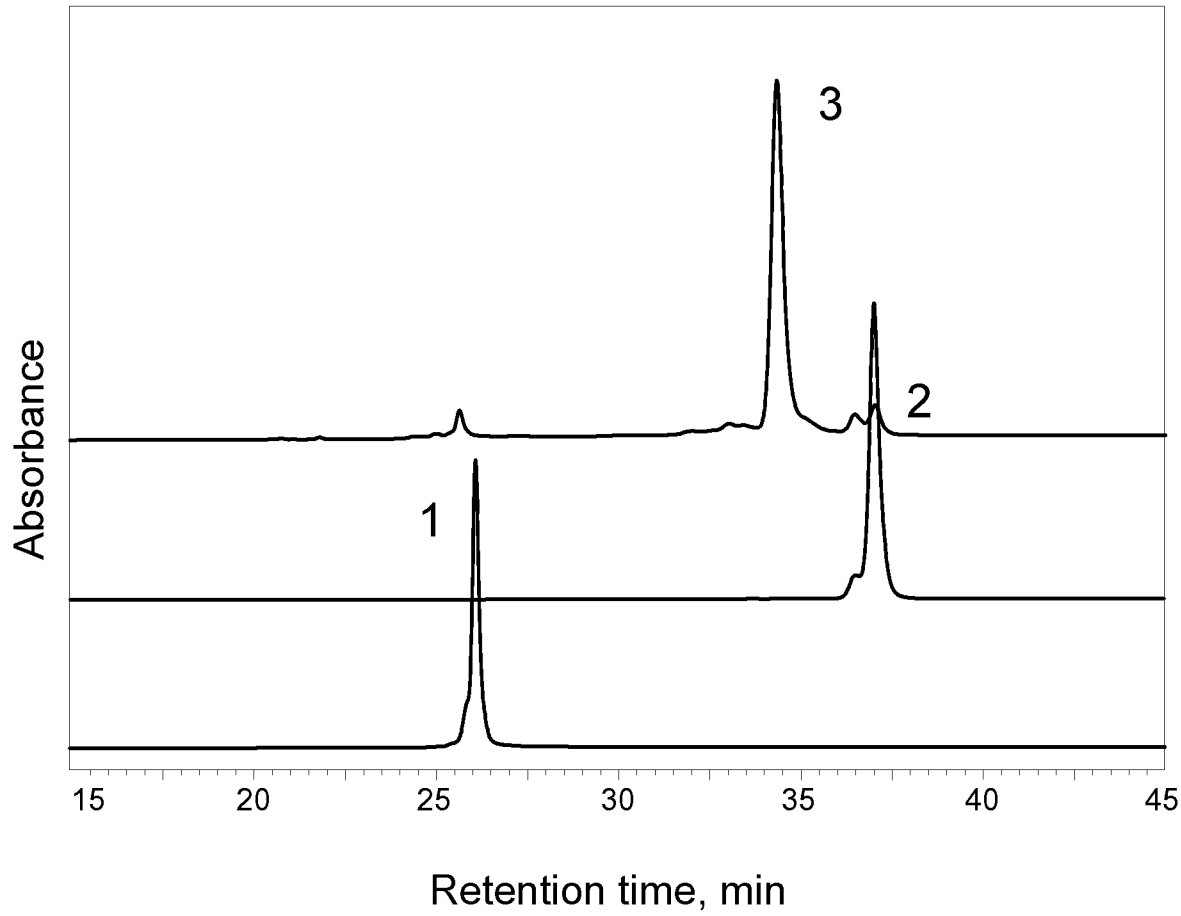


FIG. 46

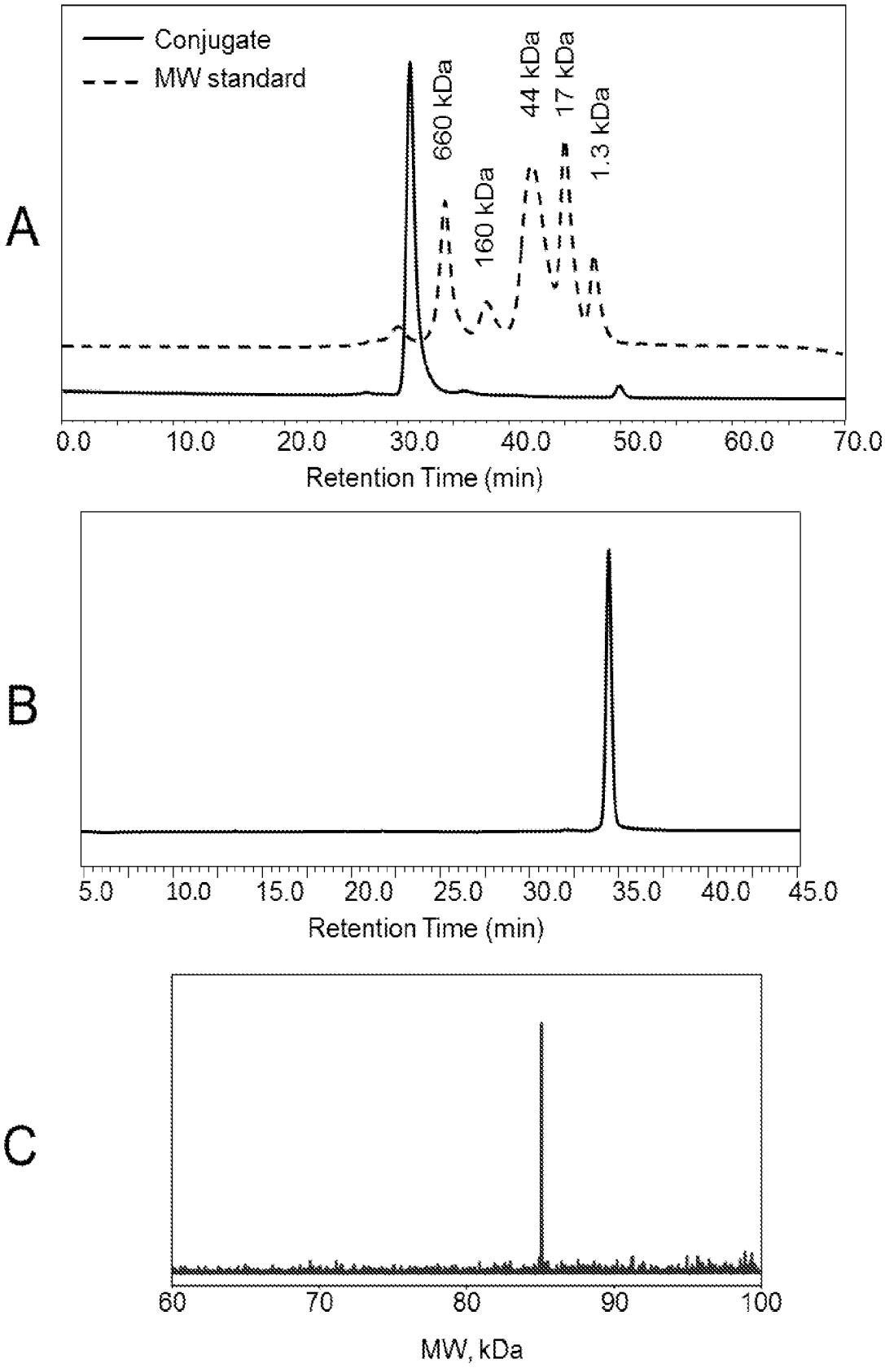

A
Conjugate
MW standard
660 kDa
160 kDa
44 kDa
17 kDa
1.3 kDa
0.0
10.0
20.0
30.0
40.0
50.0
60.0
70.0
Retention Time (min)
B
5.0
10.0
15.0
20.0
25.0
30.0
35.0
40.0
45.0
Retention Time (min)
C
60
70
80
90
100
MW, kDa

FIG. 47

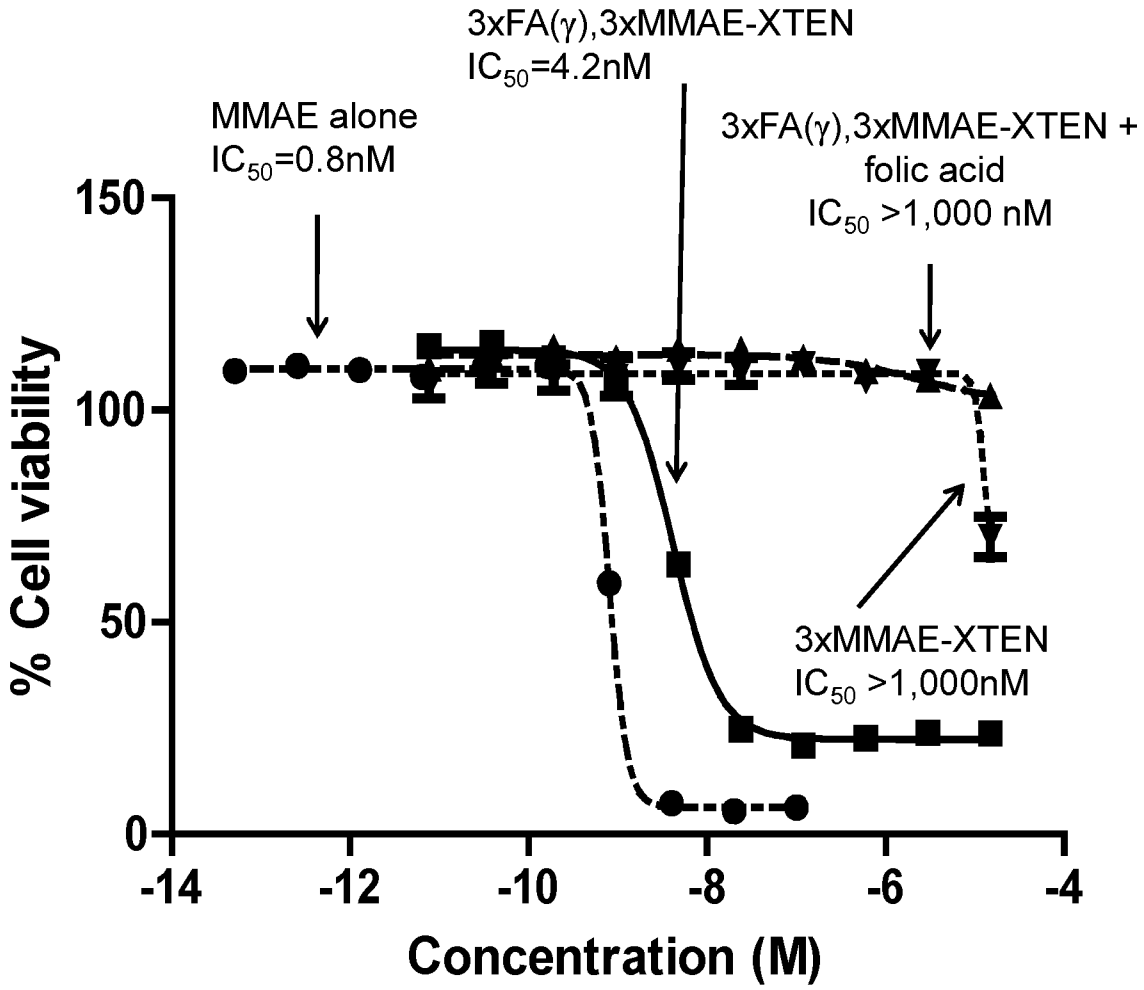
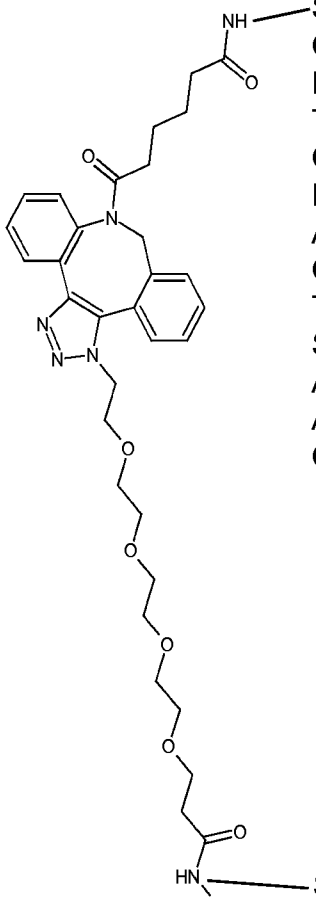

3xFA(γ),3xMMAE-XTEN
IC50=4.2nM
MMAE alone
IC50=0.8nM
3xFA(γ),3xMMAE-XTEN +
folic acid
IC50 >1,000 nM
3xMMAE-XTEN
IC50 >1,000nM
% Cell viability
150
100
50
0
-14
-12
-10
-8
-6
-4
Concentration (M)

FIG. 48

A



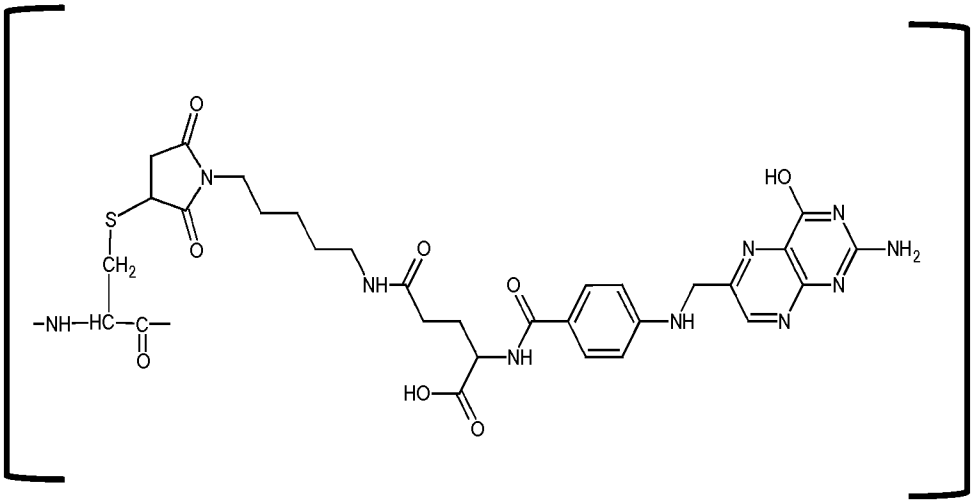
SAGSPTAEAAG**X**GTAEAAGTSESATPESGPGSEPATS
GSETPGTSESATPESGPGTSTEPSEGSAPGTSESATP
ESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTS
TEEGTSESATPESGPGTSTEPSEGSAPGTSESATPES
GPGSEPATSGSETPGTSESATPESGPGSEPATSGSET
PGTSESATPESGPGTSTEPSEGSAPTAEAAG**X**GTAEA
AGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETP
GTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEG
TSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT
SESATPESGPGSEPATSGSETPGSEPATSGSETPGSP
AGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEP
ATSGSETPTAEAA
G**X**GTAEAASASR

SAGSPTAEAAG**Z**GTAEAAGTSESATPESGPGSEPATS
GSETPGTSESATPESGPGTSTEPSEGSAPGTSESATP
ESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTS
TEEGTSESATPESGPGTSTEPSEGSAPGTSESATPES
GPGSEPATSGSETPGTSESATPESGPGSEPATSGSET
PGTSESATPESGPGTSTEPSEGSAPTAEAAG**Z**GTAEA
AGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETP
GTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEG
TSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT
SESATPESGPGSEPATSGSETPGSEPATSGSETPGSP
AGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEP
ATSGSETPTAEAAG**Z**GTAEAASASR

FIG. 49A

B

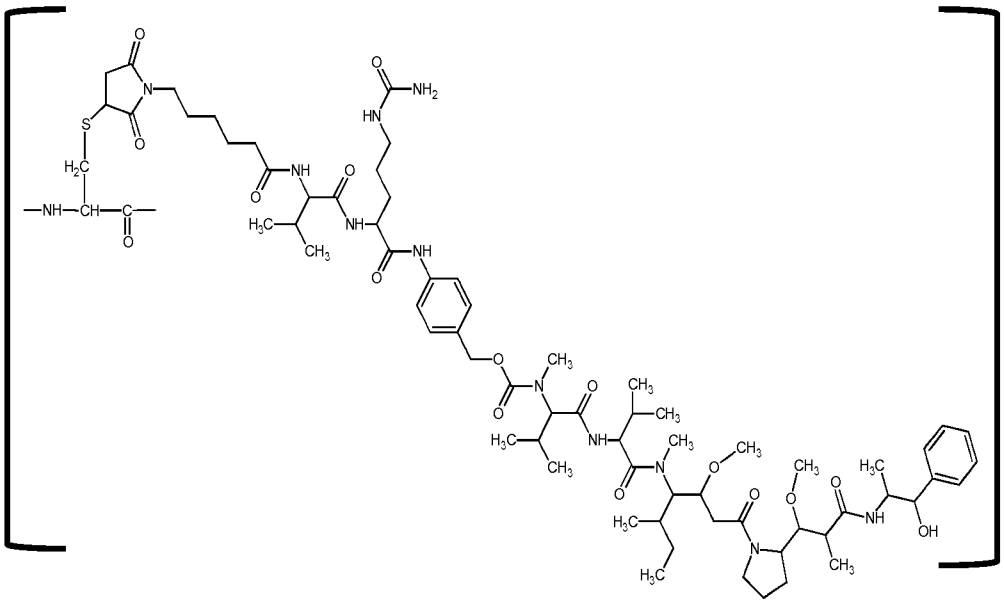
X =



Residue X = residue of Cys modified with Folate-γ-Mal

C

Z =



Residue Z = residue of Cys modified with MC-Val-Cit-PAB-MMAE

FIG. 49B, C

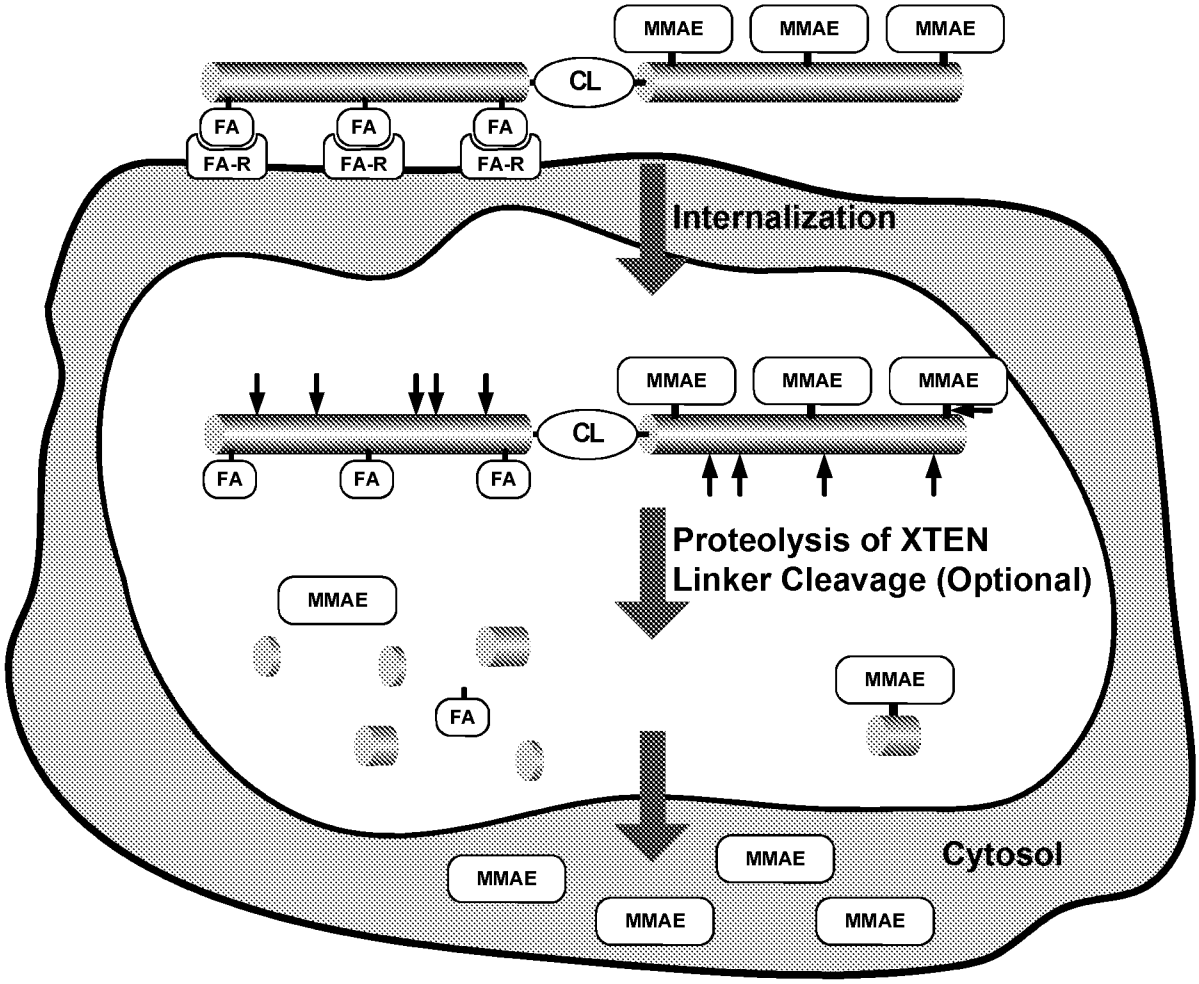

MMAE
MMAE
MMAE
CL
FA
FA
FA
FA-R
FA-R
FA-R
Internalization
MMAE
MMAE
MMAE
CL
FA
FA
FA
Proteolysis of XTEN
Linker Cleavage (Optional)
MMAE
MMAE
FA
MMAE
MMAE
MMAE
MMAE
Cytosol

FIG. 50

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 13/28117

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C07K 14/00 (2013.01)
USPC - 530/350, 402
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8) - C07K 14/00 (2013.01)
USPC - 530/350, 402

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PatBase, PubMed; auristatin, monomethylauristatin, MMAE, dolastatin, vedotin, folic, folate, XTEN, alkyne, azide, crosslink, click reaction, click chemistry

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2011/0142859 A1 (EBENS et al.) 16 June 2011 (16.06.2011) para [0011], [0113], [0302], [0641])	1-5, 15-16, 19
A	US 2005/0032081 A1 (JU et al.) 10 February 2005 (10.02.2005) para [0010], [0071]-[0074]; Fig 6, Fig 7	1-5, 15-16, 19
A	US 2010/0239554 A1 (SCHELLENBERGER et al.) 23 September 2010 (23.09.2010) para [0127], [0159]; SEQ ID NO: 213; SEQ ID NO: 909	1-5, 15-16, 19
A	US 2011/0288005 A1 (SILVERMAN et al.) 24 November 2011 (24.11.2011) SEQ ID NOs: 119, 768	1-5, 15-16
A	US 2009/0280056 A1 (DENNIS et al.) 12 November 2009 (12.11.2009) para [0026]; [0029]; [0059]; [0060]	1-5, 15-16, 19

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 May 2013 (30.05.2013)	17 JUN 2013
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 13/28117

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 6-14, 17-18, and 20-48
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.