

**PROTEASE ACTIVATABLE CYTOKINES AND METHODS FOR
MAKING AND USING THE SAME**

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of the United States Provisional Application Serial No. 63/607,898, filed December 8, 2023, the content of which is hereby incorporated by reference in its entirety.

REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

[0002] The content of the electronic sequence listing (95KG-385485-WO.xml; Size: 2,427,153 bytes; and Date of Creation: December 5, 2024) is herein incorporated by reference in its entirety.

BACKGROUND

[0003] Cytokines have the potential to treat a variety of diseases or conditions. However, the therapeutic utility of cytokines can be limited due to the cellular toxicity, short half-life, need for repetitive or frequent dosing, and the potential to elicit undesired immune response in the patients.

[0004] Most cytokine products in the clinical setting are extremely potent. Interleukins, such as IL-2 and IL-12, and IFN- α are cytokines, produced primarily by cells of the immune system to signal and organize the immune response. In cancer, cytokines facilitate the ability of the immune system to recognize tumor cells as abnormal and harmful to the host. Cytokines further increase the proliferation of, enhance the survival of, and direct a variety of immune cell types to infiltrate the tumor microenvironment and promote potent anti-tumor immune responses resulting in tumor cell killing and tumor clearance. This limits the practical applications of cytokines in a therapeutic setting, particularly in anti-cancer indications.

[0005] Interleukin-12 (IL-12) in particular, has been recognized as having potential to be an ideal payload for tumor immunotherapy. It can activate both the innate and the adaptive components of the immune system. IL-12 stimulates the production of IFN- γ and activates NK cells, as well as CD8+ and CD4+ T cells. In addition, this cytokine also induces antiangiogenic chemokines, remodeling of the tumor extracellular matrix and stimulation of MHC class I molecules expression, making it an extremely attractive anticancer candidate. However, while researchers have shown encouraging preclinical data, the severe toxicity profile of this cytokine has prevented dose escalation and significantly curbed clinical potential as an anticancer agent. Although multiple clinical trials have been on-going since the first human clinical trial of IL-12 in 1996, an FDA-approved IL-12 product remains elusive.

[0006] This presents a significant unmet need for new strategies that can overcome therapeutic index challenges for use of cytokines as anticancer agents. If the potency of cytokines like IL-12 could be safely harnessed and the toxicity challenges could be controlled, these agents could serve as powerful therapeutics for potential use against a broad spectrum of cancers.

BRIEF DESCRIPTION

[0007] The present disclosure provides, among other things, a fusion protein comprising a cytokine, a linker comprising a protease-cleavable release segment, and a mask polypeptide, wherein the linker comprising the release segment is positioned between the cytokine and the mask polypeptide, and wherein the release segment is capable of being cleaved by at least one protease that is present in a tumor. Aspects disclosed herein address a long-felt unmet need for activatable cytokines for cancer therapeutics, that have an increased therapeutic index.

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

SUMMARY

[0009] In one aspect, the disclosure provides a fusion protein comprising: a cytokine, a linker (Linker1) comprising a protease-cleavable release segment (RS1), and a mask polypeptide (Mask1), wherein the linker (Linker1) comprising the release segment (RS1) is positioned between the cytokine and the mask polypeptide (Mask1); wherein the release segment is capable of being cleaved by at least one protease that is present in a tumor; and wherein the release segment (RS1) (i) is not capable of being cleaved by legumain in human plasma, (ii) is cleaved by legumain in human plasma at a rate that is less than 25% of the rate that EAGRSANHTPAGLTGP (RSR-2295; SEQ ID NO:7048) is cleaved by legumain in human plasma, or (iii) comprises an amino acid sequence comprising the sequence EAGRSAXHTPAGLTGP, wherein X is any amino acid other than N (SEQ ID NO:7627).

[0010] In some embodiments, the cytokine is an interleukin, a transforming growth factor, an interferon, a tumor necrosis factor, a chemokine, or granulocyte macrophage-colony stimulating factor.

[0011] In some embodiments, the cytokine is an interleukin. In some embodiments, the interleukin is selected from IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, IL-21, and IL-25. In some embodiments, the interleukin is IL-12. In some embodiments, the IL-12 is a disulfide-linked heterodimer comprising a p35 polypeptide and a p40 polypeptide, and wherein the linker and mask are covalently joined to (i) the N-terminus or C-terminus of the p35 polypeptide, or (ii) the N-terminus or C-terminus of the p40 polypeptide.

[0012] In one aspect, the disclosure provides a fusion protein comprising: (a) a first polypeptide subunit comprising the following elements in an N-to-C or C-to-N terminal orientation: i. a first mask polypeptide (Mask1), ii. a first linker (Linker1) comprising a first protease-cleavable release segment (RS1), iii. an IL-12 p35 polypeptide, iv. a second linker (Linker2) comprising a second protease-cleavable release segment (RS2), and v. a second mask polypeptide (Mask2); and (b) a second polypeptide subunit comprising the following elements in an N-to-C or C-to-N terminal orientation: i. a third mask polypeptide (Mask3), ii. a third linker (Linker3) comprising a third protease-cleavable release segment (RS3), iii. an IL-12 p40 polypeptide, iv. a fourth linker (Linker4) comprising a fourth protease-cleavable release segment (RS4), and v. a fourth mask polypeptide (Mask4); wherein the IL-12 p35 polypeptide of the first polypeptide subunit and the IL-12 p40 polypeptide of the second polypeptide subunit are disulfide-linked thereby forming a disulfide-linked heterodimer; wherein each release segment (RS1, RS2, RS3, and RS4) is capable of being cleaved by at least one protease that is present in a tumor; and wherein each release segment (RS1, RS2, RS3, and RS4) (i) is not capable of being cleaved by legumain in human plasma, (ii) is cleaved by legumain in human plasma at a rate that is less than 25% of the rate that RSR-2295 (EAGRSANHTPAGLTGP; SEQ ID NO:7048) is cleaved by legumain in human plasma, or (iii) comprises an amino acid sequence comprising the sequence EAGRSAXHTPAGLTGP, wherein X is any amino acid other than N (SEQ ID NO:7627).

[0013] In some embodiments, the first mask polypeptide (Mask1) comprises a first barcode fragment (Barcode1), the second mask polypeptide (Mask2) comprises a second barcode fragment (Barcode2), the third mask polypeptide (Mask3) comprises a third barcode fragment (Barcode3), and the fourth mask polypeptide (Mask4) comprises a fourth barcode fragment (Barcode4); and wherein each barcode fragment (Barcode1, Barcode2, Barcode3, and Barcode4) is releasable from the fusion protein upon digestion with a non-mammalian protease, and each barcode fragment differs in both sequence and molecular weight from all other barcode fragments or peptide fragments that are releasable from the fusion protein upon complete digestion of the fusion protein by the non-mammalian protease.

[0014] In one aspect, the disclosure provides a fusion protein comprising: (a) a first polypeptide subunit comprising the following elements in an N-to-C or C-to-N terminal orientation: i. a first mask polypeptide (Mask1) comprising a first barcode fragment (Barcode1), ii. a first linker (Linker1) comprising a first protease-cleavable release segment (RS1), iii. an IL-12 p35 polypeptide, iv. a second linker (Linker2) comprising a second protease-cleavable release segment (RS2), and v. a second mask polypeptide (Mask2) comprising a second barcode fragment (Barcode2); and (b) a second polypeptide subunit comprising the following elements in an N-to-C or C-to-N terminal orientation: i. a third mask polypeptide (Mask3) comprising a third barcode fragment (Barcode3), ii. a third linker (Linker3) comprising a third protease-cleavable release segment (RS3), iii. an IL-12 p40 polypeptide, iv. a fourth linker (Linker4) comprising a fourth protease-cleavable release segment

(RS4), and v. a fourth mask polypeptide (Mask4) comprising a fourth barcode fragment (Barcode4); wherein the IL-12 p35 polypeptide of the first subunit and the IL-12 p40 polypeptide of the second unit are disulfide-linked thereby forming a disulfide-linked heterodimer; wherein each release segment (RS1, RS2, RS3, and RS4) is capable of being cleaved by at least one protease that is present in a tumor; and wherein each barcode fragment (Barcode1, Barcode2, Barcode3, and Barcode4) is releasable from the fusion protein upon digestion with a non-mammalian protease, and each barcode fragment differs in both sequence and molecular weight from all other barcode fragments or peptide fragments that are releasable from the fusion protein upon complete digestion of the fusion protein by the non-mammalian protease.

[0015] In some embodiments, each release segment comprises an amino acid sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity, to a sequence of Table 6a.

[0016] In some embodiments, each release segment (RS1, RS2, RS3, and RS4) (i) is not capable of being cleaved by legumain in human plasma, (ii) is cleaved by legumain in human plasma at a rate that is less than 25% of the rate that RSR-2295 (EAGRSANHTPAGLTGP; SEQ ID NO:7048) is cleaved by legumain in human plasma, or (iii) comprises an amino acid sequence comprising the sequence EAGRSAXHTPAGLTGP, wherein X is any amino acid other than N (SEQ ID NO:7627).

[0017] In some embodiments, each release segment is a substrate for at least one protease of Table 5.

[0018] In some embodiments, each release segment is capable of being cleaved by uPA, ST14, MMP2, MMP7, MMP9, and MMP14.

[0019] In some embodiments, each release segment is not capable of being cleaved by legumain in human plasma.

[0020] In some embodiments, each release segment is not capable of being cleaved by legumain in human blood, plasma, or serum.

[0021] In some embodiments, each release segment is not capable of being cleaved upon incubation with about 1nM or less legumain for about 20 hours.

[0022] In some embodiments, each release segment is cleaved by legumain in human plasma at a rate that is less than 25%, less than 15%, less than 10%, less than 5%, or less than 2.5% of the rate that RSR-2295 (EAGRSANHTPAGLTGP; SEQ ID NO:7048) is cleaved by legumain in human plasma.

[0023] In some embodiments, each release segment comprises the amino acid sequence of RSR-3213 (EAGRSASHTPAGLTGP; SEQ ID NO:7628)

[0024] In some embodiments, each mask polypeptide is an extended length non-natural polypeptide (ELNN).

[0025] In some embodiments, each ELNN is, independently, characterized in that: at least 90% of the amino acid residues of each ELNN are selected from glycine (G), alanine (A), serine (S), threonine (T), glutamate (E), proline (P), or any combination thereof; and each ELNN comprises at least 3 types of amino acid residues selected from the group consisting of G, A, S, T, E, and P.

[0026] In some embodiments, each ELNN, independently, comprises a plurality of non-overlapping sequence motifs that are each from 9 to 14 amino acid residues in length, wherein the plurality of non-overlapping sequence motifs comprises a set of non-overlapping sequence motifs, wherein each non-overlapping sequence motif of the set of non-overlapping sequence motifs is repeated at least two times in the ELNN.

[0027] In some embodiments, the plurality of non-overlapping sequence motifs comprises at least one non-overlapping sequence motif that occurs only once within the ELNN.

[0028] In some embodiments, the non-overlapping sequence motifs comprise one of or any combination of the sequence motifs listed in Table 1b.

[0029] In some embodiments, the non-overlapping sequence motifs comprise at least 2, 3, or 4 of the sequence motifs listed in Table 1b.

[0030] In some embodiments, the non-overlapping sequence motifs comprise any one of or any combination of GTSTEPSEGSAP (SEQ ID NO:189), GTSESATPESGP (SEQ ID NO:188), GSGPGTSESATP (SEQ ID NO:201), GSEPATSGSETP (SEQ ID NO:187), GSPAGSPTSTEE (SEQ ID NO:186), and GTSPSATPESGP (SEQ ID NO:202).

[0031] In some embodiments, each ELNN comprises at least 4 types of amino acid residues selected from the group consisting of G, A, S, T, E, and P.

[0032] In some embodiments, the amino acid residues of each ELNN consists of A, E, G, S, P, and/or T.

[0033] In some embodiments, each ELNN, independently, has a length of least 100 amino acid residues, at least 200 amino acid residues, or at least 250 amino acid residues.

[0034] In some embodiments, each ELNN, independently, has a length of 100 to 1000 amino acid residues, 100 to 500 amino acid residues, 200 to 100 amino acid residues, 200 to 500 amino acid residues, 200 to 300 amino acid residues, 250 to 1000 amino acid residues, 250 to 500 amino acid residues, or 250 to 300 amino acid residues.

[0035] In some embodiments, each ELNN has a length of between 250 and 300 amino acid residues.

[0036] In some embodiments, each ELNN has a length of 287 or 288 amino acid residues.

[0037] In some embodiments, each ELNN comprises an amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an amino acid sequence listed in Table 3a or 3b.

[0038] In some embodiments, the first polypeptide mask is a first ELNN (ELNN1), and wherein ELNN1 comprises an amino sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity to:

ASGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSEPATSGSETPGT
SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP
ESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT
SESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE
GSAPGSEPATSGSETPGTSESAT (SEQ ID NO: 8026).

[0039] In some embodiments, the second polypeptide mask is a second ELNN (ELNN2), and wherein ELNN2 comprises an amino sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity to:

ATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSET
PGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSES
ATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESG
PGTSASATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTE
PSEGSAPGSEPATSGSETPGEPEA (SEQ ID NO: 8028).

[0040] In some embodiments, the third polypeptide mask is a third ELNN (ELNN3), and wherein ELNN3 comprises an amino sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity to:

ASGPGTSTEPSEGSAPGTSESATPESGPGSGPATSESATPGTSESATPESGPGSEPATSGSETPGT
SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP
ESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT
SESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE
GSAPGSEPATSGSETPGTSESAT (SEQ ID NO: 8023).

[0041] In some embodiments, the fourth polypeptide mask is a fourth ELNN (ELNN4), and wherein ELNN4 comprises an amino sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity to:

ATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSA
PGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTE
PSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTE
EGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSPSATPESGPGSEPA
TSGSETPGSEPATSGSETPGSPAG (SEQ ID NO: 8025).

[0042] In some embodiments, the IL-12 p35 polypeptide is a human IL-12 p35 polypeptide and the IL12 p40 polypeptide is a human IL-12 p40 polypeptide.

[0043] In some embodiments, the human IL-12 p35 polypeptide comprises an amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity to:

RNLPVATPDPGMFPC LHHSQNLLRAVS NMLQKARQTLEFY PCTSEEIDHEDITKDKTSTVEAC
LPLELTKNESCLNSRETSFITNGSCLASRKTSFMMALCLSSIEDLKMYQVEFKTMNAKLLM
DPKRQIFLDQNMLAVIDELMQALNFNSETVPQKSSLEEPDFYKTKIKLCILLHAFRIRAVTIDR
VMSYLNAS (SEQ ID NO: 8050).

[0044] In some embodiments, the human IL-12 p35 polypeptide comprises the amino acid sequence of:

RNLPVATPDPGMFPC LHHSQNLLRAVS NMLQKARQTLEFY PCTSEEIDHEDITKDKTSTVEAC
LPLELTKNESCLNSRETSFITNGSCLASRKTSFMMALCLSSIEDLKMYQVEFKTMNAKLLM
DPKRQIFLDQNMLAVIDELMQALNFNSETVPQKSSLEEPDFYKTKIKLCILLHAFRIRAVTIDR
VMSYLNAS (SEQ ID NO: 8050).

[0045] In some embodiments, the human IL-12 p35 polypeptide is an N195 aglycosylated variant.

[0046] In some embodiments, the human IL-12 p35 polypeptide comprises the amino acid sequence of:

RNLPVATPDPGMFPC LHHSQNLLRAVS NMLQKARQTLEFY PCTSEEIDHEDITKDKT
STVEACLPLELTKNESCLNSRETSFITNGSCLASRKTSFMMALCLSSIEDLKMYQVE
FKTMNAKLLMDPKRQIFLDQNMLAVIDELMQALNFNSETVPQKSSLEEPDFYKTKIK
LCILLHAFRIRAVTIDRVMSY LXAS (SEQ ID NO: 8055), wherein X is any amino acid
residue other than asparagine (N).

[0047] In some embodiments, the human IL-12 p35 polypeptide comprises an amino acid sequence selected from:

RNLPVATPDPGMFPC LHHSQNLLRAVS NMLQKARQTLEFY PCTSEEIDHEDITKDKTSTVEAC
LPLELTKNESCLNSRETSFITNGSCLASRKTSFMMALCLSSIEDLKMYQVEFKTMNAKLLM
DPKRQIFLDQNMLAVIDELMQALNFNSETVPQKSSLEEPDFYKTKIKLCILLHAFRIRAVTIDR
VMSY LSSA (SEQ ID NO: 8052);

RNLPVATPDPGMFPC LHHSQNLLRAVS NMLQKARQTLEFY PCTSEEIDHEDITKDKTSTVEAC
LPLELTKNESCLNSRETSFITNGSCLASRKTSFMMALCLSSIEDLKMYQVEFKTMNAKLLM
DPKRQIFLDQNMLAVIDELMQALNFNSETVPQKSSLEEPDFYKTKIKLCILLHAFRIRAVTIDR
VMSY LQAS (SEQ ID NO: 8053); and

RNLPVATPDPGMFPC LHHSQNLLRAVS NMLQKARQTLEFY PCTSEEIDHEDITKDKTSTVEAC

LPLELTKNESCLNSRETSFITNGSCLASRKTSFMMALCLSSIEDLKMYQVEFKTMNAKLLM
DPKRQIFLDQNMLAVIDELMQALNFNSETVPQKSSLEEDFYKTKIKLCILLHAFRIRAVTIDR
VMSYLDAS (SEQ ID NO: 8054).

[0048] In some embodiments, the human IL-12 p40 polypeptide comprises an amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity to:

IWELKKDVYVVELDWYPDAPGEMVVLTCDTPEEDGITWTLDQSSEVLGSGKTLTIQVKEFG
DAGQYTCHKGGEVLSHSLLLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRFTCWW
LTTISTDLTFSVKSSRGSSDPQGVTCGAATLSAERVGRDNKEYEYSVECQEDSACPAAEESLPI
EVMVDAVHKLKYENYTSSFFIRDIKPDPPKNLQLKPLKNSRQVEVSWEYPDTWSTPHSYFSL
TFCVQVQGKSKREKKDRVFTDKTSATVICRKNASISVRAQDRYYSSSWSEWASVPCS (SEQ
ID NO: 8051).

[0049] In some embodiments, the human IL-12 p40 polypeptide comprises the amino acid sequence of:

IWELKKDVYVVELDWYPDAPGEMVVLTCDTPEEDGITWTLDQSSEVLGSGKTLTIQVKEFG
DAGQYTCHKGGEVLSHSLLLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRFTCWW
LTTISTDLTFSVKSSRGSSDPQGVTCGAATLSAERVGRDNKEYEYSVECQEDSACPAAEESLPI
EVMVDAVHKLKYENYTSSFFIRDIKPDPPKNLQLKPLKNSRQVEVSWEYPDTWSTPHSYFSL
TFCVQVQGKSKREKKDRVFTDKTSATVICRKNASISVRAQDRYYSSSWSEWASVPCS (SEQ
ID NO: 8051).

[0050] In some embodiments, the IL-12 p35 polypeptide is a mouse IL-12 p35 polypeptide and the IL-12 p40 polypeptide is a mouse IL-12 p40 polypeptide.

[0051] In some embodiments, each linker further comprises a spacer.

[0052] In some embodiments, each spacer is positioned between a release segment and a cytokine, an IL-12 p35 polypeptide, or an IL-12 p40 polypeptide.

[0053] In some embodiments, each spacer is, independently, characterized in that: at least 90% of the amino acid residues of each spacer are selected from glycine (G), alanine (A), serine (S), threonine (T), glutamate (E), proline (P), or any combination thereof; and each spacer comprises at least 3 types of amino acid residues selected from the group consisting of G, A, S, T, E, and P.

[0054] In some embodiments, each spacer is, independently, 5 to 14 amino acid residues in length.

[0055] In some embodiments, each spacer comprises at least 4 types of amino acid residues selected from the group consisting of G, A, S, T, E, and P.

[0056] In some embodiments, the amino acid residues of each spacer comprise only amino acid residues selected from G, A, S, T, E, and/or P.

[0057] In some embodiments, each spacer is cleavable by a non-mammalian protease.

[0058] In some embodiments, the non-mammalian protease is Glu-C.

[0059] In some embodiments, each spacer, independently, comprises an amino acid sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity, to a sequence listed in Table C.

[0060] In some embodiments, each spacer, independently, comprises an amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity to TSESATPES (SEQ ID NO:98), GTATPESGPG (SEQ ID NO:97), GTPESATPES (SEQ ID NO:99), or GTATPESAP (SEQ ID NO:100).

[0061] In some embodiments, no spacer shares 100% sequence identity with any other spacer.

[0062] In some embodiments, the fusion protein does not comprise a His-tag, wherein a His-tag is a polypeptide sequence comprising an uninterrupted chain of between 5 and 12 histidine residues.

[0063] In some embodiments, each barcode fragment is releasable from the fusion protein upon digestion with a non-mammalian protease and the non-mammalian protease is Glu-C.

[0064] In some embodiments, none of the barcode fragments include any of the following: the N-terminal amino acid of the first polypeptide subunit, the C-terminal acid of the first polypeptide subunit, the N-terminal amino acid of the second polypeptide subunit, and the C-terminal amino acid of the polypeptide subunit.

[0065] In some embodiments, the entirety of each barcode fragment is 5 to 100, 5 to 50, 10 to 100, or 10 to 50 amino acid residues from any of the following: the N-terminal amino acid of the first polypeptide subunit, the C-terminal acid of the first polypeptide subunit, the N-terminal amino acid of the second polypeptide subunit, and the C-terminal amino acid of the polypeptide subunit.

[0066] In some embodiments, each barcode fragment is, independently, at least 4 amino acid residues in length.

[0067] In some embodiments, each barcode fragment is, independently, 4 to 20, 5 to 15, 6 to 12, 7 to 10, 8 to 13, 10 to 13, or 11 to 12 amino acid residues in length.

[0068] In some embodiments, each barcode fragment comprises a glutamate at the C-terminus thereof.

[0069] In some embodiments, each barcode fragment does not include a second glutamate at a position other than the C-terminus of the barcode fragment unless the second glutamate is immediately followed by a proline.

[0070] In some embodiments, the amino acid residue that is directly N-terminal to each barcode fragment in the primary amino acid sequence of the fusion protein is a glutamate.

[0071] In some embodiments, the glutamate that is directly N-terminal to each barcode fragment is not immediately adjacent to another glutamate.

[0072] In some embodiments, none of the barcode fragments includes a glutamate that is immediately adjacent to another glutamate, if present, in the mask polypeptide that contains the barcode fragment.

[0073] In some embodiments, each barcode fragment is, independently, selected from Table 2 or Table 3a.

[0074] In some embodiments, each barcode fragment comprises only amino acid residues selected from A, E, G, S, P, and T.

[0075] In some embodiments, each barcode fragment comprises an amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity, to SGPGSGPGTSE (SEQ ID NO:78), SGPGTSASATPE (SEQ ID NO:1029), SGPGSGPATSE (SEQ ID NO:1028), or SGPGTSPSATPE (SEQ ID NO:79).

[0076] In some embodiments, each barcode fragment is selected from the following four amino acid sequences: SGPGSGPGTSE (SEQ ID NO:78), SGPGTSASATPE (SEQ ID NO:1029), SGPGSGPATSE (SEQ ID NO:1028), or SGPGTSPSATPE (SEQ ID NO:79).

[0077] In some embodiments, the first barcode fragment (Barcode1) comprises the amino acid sequence of SGPGSGPGTSE (SEQ ID NO:78).

[0078] In some embodiments, the second barcode fragment (Barcode2) comprises the amino acid sequence of SGPGTSASATPE (SEQ ID NO:1029).

[0079] In some embodiments, the third barcode fragment (Barcode3) comprises the amino acid sequence of SGPGSGPATSE (SEQ ID NO:1028).

[0080] In some embodiments, the fourth barcode fragment (Barcode4) comprises the amino acid sequence of SGPGTSPSATPE (SEQ ID NO:79).

[0081] In some embodiments, the fusion protein comprises a Glu-C cleavage site comprising one of the following amino acid sequences: ATPESGPG (SEQ ID NO:2223), SGSETPGT (SEQ ID NO:2242), and GTSESATP (SEQ ID NO:204).

[0082] In some embodiments, the fusion protein comprises at least one of the following amino acid sequences: SGPE.SGPGXnSGPE.SGPG (SEQ ID NO:2001), SGPE.SGPGXnATPE.SGPG (SEQ ID NO:2002), SGPE.SGPGXnGTSE.SATP (SEQ ID NO:2003), SGPE.SGPGXnTTPE.SGPG (SEQ ID NO:2004), SGPE.SGPGXnSTPE.SGPG (SEQ ID NO:2005), SGPE.SGPGXnGTPE.SGPG (SEQ ID

NO:2006), SGPE.SGPGXnGTPE.TPGS (SEQ ID NO:2007), SGPE.SGPGXnGTPE.TPGS (SEQ ID NO:2007), SGPE.SGPGXnSGSE.TGTP (SEQ ID NO:2008), SGPE.SGPGXnGTPE.GSAP (SEQ ID NO:2009), SGPE.SGPGXnEPSE.SATP (SEQ ID NO:2010), ATPE.SGPGXnSGPE.SGPG (SEQ ID NO:2011), ATPE.SGPGXnATPE.SGPG (SEQ ID NO:2012), ATPE.SGPGXnGTSE.SATP (SEQ ID NO:2013), ATPE.SGPGXnATSE.SATP (SEQ ID NO:2014), ATPE.SGPGXnTTPE.SGPG (SEQ ID NO:2015), ATPE.SGPGXnSTPE.SGPG (SEQ ID NO:2016), ATPE.SGPGXnGTPE.SGPG (SEQ ID NO:2017), ATPE.SGPGXnGTPE.TPGS (SEQ ID NO:2018), ATPE.SGPGXnSGSE.TGTP (SEQ ID NO:2019), ATPE.SGPGXnGTPE.GSAP (SEQ ID NO:2020), ATPE.SGPGXnEPSE.SATP (SEQ ID NO:2021), GTSE.SATPXnSGPE.SGPG (SEQ ID NO:2022), GTSE.SATPXnATPE.SGPG (SEQ ID NO:2023), GTSE.SATPXnGTSE.SATP (SEQ ID NO:2024), GTSE.SATPXnTTPE.SGPG (SEQ ID NO:2025), GTSE.SATPXnSTPE.SGPG (SEQ ID NO:2026), GTSE.SATPXnGTPE.SGPG (SEQ ID NO:2027), GTSE.SATPXnGTPE.TPGS (SEQ ID NO:2028), GTSE.SATPXnSGSE.TGTP (SEQ ID NO:2029), GTSE.SATPXnGTPE.GSAP (SEQ ID NO:2030), GTSE.SATPXnEPSE.SATP (SEQ ID NO:2031), TTPE.SGPGXnSGPE.SGPG (SEQ ID NO:2032), TTPE.SGPGXnATPE.SGPG (SEQ ID NO:2033), TTPE.SGPGXnGTSE.SATP (SEQ ID NO:2034), TTPE.SGPGXnTTPE.SGPG (SEQ ID NO:2035), TTPE.SGPGXnSTPE.SGPG (SEQ ID NO:2036), TTPE.SGPGXnGTPE.SGPG (SEQ ID NO:2037), TTPE.SGPGXnGTPE.TPGS (SEQ ID NO:2038), TTPE.SGPGXnSGSE.TGTP (SEQ ID NO:2039), TTPE.SGPGXnGTPE.GSAP (SEQ ID NO:2040), TTPE.SGPGXnEPSE.SATP (SEQ ID NO:2041), STPE.SGPGXnSGPE.SGPG (SEQ ID NO:2042), STPE.SGPGXnATPE.SGPG (SEQ ID NO:2043), STPE.SGPGXnGTSE.SATP (SEQ ID NO:2044), STPE.SGPGXnTTPE.SGPG (SEQ ID NO:2045), STPE.SGPGXnSTPE.SGPG (SEQ ID NO:2046), STPE.SGPGXnGTPE.SGPG (SEQ ID NO:2047), STPE.SGPGXnGTPE.TPGS (SEQ ID NO:2048), STPE.SGPGXnSGSE.TGTP (SEQ ID NO:2049), STPE.SGPGXnGTPE.GSAP (SEQ ID NO:2050), STPE.SGPGXnEPSE.SATP (SEQ ID NO:2051), GTPE.SGPGXnSGPE.SGPG (SEQ ID NO:2052), GTPE.SGPGXnATPE.SGPG (SEQ ID NO:2053), GTPE.SGPGXnGTSE.SATP (SEQ ID NO:2054), GTPE.SGPGXnTTPE.SGPG (SEQ ID NO:2055), GTPE.SGPGXnSTPE.SGPG (SEQ ID NO:2056), GTPE.SGPGXnGTPE.SGPG (SEQ ID NO:2057), GTPE.SGPGXnGTPE.TPGS (SEQ ID NO:2058), GTPE.SGPGXnSGSE.TGTP (SEQ ID NO:2059), GTPE.SGPGXnGTPE.GSAP (SEQ ID NO:2060), GTPE.SGPGXnEPSE.SATP (SEQ ID NO:2061), GTPE.TPGSXnSGPE.SGPG (SEQ ID NO:2062), GTPE.TPGSXnATPE.SGPG (SEQ ID NO:2063), GTPE.TPGSXnGTSE.SATP (SEQ ID NO:2064), GTPE.TPGSXnTTPE.SGPG (SEQ ID NO:2065), GTPE.TPGSXnSTPE.SGPG (SEQ ID NO:2066), GTPE.TPGSXnGTPE.SGPG (SEQ ID NO:2067), GTPE.TPGSXnGTPE.TPGS (SEQ ID NO:2068), GTPE.TPGSXnSGSE.TGTP (SEQ ID NO:2069), GTPE.TPGSXnGTPE.GSAP (SEQ ID NO:2070), GTPE.TPGSXnEPSE.SATP (SEQ ID NO:2071), SGSE.TGTPXnSGPE.SGPG (SEQ ID NO:2072), SGSE.TGTPXnATPE.SGPG (SEQ ID NO:2073), SGSE.TGTPXnGTSE.SATP (SEQ ID NO:2074), SGSE.TGTPXnTTPE.SGPG (SEQ ID NO:2075), SGSE.TGTPXnSTPE.SGPG (SEQ ID NO:2076), SGSE.TGTPXnGTPE.SGPG (SEQ ID NO:2077), SGSE.TGTPXnGTPE.TPGS (SEQ ID NO:2078), SGSE.TGTPXnSGSE.TGTP (SEQ ID

NO:2079), SGSE.TGTPX_nGTPE.GSAP (SEQ ID NO:2080), SGSE.TGTPX_nEPSE.SATP (SEQ ID NO:2081), GTPE.GSAPX_nSGPE.SGPG (SEQ ID NO:2082), GTPE.GSAPX_nATPE.SGPG (SEQ ID NO:2083), GTPE.GSAPX_nGTSE.SATP (SEQ ID NO:2084), GTPE.GSAPX_nTTPE.SGPG (SEQ ID NO:2085), GTPE.GSAPX_nSTPE.SGPG (SEQ ID NO:2086), GTPE.GSAPX_nGTPE.SGPG (SEQ ID NO:2087), GTPE.GSAPX_nGTPE.TPGS (SEQ ID NO:2088), GTPE.GSAPX_nSGSE.TGTP (SEQ ID NO:2089), GTPE.GSAPX_nGTPE.GSAP (SEQ ID NO:2090), GTPE.GSAPX_nEPSE.SATP (SEQ ID NO:2091), EPSE.SATPX_nSGPE.SGPG (SEQ ID NO:2092), EPSE.SATPX_nATPE.SGPG (SEQ ID NO:2093), EPSE.SATPX_nGTSE.SATP (SEQ ID NO:2094), EPSE.SATPX_nTTPE.SGPG (SEQ ID NO:2095), EPSE.SATPX_nSTPE.SGPG (SEQ ID NO:2096), EPSE.SATPX_nGTPE.SGPG (SEQ ID NO:2097), EPSE.SATPX_nGTPE.TPGS (SEQ ID NO:2098), EPSE.SATPX_nSGSE.TGTP (SEQ ID NO:2099), EPSE.SATPX_nGTPE.GSAP (SEQ ID NO:2100), or EPSE.SATPX_nEPSE.SATP (SEQ ID NO:2101), wherein each “.” is a Glu-C cleavage site and n is any integer from 0 to 50.

[0083] In some embodiments, the fusion protein comprises at least one of the following amino acid sequences: ATPE.SGPGX_nATPE.SGPG (SEQ ID NO:2102), ATPE.SGPGX_nGTSE.SATP (SEQ ID NO:2103), ATPE.SGPGX_nATSE.SATP (SEQ ID NO:2104), GPE.SGPGX_nATPE.SGPG (SEQ ID NO:2116), ATPE.SGPGX_nGTSE.SATP (SEQ ID NO:2103), ATPE.SGPGX_nTTPE.SGPG (SEQ ID NO:2106), ATPE.SGPGX_nSTPE.SGPG (SEQ ID NO:2107), ATPE.SGPGX_nATPE.SGPG (SEQ ID NO:2102), ATPE.SGPGX_nGTPE.SGPG (SEQ ID NO:2108), ATPE.SGPGX_nGTPE.SGPG (SEQ ID NO:2108), ATPE.SGPGX_nATPE.SGPG (SEQ ID NO:2102), GTPE.SGPGX_nGTPE.SGPG (SEQ ID NO:2109), GTPE.SGPGX_nSTPE.SGPG (SEQ ID NO:2110), GTPE.SGPGX_nTTPE.SGPG (SEQ ID NO:2111), GTPE.SGPGX_nSTPE.SGPG (SEQ ID NO:2110), GTPE.TPGSX_nSGSE.TGTP (SEQ ID NO:2112), GTPE.GSAPX_nEPSE.SATP (SEQ ID NO:2113), ATPE.SGPGX_nGTPE.SGPG (SEQ ID NO:2108), ATPE.SGPGX_nGTPE.SGPG (SEQ ID NO:2108), ATPE.SGPGX_nATPE.SGPG (SEQ ID NO:2102), ATPE.SGPGX_nGTPE.SGPG (SEQ ID NO:2108), TTPE.SGPGX_nTTPE.SGPG (SEQ ID NO:2114), or STPE.SGPGX_nSTPE.SGPG (SEQ ID NO:2115), wherein each “.” is a Glu-C cleavage site and n is any integer from 0 to 30.

[0084] In some embodiments, the fusion protein comprises at least one of the following amino acid sequence ATPE.SGPGX_nATSE.SATP, wherein “.” is a Glu-C cleavage site and n is any integer from 0 to 30 (SEQ ID NO:2104).

[0085] In some embodiments, n is any integer from 1 to 20, 5 to 15, 3 to 7, 2 to 5, or 5 to 10. In some embodiments, n is 9. In some embodiments, n is 4. In some embodiments, n is 3.

[0086] In some embodiments, X_n is PGTGTSAT (SEQ ID NO:2233), PGSGPGT (SEQ ID NO:2221), PGTTPTGT (SEQ ID NO:2241), PGTPPTST (SEQ ID NO:2235), PGTSPSAT (SEQ ID NO:2238), PGTGSAGT (SEQ ID NO:2230), PGTGGAGT (SEQ ID NO:2228), PGTSPGAT (SEQ ID NO:2237), PGTSGSGT (SEQ ID NO:2236), PGTSSAST (SEQ ID NO:2239), PGTGAGTT (SEQ

ID NO:2227), PGTGSTST (SEQ ID NO:2232), GSEPATSG (SEQ ID NO:2224), APGTSTEP (SEQ ID NO:2222), PGTAGSGT (SEQ ID NO:2226), PGTSSGGT (SEQ ID NO:2240), PGTAGPAT (SEQ ID NO:2225), PGTPGTGT (SEQ ID NO:2234), PGTGGPTT (SEQ ID NO:2229), or PGTGSGST (SEQ ID NO:2231).

[0087] In some embodiments, Xn is TSAS (SEQ ID NO:2214), TGTS (SEQ ID NO:2211), SGP, TTPG (SEQ ID NO:2220), TPPT (SEQ ID NO:2213), TSPS (SEQ ID NO:2217), TGSA (SEQ ID NO:2208), TGGA (SEQ ID NO:2206), TSPG (SEQ ID NO:2216), TSGS (SEQ ID NO:2215), TSSA (SEQ ID NO:2218), TGAG (SEQ ID NO:2205), TGST (SEQ ID NO:2210), EPAT (SEQ ID NO:2201), GTST (SEQ ID NO:2202), TAGS (SEQ ID NO:2204), TSSG (SEQ ID NO:2219), TAGP (SEQ ID NO:2203), TPGT (SEQ ID NO:2212), TGGP (SEQ ID NO:2207), or TGSG (SEQ ID NO:2209). In some embodiments, Xn is SGP.

[0088] In some embodiments, the first polypeptide subunit comprises an amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity, to a sequence listed in Table A1 (SEQ ID NOs: 1000-1005).

[0089] In some embodiments, the first polypeptide subunit comprises an amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity, to:

ASGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSEPATSGSETPGT
SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP
ESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT
SESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE
GSPAGSEPATSGSETPGTSESATE**EAGRSASHTPAGLTGPGT**PESATPESRNLPVATPDPMFP
CLHHSQNLLRAVSNMLQKAROTLEFYPTCTSEEIDHEDITKDKTSTVEACLPLELTKNESCLNS
RETSFITNGSCLASRKTSFMMLCLSSIIYEDLKMYQVEFKTMNAKLLMDPKRQIFLDQNMLA
VIDELMQALNFNSETVPQKSSLEEPDFYKTKIKLCILLHAFRIRAVTIDRVMSYLNASGTATPE
SAPE**EAGRSASHTPAGLTGP**ATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGT
SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATP
ESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGT
SESATPESGPGTSESATPESGPGTSASATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPT
STEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGEPEA (SEQ ID NO: 1000).

[0090] In some embodiments, the second polypeptide subunit comprises an amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity, to a sequence listed in Table A2 (SEQ ID NOs: 1033-1036).

[0091] In some embodiments, the second polypeptide subunit comprises an amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100%

identity, to:

ASGPGTSTEPSEGSAPGTSESATPESGPGSGPATSESATPGTSESATPESGPGSEPATSGSETPGT
SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP
ESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT
SESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE
GSAPGSEPATSGSETPGTSESAT**EAGRSASHTPAGLTGPT**SESATPESIWELKKDVYVVELD
WYPDAPGEMVVLTCDTPEEDGITWTLDOSSSEVLGSGKTLTIQVKEFGDAGQYTCHKGGEVL
SHSLLLLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRFTCWWTITISTDLTFSVKSSR
GSSDPQGVTCGAATLSAERVGRDNKEYEYSVECOEDSACPAAEESLPIEVMVDAVHKLKYE
NYTSSFFIRDIIKPDPPKNLQLKPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCVQVQGKSKRE
KKDRVFTDKTSATVICRKNASISVRAQDRYSSSWSEWASVPCSGTATPESGPGE**EAGRSASH**
TPAGLTGPATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGT
STEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATP
ESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGS
PAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSPSATP
ESGPGSEPATSGSETPGSEPATSGSETPGSPAG (SEQ ID NO: 1033).

[0092] In some embodiments, the first polypeptide subunit comprises the amino acid sequence of:

ASGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSEPATSGSETPGT
SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP
ESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT
SESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE
GSAPGSEPATSGSETPGTSESAT**EAGRSASHTPAGLTGP**GPESATPESRNLPVATPDPMFP
CLHHSQNLLRAVSNMLQKAROTLEFYPTCTSEEIDHEDITKDKTSTVEACLPLELTKNESCLNS
RETSFITNGSCLASRKTSFMMALCLSSIEDLKMVQVEFKTMNAKLLMDPKROIFLDONMLA
VIDELMQALNFNSETVPQKSSLEEPDFYKTKIKLCILLHAFRIRAVTIDRVMSYLNASGTATPE
SAPEAGRSASHTPAGLTGP****ATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGT
SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATP
ESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGT
SESATPESGPGTSESATPESGPGTSASATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPT
STEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGEPEA (SEQ ID NO: 1000), and

wherein the second polypeptide subunit comprises the amino acid sequence of:

ASGPGTSTEPSEGSAPGTSESATPESGPGSGPATSESATPGTSESATPESGPGSEPATSGSETPGT
SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP
ESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT
SESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE
GSAPGSEPATSGSETPGTSESAT**EAGRSASHTPAGLTGPT**SESATPESIWELKKDVYVVELD

WYPDAPGEMVVLTCDTPEEDGITWTLDOQSSEVLGSGKTLTIQVKEFGDAGQYTCHKGGEVL
SHSLLLLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRFTCWWLTTISTDLTFSVKSSR
GSSDPQGVTCGAATLSAERVVRGDNKEYEYSVECOEDSACPAAEESLPIEVMVDAVHKLKYE
NYTSSFFIRDIIKPDPPKNLQLKPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCVQVQGKSKRE
KKDRVFTDKTSATVICRKNASISVRAQDRYYSSSWSEWASVPCSGTATPESGPGSEAGRSASH
TPAGLTGPATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGT
STEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATP
ESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGS
PAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSPSATP
ESGPGSEPATSGSETPGSEPATSGSETPGSPAG (SEQ ID NO: 1033).

[0093] In one aspect, the disclosure provides a fusion protein comprising a first polypeptide subunit and a second polypeptide subunit, wherein the first polypeptide subunit and the second polypeptide subunit are disulfide-linked thereby forming a disulfide-linked heterodimer, wherein the first polypeptide subunit comprises an amino acid sequence having 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity, to:

ASGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSEPATSGSETPGT
SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP
ESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT
SESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE
GSAPGSEPATSGSETPGTSESATEAGRSASHTPAGLTGPGTPESATPESRNLPAVTPDPMFPC
LHHSQNLLRAVSNNMLQKARQTLEFYPTSEEIDHEDITKDKTSTVEACLPLELTKNESCLNSR
ETSFITNGSCLASRKTSFMMALCLSSYEDLKMYQVEFKTMNAKLLMDPKRQIFLDQNMLAV
IDELMQALNFNSETVPQKSSLEEDFYKTKIKLCILLHAFRIRAVTIDRVMSYLNASGTATPES
APEAGRSASHTPAGLTGPATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSE
SATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPES
GPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSE
SATPESGPGTSESATPESGPGTSASATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTST
EEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGSEPA (SEQ ID NO: 1000), and wherein
the second polypeptide subunit comprises an amino acid sequence having 85%, 90%, 91%, 92%,
93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity, to:

ASGPGTSTEPSEGSAPGTSESATPESGPGSPATSESATPGTSESATPESGPGSEPATSGSETPGT
SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP
ESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT
SESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE
GSAPGSEPATSGSETPGTSESATE**AGRSASHTPAGLTGPT**SESATPES**IWELKKDVVVVELD**
WYPDAPGEMVVLTCDTPEEDGITWTLDOQSSEVLGSGKTLTIQVKEFGDAGQYTCHKGGEVL

SHSLLLLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRFTCWWLTTISTDLTFSVKSSR
GSSDPQGVTCGAATLSAERVVRGDNKEYEYSVEQCQEDSACPAAEESLPIEVMVDAVHKLKYE
NYTSSFFIRDIKPDPPKNLQKPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCVQVQGKSKRE
KKDRVFTDKTSATVICRKNASISVRAQDRYYSSSWSEWASVPCSGTATPESGPG**EAGRSASH**
TPAGLTGPATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGT
STEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATP
ESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGS
PAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSPSATP
ESGPGSEPATSGSETPGSEPATSGSETPGSPAG (SEQ ID NO: 1033).

[0094] In some embodiments, the first polypeptide subunit comprises the amino acid sequence of:
ASGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSEPATSGSETPGT
SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP
ESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT
SESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE
GSAPGSEPATSGSETPGTSESAT**EAGRSASHTPAGLTG**PTSESATPES**RNL**PVATPD**PGMFP**
CLHHSQNLLRAVSNMLQKARQTLEFYPCTSEEIDHEDITKDKTSTVEACLPLELTKNESCLNS
RETSFITNGSCLASRKTSFMMALCLSSIEDLKMYQVEFKTMNAKLLMDPKRQIFLDQNMLA
VIDELMQALNFNSETVPQKSSLEEDFYKTKIKLCILLHAFRIRAVTIDRVMSYLNASGTATPE
SAPEAGRSASHTPAGLTGPATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGT
SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATP
ESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGT
SESATPESGPGTSESATPESGPGTSASATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPT
STEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGPEA (SEQ ID NO: 1000), and

wherein the second polypeptide subunit comprises the amino acid sequence of:

ASGPGTSTEPSEGSAPGTSESATPESGPGSGPATSESATPGTSESATPESGPGSEPATSGSETPGT
SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP
ESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT
SESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE
GSAPGSEPATSGSETPGTSESAT**EAGRSASHTPAGLTG**PTSESATPES**IWEL**KKDVYVVELD
WYPDAPGEMVVLTCDTPEEDGITWTLDQSSEVLGSGKTLTIQVKEFGDAGQYTCHKGGEVL
SHSLLLLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRFTCWWLTTISTDLTFSVKSSR
GSSDPQGVTCGAATLSAERVVRGDNKEYEYSVEQCQEDSACPAAEESLPIEVMVDAVHKLKYE
NYTSSFFIRDIKPDPPKNLQKPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCVQVQGKSKRE
KKDRVFTDKTSATVICRKNASISVRAQDRYYSSSWSEWASVPCSGTATPESGPG**EAGRSASH**
TPAGLTGPATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGT
STEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATP

ESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGS
PAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSPSATP
ESGPGSEPATSGSETPGSEPATSGSETPGSPAG (SEQ ID NO: 1033).

[0095] In one aspect, the disclosure provides a pharmaceutical composition comprising the fusion disclosed herein and at least one pharmaceutically acceptable excipient.

[0096] In one aspect, the disclosure provides a polynucleotide sequence encoding the fusion protein disclosed herein.

[0097] In one aspect, the disclosure provides an expression vector comprising the polynucleotide sequence described herein.

[0098] In one aspect, the disclosure provides a host cell comprising the expression vector described herein.

[0099] In one aspect, the disclosure provides a method of producing the fusion protein described herein. In some embodiments, the method further comprises isolating the fusion protein from a host cell.

[0100] In one aspect, the disclosure provides a method of treating cancer in a subject in need thereof, the method comprising administering an effective amount of the fusion protein described herein or the pharmaceutical composition described herein to the subject.

[0101] In some embodiments, the cancer comprises a solid tumor.

BRIEF DESCRIPTION OF THE DRAWINGS

[0102] FIG. 1 depicts a schematic of a protease-activatable IL-12 fusion protein.

[0103] FIG. 2 depicts IL-12R activity assay curves of various protease-activatable cytokine fusion proteins based on the ability to bind to and activate IL-12R in HEK-Blue IL-12 reporter cells. "X.AP####" refers to the masked fusion protein and "AP####" refers to the MMP9 unmasked or partially unmasked fusion protein.

[0104] FIG. 3 depicts IL-12R activity assay curves of various protease-activatable cytokine fusion proteins based on the ability to bind to and activate IL-12R in HEK-Blue IL-12 reporter cells. "X.AP####" refers to the masked fusion protein and "AP####" refers to the MMP9 unmasked or partially unmasked fusion protein.

[0105] FIG. 4 depicts IL-12R activity assay curves of various protease-activatable cytokine fusion proteins based on the ability to bind to and activate IL-12R in HEK-Blue IL-12 reporter cells.

“X.AP####” refers to the masked fusion protein and “AP####” refers to the MMP9 unmasked or partially unmasked fusion protein. Recombinant murine IL-12 is used as a reference.

[0106] **FIG. 5** depicts an IL-12R activity assay curve of a quadruple masked protease-activatable cytokine fusion protein, and its unmasked counterpart, based on the ability to bind to and activate IL-12R in HEK-Blue IL-12 reporter cells. Recombinant murine IL-12 is used as a reference.

[0107] **FIG. 6** depicts mouse splenocyte *in vitro* assay data comparing protease-activatable cytokine fusion proteins: P26M (having 4 x 210 amino acid ELNNs) to P24M (having 4 x 288 amino acid ELNNs).

[0108] **FIG. 7** depicts the mean tumor volume of B16F10 tumor model mice treated with P26M (having 4 x 210 amino acid ELNNs) or P24M (having 4 x 288 amino acid ELNNs) at 3 µg QW, as compared to diluent. A scatter plot of mean tumor volume at Day 21 from the same experiment is also provided.

[0109] **FIG. 8** depicts a schematic of a protease-activatable IL-12 fusion protein with barcode fragments (C-35 BC; N-35 BC; C-40 BC, and N-40 BC) incorporated into the ELNNs.

[0110] **FIG. 9A** depicts an alignment of the RSR-2295 and RSR-3213 amino acid sequences and proteases capable of cleaving them. **FIG. 9B** depicts *in vitro* protease digestion of protease activatable T cell engagers (paTCEs) fusion proteins employing RSR-2295 or RSR-3213. The RSR-3213 sequence is modified to substantially reduce cleavage by legumain. paTCE refers to the intact fusion protein having two ELNNs (N-terminal and C-terminal) flanking a bispecific TCE core. Unmasked TCE (uTCE) refers to the unmasked TCE core. 1x-C and 1x-N refer to the C-terminal or N-terminal single-masked metabolites, respectively.

[0111] **FIG. 10A** and **FIG. 10B** depict relative plasma stability of paTCEs employing RSR-2295 or RSR-3213, measured at Day 0 and Day 7. In **FIG. 10A**, RSR-2295 employed the SCy5.5 fluorophore and RSR-3213 employed the Scy7.5 fluorophore. In **FIG. 10B**, the RSR-2295 employed the Scy7.5 fluorophore and RSR-3213 employed the Scy5.5 fluorophore. **FIG. 10C** depicts the observed cleavability *in vivo* from tumor homogenates from 3 different mouse tumor models. For each set of bar graphs (i.e., % 1x-C, % 1x-N, % uTCE), each bar from left to right represents B1, B2, B3, B4, A1, A2, A3, A4, 43-1, 43-2, 43-3, and 43-4. B1-B4 represent 4 different mice from a first tumor model (NCI-N87). A1-A4 represent 4 different mice from a second tumor model (HT-29). 43-1 – 43-4 represent 4 different mice from a third tumor model (HT-55). **FIG. 10D** depicts the % of total for the 3 metabolites plus the paTCE (paTCE, 1x-N, 1x-C, and uTCE) when employing RSR-2295 or RSR-3213.

[0112] **FIG. 11** depicts relative tumor uptake of paTCEs employing RSR-2295 or RSR-3213. The plasma:tumor ratio was calculated in 3 different mouse tumor models (4 mice per tumor model).

There is a “Mouse 1” for each of the 3 different tumor models, a “Mouse 2” for each of the 3 different tumor models, a “Mouse 3” for each of the different tumor models, and a “Mouse 4” for each of the 3 different tumor models.

[0113] FIG. 12 depicts mouse splenocyte *in vitro* assay data comparing protease-activatable cytokine fusion proteins: P24M (having RSR-3213) to P01M (having RSR-2295).

[0114] FIG. 13 depicts the mean tumor volume of B16F10 tumor model mice treated with P24M (having RSR-3213) or P01M (having RSR-2295) at 3 µg QW, as compared to diluent. A scatter plot of mean tumor volume at Day 21 from the same experiment is also provided.

[0115] FIG. 14 depicts a plot of IFN γ induction in a mouse splenocyte assay. Masked and unmasked version of protease-activatable cytokine fusion proteins P71M (mouse surrogate for P72H) and P70M (mouse surrogate from P74H) were tested at various concentrations. Recombinant mouse IL-12 was included as reference.

[0116] FIG. 15 depicts pharmacokinetic data of masked and unmasked versions of protease-activatable cytokine fusion proteins P70M (mouse surrogate for P72H) and P71M (mouse surrogate for P74H) in C57BL/6 mice following IV administration of each fusion protein.

[0117] FIG. 16 depicts efficacy data (as measured by tumor volume) and toxicity data (as measured by survival) in a MC38 colon carcinoma C57BL/6 mouse model. Protease-activatable cytokine fusion proteins P70M (mouse surrogate for P72H) and P71M (mouse surrogate for P74H) in masked versions only were tested.

[0118] FIG. 17 depicts IFN γ induction from PBMCs incubated with increasing concentrations of protease-activatable cytokine fusion proteins P72H or P74H, in masked and unmasked versions.

DETAILED DESCRIPTION

[0119] There is a significant unmet need in cancer therapeutics for an activatable cytokine that is efficacious against cancer. The fusion proteins of the disclosure comprise a cytokine to which one or more (e.g., two, three, or four) unstructured polypeptide masks are attached. In some embodiments, these unstructured polypeptide masks sterically reduce target engagement of the cytokine to said cytokine target receptor (e.g., IL-12 being masked to reduce binding to IL-12 receptor), and also extend protein half-life. In some embodiments, the unstructured polypeptide masks are extended length non-natural polypeptides (ELNNs).

[0120] In some embodiments, the properties of ELNNs also minimize the potential for immunogenicity, as their lack of stable tertiary structures disfavors cytokine binding, and the absence of hydrophobic, aromatic, and positively charged residues that serve as anchor residues for peptide MHC II binding reduces the potential for T cell epitopes.

[0121] In some embodiments, protease cleavage sites at the base of the ELNN or ELNNs enable proteolytic activation of fusion proteins in the tumor microenvironment, unleashing smaller, highly potent cytokines that are capable of binding to said cytokine's target receptor. In some embodiments, in healthy tissues, where protease activity is tightly regulated, fusion proteins remain predominantly inactive, thus expanding the therapeutic index compared to unmasked cytokines.

[0122] In some embodiments, in addition to localized activation, the short half-life of the unmasked cytokines further widens the therapeutic index while providing the potency of T-cell immunity to improve the eradication of solid tumors. In some embodiments, the release sites used in the fusion proteins can be cleaved across a broad array of tumors by proteases that are collectively involved in every cancer hallmark (growth; survival and death; angiogenesis; invasion and metastasis; inflammation; and immune evasion). Thus, cytokine activity of the fusion proteins is localized to tumors by exploiting the enhanced protease activity that is upregulated in all stages of cancer and tumor development but is tightly regulated in healthy tissues.

TERMINOLOGY

[0123] As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

[0124] 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, unless the context clearly dictates otherwise.

[0125] Furthermore, "and/or" where used herein is to be taken as specific disclosure of each of the two specified features or components with or without the other. Thus, the term "and/or" as used in a phrase such as "A and/or B" herein is intended to include "A and B," "A or B," "A" (alone), and "B" (alone). Likewise, the term "and/or" as used in a phrase such as "A, B, and/or C" is intended to encompass each of the following aspects: "A, B, and C"; "A, B, or C"; "A or C"; "A or B"; "B or C"; "A and C"; "A and B"; "B and C"; "A" (alone); "B" (alone); and "C" (alone).

[0126] It is understood that wherever aspects are described herein with the language "comprising," otherwise analogous aspects described in terms of "consisting of" and/or "consisting essentially of" are also provided.

[0127] Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, amino acid sequences are written left to right in amino to carboxy orientation. The headings provided herein are not limitations of the various aspects of the disclosure. Accordingly, the terms defined immediately below are more fully defined by reference to the specification in its entirety.

[0128] The term "about" is used herein to mean approximately, roughly, around, or in the regions of. When the term "about" is used in conjunction with a numerical range, it modifies that range by

extending the boundaries above and below the numerical values set forth. In general, the term “about” can modify a numerical value above and below the stated value by a variance of, *e.g.*, 10 percent, up or down (higher or lower). In some embodiments, the term indicates deviation from the indicated numerical value by $\pm 10\%$, $\pm 5\%$, $\pm 4\%$, $\pm 3\%$, $\pm 2\%$, $\pm 1\%$, $\pm 0.9\%$, $\pm 0.8\%$, $\pm 0.7\%$, $\pm 0.6\%$, $\pm 0.5\%$, $\pm 0.4\%$, $\pm 0.3\%$, $\pm 0.2\%$, $\pm 0.1\%$, $\pm 0.05\%$, or $\pm 0.01\%$. In some embodiments, “about” indicates deviation from the indicated numerical value by $\pm 10\%$. In some embodiments, “about” indicates deviation from the indicated numerical value by $\pm 5\%$. In some embodiments, “about” indicates deviation from the indicated numerical value by $\pm 4\%$. In some embodiments, “about” indicates deviation from the indicated numerical value by $\pm 3\%$. In some embodiments, “about” indicates deviation from the indicated numerical value by $\pm 2\%$. In some embodiments, “about” indicates deviation from the indicated numerical value by $\pm 1\%$. In some embodiments, “about” indicates deviation from the indicated numerical value by $\pm 0.9\%$. In some embodiments, “about” indicates deviation from the indicated numerical value by $\pm 0.8\%$. In some embodiments, “about” indicates deviation from the indicated numerical value by $\pm 0.7\%$. In some embodiments, “about” indicates deviation from the indicated numerical value by $\pm 0.6\%$. In some embodiments, “about” indicates deviation from the indicated numerical value by $\pm 0.5\%$. In some embodiments, “about” indicates deviation from the indicated numerical value by $\pm 0.4\%$. In some embodiments, “about” indicates deviation from the indicated numerical value by $\pm 0.3\%$. In some embodiments, “about” indicates deviation from the indicated numerical value by $\pm 0.1\%$. In some embodiments, “about” indicates deviation from the indicated numerical value by $\pm 0.05\%$. In some embodiments, “about” indicates deviation from the indicated numerical value by $\pm 0.01\%$.

[0129] With respect to naturally occurring compounds, the term “isolated” refers to a compound (*i.e.*, a polypeptide or polynucleotide) that is not in its native state (*e.g.*, free to varying degrees from components that naturally accompany the compound in nature). No particular level of purification is required. For example, an isolated polypeptide can simply be removed from its native or natural environment. Recombinantly produced polypeptides and proteins expressed in host cells are considered isolated for the purpose of the disclosure, as are native or recombinant polypeptides which have been separated, fractionated, or partially or substantially purified by any suitable technique. “Isolate” and “isolated” may also denote a degree of separation from an original source or surrounding, depending on context.

[0130] The term “polypeptide” refers to any polymer of two or more amino acids. Thus, the terms peptide, dipeptide, tripeptide, oligopeptide, protein, amino acid chain, or any other term used to refer to a chain of two or more amino acids, is included within the definition of “polypeptide.” The term “polypeptide” also encompasses an amino acid polymer that has been modified (*e.g.*, by post-translational modification), for example, by disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation, such as conjugation with a labeling

component. Depending on context, the term “polypeptide” may also be used to refer to a protein comprising two or more polymers of two or more amino acids.

[0131] A “host cell” includes an individual cell (e.g., in culture) which that comprises an exogenous polynucleotide. Host cells may include progeny of a single host cell. The progeny may not necessarily be completely identical (in morphology or in genomic or total DNA complement) to the original parent cell due to naturally occurring or genetically engineered variation.

[0132] A “fusion” or “chimeric” polypeptide or protein comprises a first polypeptide portion linked to a second polypeptide portion with which it is not naturally linked in nature. In some embodiments, the portions may normally exist in separate proteins and are brought together in the fusion polypeptide; they may normally exist in the same protein but are placed in a new arrangement in the fusion polypeptide; or the portions may be brought together from different sources. In some embodiments, a fusion or chimeric protein comprises two or more moieties that do not occur in nature (e.g., are created, designed, or otherwise generated by humans, such as binding domains, masks, linkers, barcodes, and other polypeptides provided herein). A chimeric protein may be created, for example, by chemical synthesis, or by recombinant expression (e.g., comprising creating and translating a polynucleotide in which the peptide regions are encoded in the desired relationship).

[0133] “Conjugated”, “linked,” “fused,” and “fusion” may be used interchangeably herein, depending on context. These terms may refer to the covalent joining together of two more chemical (e.g., polypeptide) elements or components, by whatever means including chemical conjugation or recombinant means.

[0134] The term “cytokine” as used herein refers to any of a class of immunoregulatory proteins that are secreted by cells especially of the immune system and are immunomodulators. Cytokine polypeptides that can be used in the fusion proteins disclosed herein include, but are not limited to interleukins, such as IL-1, IL-1 alpha, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12 (e.g., IL-12p35 and IL-12p40), IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, IL-21 and IL-25, transforming growth factors, such as TGF-alpha, and TGF-beta (e.g., TGF-beta1, TGF-beta2, TGF-beta3); interferons, such as interferon-alpha, interferon-beta, interferon-gamma, interferon-kappa and interferon-omega; tumor necrosis factors, such as tumor necrosis factor alpha and lymphotoxin; chemokines (e.g., C-X-C motif chemokine 10 (CXCL10), CCL19, CCL20, CCL21), and granulocyte macrophage-colony stimulating factor (GM-CS), as well as functional fragments thereof that retain receptor agonist activity.

[0135] The term “chemokine” refers to any of a family of small cytokines with the ability to induce directed chemotaxis in nearby responsive cells.

[0136] As used herein, the terms “activatable,” “activate,” “induce,” and “inducible” refer to the ability of a protein, i.e., a cytokine, that is part of a fusion protein, to bind its receptor and effectuate activity upon cleavage of a mask polypeptide from the fusion protein.

[0137] As known in the art, “sequence identity” between two polypeptides is determined by comparing the amino acid sequence of one polypeptide to the sequence of a second polypeptide. Similarly, “sequence identity” between two polynucleotides is determined by comparing the nucleotide sequence of one polynucleotide to the sequence of a second polynucleotide. The terms “% identical”, “% identity” or similar terms are intended to refer, in particular, to the percentage of nucleotides or amino acids (as applicable) which are identical in an optimal alignment between the sequences to be compared. Said percentage may be purely statistical, and the differences between the two sequences may be but are not necessarily randomly distributed over the entire length of the sequences to be compared. Comparisons of two sequences are usually carried out by comparing the sequences, after optimal alignment, with respect to a segment or “window of comparison”, in order to identify local regions of corresponding sequences. For example, the optimal alignment for a comparison may be carried out manually or with the aid of the local homology algorithm by Smith and Waterman, 1981, *Adv. App. Math.* 2, 482, with the aid of the local homology algorithm by Needleman and Wunsch, 1970, *J. Mol. Biol.* 48, 443, with the aid of the similarity search algorithm by Pearson and Lipman, 1988, *Proc. Natl Acad. Sci. USA* 88, 2444, or with the aid of computer programs using the algorithms (GAP, BESTFIT, FASTA, BLAST P, BLAST N and TFASTA in Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Drive, Madison, Wis.). In some embodiments, percent identity of two sequences is determined using the BLASTN or BLASTP algorithm, as available on the United States National Center for Biotechnology Information (NCBI) website (e.g. at blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&BLAST_SPEC=blast2seq&LINK_LOC=align2seq). In some embodiments, the algorithm parameters used for BLASTN algorithm on the NCBI website include: (i) Expect Threshold set to 10; (ii) Word Size set to 28; (iii) Max matches in a query range set to 0; (iv) Match/Mismatch Scores set to 1, -2; (v) Gap Costs set to Linear; and (vi) the filter for low complexity regions being used. In some embodiments, the algorithm parameters used for BLASTP algorithm on the NCBI website include: (i) Expect Threshold set to 10; (ii) Word Size set to 3; (iii) Max matches in a query range set to 0; (iv) Matrix set to BLOSUM62; (v) Gap Costs set to Existence: 11 Extension: 1; and (vi) conditional compositional score matrix adjustment. When discussed herein, whether any particular polypeptide is at least about 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 99%, or 100% identical to another polypeptide can be determined using methods and computer programs/software known in the art such as, but not limited to, the BESTFIT program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711). BESTFIT uses the local

homology algorithm of Smith and Waterman, *Advances in Applied Mathematics* 2:482-489 (1981), to find the best segment of homology between two sequences. When using BESTFIT or any other sequence alignment program to determine whether a particular sequence is, for example, 95% identical to a reference sequence according to the present disclosure, the parameters are set, of course, such that the percentage of identity is calculated over the full-length of the reference polypeptide sequence and that gaps in homology of up to 5% of the total number of amino acids in the reference sequence are allowed.

[0138] As used herein, the terms “mask polypeptide”, “mask”, and “masking moiety” refer to a polypeptide that is capable of reducing the binding of a cytokine to the target cytokine receptor in the context of a fusion protein (such as a chimeric polypeptide) provided herein. Exemplary mask polypeptides include, but are not limited to, the ELNN polypeptides described herein. Additional mask polypeptides include albumin, polypeptides consisting of proline, serine and alanine, coiled-coil domains, albumin binding domains, Fc domains, and binding domains with specificity to conserved regions of an antibody variable domain. Mask polypeptides are described in further detail in Lucchi et al. (*ACS Cent Sci.* 2021 May 26; 7(5): 724–738).

[0139] As used herein, the terms “ELNN polypeptides” and “ELNNs” are synonymous and refer to extended length polypeptides comprising non-naturally occurring, substantially non-repetitive sequences (*e.g.*, polypeptide motifs) that are composed mainly of small hydrophilic amino acids, with the sequence having a low degree or no secondary or tertiary structure under physiologic conditions. ELNN polypeptides include unstructured hydrophilic polypeptides comprising repeating motifs of 6 natural amino acids (G, A, P, E, S, and/or T). In some embodiments, an ELNN polypeptide comprises multiple motifs of 6 natural amino acids (G, A, P, E, S, T), wherein the motifs are the same or comprise a combination of different motifs. In some embodiments, ELNN polypeptides can confer certain desirable pharmacokinetic, physicochemical, and pharmaceutical properties when linked to proteins, including T-cell engagers as disclosed herein. Such desirable properties may include but are not limited to enhanced pharmacokinetic parameters and solubility characteristics, as well as improved therapeutic index. ELNN polypeptides are known in the art, and non-limiting descriptions relating to and examples of ELNN polypeptides known as XTEN[®] polypeptides are available in Schellenberger et al., (2009) *Nat Biotechnol* 27(12):1186-90; Brandl et al., (2020) *Journal of Controlled Release* 327:186-197; and Radon et al., (2021) *Advanced Functional Materials* 31, 2101633 (pages 1-33), the entire contents of each of which are incorporated herein by reference.

[0140] In some embodiments, the repetitiveness of an ELNN sequence refers to the 3-mer repetitiveness and can be measured by computer programs or algorithms or by other means known in the art. In some embodiments, the 3-mer repetitiveness of an ELNN may be assessed by determining the number of occurrences of the overlapping 3-mer sequences within the polypeptide. For example, a polypeptide of 200 amino acid residues has 198 overlapping 3-amino acid sequences (3-mers), but the

number of unique 3-mer sequences will depend on the amount of repetitiveness within the sequence. In some embodiments, the score can be generated (hereinafter “subsequence score”) that is reflective of the degree of repetitiveness of the 3-mers in the overall polypeptide sequence. In this context, “subsequence score” means the sum of occurrences of each unique 3-mer frame across a 200 consecutive amino acid sequence of the polypeptide divided by the absolute number of unique 3-mer subsequences within the 200 amino acid sequence. Examples of such subsequence scores derived from the first 200 amino acids of repetitive and non-repetitive polypeptides are presented in Example 73 of International Patent Application Publication No. WO 2010/091122 A1, which is incorporated by reference in its entirety.

[0141] In some embodiments, and in the context of ELNNs, a “substantially non-repetitive sequence,” refers to an ELNN sequence, wherein (1) there are few or no instances of four identical amino acids in a row in the ELNN sequence and wherein (2) the ELNN has a subsequence score (defined in the preceding paragraph herein) of 12, or 10 or less or that there is not a pattern in the order, from N- to C-terminus, of the sequence motifs that constitute the polypeptide sequence.

[0142] A “vector” is a nucleic acid molecule that transfers an inserted nucleic acid molecule into and/or between host cells. In some embodiments, a vector self-replicates in an appropriate host. 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 used for the transcription of mRNA that is translated into a polypeptide(s). In some embodiments, an “expression system” is a suitable host cell comprising an expression vector that can function to yield a desired expression product.

[0143] The terms “treatment” or “treating,” and “ameliorating” may be used interchangeably herein. These terms refer to an approach for obtaining beneficial or desired results including but not limited to a therapeutic benefit. By “therapeutic benefit” is meant eradication or amelioration of the underlying disorder being treated. In some embodiments, a therapeutic benefit is achieved with the eradication or amelioration of one or more of the physiological symptoms associated with the underlying disease condition such that an improvement is observed in the subject, notwithstanding that the subject may still be afflicted with the underlying disorder. In some embodiments, a therapeutic benefit comprises slowing or halting the growth of one or more tumors. In some embodiments, a therapeutic benefit comprises reducing the size of one or more tumors. In some embodiments, a therapeutic benefit comprises eradicating one or more tumors from a subject. In some embodiments, a therapeutic benefit comprises effecting the death of cancer cells.

[0144] As used herein, the term “therapeutically effective amount” refers to an amount of a biologically active agent (such as a fusion protein provided herein, e.g., as part of a pharmaceutical composition), that is capable of having any detectable, beneficial effect on any symptom, aspect, measured parameter or characteristics of a disease state or condition when administered in one or repeated doses to a subject. Such effect need not be absolute to be beneficial. The disease condition can refer to a disorder or a disease, e.g., cancer or a symptom of cancer.

CYTOKINES, CLEAVAGE SEQUENCES, BARCODE FRAGMENTS, AND FUSION POLYPEPTIDES

[0145] The present disclosure provides, *inter alia*, new and useful fusion proteins comprising (i) one or more mask polypeptides (such as ELNNs), (ii) a cytokine (e.g., IL-12) linked to the mask polypeptide(s), and (iii) one or more protease-cleavable release segments (RS), wherein an RS is positioned between the mask polypeptide(s) and the cytokine.

[0146] In some embodiments, the fusion polypeptide comprises a first ELNN (such as an ELNN described herein). In some embodiments, the polypeptide further comprises a second ELNN (such as an ELNN described herein). In some embodiments, the polypeptide comprises an ELNN at or near its N-terminus (an “N-terminal ELNN”). In some embodiments, the polypeptide comprises an ELNN at or near its C-terminus (a “C-terminal ELNN”). In some embodiments, the polypeptide comprises both an N-terminal ELNN and a C-terminal ELNN.

[0147] In some embodiments, a fusion polypeptide comprises a cytokine and a first ELNN is attached to the N-terminus of the cytokine by a first RS and a second ELNN is attached to the C-terminus of the cytokine by a second RS. In some embodiments, each RS is cleavable by a protease mentioned herein. In some embodiments, each RS comprises an RS sequence disclosed herein. In some embodiments, the fusion polypeptide is a cytokine fusion.

[0148] Included herein are polypeptide sequences that may be used, e.g., to link one polypeptide moiety to another within a fusion protein. For example, useful linkers are provided that are cleaved by multiple proteases but not legumain. In some embodiments, such linkers may be used outside the context of cytokines such as those described herein.

[0149] In some embodiments, a fusion can comprise one or more barcode fragments (e.g., as described herein) releasable (e.g., configured to be released) the fusion polypeptide upon cleavage or digestion of the fusion polypeptide by a protease. In some embodiments, the protease is a non-mammalian protease. In some embodiments, each barcode fragment differs in sequence and molecular weight from all other peptide fragments (including all other barcode fragments if present) that are releasable from the polypeptide upon complete digestion of the polypeptide by the protease, thereby making it unique and making its presence detectable through techniques such as mass spectrometry.

CYTOKINES

[0150] In general, the therapeutic use of cytokines is strongly limited by their systemic toxicity. TNF, for example, was originally discovered for its capacity of inducing the hemorrhagic necrosis of some tumors, and for its in vitro cytotoxic effect on different tumoral lines, but it subsequently proved to have strong pro-inflammatory activity, which can, in case of overproduction conditions, dangerously affect the human body. As the systemic toxicity is a fundamental problem with the use of pharmacologically active amounts of cytokines in humans, novel derivatives and therapeutic strategies are now under evaluation, aimed at reducing the toxic effects of this class of biological effectors while keeping their therapeutic efficacy.

[0151] One cytokine for use in production of the fusion proteins of the disclosure is Interleukin-12 (IL-12). IL-12 is a disulfide-linked heterodimer of two separately encoded subunits (p35 and p40), which are linked covalently to give rise to the so-called bioactive heterodimeric (p70) molecule. Apart from forming heterodimers (IL-12 and IL-23), the p40 subunit is also secreted as a monomer (p40) and a homodimer (p40₂). It is known in the art that synthesis of the heterodimer as a single chain with a linker connecting the p35 to the p40 subunit preserves the full biological activity of the heterodimer. IL-12 plays a critical role in the early inflammatory response to infection and in the generation of Th1 cells, which favor cell-mediated immunity. It has been found that overproduction of IL-12 can be dangerous to the host because it is involved in the pathogenesis of a number of autoimmune inflammatory diseases (e.g., MS, arthritis, type 1 diabetes).

[0152] IL-12 is a pleiotropic cytokine, that creates an interconnection between the innate and adaptive immunity. IL-12 was first described as a factor secreted from PMA-induced EBV-transformed B-cell lines. Based on its actions, IL-12 has been designated as cytotoxic lymphocyte maturation factor and natural killer cell stimulatory factor. Due to bridging the innate and adaptive immunity and potently stimulating the production of IFN γ , a cytokine coordinating natural mechanisms of anticancer defense, IL-12 seemed ideal candidate for tumor immunotherapy in humans. However, severe side effects associated with systemic administration of IL-12 in clinical investigations and the very narrow therapeutic index of this cytokine markedly hampered the use of this cytokine in cancer patients. Approaches to IL-12 therapy in which delivery of the cytokine is tumor-targeted, which may diminish some of the previous issues with IL-12 therapy, are currently in clinical trials for cancers.

[0153] The IL-12 receptor (IL-12R) is a heterodimeric complex consisting of IL-12R β 1 and IL-12R β 2 chains expressed on the surface of activated T-cells and natural killer cells. The IL-12R β 1 chain binds to the IL-12p40 subunit, whereas IL-12p35 in association with IL-12R β 2 confers an intracellular signaling ability. Signal transduction through IL-12R induces phosphorylation of Janus kinase (Jak2) and tyrosine kinase (Tyk2), that phosphorylate and activate signal transducer and

activator of transcription (STAT)1, STAT3, STAT4, and STATS. The specific cellular effects of IL-12 are due mainly to activation of STAT4. IL-12 induces natural killer and T-cells to produce cytokines, in particular interferon (IFN) γ , that mediate many of the proinflammatory activities of IL-12, including CD4+ T-cell differentiation toward the Th1 phenotype.

[0154] Disclosed herein are compositions and methods comprising conditionally active IL-12 and other cytokines designed to address the risks associated with conventional cytokine therapy and provide much needed immunomodulatory therapeutics.

[0155] The amino acid sequence for exemplary IL-12 p35 and p40 subunits are provided in Table 1a.

Table 1a. IL-12 amino acid sequences

Subunit	Sequence
mIL12 p35 WT	MWELEKDVYVVEVDWTPDAPGETVNLTCDTPEEDDITWTSQQRHGVIGSGKTLTITVKEFL DAGQYTCHKGGETLSHSHLLHKKENGIWSTEILKNFNKNTFLKCEAPNYSGRFTCSWLTVQ RNMDLKFNKSSSSSPDSRAVTCGMASLSAEKVTLQDRDYEKYSVSCQEDVTCPTAEETLP IELALEARQQNKYENYSTSFFIRDIIKPDPPKNLQMKPLKNSQVEVSWEYPDSWSTPHSYF SLKFFVRIQRKKEKMKETEEGCNQKGAFLEKTSSTEVQCKGGNVCVQAQDRYYNSSCSKWA CVPCRVRS (SEQ ID NO: 8056)
mIL12 p40 WT	RVIPVSGFARCLSQSRNLLKTTDDMVKTAREKLKHYSCTAEDIDHEDITRDQTSTLKTCLP LELHKNESCLATRETSSTTRGSLPPQKTSMMTLCLGSIYEDLKMYQTEFQAINAALQNH NHQQIILDKGMLVAIDELMQSLNHNGETLRQKPPVGEADPYRVKMKLCILLHAFSTRVVTI NRVMGYLSSA (SEQ ID NO: 8057)
hIL-12 p35 WT	RNLPVATPDPGMFPC LHHSQNLLRAVSNMLQKARQTFEYFCTSEEIDHEDITKDKTSTVE ACLPLELTKNESCLNSRETSFITNGSCLASRKTSFMMALCLSSIYEDLKMYQVEFKTMNAK LLMDPKRQIFLDQNMLAVIDELMQALNFNSETVPQKSSLEEDFYKTKIKLCILLHAFRIR AVTIDRVMSYLSA (SEQ ID NO: 8050)
hIL-12 p40 WT	IWELKKDVYVVELDWYDPAGEMVVLTCDTPEEDGITWTLQSSSEVLGSGKTLTIQVKEFG DAGQYTCHKGGEVLSHSLLLLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRFTCWW LTTISTDLTFSVKSSRGSSDPQGVTCGAATLSAERVGRDNKEYEYSVEQEDSACPAEES LPIEVMDAVHKLKYENYTSSFFIRDIIKPDPPKNLQKPLKNSRQVEVSWEYPDTWSTPH SYFSLTFCVQVQGKSKREKKDRVFTDKTSATVICRKNASISVRAQDRYYSSSWSEWASVPC S (SEQ ID NO: 8051)
hIL-12 p35 SSA variant	RNLPVATPDPGMFPC LHHSQNLLRAVSNMLQKARQTFEYFCTSEEIDHEDITKDKTSTVE ACLPLELTKNESCLNSRETSFITNGSCLASRKTSFMMALCLSSIYEDLKMYQVEFKTMNAK LLMDPKRQIFLDQNMLAVIDELMQALNFNSETVPQKSSLEEDFYKTKIKLCILLHAFRIR AVTIDRVMSYLSA (SEQ ID NO: 8052)
hIL-12 p35 QAS variant (N195Q)	RNLPVATPDPGMFPC LHHSQNLLRAVSNMLQKARQTFEYFCTSEEIDHEDITKDKTSTVE ACLPLELTKNESCLNSRETSFITNGSCLASRKTSFMMALCLSSIYEDLKMYQVEFKTMNAK LLMDPKRQIFLDQNMLAVIDELMQALNFNSETVPQKSSLEEDFYKTKIKLCILLHAFRIR AVTIDRVMSYLSA (SEQ ID NO: 8053)
hIL-12 p35 DAS variant (N195D)	RNLPVATPDPGMFPC LHHSQNLLRAVSNMLQKARQTFEYFCTSEEIDHEDITKDKTSTVE ACLPLELTKNESCLNSRETSFITNGSCLASRKTSFMMALCLSSIYEDLKMYQVEFKTMNAK LLMDPKRQIFLDQNMLAVIDELMQALNFNSETVPQKSSLEEDFYKTKIKLCILLHAFRIR AVTIDRVMSYLDAS (SEQ ID NO: 8054)
hIL-12 p35 XAS variant (N195X)	RNLPVATPDPGMFPC LHHSQNLLRAVSNMLQKARQTFEYFCTSEEIDHEDITKDKTSTVE ACLPLELTKNESCLNSRETSFITNGSCLASRKTSFMMALCLSSIYEDLKMYQVEFKTMNAK LLMDPKRQIFLDQNMLAVIDELMQALNFNSETVPQKSSLEEDFYKTKIKLCILLHAFRIR

	AVTIDRVMSYLXAS, wherein X is any amino acid residue other than N (SEQ ID NO: 8055)
--	--

[0156] In some embodiments, the fusion protein disclosed herein comprises a first polypeptide subunit comprising a wild-type p35 amino acid sequence, such as SEQ ID NO: 8050 and a second polypeptide subunit comprising a wild-type p40 amino acid sequence, such as SEQ ID NO: 8051.

[0157] In some embodiments, the fusion protein disclosed herein comprises a first polypeptide subunit comprising a p35 amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity to SEQ ID NO: 8050, and a second polypeptide subunit comprising a p40 amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity to SEQ ID NO: 8051.

[0158] In some embodiments, the fusion protein disclosed herein comprises a first polypeptide subunit comprising a wild-type p35 amino acid sequence, such as SEQ ID NO: 8056 and a second polypeptide subunit comprising a wild-type p40 amino acid sequence, such as SEQ ID NO: 8057.

[0159] In some embodiments, the fusion protein disclosed herein comprises a first polypeptide subunit comprising a p35 amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity to SEQ ID NO: 8056, and a second polypeptide subunit comprising a p40 amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity to SEQ ID NO: 8057.

[0160] In some embodiments, the fusion protein disclosed herein comprises a first polypeptide subunit comprising an aglycosylation mutant variant p35 amino acid sequence. The aglycosylation mutant variant p35 amino acid sequence comprises at least one amino acid substitution of the NAS amino acid sequence located at the C-terminus of the wild-type p35 amino acid sequence. In some embodiments, the fusion protein disclosed herein comprises a first polypeptide subunit comprising an aglycosylation mutant variant p35 amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity to any one of SEQ ID NOs: 8052-8055. In some embodiments, the fusion protein disclosed herein comprises a first polypeptide subunit comprising an aglycosylation mutant variant p35 amino acid sequence of any one of SEQ ID NOs: 8052-8055, and a second polypeptide subunit comprising a wild-type p40 amino acid sequence, such as SEQ ID NO: 8051.

[0161] In some embodiments, the cytokine polypeptide in the fusion protein is Interleukin-2 (IL-2). IL-2 exerts both stimulatory and regulatory functions in the immune system and is, along with other members of the common γ -chain cytokine family, central to immune homeostasis. IL-2 mediates its action by binding to IL-2 receptors (IL-2R), consisting of either trimeric receptors made of IL-2R α (CD25), IL-2R β (CD122), and IL-2R- γ (γ -c, CD132) chains or dimeric β γ IL-2Rs.

[0162] In some embodiments, the cytokine polypeptide in the fusion protein is Interleukin-15 (IL-15). IL-15, a member of the 4-alpha-helix bundle family of cytokines, is an immunomodulator for the treatment of cancer. IL-15 mediates its action by binding to IL-15R α , which is expressed on antigen-presenting dendritic cells, monocytes and macrophages.

[0163] In some embodiments, the cytokine polypeptide in the fusion protein is Interleukin-7 (IL-7). IL-7, also of the IL-2/IL-15 family, is expressed by stromal cells, epithelial cells, endothelial cells, fibroblasts, smooth muscle cells and keratinocytes, and following activation, by dendritic cells (Alpdogan et al., 2005).

[0164] While the present application is exemplified using IL-12 as the exemplary cytokine, those of skill in the art will understand that the teachings provided herein may readily be adapted for and describe and enable the use of fusion proteins formed from other cytokines, fragments and muteins, such as IL-2, IL-7, IL-12, IL-15, IL-18, IL-21 IL-23, IFN-alpha, IFN-beta, IFN-gamma, TNF-alpha, lymphotoxin, TGF-beta1, TGFbeta2, TGFbeta3, GM-CSF, CXCL10, CCL19, CCL20, CCL21 and functional fragments or muteins of any of the foregoing. A description of cytokines that may be incorporated into the fusion proteins disclosed herein is provided in WO2021262985A1, incorporated herein by reference.

[0165] Various elements ensure the delivery and activity of the cytokine in the fusion proteins of the invention preferentially at the site of desired cytokine activity and to severely limit systemic exposure to the cytokine via linkage to an ELNN described herein, which allows serum half-life extension for the cytokine of interest. In this serum half-life extension strategy, the fusion protein may circulate for extended times (preferentially 1-2 or more weeks) but the activated version from which the ELNN sequence has been cleaved has the typical serum half-life of the cytokine.

[0166] By comparison to a fusion protein, the serum half-life of the underlying cytokine administered intravenously is only about 10 minutes due to distribution into the total body extracellular space. Subsequently, the cytokine is metabolized by the kidneys with a half-life of 2.5 hours.

[0167] In some embodiments of this invention, the fusion protein comprises a release segment which is cleaved at the site of action (e.g., by inflammation-specific or tumor-specific proteases) thereby releasing the cytokine's full activity at the desired site and also separating it from the half-life extension of the uncleaved version. In such embodiments, the fully active and free cytokine would have very different pharmacokinetic (pK) properties--a half-life of hours instead of weeks. In addition, exposure to active cytokine is limited to the site of desired cytokine activity (e.g., an inflammatory site or the tumor microenvironment) and systemic exposure to active cytokine, and associated toxicity and side effects, are reduced.

[0168] Creating a fusion protein from cytokines is an elegant mechanism by which to improve the use of cytokines, as immunostimulatory agents, for example for treating cancer. For example, in this aspect, the pharmacokinetics and/or pharmacodynamics of the cytokine (e.g., IL-2, IL-7, IL-12, IL-15, IL-18, IL-21, IL-23, IFN-alpha, IFN-beta and IFN-gamma, TNF-alpha, lymphotoxin, TGF-beta1, TGF-beta2, TGF-beta3, GM-CSF, CXCL10, CCL19, CCL20, and CCL21) can be tailored to maximally activate effector cells (e.g., effect T cells, NK cells) and/or cytotoxic immune response promoting cells (e.g., induce dendritic cell maturation) at a site of desired activity, such as in a tumor or tumor microenvironment, but preferably not systemically.

[0169] Thus, provided herein are pharmaceutical compositions comprising fusion proteins that are comprised of at least one cytokine polypeptide, such as interleukins (e.g., IL-2, IL-7, IL-12, IL-15, IL-18, IL-21, IL-23), interferons (IFNs, including IFN-alpha, IFN-beta and IFN-gamma), tumor necrosis factors (e.g., TNF-alpha, lymphotoxin), transforming growth factors (e.g., TGF-beta1, TGF-beta2, TGF-beta3), chemokines (e.g., CXCL10, CCL19, CCL20, CCL21) and granulocyte macrophage-colony stimulating factor (GM-CS) or a functional fragment or mutein of any of the foregoing.

[0170] In some embodiments, the cytokine polypeptides (including functional fragments) that are included in the fusion proteins disclosed herein are not mutated or engineered to alter the properties of the naturally occurring cytokine, including receptor binding affinity and specificity or serum half-life.

EXTENDED RECOMBINANT POLYPEPTIDES (ELNNs)

Chain Length and Amino Acid Composition

[0171] In some embodiments, an ELNN comprises at least 100, or at least 150 amino acids. In some embodiments, an ELNN is from 100 to 3,000, or from 150 to 3,000 amino acids in length. In some embodiments, an ELNN is from 100 to 1,000, or from 150 to 1,000 amino acids in length. In some embodiments, an ELNN is at least (about) 100, at least (about) 150, at least (about) 200, at least (about) 250, at least (about) 300, at least (about) 350, at least (about) 400, at least (about) 450, at least (about) 500, at least (about) 550, at least (about) 600, at least (about) 650, at least (about) 700, at least (about) 750, at least (about) 800, at least (about) 850, at least (about) 900, at least (about) 950, at least (about) 1,000, at least (about) 1,100, at least (about) 1,200, at least (about) 1,300, at least (about) 1,400, at least (about) 1,500, at least (about) 1,600, at least (about) 1,700, at least (about) 1,800, at least (about) 1,900, or at least (about) 2,000 amino acids in length. In some embodiments, an ELNN is at most (about) 100, at most (about) 150, at most (about) 200, at most (about) 250, at most (about) 300, at most (about) 350, at most (about) 400, at most (about) 450, at most (about) 500, at most (about) 550, at most (about) 600, at most (about) 650, at most (about) 700, at most (about) 750, at most (about) 800, at most (about) 850, at most (about) 900, at most (about) 950, at most (about) 1,000, at most (about) 1,100, at most (about) 1,200, at most (about) 1,300, at most (about) 1,400, at most (about) 1,500, at most (about) 1,600, at most (about) 1,700, at most (about) 1,800, at most (about) 1,900, or at most (about) 2,000 amino acids in length. In some embodiments, an ELNN has

(about) 100, (about) 150, (about) 200, (about) 250, (about) 300, (about) 350, (about) 400, (about) 450, (about) 500, (about) 550, (about) 600, (about) 650, (about) 700, (about) 750, (about) 800, (about) 850, (about) 900, (about) 950, (about) 1,000, (about) 1,100, (about) 1,200, (about) 1,300, (about) 1,400, (about) 1,500, (about) 1,600, (about) 1,700, (about) 1,800, (about) 1,900, or (about) 2,000 amino acids in length, or of a range between any two of the foregoing. In some embodiments, at least 90% of the amino acid residues of the ELNN are glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) or proline (P). In some embodiments, at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% of the amino acid residues of the ELNN are glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) or proline (P). In some embodiments, an ELNN comprises at least 3 different types of amino acids selected from the group consisting of G, A, S, T, E, and P. In some embodiments, an ELNN comprises at least 4 different types of amino acids selected from the group consisting of G, A, S, T, E, and P. In some embodiments, an ELNN comprises at least 5 different types of amino acids selected from the group consisting of G, A, S, T, E, and P. In some embodiments, an ELNN consists of amino acids selected from the group consisting of G, A, S, T, E, and P. In some embodiments, an ELNN comprises G, A, S, T, E, or P amino acids. In some embodiments, an ELNN (e.g., ELNN1, ELNN2, etc.) is characterized in that: (i) it comprises at least 100, or at least 150 amino acids; (ii) at least 90% of the amino acid residues of the ELNN are glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) or proline (P); and (iii) it comprises at least 4 different types of the amino acids from G, A, S, T, E, or P. As used herein, the term “glutamate” is a synonym for “glutamic acid,” and refers to the glutamic acid residue whether or not the side-chain carboxyl is deprotonated. In some embodiments, the ELNN-containing fusion polypeptide comprises a first ELNN and a second ELNN. In some embodiments, the sum of the total number of amino acids in the first ELNN and the total number of amino acids in the second ELNN is at least 300, at least 350, at least 400, at least 500, at least 600, at least 700, or at least 800 amino acids.

Non-Overlapping Sequence Motif

[0172] In some embodiments, the ELNN comprises, or is formed from, a plurality of non-overlapping sequence motifs. In some embodiments, at least one of the non-overlapping sequence motifs is recurring (or repeated at least two times in the ELNN). In some embodiments, the ELNN comprises at least one other non-overlapping sequence motif that is non-recurring (or found only once within the ELNN). In some embodiments, the plurality of non-overlapping sequence motifs comprises (a) a set of (recurring) non-overlapping sequence motifs, wherein each non-overlapping sequence motif of the set of non-overlapping sequence motifs is repeated at least two times in the ELNN; and (b) a non-overlapping (non-recurring) sequence motif that occurs (or is found) only once within the ELNN. In some embodiments, each non-overlapping sequence motif is from 9 to 14 (or 10 to 14, or 11 to 13) amino acids in length. In some embodiments, each non-overlapping sequence motif is 12 amino acids in length. In some embodiments, the plurality of non-overlapping sequence motifs

comprises a set of non-overlapping (recurring) sequence motifs, wherein each non-overlapping sequence motif of the set of non-overlapping sequence motifs is (1) repeated at least two times in the ELNN; and (2) is between 9 and 14 amino acids in length. In some embodiments, the set of (recurring) non-overlapping sequence motifs comprises 12-mer sequence motifs identified herein by SEQ ID NOs: 179-200 and 1715-1722 in Table 1b. In some embodiments, the set of (recurring) non-overlapping sequence motifs comprise 12-mer sequence motifs identified herein by SEQ ID NOs: 186-189 in Table 1b. In some embodiments, the set of (recurring) non-overlapping sequence motifs comprise at least two, at least three, or all four of 12-mer sequence motifs of SEQ ID NOs: 186-189 in Table 1b. In some embodiments, an ELNN further comprises a sequence other than a 12-mer sequence motif shown in Table 1. In some embodiments, an ELNN comprises a sequence that is not in Table 1b such as ASSATPESGP (SEQ ID NO:203), GSGPGTSESATP (SEQ ID NO:201), or GTSESATP (SEQ ID NO:204). In some embodiments, an ELNN comprises a sequence that is not in Table 1b such as ATPESGP (SEQ ID NO:205), GTSPSATPESGP (SEQ ID NO:202), or GTSESAGEPEA (SEQ ID NO:206). In some embodiments, an ELNN comprises a barcode sequence.

Table 1b. Exemplary 12-Mer Sequence Motifs for Construction of ELNNs

Motif Family*	Amino Acid Sequence	SEQ ID NO:
AD	GESP GGSSGSES	182
AD	GSEGSSGPGESS	183
AD	GSSESGSSEGGP	184
AD	GSGGEPSESGSS	185
AE, AM	GSPAGSPTSTEE	186
AE, AM, AQ	GSEPATSGSETP	187
AE, AM, AQ	GTSESATPESGP	188
AE, AM, AQ	GTSTEPSEGSAP	189
AF, AM	GSTSESPSGTAP	190
AF, AM	GTSTPESGSASP	191
AF, AM	GTSPSGESSTAP	192
AF, AM	GSTSSTAESP GP	193
AG, AM	GTPGSGTASSSP	194
AG, AM	GSSTPSGATGSP	195
AG, AM	GSSPSASTGTGP	196
AG, AM	GASPGTSSTGSP	197
AQ	GEPAGSPTSTSE	198
AQ	GTGEPSTPASE	199
AQ	GSGPSTESAPTE	200
AQ	GSETPSGPSETA	179
AQ	GPSETSTSEPGA	180
AQ	GSPSEPTEGTSA	181
BC	GSGASEPTSTEP	1715
BC	GSEPATSGTEPS	1716
BC	GTSEPSTSEPGA	1717
BC	GTSTEPSEPGSA	1718

Motif Family*	Amino Acid Sequence	SEQ ID NO:
BD	GSTAGSETSTEA	1719
BD	GSETATSGSETA	1720
BD	GTSESATSESGA	1721
BD	GTSTEASEGSAS	1722

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

Unstructured Polypeptide Confirmation

[0173] In various embodiments, an ELNN component (or the ELNN components) of a fusion protein has an unstructured conformation under physiological conditions, regardless of the length (e.g., extended length) of the polymer. For example, the ELNN is characterized by a large conformational freedom of the peptide backbone. In some embodiments, the ELNN is characterized by a lack of long-range interactions as determined by NMR. In some embodiments, the present disclosure provides ELNNs that, under physiologic conditions, resemble the structure of denatured sequences largely devoid in secondary structure. In some embodiments, the ELNNs can be substantially devoid of secondary structure under physiologic conditions. “Largely devoid,” as used in this context, means that less than 50% of the ELNN amino acid residues of the ELNN 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 at least about 99% of the ELNN amino acid residues of the ELNN sequence do not contribute to secondary structure, as measured or determined by the means described herein.

[0174] A variety of methods have been established in the art to discern the presence or absence of secondary and tertiary structures in a given polypeptide. In some embodiments, ELNN 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. 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 (“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 (the entire contents of which are incorporated herein by reference). 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).

[0175] In some embodiments, the ELNNs used in a fusion protein composition can have an alpha-helix percentage ranging from 0% to less than about 5% as determined by a Chou-Fasman algorithm.

In some embodiments, the ELNNs of the fusion protein compositions can have a beta-sheet percentage ranging from 0% to less than about 5% as determined by a Chou-Fasman algorithm. In some embodiments, the ELNNs of the fusion protein compositions can 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 a Chou-Fasman algorithm. In some embodiments, the ELNNs of the fusion protein compositions will have an alpha-helix percentage less than about 2% and a beta-sheet percentage less than about 2%. In some embodiments, the ELNNs of the fusion protein compositions can have a high degree of random coil percentage, as determined by a GOR algorithm. In some embodiments, an ELNN can have 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 a GOR algorithm.

Net Charge

[0176] In some embodiments, the ELNN polypeptides can have an unstructured characteristic imparted by incorporation of amino acid residues with a net charge and/or reducing the proportion of hydrophobic amino acids in the ELNN sequence. The overall net charge and net charge density may be controlled, e.g., by modifying the content of charged amino acids in the ELNNs. In some embodiments, the net charge density of the ELNN of the compositions may be above +0.1 or below -0.1 charges/residue. In some embodiments, the net charge of a ELNN 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.

[0177] Since most tissues and surfaces in a human or animal have a net negative charge, the ELNNs can optionally be designed to have a net negative charge to minimize non-specific interactions between the ELNN containing compositions and various surfaces such as blood vessels, healthy tissues, or various receptors. Not to be bound by a particular theory, an ELNN may adopt open conformations due to electrostatic repulsion between individual amino acids of the ELNN polypeptide that individually carry a high net negative charge and that are distributed across the sequence of the ELNN polypeptide. Such a distribution of net negative charge in the extended sequence lengths of ELNN can lead to an unstructured conformation that, in turn, can result in an effective increase in hydrodynamic radius. Accordingly, in some embodiments the ELNNs contain glutamic acid such that the glutamic acid is at about 8, 10, 15, 20, 25, or even about 30% of the amino acids in the sequences. The ELNN of the compositions of the present disclosure generally have no or a low content of positively charged amino acids. In some embodiments the ELNN may have less than about 10% amino acid residues with a positive charge, or less than about 7%, or less than about 5%, or less than

about 2% amino acid residues with a positive charge. However, the present disclosure contemplates polypeptides where a limited number of amino acids with a positive charge, such as lysine, may be incorporated into an ELNN, e.g., to permit conjugation between the epsilon amine of the lysine and a reactive group on a peptide, a linker bridge, or a reactive group on a drug or small molecule to be conjugated to the ELNN backbone.

[0178] In some embodiments, an ELNN may comprise charged residues separated by other residues such as serine or glycine, which may lead to better expression or purification behavior. Based on the net charge, ELNNs of the subject compositions may 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 some embodiments, the ELNN will have an isoelectric point between 1.5 and 4.5. In some embodiments, an ELNN incorporated into a fusion protein carries a net negative charge under physiologic conditions contributes to the unstructured conformation and reduced binding of the ELNN component to mammalian proteins and tissues.

[0179] As hydrophobic amino acids can impart structure to a polypeptide, in some embodiments the content of hydrophobic amino acids in the ELNN is less than 5%, or less than 2%, or less than 1% hydrophobic amino acid content. In some embodiments, an ELNN has no hydrophobic amino acids. In some embodiments, the amino acid content of methionine and tryptophan in the ELNN component of a fusion protein is less than 5%, or less than 2%, and most preferably less than 1%. In some embodiments, the ELNN has a sequence that has less than 10% amino acid residues with a positive charge, or less than about 7%, or less than 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 10% of the total ELNN sequence. In some embodiments, the ELNN has no methionine or tryptophan residues.

Increased Hydrodynamic Radius

[0180] In some embodiments, the ELNN can have a high hydrodynamic radius, conferring a corresponding increased Apparent Molecular Weight to the fusion protein which incorporates the ELNN. The linking of ELNNs to cytokine sequences can result in the fusion protein compositions that can have increased hydrodynamic radii, increased Apparent Molecular Weight, and increased Apparent Molecular Weight Factor compared to cytokines not linked to an ELNN. For example, in some therapeutic applications in which prolonged half-life is desired, one or more ELNNs with a high hydrodynamic radius are incorporated into a fusion protein comprising a cytokine to effectively enlarge the hydrodynamic radius of the fusion protein 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. In some embodiments, the hydrodynamic radius of a protein is determined by its molecular weight as well as by its structure, including shape and compactness. Not to be bound by a particular theory, the

ELNN may adopt open conformations due to electrostatic repulsion between individual charges of the peptide or the inherent flexibility imparted by the particular amino acids in the sequence that lack potential to confer secondary structure. In some embodiments, the open, extended and unstructured conformation of the ELNN polypeptide has a greater proportional hydrodynamic radius compared to polypeptides of a comparable sequence length and/or molecular weight that have secondary and/or tertiary structure, such as typical globular proteins. Methods for determining the hydrodynamic radius are well known in the art, such as by the use of size exclusion chromatography (SEC), as described in U.S. Patent Nos. 6,406,632 and 7,294,513. In some embodiments, the addition of increasing lengths of ELNN results in proportional increases in the parameters of hydrodynamic radius, Apparent Molecular Weight, and Apparent Molecular Weight Factor, permitting the tailoring of the fusion to desired characteristic cut-off Apparent Molecular Weights or hydrodynamic radii. Accordingly, in some embodiments, the fusion protein can be configured with an ELNN such that the fusion protein can have a hydrodynamic radius of at least about 5 nm, or at least about 8 nm, or at least about 10 nm, or 12 nm, or at least about 15 nm. In some embodiments, the large hydrodynamic radius conferred by the ELNN in a fusion protein can lead to reduced renal clearance of the resulting fusion protein, leading to a corresponding increase in terminal half-life, an increase in mean residence time, and/or a decrease in renal clearance rate.

[0181] In some embodiments, an ELNN (or multiple ELNNs, such as two ELNNs) of a chosen length and sequence can be selectively incorporated into a cytokine to create a fusion protein that will have, under physiologic conditions, an Apparent Molecular Weight of at least about 150 kDa, or at least about 300 kDa, or at least about 400 kDa, or at least about 500 kDa, or at least about 600 kDa, or at least about 700 kDa, or at least about 800 kDa, or at least about 900 kDa, or at least about 1000 kDa, or at least about 1200 kDa, or at least about 1500 kDa, or at least about 1800 kDa, or at least about 2000 kDa, or at least about 2300 kDa or more. In some embodiments, an ELNN (or multiple ELNNs, such as two ELNNs) of a chosen length and sequence can be selectively linked to a cytokine to result in a fusion protein that has, under physiologic conditions, an Apparent Molecular Weight Factor of at least 3, alternatively of at least 4, alternatively of at least 5, alternatively of at least 6, alternatively of at least 7, alternatively of at least 8, alternatively of at least 9, alternatively of at least 10, alternatively of at least 15, or an Apparent Molecular Weight Factor of at least 20 or greater. In some embodiments, the fusion protein has, under physiologic conditions, an Apparent Molecular Weight Factor that is about 4 to about 20, or is about 6 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 fusion protein. In some embodiments, the fusion polypeptide exhibits an apparent molecular weight factor under physiological conditions that is greater than about 6.

Increased Terminal Half-Life

[0182] In some embodiments, a fusion polypeptide comprising an ELNN has a terminal half-life that is at least two-fold longer, or at least three-fold longer, or at least four-fold longer, or at least five-fold longer, compared to a corresponding biologically active polypeptide that is not linked to the ELNN. In some embodiments, the (fusion) polypeptide has a terminal half-life that is at least two-fold longer compared to the biologically active polypeptide not linked to the ELNN.

[0183] In some embodiments, administration of a therapeutically effective amount of a fusion protein to a subject in need thereof results in a gain in time of at least two-fold, or at least three-fold, or at least four-fold, or at least five-fold or more spent within a therapeutic window for the fusion protein compared to the corresponding cytokine not linked to the ELNN(s) when administered at a comparable dose to a subject.

[0184] In some embodiments, a cytokine released from a fusion protein upon protease cleavage comprises one or more short polypeptides (e.g., about 30, 25, 20, 15, 14, 13, 12, 11, 10, or less amino acids in length) that has no amino acids other than G, A, P, E, S, and/or T. For example, a short polypeptide that has no amino acids other than G, A, P, E, S, and/or T might be incorporated into one or more spacer or linker sequences of the cytokine, and/or a portion of one or more spacers or linkers that remain part of the cytokine after cleavage. In some embodiments, a cytokine that is released from a fusion protein comprises a GTSESATPES (SEQ ID NO:96) on the N-terminal side (e.g., the closest amino acid of the sequence is within 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 amino acid positions of the N-terminal amino acid or the sequence includes the N-terminus) of the cytokine. In some embodiments, a cytokine that is released from a fusion protein comprises a GTATPESGPG on the C-terminal side (e.g., the closest amino acid of the sequence is within 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 amino acid positions of the N-terminal amino acid or the sequence includes the N-terminus) of the cytokine.

Low Immunogenicity

[0185] In some embodiments, the present disclosure provides compositions in which the ELNNs have a low degree of immunogenicity or are substantially non-immunogenic. Several factors can contribute to the low immunogenicity of an ELNN, e.g., the substantially 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 ELNN.

[0186] One of ordinary skill in the art will understand that, in general, polypeptides having highly repetitive short amino acid sequences (e.g., wherein a 200 amino acid-long sequence contain on average 20 repeats or more of a limited set of 3- or 4-mers) and/or having contiguous repetitive amino acid residues (e.g., wherein 5- or 6-mer sequences have identical amino acid residues) have a

tendency to aggregate or form higher order structures or form contacts resulting in crystalline or pseudo-crystalline structures.

[0187] In some embodiments, a ELNN sequence is substantially non-repetitive, wherein (1) the ELNN sequence has no three contiguous amino acids that are identical amino acid types, unless the amino acid is serine, in which case no more than three contiguous amino acids can be serine residues; and wherein (2) the ELNN contains no 3-amino acid sequences (3-mers) that occur more than 16, more than 14, more than 12, or more than 10 times within an at least 200 amino acid-long sequence of the ELNN (e.g., the entire span of an ELNN that is at least amino acids long). Without being bound by any scientific theory, such substantially non-repetitive sequences have less tendency to aggregate and, thus, enable the design of long-sequence ELNNs with a relatively low frequency of charged amino acids that would be likely to aggregate if the sequences or amino acid residues were otherwise more repetitive.

[0188] Conformational epitopes can be formed by regions of protein surfaces that are composed of multiple discontinuous amino acid sequences of a protein antigen. Without being bound by any scientific theory, the precise folding of the protein may bring these sequences into 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 and/or triggering 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 an 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 may lead 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.

[0189] Without being bound by any scientific theory, the ability of a peptide to bind a given MHC Class II molecule for presentation on the surface of an APC (antigen presenting cell) may depend on a number of factors; most notably its primary sequence. In some embodiments, a lower degree of immunogenicity may be achieved by designing ELNNs that resist antigen processing in antigen presenting cells, and/or choosing sequences that do not bind MHC receptors well. In some embodiments, ELNN-containing fusion proteins have substantially non-repetitive ELNN polypeptides designed to reduce binding with MHC II receptors, as well as to avoid formation of epitopes for T-cell receptor or antibody binding, resulting in a low degree of immunogenicity. Without being bound by any scientific theory, avoidance of immunogenicity is, in part, a direct result of the conformational flexibility of ELNNs; 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 adopt compactly folded conformations in aqueous solution or under physiologic conditions that could result in

conformational epitopes. The administration of fusion proteins comprising ELNNs, using conventional therapeutic practices and dosing, would generally not result in the formation of neutralizing antibodies to the ELNNs, and may also reduce the immunogenicity of cytokine fusion partners in the fusion protein compositions.

[0190] In some embodiments, the ELNNs utilized in the subject fusion proteins 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 can be 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 ELNN sequence to eliminate or disrupt the epitope sequence. In some embodiments, the ELNNs are substantially non-immunogenic by the restriction of the numbers of epitopes of the ELNN 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 74 of International Patent Application Publication No. WO 2010/144502 A2, which is incorporated by reference in its entirety. Aspects of the TEPITOPE score of a given peptide frame within a protein are 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 can serve as anchor residues during peptide display on MHC, such as M, I, L, V, or F. In some embodiments, an ELNN component incorporated into a fusion protein does not have a predicted T-cell epitope at a TEPITOPE score of about -5 or greater, or -6 or greater, or -7 or greater, or -8 or greater, or at a TEPITOPE score of -9 or greater. As used herein, a score of “-9 or greater” would encompass TEPITOPE scores of 10 to -9, inclusive, but would not encompass a score of -10, as -10 is less than -9.

[0191] In some embodiments, the ELNNs, including those incorporated into the subject fusion proteins, can be rendered substantially non-immunogenic by the restriction of known proteolytic sites from the sequence of the ELNN, reducing the processing of ELNN into small peptides that can bind to MHC II receptors. In some embodiments, the ELNN sequence can be rendered substantially non-immunogenic by the use a sequence that is substantially devoid of secondary structure, conferring resistance to many proteases due to the high entropy of the structure. Accordingly, the reduced

TEPITOPE score and elimination of known proteolytic sites from the ELNN may render the ELNN compositions, including the ELNN of the fusion protein compositions, substantially unable to be bound by mammalian receptors, including those of the immune system. In some embodiments, an ELNN of a fusion protein can have >100 nM K_D binding to a mammalian receptor, or greater than 500 nM K_D , or greater than 1 μ M K_D towards a mammalian cell surface or circulating polypeptide receptor.

[0192] Additionally, the substantially non-repetitive sequence and corresponding lack of epitopes of such embodiments of ELNNs can limit the ability of B cells to bind to or be activated by the ELNNs. In some embodiments, while an ELNN can make contacts with many different B cells over its extended sequence, each individual B cell may only make one or a small number of contacts with an individual ELNN. As a result, ELNNs typically may have a much lower tendency to stimulate proliferation of B cells and thus an immune response. In some embodiments, the fusion protein may have reduced immunogenicity as compared to the corresponding cytokine that is not fused to a mask polypeptide such as an ELNN. In some embodiments, the administration of up to three parenteral doses of a fusion protein to a mammal may result in detectable anti-fusion protein IgG at a serum dilution of 1:100 but not at a dilution of 1:1000. In some embodiments, the administration of up to three parenteral doses of a fusion protein to a mammal may result in detectable anti-cytokine IgG at a serum dilution of 1:100 but not at a dilution of 1:1000. In some embodiments, the administration of up to three parenteral doses of a fusion protein to a mammal may result in detectable anti-ELNN IgG at a serum dilution of 1:100 but not at a dilution of 1:1000. In some embodiments, the mammal can be, e.g., a mouse, a rat, a rabbit, cynomolgus monkey, or human. In some embodiments, the mammal is a human.

[0193] An additional feature of certain ELNNs with substantially non-repetitive sequences relative to those less non-repetitive sequences (such as one having three contiguous amino acids that are identical) can be that non-repetitive ELNNs form weaker contacts with cytokines (e.g., monovalent interactions), thereby resulting in less likelihood of immune clearance such that the fusion protein compositions can remain in circulation for an increased period of time.

[0194] In some embodiments, a biologically active polypeptide (such as a cytokine) comprising an ELNN is less immunogenic compared to the fusion polypeptide not linked to any ELNN, wherein immunogenicity is ascertained by measuring production of IgG antibodies that selectively bind to the biologically active polypeptide after administration of comparable doses to a subject.

BARCODE FRAGMENT

[0195] In some embodiments, a polypeptide (e.g., a fusion polypeptide or a portion thereof such as an ELNN) comprises one or more barcode fragments (e.g., a first, second, or third barcode fragment) releasable from the polypeptide upon digestion by a protease. In some embodiments, the protease is a

non-mammalian protease. In some embodiments, the protease is a prokaryotic protease. As used herein, the term “barcode fragment” (or “barcode,” or “barcode sequence”) can refer to either the portion of the polypeptide cleavably fused within the polypeptide, or the resulting peptide fragment released from the polypeptide.

[0196] In some embodiments, a barcode fragment (1) is a portion of an ELNN that includes at least part of the (non-recurring, non-overlapping) sequence motif that occurs (or is found) only once within the ELNN; and (2) differs in sequence and molecular weight from all other peptide fragments that are releasable from the polypeptide upon cleavage or complete digestion of the polypeptide by the protease.

[0197] In some embodiments, a barcode fragment does not include the N-terminal amino acid or the C-terminal amino acid of the fusion polypeptide. As described herein, in some embodiments, a barcode fragment is releasable (e.g., configured to be released) upon Glu-C digestion of the fusion polypeptide. In some embodiments, a barcode fragment is in an ELNN and does not include a glutamic acid that is immediately adjacent to another glutamic acid, if present, in the ELNN. In some embodiments, a barcode fragment has a glutamic acid at its C-terminus. One of ordinary skill in the art will understand that the C-terminus of a barcode fragment can refer to the “last” (or the most C-terminal) amino acid residue within the barcode fragment, when cleavably fused within a polypeptide (such as an ELNN), even if other non-barcode amino acid residues are positioned C-terminal to the barcode fragment within the polypeptide (e.g., ELNN). In some embodiments, a barcode fragment has an N-terminal amino acid that is immediately preceded by a glutamic acid residue. In some embodiments, the glutamic acid residue that precedes the N-terminal amino acid is not immediately adjacent to another glutamic acid residue. In some embodiments, a barcode fragment does not include a (second) glutamic acid residue at a position other than the C-terminus of the barcode fragment unless the glutamic acid is immediately followed by a proline. In some embodiments, a barcode fragment is positioned a distance from either the N-terminus of the polypeptide or the C-terminus of the polypeptide, wherein the distance is from 10 to 150, or 10 to 125 amino acids. In some embodiments, a barcode fragment is positioned within, or at a location of, 300, 280, 260, 250, 240, 220, 200, 190, 180, 170, 160, 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 48, 40, 36, 30, 24, 20, 12, or 10 amino acids from the N-terminus of the polypeptide, or at a location in a range between any of the foregoing. In some embodiments, a barcode fragment is positioned within 200, within 150, within 100, or within 50 amino acids of the N-terminus of the polypeptide. In some embodiments, a barcode fragment is positioned at a location that is between 10 and 200, between 30 and 200, between 40 and 150, or between 50 and 100 amino acids from the N-terminus of the polypeptide. In some embodiments, a barcode fragment is positioned within, or at a location of, 300, 280, 260, 250, 240, 220, 200, 190, 180, 170, 160, 150, 140, 130, 120, 110, 100, 90, 80, 70, 60, 50, 48, 40, 36, 30, 24, 20, 12, or 10 amino acids from the C-terminus of the polypeptide, or at a location in a range between any

of the foregoing. In some embodiments, a barcode fragment is positioned within 200, within 150, within 100, or within 50 amino acids of the C-terminus of the polypeptide. In some embodiments, a barcode fragment is positioned at a location that is between 10 and 200, between 30 and 200, between 40 and 150, or between 50 and 100 amino acids from the C-terminus of the polypeptide. In some embodiments, a barcode fragment (BAR) is characterized in that: (i) it does not include a glutamic acid that is immediately adjacent to another glutamic acid, if present, in the ELNN; (ii) it has a glutamic acid at its C-terminus; (iii) it has an N-terminal amino acid that is immediately preceded by a glutamic acid residue; and (iv) it is positioned a distance from either the N-terminus of the polypeptide or the C-terminus of the polypeptide, wherein the distance is from 10 to 150 amino acids, or from 10 to 125 amino acids in length. In some embodiments, a barcode fragment is in an ELNN and (i) does not include the N-terminal amino acid or the C-terminal amino acid of the polypeptide; (ii) does not include a glutamic acid that is immediately adjacent to another glutamic acid in the ELNN; (iii) has a glutamic acid at its C-terminus; (iv) has an N-terminal amino acid that is immediately preceded by a glutamic acid residue; and (v) is positioned a distance from either the N-terminus of the polypeptide or the C-terminus of the polypeptide, wherein the distance is from 10 to 150, or 10 to 125 amino acids in length. In some embodiments, the glutamic acid residue that precedes the N-terminal amino acid is not immediately adjacent to another glutamic acid residue. In some embodiments, a barcode fragment does not include a glutamic acid residue at a position other than the C-terminus of the barcode fragment unless the glutamic acid is immediately followed by a proline. Depending on context herein and when referring to placement within a polypeptide sequence, the term “distance” can refer to the number of amino acid residues from the N-terminus of the polypeptide to the most N-terminal amino acid residue of the barcode fragment, or from the C-terminus of the polypeptide to the most C-terminal amino acid residue of the barcode fragment. In some embodiments, for a barcoded ELNN fused to a biologically active polypeptide, at least one barcode fragment (or at least two barcode fragments, or three barcode fragments) contained in the barcoded ELNN is positioned at least 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300 amino acids from the biologically active polypeptide. In some embodiments, a barcode fragment is at least 4, at least 5, at least 6, at least 7, or at least 8 amino acids in length. In some embodiments, a barcode fragment is at least 4 amino acids in length. In some embodiments, a barcode fragment is 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 amino acids in length, or in a range between any of the foregoing values. In some embodiments, a barcode fragment is between 4 and 20, between 5 and 15, between 6 and 12, or between 7 and 10 amino acids in length. In some embodiments, a barcode fragment comprises an amino acid sequence identified herein by SEQ ID NOs: 68-79 and SEQ ID NOs: 1010-1029 in Table 2.

Table 2. Exemplary Barcode Fragments Releasable Upon Glu-C Digest

Amino Acid Sequence	SEQ ID NO:
SPATSGSTPE	68
GSAPATSE	69
GSAPGTATE	70
GSAPGTE	71
PATSGPTE	72
SASPE	73
PATSGSTE	74
GSAPGTSAE	75
SATSGSE	76
SGPGSTPAE	77
SGPGSGPGTSE	78
SGPGTSPSATPE	79
SGPGTGTSATPE	1010
SGPGTTPGTPE	1011
SGPGTPPTSTPE	1012
SGPGTGSAGTPE	1013
SGPGTGGAGTPE	1014
SGPGTSPGATPE	1015
SGPGTSGSGTPE	1016
SGPGTSSASTPE	1017
SGPGTGAGTPE	1018
SGPGTGSTSTPE	1019
TPGSEPATSGSE	1020
GSAPGTSTEPSE	1021
SGPGTAGSGTPE	1022
SGPGTSSGGTPE	1023
SGPGTAGPATPE	1024
SGPGTPGTGTPE	1025
SGPGTGGPTTPE	1026
SGPGTGSGSTPE	1027
SGPGSGPATSE	1028
SGPGTSASATPE	1029

[0198] In some embodiments, each barcode fragment differs in both sequence and molecular weight from all other peptide fragments that are releasable from the chimeric polypeptides described herein upon complete digestion the chimeric polypeptide by a non-mammalian protease. In some embodiments, the non-mammalian protease is Glu-C.

[0199] In some embodiments, the chimeric polypeptides disclosed herein comprises a Glu-C cleavage site comprising one of the following amino acid sequences: ATPESGPG (SEQ ID NO:2223), SGSETPGT (SEQ ID NO:2242), and GTSESATP (SEQ ID NO:204).

[0200] In some embodiments, the chimeric polypeptides disclosed herein comprises at least one of the following amino acid sequences: PE.GSX_nPE.SG (SEQ ID NO:2243), PE.GSX_nSE.GG (SEQ ID NO:2244), PE.GSX_nSE.TG (SEQ ID NO:2246), PE.GSX_nSE.SA (SEQ ID NO:2245), PE.SGX_nPE.SG (SEQ ID NO:2247), PE.SGX_nSE.GG (SEQ ID NO:2248), PE.SGX_nSE.TG (SEQ ID NO:2250), PE.SGX_nSE.SA (SEQ ID NO:2249), and PE.TPX_nPE.SG (SEQ ID NO:2251), PE.TPX_nSE.GG (SEQ ID NO:2252), PE.TPX_nSE.TG (SEQ ID NO:2254), PE.TPX_nSE.SA (SEQ ID NO:2253), wherein each “.” is a Glu-C cleavage site and n is any integer from 0 to 50. In some embodiments, the chimeric polypeptides disclosed herein comprises at least one of the following amino acid sequences: PE.SGX_nPE.SG (SEQ ID NO:2247), PE.GSX_nSE.GG (SEQ ID NO:2244), PE.TPX_nSE.TG (SEQ ID NO:2254), PE.SGX_nSE.SA (SEQ ID NO:2249). In some embodiments, n is any integer from 1 to 20. In some embodiments, n is any integer from 5 to 15. In some embodiments, n is any integer from 5 to 10. In some embodiments, n is 9. In some embodiments, n is any integer from 5 to 15. In some embodiments, X_n is SGPGTGT SATPE (SEQ ID NO:1010), SGPGSGPGTSE (SEQ ID NO:78), SGPGTTPGTPE (SEQ ID NO:1011), SGPGTPPTSTPE (SEQ ID NO:1012), SGPGTSPSATPE (SEQ ID NO:79), SGPGTGSAGTPE (SEQ ID NO:1013), SGPGTGGAGTPE (SEQ ID NO:1014), SGPGTSPGATPE (SEQ ID NO:1015), SGPGTSGSGTPE (SEQ ID NO:1016), SGPGTSSASTPE (SEQ ID NO:1017), SGPGTGAGTTPE (SEQ ID NO:1018), SGPGTGSTSTPE (SEQ ID NO:1019), TPGSEPATSGSE (SEQ ID NO:1020), GSAPGTSTEPSE (SEQ ID NO:1021), SGPGTAGSGTPE (SEQ ID NO:1022), SGPGTSSGGTPE (SEQ ID NO:1023), SGPGTAGPATPE (SEQ ID NO:1024), SGPGTPGTGTPE (SEQ ID NO:1025), SGPGTGGPTTPE (SEQ ID NO:1026), or SGPGTGSSTPE (SEQ ID NO:1027).

[0201] In some embodiments, a chimeric polypeptide comprises at least one of the following amino acid sequences:

SGPE.SGPGX_nSGPE.SGPG (SEQ ID NO:2001), SGPE.SGPGX_nATPE.SGPG (SEQ ID NO:2002), SGPE.SGPGX_nGTSE.SATP (SEQ ID NO:2003), SGPE.SGPGX_nTTPE.SGPG (SEQ ID NO:2004), SGPE.SGPGX_nSTPE.SGPG (SEQ ID NO:2005), SGPE.SGPGX_nGTPE.SGPG (SEQ ID NO:2006), SGPE.SGPGX_nGTPE.TPGS (SEQ ID NO:2007), SGPE.SGPGX_nGTPE.TPGS (SEQ ID NO:2007), SGPE.SGPGX_nSGSE.TGTP (SEQ ID NO:2008), SGPE.SGPGX_nGTPE.GSAP (SEQ ID NO:2009), SGPE.SGPGX_nEPSE.SATP (SEQ ID NO:2010), ATPE.SGPGX_nSGPE.SGPG (SEQ ID NO:2011), ATPE.SGPGX_nATPE.SGPG (SEQ ID NO:2012), ATPE.SGPGX_nGTSE.SATP (SEQ ID NO:2013), ATPE.SGPGX_nATSE.SATP (SEQ ID NO:2014), ATPE.SGPGX_nTTPE.SGPG (SEQ ID NO:2015), ATPE.SGPGX_nSTPE.SGPG (SEQ ID NO:2016), ATPE.SGPGX_nGTPE.SGPG (SEQ ID NO:2017), ATPE.SGPGX_nGTPE.TPGS (SEQ ID NO:2018), ATPE.SGPGX_nSGSE.TGTP (SEQ ID NO:2019), ATPE.SGPGX_nGTPE.GSAP (SEQ ID NO:2020), ATPE.SGPGX_nEPSE.SATP (SEQ ID NO:2021), GTSE.SATPX_nSGPE.SGPG (SEQ ID NO:2022), GTSE.SATPX_nATPE.SGPG (SEQ ID NO:2023), GTSE.SATPX_nGTSE.SATP (SEQ ID NO:2024), GTSE.SATPX_nTTPE.SGPG (SEQ ID NO:2025).

GTSE.SATPX_nSTPE.SGPG (SEQ ID NO:2026), GTSE.SATPX_nGTPE.SGPG (SEQ ID NO:2027),
GTSE.SATPX_nGTPE.TPGS (SEQ ID NO:2028), GTSE.SATPX_nSGSE.TGTP (SEQ ID NO:2029),
GTSE.SATPX_nGTPE.GSAP (SEQ ID NO:2030), GTSE.SATPX_nEPSE.SATP (SEQ ID NO:2031),
TTPE.SGPGX_nSGPE.SGPG (SEQ ID NO:2032), TTPE.SGPGX_nATPE.SGPG (SEQ ID NO:2033),
TTPE.SGPGX_nGTSE.SATP (SEQ ID NO:2034), TTPE.SGPGX_nTTPE.SGPG (SEQ ID NO:2035),
TTPE.SGPGX_nSTPE.SGPG (SEQ ID NO:2036), TTPE.SGPGX_nGTPE.SGPG (SEQ ID NO:2037),
TTPE.SGPGX_nGTPE.TPGS (SEQ ID NO:2038), TTPE.SGPGX_nSGSE.TGTP (SEQ ID NO:2039),
TTPE.SGPGX_nGTPE.GSAP (SEQ ID NO:2040), TTPE.SGPGX_nEPSE.SATP (SEQ ID NO:2041),
STPE.SGPGX_nSGPE.SGPG (SEQ ID NO:2042), STPE.SGPGX_nATPE.SGPG (SEQ ID NO:2043),
STPE.SGPGX_nGTSE.SATP (SEQ ID NO:2044), STPE.SGPGX_nTTPE.SGPG (SEQ ID NO:2045),
STPE.SGPGX_nSTPE.SGPG (SEQ ID NO:2046), STPE.SGPGX_nGTPE.SGPG (SEQ ID NO:2047),
STPE.SGPGX_nGTPE.TPGS (SEQ ID NO:2048), STPE.SGPGX_nSGSE.TGTP (SEQ ID NO:2049),
STPE.SGPGX_nGTPE.GSAP (SEQ ID NO:2050), STPE.SGPGX_nEPSE.SATP (SEQ ID NO:2051),
GTPE.SGPGX_nSGPE.SGPG (SEQ ID NO:2052), GTPE.SGPGX_nATPE.SGPG (SEQ ID NO:2053),
GTPE.SGPGX_nGTSE.SATP (SEQ ID NO:2054), GTPE.SGPGX_nTTPE.SGPG (SEQ ID NO:2055),
GTPE.SGPGX_nSTPE.SGPG (SEQ ID NO:2056), GTPE.SGPGX_nGTPE.SGPG (SEQ ID NO:2057),
GTPE.SGPGX_nGTPE.TPGS (SEQ ID NO:2058), GTPE.SGPGX_nSGSE.TGTP (SEQ ID NO:2059),
GTPE.SGPGX_nGTPE.GSAP (SEQ ID NO:2060), GTPE.SGPGX_nEPSE.SATP (SEQ ID NO:2061),
GTPE.TPGSX_nSGPE.SGPG (SEQ ID NO:2062), GTPE.TPGSX_nATPE.SGPG (SEQ ID NO:2063),
GTPE.TPGSX_nGTSE.SATP (SEQ ID NO:2064), GTPE.TPGSX_nTTPE.SGPG (SEQ ID NO:2065),
GTPE.TPGSX_nSTPE.SGPG (SEQ ID NO:2066), GTPE.TPGSX_nGTPE.SGPG (SEQ ID NO:2067),
GTPE.TPGSX_nGTPE.TPGS (SEQ ID NO:2068), GTPE.TPGSX_nSGSE.TGTP (SEQ ID NO:2069),
GTPE.TPGSX_nGTPE.GSAP (SEQ ID NO:2070), GTPE.TPGSX_nEPSE.SATP (SEQ ID NO:2071),
SGSE.TGTPX_nSGPE.SGPG (SEQ ID NO:2072), SGSE.TGTPX_nATPE.SGPG (SEQ ID NO:2073),
SGSE.TGTPX_nGTSE.SATP (SEQ ID NO:2074), SGSE.TGTPX_nTTPE.SGPG (SEQ ID NO:2075),
SGSE.TGTPX_nSTPE.SGPG (SEQ ID NO:2076), SGSE.TGTPX_nGTPE.SGPG (SEQ ID NO:2077),
SGSE.TGTPX_nGTPE.TPGS (SEQ ID NO:2078), SGSE.TGTPX_nSGSE.TGTP (SEQ ID NO:2079),
SGSE.TGTPX_nGTPE.GSAP (SEQ ID NO:2080), SGSE.TGTPX_nEPSE.SATP (SEQ ID NO:2081),
GTPE.GSAPX_nSGPE.SGPG (SEQ ID NO:2082), GTPE.GSAPX_nATPE.SGPG (SEQ ID NO:2083),
GTPE.GSAPX_nGTSE.SATP (SEQ ID NO:2084), GTPE.GSAPX_nTTPE.SGPG (SEQ ID NO:2085),
GTPE.GSAPX_nSTPE.SGPG (SEQ ID NO:2086), GTPE.GSAPX_nGTPE.SGPG (SEQ ID NO:2087),
GTPE.GSAPX_nGTPE.TPGS (SEQ ID NO:2088), GTPE.GSAPX_nSGSE.TGTP (SEQ ID NO:2089),
GTPE.GSAPX_nGTPE.GSAP (SEQ ID NO:2090), GTPE.GSAPX_nEPSE.SATP (SEQ ID NO:2091),
EPSE.SATPX_nSGPE.SGPG (SEQ ID NO:2092), EPSE.SATPX_nATPE.SGPG (SEQ ID NO:2093),
EPSE.SATPX_nGTSE.SATP (SEQ ID NO:2094), EPSE.SATPX_nTTPE.SGPG (SEQ ID NO:2095),
EPSE.SATPX_nSTPE.SGPG (SEQ ID NO:2096), EPSE.SATPX_nGTPE.SGPG (SEQ ID NO:2097),
EPSE.SATPX_nGTPE.TPGS (SEQ ID NO:2098), EPSE.SATPX_nSGSE.TGTP (SEQ ID NO:2099),

EPSE.SATPX_nGTPE.GSAP (SEQ ID NO:2100), or EPSE.SATPX_nEPSE.SATP (SEQ ID NO:2101), wherein each “.” is a Glu-C cleavage site and n is any integer from 0 to 50. In some embodiments, the chimeric polypeptide comprises at least one of the following amino acid sequences:

SGPE.SGPGX_nATPE.SGPG (SEQ ID NO:2105), ATPE.SGPGX_nGTSE.SATP (SEQ ID NO:2103), ATPE.SGPGX_nATSE.SATP (SEQ ID NO:2104), ATPE.SGPGX_nTTPE.SGPG (SEQ ID NO:2106), ATPE.SGPGX_nSTPE.SGPG (SEQ ID NO:2107), ATPE.SGPGX_nATPE.SGPG (SEQ ID NO:2102), ATPE.SGPGX_nGTPE.SGPG (SEQ ID NO:2108), ATPE.SGPGX_nGTPE.SGPG (SEQ ID NO:2108), ATPE.SGPGX_nATPE.SGPG (SEQ ID NO:2102), GTPE.SGPGX_nGTPE.SGPG (SEQ ID NO:2109), GTPE.SGPGX_nSTPE.SGPG (SEQ ID NO:2110), GTPE.SGPGX_nTTPE.SGPG (SEQ ID NO:2111), GTPE.SGPGX_nSTPE.SGPG (SEQ ID NO:2110), GTPE.TPGSX_nSGSE.TGTP (SEQ ID NO:2112), GTPE.GSAPX_nEPSE.SATP (SEQ ID NO:2113), ATPE.SGPGX_nGTPE.SGPG (SEQ ID NO:2108), ATPE.SGPGX_nGTPE.SGPG (SEQ ID NO:2108), ATPE.SGPGX_nATPE.SGPG (SEQ ID NO:2102), ATPE.SGPGX_nGTPE.SGPG (SEQ ID NO:2108), TTPE.SGPGX_nTTPE.SGPG (SEQ ID NO:2114), or STPE.SGPGX_nSTPE.SGPG (SEQ ID NO:2115), wherein each “.” is a Glu-C cleavage site and n is any integer from 0 to 30. In some embodiments, n is any integer from 1 to 20. In some embodiments, n is any integer from 5 to 15. In some embodiments, n is any integer from 3 to 7. In some embodiments, n is any integer from 5 to 10. In some embodiments, n is 9. In some embodiments, n is 4. In some embodiments, n is 3. In some embodiments, n is any integer from 5 to 15. In some embodiments, wherein X_n is PGTGTSAT (SEQ ID NO:2233), PGSGPGT (SEQ ID NO:2221), PGTPGT (SEQ ID NO:2241), PGTPPTST (SEQ ID NO:2235), PGTSPSAT (SEQ ID NO:2238), PGTGSAGT (SEQ ID NO:2230), PGTGGAGT (SEQ ID NO:2228), PGTSPGAT (SEQ ID NO:2237), PGTSGSGT (SEQ ID NO:2236), PGTSSAST (SEQ ID NO:2239), PGTGAGTT (SEQ ID NO:2227), PGTGSTST (SEQ ID NO:2232), GSEPATSG (SEQ ID NO:2224), APGTSTEP (SEQ ID NO:2222), PGTAGSGT (SEQ ID NO:2226), PGTSSGGT (SEQ ID NO:2240), PGTAGPAT (SEQ ID NO:2225), PGTPGTGT (SEQ ID NO:2234), PGTGGPPT (SEQ ID NO:2229), or PGTGSGST (SEQ ID NO:2231). In some embodiments, X_n is TSAS (SEQ ID NO:2214), TGTS (SEQ ID NO:2211), SGP, TTPG (SEQ ID NO:2220), TPPT (SEQ ID NO:2213), TSPS (SEQ ID NO:2217), TGSA (SEQ ID NO:2208), TGGA (SEQ ID NO:2206), TSPG (SEQ ID NO:2216), TSGS (SEQ ID NO:2215), TSSA (SEQ ID NO:2218), TGAG (SEQ ID NO:2205), TGST (SEQ ID NO:2210), EPAT (SEQ ID NO:2201), GTST (SEQ ID NO:2202), TAGS (SEQ ID NO:2204), TSSG (SEQ ID NO:2219), TAGP (SEQ ID NO:2203), TPGT (SEQ ID NO:2212), TGGP (SEQ ID NO:2207), or TGSG (SEQ ID NO:2209).

[0202] In some embodiments, barcodes are designed to have improved analytical properties. In some embodiments, such barcodes can be released with relatively modest concentrations of a non-mammalian protease such as Glu-C. This facilitates better detection, e.g., through LC/MS, and also

allows measurement of peptides that are generated from the cleavable linker thereby allowing a measurement of cleavage products using, e.g., LC/MS.

[0203] In some embodiments of fusion proteins comprising an ELNN, the fusion protein has a single polypeptide chain, and the polypeptide chain comprises a barcode fragment that is at a position within the polypeptide chain that is from 10 to 200 amino acids or from 10 to 125 amino acids from the N-terminus or the C-terminus of the polypeptide chain. In some embodiments, a fusion protein comprises a first ELNN and a second ELNN, the first ELNN is at the N-terminal side of the cytokine, and the first barcode fragment is positioned within 200, 150, 100, or 50 amino acids of the N-terminus of the fusion protein. In some embodiments, the second ELNN is at the C-terminal side of the cytokine, and the second barcode fragment is positioned within 200, 150, 100, or 50 amino acids of the C-terminus of the chimeric polypeptide.

[0204] In some embodiments, an ELNN further comprises one or more additional barcode fragments, wherein the one or more additional barcode fragments each differs in sequence and molecular weight from all other peptides fragments that are releasable from the polypeptide upon complete digestion of the polypeptide by the protease. In some embodiments, a barcoded ELNN comprises only one barcode fragment. In some embodiments, a barcoded ELNN comprises a set of barcode fragments, comprising a first barcode fragment, such as those described herein. In some embodiments, the set of barcode fragments comprises a second barcode fragment (or a further barcode fragment), such as those described herein. In some embodiments, the set of barcode fragments comprises a third barcode fragment, such as those described herein.

[0205] A set of barcode fragments fused within an N-terminal ELNN can be referred to as an N-terminal set of barcodes (an “N-terminal set”). A set of barcode fragments fused within a C-terminal ELNN can be referred to as a C-terminal set of barcodes (a “C-terminal set”). In some embodiments, the N-terminal set comprises a first barcode fragment and a second barcode fragment. In some embodiments, the N-terminal set further comprises a third barcode fragment. In some embodiments, the C-terminal set comprises a first barcode fragment and a second barcode fragment. In some embodiments, the C-terminal set further comprises a third barcode fragment. In some embodiments, the polypeptide comprises a set of barcode fragments that includes a first barcode fragment, a further (second) barcode fragment, and at least one additional barcode fragment, wherein each barcode fragment of the set of barcode fragments (1) is a portion of the second ELNN and (2) differs in sequence and molecular weight from all other peptides fragments that are releasable from the polypeptide upon complete digestion of the polypeptide by the protease.

[0206] Included herein is a mixture comprising a plurality of polypeptides of varying length; the mixture comprising a first set of polypeptides and a second set of polypeptides. In some embodiments, each polypeptide of the first set of polypeptides comprises a barcode fragment that (a) is releasable

from the polypeptide by digestion with a protease and (b) has a sequence and molecular weight that differs from the sequence and molecular weight of all other fragments that are releasable from the first set of polypeptides. In some embodiments, the second set of polypeptides lack the barcode fragment of the first set of polypeptides (e.g., due to truncation). In some embodiments, both the first set of polypeptides and the second set of polypeptides each comprise a reference fragment that (a) is common to the first set of polypeptides and the second set of polypeptides and (b) releasable by digestion with the protease. In some embodiments, the ratio of the first set of polypeptides to polypeptides comprising the reference fragment is greater than 0.70. In some embodiments, the ratio of the first set of polypeptides to polypeptides comprising the reference fragment is greater than 0.80, 0.90, 0.95, or 0.98. In some embodiments, the reference fragment occurs no more than once in each polypeptide of the first set of polypeptides and the second set of polypeptides. In some embodiments, the protease is a protease that cleaves on the C-terminal side of glutamic acid residues. In some embodiments, the protease is a Glu-C protease. In some embodiments, the protease is not trypsin. In some embodiments, the polypeptides of varying lengths comprise polypeptides comprising at least one ELNN, such as any described herein. In some embodiments, the first set of polypeptides comprises a full-length polypeptide, wherein the barcode fragment is a portion of the full-length polypeptide. In some embodiments, the full-length polypeptide is a (fusion) polypeptide, such as any described hereinabove or described anywhere else herein. In some embodiments, the polypeptides of varying lengths in a mixture differ from one another due to N-terminal truncation, C-terminal truncation, or both N- and C-terminal truncation of a full-length polypeptide. In some embodiments, the first set of polypeptides and the second set of polypeptides may differ in one or more pharmacological properties.

[0207] The present disclosure also provides methods for assessing, in a mixture comprising polypeptides of varying length, a relative amount of a first set of polypeptides in the mixture to a second set of polypeptides in the mixture, wherein (1) each polypeptide of the first set of polypeptides shares a barcode fragment that occurs once and only once in the polypeptide and (2) each polypeptide of the second set of polypeptides lacks the barcode fragment that is shared by polypeptides of the first set, wherein individual polypeptides of both the first of polypeptides and the second set of polypeptides each comprises a reference fragment. In some embodiments, the methods comprise contacting the mixture with a protease to produce a plurality of proteolytic fragments that result from cleavage of the first set of polypeptides and the second set of polypeptides, wherein the plurality of proteolytic fragments comprise a plurality of reference fragments, and a plurality of barcode fragments. In some embodiments, the methods can further comprise determining a ratio of the amount of barcode fragments to the amount of reference fragments, thereby assessing the relative amounts of the first set of polypeptides to the second set of polypeptides. In some embodiments, the barcode fragment occurs no more than once in each polypeptide of the first set of polypeptides. In some

embodiments, the reference fragment occurs no more than once in each polypeptide of the first set of polypeptides and the second set of polypeptides. In some embodiments, the plurality of proteolytic fragments comprises a plurality of reference fragments, and a plurality of barcode fragments. In some embodiments, the protease cleaves the first and second sets of polypeptides (or the polypeptides of varying length) on the C-terminal side of glutamic acid residues that are not followed by a proline residue. In some embodiments, the protease is a Glu-C protease. In some embodiments, the protease is not trypsin. In some embodiments, the step of determining a ratio of the amount of barcode fragments to the amount of reference fragments comprises identifying barcode fragments and reference fragments from the mixture after it has been contacted with the protease. In some embodiments, the barcode fragments and the reference fragments are identified based on their respective masses. In some embodiments, the barcode fragments and the reference fragments are identified via mass spectrometry.

[0208] In some embodiments, the barcode fragments and reference fragments are identified via liquid chromatography-mass spectrometry (LC-MS). In some embodiments, the step of determining a ratio of the barcode fragments to the reference fragments comprises isobaric labeling. In some embodiments, the step of determining a ratio of the barcode fragments to the reference fragments comprises spiking the mixture with one or both of an isotope-labeled reference fragment and an isotope labeled barcode fragment. In some embodiments, the polypeptides of varying lengths comprise polypeptides that comprise at least one ELNN, as described hereinabove or described anywhere else herein. In some embodiments, the ELNN is characterized in that (i) it comprises at least 100, or at least 150 amino acids; (ii) at least 90% of the amino acid residues of the ELNN are glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) or proline (P); and (iii) it comprises at least 4 different types of amino acids that are G, A, S, T, E, or P. In some embodiments, the barcode fragment, when present, is a portion of the ELNN. In some embodiments, the mixture of polypeptides of varying lengths comprises a polypeptide as any described hereinabove or described anywhere else herein. In some embodiments, the polypeptides of varying length comprise a full-length polypeptide and truncated fragments thereof. In some embodiments, the polypeptides of varying length consist essentially of the full-length polypeptide and truncated fragments thereof. In some embodiments, the polypeptides of varying lengths in a mixture differ from one another due to N-terminal truncation, C-terminal truncation, or both N- and C-terminal truncation of a full-length polypeptide. In some embodiments, the full-length polypeptide is a polypeptide as described hereinabove or described anywhere else herein. In some embodiments, the ratio of the amount of barcode fragments to reference fragments is greater than 0.50, 0.60, 0.70, 0.80, 0.90, 0.95, 0.98, or 0.99.

SEQ ID NO:	ELNN Type	# of Barcode(s)	Amino Acid Sequence	Total # of AAs
			GSETPGSPAGSPTSTEEGTSTEPSE <u>GSAPGTE</u> STPS EGSAPGSEPATSGSETPGTSESATPESGPGTSTEPS EGSAPGEPEA	
8002	C-terminal ELNN	1	PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS PGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGS PGTSESATPESGPGSEPATSGSETPGSEPATSGSET PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS PGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGS PGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGS PGTSESATPESGPGSEPATSGSETPGTSTEPSEGS PGTSTEPSEGSAPGTSESATPESGPGTSESATPESG PGSPAGSPTSTEEGTSESATPESGPGSEPATSGSET PGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGS PGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGS PGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGS PGTSESATPESGPGSEPATSGSETPGTSESATPESG PGSEPATSGSETPGTSESATPESGPGTSTEPSEGS PGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTE EGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS PGTSESATPESGPGSEPATSGSETPGTSESATPESG PGSEPATSGSETPGTSESATPESGPGTSTEPSEGS PGSPAGSPTSTEEGTSESATPESGPGSEPATSGSET PGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTE EGTSTEPSEGSAPGTSESATPESGPGTSESATPESG PGTSESATPESGPGSE <u>PATSGPTE</u> SGSEPATSGSET PGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGS PGSEPATSGSETPGTSESATPESGPGTSTEPSEGS PGEPEA	864
8003	C-terminal ELNN	1	PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS PGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGS PGTSESATPESGPGSEPATSGSETPGSEPATSGSET PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS PGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGS PGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGS PGTSESATPESGPGSEPATSGSETPGTSTEPSEGS PGTSTEPSEGSAPGTSESATPESGPGTSESATPESG PGSPAGSPTSTEEGTSESATPESGPGSEPATSGSET PGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGS PGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGS PGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGS PGTSESATPESGPGSEPATSGSETPGTSESATPESG PGSEPATSGSETPGTSESATPESGPGTSTEPSEGS PGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTE EGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGS PGTSESATPESGPGSEPATSGSETPGTSESATPESG PGSEPATSGSETPGTSESATPESGPGTSTEPSEGS PGSPAGSPTSTEEGTSESATPESGPGSEPATSGSET PGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTE EGTSTEPSEGSAPGTSESATPESGPGTSESATPESG PGTSESATPESGPGSEPATSGSETPGSEPATSGSET PGSPAGSPTSTEEGTSTEPSE <u>GSAPGTE</u> STEPSEGS PGSEPATSGSETPGTSESATPESGPGTSTEPSEGS PGEPEA	864
8004	N-terminal ELNN	2	ASSPAGSPTSTESGTSESATPESGPGTETEPSEGS PGTSESATPESGPGSEPATSGSETPGTSESATPES <u>SG</u> <u>PGSTPAE</u> SGSETPGTSESATPESGPGTSTEPSEGS PGSPAGSPTSTEEGTSESATPESGPGE <u>SPATSGSTP</u> <u>EGT</u> SESATPESGPGSPAGSPTSTEEGSPAGSPTSTE EGTSTEPSEGSAPGTSESATPESGPGTSESATPESG	288

SEQ ID NO:	ELNN Type	# of Barcode(s)	Amino Acid Sequence	Total # of AAs
			PGTSESATPESGPGSEPATSGSETPGTSESATPESG PGSEPATSGSETPGTSESATPESGPGTSTEPSEGS PGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTE EGS PAGESPTSTEEGTSESATPESGPGTSTEPSEGS PGTSESATPESGPGSEPATSGSETPGTSESATPESG PGSEPATSGSETPGTSESATPESGPGTSTEPSEGS PGS PAGESPTSTEEGTSESATPESGPGSEPATSGSET PGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTE EGTSTEPSEGSAPGTSESATPESGPGTSESATPESG PGTSESATPESGPGSEPATSGSETPGSEPATSGSET PGS PAGESPTSTEEGTSTEPSEGSAPGTSTEPSEGS PGSEPATSGSETPGTSESATPESGPGTSTEPSEGS PG	
8010	N-terminal ELNN	2	SAGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPS EGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGTSESATPESGPGSTPAESGSETPGSEPAT GSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPS EGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPS EGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPS EGSAPGTSESATPESGPGSEPATSGSTETPGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESA TPESGPGSPAGSPTSTEEGTSESATPESGPGSEPAT SGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEP SEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGSEPATSGSETPGTSESATPESGPGTSTEP SEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGS PTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEP SEGSAPGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGSEPATSGSETPGTSESATPESGPGTSTEP SEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPAT SGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGS PTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESA TPESGPGTSESATPESGPGSEPATSGSETPGSEPAT SGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEP SEGSAPGSEPATSGSETPGTSESATPESGPGTSTEP SEGSAPGSPAGSPTSTEEGTSESATPESGPGTSTEP SEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEP SEGSAPGTSESATPESGPGSEPATSGSETPGSEPAT SGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEP SEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEP SEGSAPGTSESATPESGPGSEPATSGSETPGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGTESAS	1152
8011	C-terminal ELNN	3	SAGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPS EGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGTSESATPESGPGSEPATSGSETPGSEPAT GSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPS EGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPS EGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPS EGSAPGTSESATPESGPGSEPATSGSETPGTSTEPS EGSAPGTSTEPSEGSAPGTSESATPESGPGTSESA TPESGPGSPAGSPTSTEEGTSESATPESGPGSEPAT GSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPS EGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPS EGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPS EGSAPGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGSEPATSGSETPGTSESATPESGPGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESA TPESGPGTSESATPESGPGSEPATSGSETPGSEPAT SGSETPGSPAGSPTSTEEGTSESATPESGPGTSTEP SEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEP SEGSAPGTSESATPESGPGSEPATSGSETPGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGTESAS	1152

SEQ ID NO:	ELNN Type	# of Barcode(s)	Amino Acid Sequence	Total # of AAs
			EGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGSPAGSPSTSTEEGTSESATPESGPGTSTEPS EGSAPGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGTSTEPS EGSAPGSPAGSPSTSTEEGTSESATPESGPGSEPATSG SETPGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGTSTEPSEGSAPGTSESATPESGPGTSESAT PESGPGTSESATPESGPGSEPATSGSETPGSEPATSG SETPGSPAGSPSTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGSEPATSGSETPGTSESATPESGPGTSTEPS EGSAPGSPAGSPSTSTEEGTSESATPESGPGTSTEPS EGSAPGSPAGSPSTSTEEGTSTEPSEGSAPGTSTEPS EGSAPGTSESATPESGPGSEPATSGSETPGSE PATS GSTE TPGSPAGSPSTSTEEGTSESATPESGPGTSTEP SEGSAPGTSTEPSEGSAPGSPAGSPSTSTEEGTSTEP SE GSAPGTATE SPESGAPGTSESATPESGPGTSTEP SE GSAPGTSAE SATPESGPGSEPATSGSETPGTSTE PSEGSAPGTSTEPSEGSAPGTSESATPESGPGTESA S	
8012	N-terminal ELNN	2	GSPAGSPSTSTEEGTSESATPESGPGTSTEPSEGSAP GSPAGSPSTSTEEGTSTEPSEGSAPGTSTEPSE GSAP ATSE SATPESGPGSEPATSGSETPGSEPATSGSETP GSPAGSPSTSTEEGTSE SASPE SGPGTSTEPSEGSAP GTSTEPSEGSAPGSPAGSPSTSTEEGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGSEPATSGSETPGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGTSESATPESGP GSPAGSPSTSTEEGTSESATPESGPGSEPATSGSETP GTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAP GTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAP GTSTEPSEGSAPGSPAGSPSTSTEEGTSTEPSEGSAP GTSESATPESGPGSEPATSGSETPGTSESATPESGP GSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGSPAGSPSTSTEEGSPAGSPSTTEE GSPAGSPSTSTEEGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGSEPATSGSETPGTSESATPESGP GSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP GSPAGSPSTSTEEGTSESATPESGPGSEPATSGSETP GTSESATPESGPGSPAGSPSTSTEEGSPAGSPSTTEE GTSTEPSEGSAPGTSESATPESGPGTSESATPESGP GTSESATPESGPGSEPATSGSETPGSEPATSGSETP GSPAGSPSTSTEEGTSTEPSEGSAPGTSTEPSEGSAP GSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP	864
8013	N-terminal ELNN	1	GSPAGSPSTSTEEGTSESATPESGPGTSTEPSEGSAP GSPAGSPSTSTEEGTSTEPSEGSAPGTSTEPSEGSAP GTSESATPESGPGSE SATSGSE TPGSEPATSGSETP GSPAGSPSTSTEEGTSESATPESGPGTSTEPSEGSAP GTSTEPSEGSAPGSPAGSPSTSTEEGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGSEPATSGSETPGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGTSESATPESGP GSPAGSPSTSTEEGTSESATPESGPGSEPATSGSETP GTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAP GTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAP GTSTEPSEGSAPGSPAGSPSTSTEEGTSTEPSEGSAP GTSESATPESGPGSEPATSGSETPGTSESATPESGP GSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGSPAGSPSTSTEEGSPAGSPSTTEE GSPAGSPSTSTEEGTSESATPESGPGTSTEPSEGSAP	864

SEQ ID NO:	ELNN Type	# of Barcode(s)	Amino Acid Sequence	Total # of AAs
			GTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGSEPATSGSETPGTSESATPESGP GTSTEPSEGSAPGTSE SASPE SGPGSPAGSPTSTEE GSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGP GTSTEPSEGSAPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGSEPATSGSETPGTSESATPESGP	
8018	C-terminal ELNN	1	GSETPGSPAGSPTSTEEGTSESATPESGPGTSTEPS EGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPS EGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPS EGSAPGTSESATPESGPGSEPATSGSETPGTSTEPS EGSAPGTSTEPSEGSAPGTSESATPESGPGTSESAT PESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSG GSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPS EGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPS EGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPS EGSAPGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGTSTEPS EGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPS EGSAPGTSESATPESGPGSE PATSGSTE TGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGTSTEPS EGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSG	576
8019	C-terminal ELNN	1	EGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPS EGSAPGTSESATPESGPGSEPATSGSETPGTSTEPS EGSAPGTSTEPSEGSAPGTSESATPESGPGTSESAT PESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSG GSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPS EGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPS EGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPS EGSAPGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGTSTEPS EGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPS EGSAPGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGSEPATSGSETPGTSE SASPE SGPGTSTEPS EGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSG GSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGTSTEPSEGSAPGTSESATPESGPGTSESAT	576
8020	N-terminal ELNN	1	ASSPAGSPTSTESGTSESATPESGPGTSTEPSEGS PGTSESATPESGPGSEPATSGSETPGTSESATPE SG PGSTPA ESGSETPGTSESATPESGPGTSTEPSEGS PGSPAGSPTSTEEGTSESATPESGPGSEPATSGSET PGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEE EGTSTEPSEGSAPGTSESATPESGPGTSESATPESG PGTSESATPESGPGSEPATSGSETPGSEPATSGSET PGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGS PGGSAP	294
8021	N-terminal ELNN	1	ASSATPESGPGTSTEPSEGSAPGTSESATPE SGPGS GPSTSE SATPGTSESATPESGPGSEPATSGSETPGT SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT SESATPESGPGSEPATSGSETPGTSESATPESGPGS PAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGT SESATPESGPGTSESATPESGPGTSESATPESGPGS EPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGT STEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGT SESATP	294
8022	C-terminal ELNN	1	ATPESGPGTSESATPESGPGSPAGSPTSTEEGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGTSTE	582

SEQ ID NO:	ELNN Type	# of Barcode(s)	Amino Acid Sequence	Total # of AAs
			SPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSES ATPESGPGTSESATPESGPGTSASATPESGPGSEPA TSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTE PSEGSAPGTSTEPSEGSAPGSEPATSGSETPGPEPA	
8029	N-terminal ELNN		GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP GSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAP GTSESATPESGPGSEPATSGSETPGSEPATSGSETP GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAP GTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAP GTSESATPESGPGSEPATSGSETPGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGTSESATPESGP	288
8030	C-terminal ELNN		GTSESATPESGPGSEPATSGSETPGTSESATPESGP GSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP GSPAGSPTSTEEGTSESATPESGPGSEPATSGSETP GTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEE GTSTEPSEGSAPGTSESATPESGPGTSESATPESGP GTSESATPESGPGSEPATSGSETPGSEPATSGSETP GSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAP GSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP	288
8031	N-terminal ELNN		ASSATPESGPGTSTEPSEGSAPGTSESATPESGPGS GPGTSESATPGTSESATPESGPGSEPATSGSETPGT SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT SESATPESGPGSEPATSGSETPGTSESATPESGPGS PAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGT SESATPESGPGTSESATPESGPGTSESATP	210
8032	C-terminal ELNN		ATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGSPAG SPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSES ATPESGPGTSESATPESGPGTSASATPESGPGSEPA TSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTE PSEGSAPGTSTEPSEGSAPGSEPATGEPEA	210
8033	N-terminal ELNN		ASSATPESGPGTSTEPSEGSAPGTSESATPESGPGS GPATSESATPGTSESATPESGPGSEPATSGSETPGT SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT SESATPESGPGSEPATSGSETPGTSESATPESGPGS PAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGT SESATPESGPGTSESATPESGPGTSESATP	210
8034	C-terminal ELNN		APGTSESATPESGPGTSTEPSEGSAPGTSESATPES GPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGS APGTSESATPESGPGPSESATPESGPGSPAGSPTST EEGTSESATPESGPGSEPATSGSETPGTSESATPES GPGT	148
8035	C-terminal ELNN		ATPESGPGSEPATSGSETPGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAG SPTSTEEGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTE PSEGSAPGTSESATPESGPGTSESATPESGPGTSPS ATPESGPGSEPATSGSETPGSEPATSGSET	210

[0212] In some embodiments, a barcoded ELNN can be obtained by making one or more mutations to existing ELNN, such as any listed in Table 3b, according to one or more of the following criteria: to minimize the sequence change in the ELNN, to minimize the amino acid composition change in the ELNN, to substantially maintain the net charge of the ELNN, to substantially maintain (or improve) low immunogenicity of the ELNN, and to substantially maintain (or improve) the pharmacokinetic

properties of the ELNN. In some embodiments, the ELNN sequence has at least 90%, at least 92%, at least 95%, at least 98%, at least 99%, or 100% sequence identity to any one of SEQ ID NOs: 601-659 listed in Table 3b. In some embodiments, the ELNN sequence, having at least 90% (e.g., at least 92%, at least 95%, at least 98%, or at least 99%) but less than 100% sequence identity to any of SEQ ID NOs: 601-659 listed in Table 3b, is obtained by one or more mutations (e.g., less than 10, less than 8, less than 6, less than 5, less than 4, less than 3, less than 2 mutations) of the corresponding sequence from Table 3b. In some embodiments, the one or more mutations comprise deletion of a glutamic acid residue, insertion of a glutamic acid residue, substitution of a glutamic acid residue, or substitution for a glutamic acid residue, or any combination thereof. In some embodiments, where the ELNN sequence differs from, but has at least 90% (e.g., at least 92%, at least 95%, at least 98%, or at least 99%) sequence identity to, any one of SEQ ID NOs: 601-659 listed in Table 3b, at least 80%, at least 90%, at least 95%, at least 97%, or about 100% of the difference between the ELNN sequence and the corresponding sequence of Table 3b involve deletion of a glutamic acid residue, insertion of a glutamic acid residue, substitution of a glutamic acid residue, or substitution for a glutamic acid residue, or any combination thereof. In some such embodiments, at least 80%, at least 90%, at least 95%, at least 97%, or about 100% of the difference between the ELNN sequence and the corresponding sequence of Table 3b involve a substitution of a glutamic acid residue, or a substitution for a glutamic acid residue, or both.

[0213] The “a substitution of a first amino acid,” as used herein, refers to replacement of the first amino acid residue with a second amino acid residue, resulting in the second amino acid residue taking its place at the substitution position in the obtained sequence. For example, “a substitution of glutamic acid” refers to replacement of the glutamic acid (E) residue for a non-glutamic acid residue (e.g., serine (S)).

Table 3b. Exemplary Existing ELNNs for Engineering into Barcoded ELNN(s)

ELNN Name	Amino Acid Sequence	SEQ ID NO:
AE144	GSEPATSGSETPGTSESATPESGPGSEPATSGSETPGSPAGSPTSTEEGTSTEPS EGSAPGSEPATSGSETPGSEPATSGSETPGSEPATSGSETPGTSTEPSEGSAPGT SESATPESGPGSEPATSGSETPGTSTEPSEGSAP	601
AE144_1A	SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPG	602
AE144_2A	TSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATP ESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTS TEPSEGSAPGTSESATPESGPGTSESATPESGPG	603
AE144_2B	TSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATP ESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTS TEPSEGSAPGTSESATPESGPGTSESATPESGPG	604
AE144_3A	SPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSE GSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTS TEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG	605

ELNN Name	Amino Acid Sequence	SEQ ID NO:
	TEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTST EEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGS PTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGS EPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE SAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP	
AE865	GGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPG SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPT STEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGT ESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPES GPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAP GTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGT STEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPT TEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTE EGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGS PTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPG SEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE GSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP	616
AE866	PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPG SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPT STEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGT ESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPES GPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAP GTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGT STEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPT TEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTE EGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGS PTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPG SEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE GSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG	617
AE1152	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPS EGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGS PAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPT TEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSE SATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESG PGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATP ESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGT TEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTST EEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGS PTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPG EPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE SAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSE PATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGS PGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGS PTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPG	618

ELNN Name	Amino Acid Sequence	SEQ ID NO:
	TSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSE GSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP	
AE144A	STEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSE SATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGS	619
AE144B	SEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSE GSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSP AGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPG	620
AE180A	TSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGS PAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPE SGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTST EPSEGSAPGSEPATSG	621
AE216A	PESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGT SESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPE SGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESAT	622
AE252A	ESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT ESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTST EEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPA TSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAP GSEPATSGSETPGTSESATPESGPGTSTEPSE	623
AE288A	TPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPG SEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATP ESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGT STEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSE TPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPA TSGSETPGTSESA	624
AE324A	PESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGT STEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSE ATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGS PGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPA TSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSG	625
AE360A	PESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGS PAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPA GSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTST EEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESA TPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAP GTSTEPSEGSAPGSEPATSGSETPGTSESAT	626
AE396A	PESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGS PAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSE SAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSET PGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESA TPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPG SPAGSPTSTEEGTSTEPSEGSAPGTSTEP	627
AE432A	EGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGT SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPE SGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTST EPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSET PGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAP TSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSG SETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSG	628
AE468A	EGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGT SESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPE SGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPA TSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGT	629

ELNN Name	Amino Acid Sequence	SEQ ID NO:
	GSPTSTEEGTSESATPESGPGPTSTPESEGSAPGTSESATPESGPGSEPATSGSET PGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGS PTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPSTSTEEG SPAGSPSTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATP ESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPSTSTEEGTSTEPSEGSAPGTS TEPSEGSAPGSEPATSGSETPGTSESAT	
AE504A	EGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGS PAGSPSTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPE SGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPA GSPSTSTEEGSPAGSPSTSTEEGSPAGSPSTSTEEGTSESATPESGPGTSTEPSEGA PGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGTSTEPSEGSAPGSPAGSPSTSTEEGTSESATPESGPGSEPATSGSETPG TSESATPESGPGSPAGSPSTSTEEGSPAGSPSTSTEEGTSTEPSEGSAPGTSESATP ESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSP AGSPSTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPES GPGTSTEPS	630
AE540A	TPESGPGSPAGSPSTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPG TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSE GSPGTSTEPSEGSAPGSPAGSPSTSTEEGTSTEPSEGSAPGTSESATPESGPGSE PATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGS APGTSESATPESGPGSPAGSPSTSTEEGSPAGSPSTSTEEGSPAGSPSTSTEEGTSES ATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGP GSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPSTSTEEGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSPAGSPSTSTEEGSPAGSPSTSTEEGT STEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGS ETPGSEPATSGSETPGSPAGSPSTSTEEGTSTEPSEGSAPGTSTEP	631
AE576A	TPESGPGTSESATPESGPGSPAGSPSTSTEEGTSESATPESGPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSE GSPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPSTSTEEGTSTEPSEGSAPGTS ESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPES PGTSTEPSEGSAPGTSESATPESGPGSPAGSPSTSTEEGSPAGSPSTSTEEGSPAG SPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSP TSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPSTSTEEGS PAGSPSTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSTEPSE SGPGSEPATSGSETPGSEPATSGSETPGSPAGSPSTSTEEGTSTEPSEGSAPGTST EPSEGSAPGSEPATSGSETPGTSESA	632
AE612A	GSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGS PAGSPSTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSE GAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGSPAGSPSTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSET PGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESA TPESGPGSPAGSPSTSTEEGSPAGSPSTSTEEGSPAGSPSTSTEEGTSESATPESGP TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSG SETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPSTSTEEGTSESATPESGPGSE PATSGSETPGTSESATPESGPGSPAGSPSTSTEEGSPAGSPSTSTEEGTSTEPSEGS APGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPA TSGSETPGSPAGSPSTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETP GTSESAT	633
AE648A	PESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGT STEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGSPSTSTEEGTSESATPE SGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPSTSE EGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPAT SGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSPAGSPSTSTEEG SPAGSPSTSTEEGSPAGSPSTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATP ESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTS TEPSEGSAPGSPAGSPSTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPES GPGSPAGSPSTSTEEGSPAGSPSTSTEEGTSTEPSEGSAPGTSESATPESGPGTSES ES	634

ELNN Name	Amino Acid Sequence	SEQ ID NO:
AE828A	PESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGT SESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSESATPE SGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTST EPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSET PGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSPAGS PTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSE GSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTS ESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATPES GPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTST PSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSG GSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGT SESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGS ETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSE SAT	639
AE869	GSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTST EPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSET PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGS PTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPG TSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATP ESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTS ESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGS APGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSG SETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGS PTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGT SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTST EEGTSSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPA GSPPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESG PGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEP SEGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGR	640
AE144_R1	SAGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT STEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGS ETPGSPAGSPTSTEEGTSESATPESGPGTESASR	641
AE288_R1	SAGSPGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESG PGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPG TSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPT STEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESGPGTS TEPSEGSAPASR	642
AE432_R1	SAGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT STEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGS ETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSA PGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESA TPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSE GSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSESASR	643
AE576_R1	SAGSPTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSA PGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEP SEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPG TSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGTSESATP ESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGT TEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAP GTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSG	644

ELNN Name	Amino Acid Sequence	SEQ ID NO:
	GSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGT SESATPESGPGTSTEPSEGSAPSASR	
AE864_R1	SAGSPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT STEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGS ETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGS PGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESA TPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSE GSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGT ESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAG SPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSP TSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEG PAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPE SGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTST EPSEGSAPGSEPATSGSETPGTSESATPESGPGTESASR	645
AE712	PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPG SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPT STEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGT ESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPES GPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAP GTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGT STEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPT TEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTE EGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEAHHH	646
AE864_R2	GSFAGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT STEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGS ETPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGS PGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESA TPESGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPG TSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSE GSAPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGT ESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAG SPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETP GTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSP TSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEG PAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPE SGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTST EPSEGSAPGSEPATSGSETPGTSESATPESGPGTESASR	647
AE288_3	SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSG SETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSP AGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTST EEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSES ATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAP GTSTEPSEGSAPG	648
AE284	GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESAT PESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGT SESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPE SGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPA GSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPESG PGTSTEPSE	649

ELNN Name	Amino Acid Sequence	SEQ ID NO:
AE292	SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSG SETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSP AGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTST EEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSES ATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAP GTSTEPSEGSAPGGSAP	650
AE864_2	AGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGS APGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAG SPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESAT PESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGT SESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPE SGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTST EPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESG PGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEP SEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEG TSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATP ESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT ESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTST EEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPA TSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAP GSEPATSGSETPGTSESATPESGPGTSTEPSEGAEEPEA	651
AE867	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP EGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPG PAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPT TEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSE SATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESG PGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPG TSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATP ESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGT TEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTST EEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEE GTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSP TSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPG EPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE SAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGAEEPEA	652
AE867_2	SPGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTST PSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETP GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGT SESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPE SGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGS APGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPG TSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSP TSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGT ESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTST EEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAG SPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGP GSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEP EGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG	653
AE868	PGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEP SEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPG SPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSP TSTEEGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGT SESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPE SGPGTSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGSAPGTSTEPSEGS APGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPG TSTEPSEGSAPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSP TSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGT ESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTST EEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAG SPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGP GSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEP EGSAPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPG	654

ELNN Name	Amino Acid Sequence	SEQ ID NO:
	GPGTSESATPESGPGSPAGSPSTSTEEGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGTSTPESEGSAPGTSTPESEGSAPGTSTPESEGSAPGTSTPESEGSAP GTSTPESEGSAPGTSTPESEGSAPGSPAGSPSTSTEEGTSTPESEGSAPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGT STPESEGSAPGTSESATPESGPGSPAGSPSTSTEEGSPAGSPSTSTEEGSPAGSPST TEEGTSESATPESGPGTSTPESEGSAPGTSESATPESGPGSEPATSGSETPGTSE SATPESGPGSEPATSGSETPGTSESATPESGPGTSTPESEGSAPGSPAGSPSTST EGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPSTSTEEGSPAGS PTSTEEGTSTPESEGSAPGTSESATPESGPGTSTPESEGSAPGSPAGSPSTST SEPATSGSETPGSEPATSGSETPGSPAGSPSTSTEEGTSTPESEGSAPGTSTPESE GSAPGSEPATSGSETPGTSESATPESGPGTSTPESEGAEEPEA	
AE144_7A	GSPAGSPSTSTEEGTSESATPESGPGTSTPESEGSAPGSPAGSPSTSTEEGTSTPE EGSAPGTSTPESEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGS PAGSPSTSTEEGTSESATPESGPGTSTPESEGSAP	655
AE292	SPAGSPSTSTEEGTSESATPESGPGTSTPESEGSAPGTSESATPESGPGSEPATSG SETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTPESEGSAPGSP AGSPSTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPSTST EEGSPAGSPSTSTEEGTSTPESEGSAPGTSESATPESGPGTSTPESEGSAPGSP ATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPSTSTEEGTSTPESEGSAP GTSTPESEGSAPGGSAP	656
AE293	PGSPAGSPSTSTEEGTSESATPESGPGTSTPESEGSAPGSPAGSPSTSTEEGTSTPE SEGSAPGTSTPESEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPG SPAGSPSTSTEEGTSESATPESGPGTSTPESEGSAPGTSTPESEGSAPGSPAGSP STEEGTSTPESEGSAPGTSTPESEGSAPGTSESATPESGPGTSTPESEGSAPGT ESATPESGPGSEPATSGSETPGTSTPESEGSAPGTSTPESEGSAPGTSESATPES GPGTSESATPEGAEEPEA	657
AE300	PGSPAGSPSTSTEEGTSESATPESGPGTSTPESEGSAPGSPAGSPSTSTEEGTSTPE SEGSAPGTSTPESEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPG SPAGSPSTSTEEGTSESATPESGPGTSTPESEGSAPGTSTPESEGSAPGSPAGSP STEEGTSTPESEGSAPGTSTPESEGSAPGTSESATPESGPGTSTPESEGSAPGT ESATPESGPGSEPATSGSETPGTSTPESEGSAPGTSTPESEGSAPGTSESATPES GPGTSESATPESGPGSPAGAAEPEA	658
AE584	PGSPAGSPSTSTEEGTSESATPESGPGTSTPESEGSAPGSPAGSPSTSTEEGTSTPE SEGSAPGTSTPESEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPG SPAGSPSTSTEEGTSESATPESGPGTSTPESEGSAPGTSTPESEGSAPGSPAGSP STEEGTSTPESEGSAPGTSTPESEGSAPGTSESATPESGPGTSTPESEGSAPGT ESATPESGPGSEPATSGSETPGTSTPESEGSAPGTSTPESEGSAPGTSESATPES GPGTSESATPESGPGSPAGSPSTSTEEGTSESATPESGPGSEPATSGSETPGTSES ATPESGPGTSTPESEGSAPGTSTPESEGSAPGTSTPESEGSAPGTSTPESEGSAP GTSTPESEGSAPGTSTPESEGSAPGSPAGSPSTSTEEGTSTPESEGSAPGTSESAT PESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGT STPESEGSAPGTSESATPESGPGSPAGSPSTSTEEGSPAGSPSTSTEEGSPAGSP TSTEEGTSESATPESGPGTSTPESEGSAPGAAEPEA	659

[0214] In some embodiments, for constructing the sequence of a barcoded ELNN, amino-acid mutations are performed on ELNN of intermediate lengths to those of Table 3b, as well as ELNN of longer lengths than those of Table 3b, such as those in which one or more 12-mer motifs of Table 1b are added to the N- or C- terminus of a general-purpose ELNN of Table 3b.

[0215] Additional examples of existing ELNNs that can be used according to the present disclosure are disclosed in U.S. Patent Publication Nos. 2010/0239554, 2010/0323956, 2011/0046060, 2011/0046061, 2011/0077199, or 2011/0172146, or International Patent Publication Nos. WO 2010091122, WO 2010144502, WO 2010144508, WO 2011028228, WO 2011028229, WO

2011028344, WO 2014/011819, WO 2015/023891, WO 2017/040344, WO 2019/126576, WO 2020/264200; WO 2020/264208; WO 2021/097186, WO 2021/262985, or WO 2021/263058.

[0216] In some embodiments, a barcoded ELNN fused within a polypeptide chain adjacent to the N-terminus of the polypeptide chain ("N-terminal ELNN") can be attached to a His tag of HHHHHH (SEQ ID NO: 48) or HHHHHHHH (SEQ ID NO: 49) at the N-terminus to facilitate the purification of the fusion polypeptide. In some embodiments, a barcoded ELNN fused within a polypeptide chain at the C-terminus of the polypeptide chain ("C-terminal ELNN") can be comprise or be attached to the sequence EPEA at the C-terminus to facilitate the purification of the fusion polypeptide. In some embodiments, the fusion polypeptide comprises both an N-terminal barcoded ELNN and a C-terminal barcoded ELNN, wherein the N-terminal barcoded ELNN is attached to a His tag of HHHHHH (SEQ ID NO: 48) or HHHHHHHH (SEQ ID NO: 49) at the N-terminus; and wherein the C-terminal barcoded ELNN is attached to the sequence EPEA at the C-terminus, thereby facilitating purification of the fusion polypeptide, for example, to at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or at least 99% purity by chromatography methods known in the art, including but not limited to IMAC chromatography, C-tagXL affinity matrix, and other such methods.

[0217] A barcode fragment, as described herein, can be cleavably fused within the ELNN and releasable (i.e., configured to be released) from the ELNN upon digestion of the polypeptide by a protease. In some embodiments, the protease is a Glu-C protease. In some embodiments, the protease cleaves on the C-terminal side of glutamic acid residues that are not followed by proline. In some embodiments, a barcoded ELNN (an ELNN that contains barcode fragment(s) therewithin) is designed to achieve high efficiency, precision and accuracy of the protease digestion. For example, in some embodiments, adjacent Glu-Glu (EE) residues in an ELNN sequence can result in varying cleavage patterns upon Glu-C digestion. Accordingly, when Glu-C protease is used for barcode release, the barcoded ELNN or the barcode fragment(s) may not contain any Glu-Glu (EE) sequence. Additionally, a di-peptide Glu-Pro (EP) sequence, if present in the fusion polypeptide, may not be cleaved by Glu-C protease during the barcode release process.

ACTIVATABLE CYTOKINE FUSION PROTEINS

[0218] In one aspect, the disclosure provides a fusion protein comprising: a cytokine (e.g., an interleukin, a transforming growth factor, an interferon, a tumor necrosis factor, a chemokine, or granulocyte macrophage-colony stimulating factor), a linker (Linker1) comprising a protease-cleavable release segment (RS1), and a mask polypeptide (Mask1), wherein the linker (Linker1) comprising the release segment (RS1) is positioned between the cytokine and the mask polypeptide (Mask1); wherein the release segment is capable of being cleaved by at least one protease that is present in a tumor; and wherein the release segment (RS1) is not capable of being cleaved by legumain in human plasma,.

[0219] In another aspect, the disclosure provides a fusion protein comprising: a cytokine (e.g., an interleukin, a transforming growth factor, an interferon, a tumor necrosis factor, a chemokine, or granulocyte macrophage-colony stimulating factor), a linker (Linker1) comprising a protease-cleavable release segment (RS1), and a mask polypeptide (Mask1), wherein the linker (Linker1) comprising the release segment (RS1) is positioned between the cytokine and the mask polypeptide (Mask1); wherein the release segment is capable of being cleaved by at least one protease that is present in a tumor; and wherein the release segment (RS1) is cleaved by legumain in human plasma at a rate that is less than 25% of the rate that EAGRSANHTPAGLTGP (RSR-2295; SEQ ID NO:7048) is cleaved by legumain in human plasma.

[0220] In yet another aspect, the disclosure provides a fusion protein comprising: a cytokine (e.g., an interleukin, a transforming growth factor, an interferon, a tumor necrosis factor, a chemokine, or granulocyte macrophage-colony stimulating factor), a linker (Linker1) comprising a protease-cleavable release segment (RS1), and a mask polypeptide (Mask1), wherein the linker (Linker1) comprising the release segment (RS1) is positioned between the cytokine and the mask polypeptide (Mask1); wherein the release segment is capable of being cleaved by at least one protease that is present in a tumor; and wherein the release segment (RS1) comprises an amino acid sequence comprising the sequence EAGRSAXHTPAGLTGP, wherein X is any amino acid other than N (SEQ ID NO:7627).

[0221] In some embodiments, the linker is attached to the N-terminus of the cytokine. In some embodiments, the linker is attached to the C-terminus of the cytokine.

[0222] In another aspect, the disclosure provides a fusion protein comprising the following elements in an N-to-C or C-to-N terminal orientation: a first mask polypeptide (Mask1), a first linker (Linker1) comprising a first protease-cleavable release segment (RS1), a cytokine (e.g., an interleukin, a transforming growth factor, an interferon, a tumor necrosis factor, a chemokine, or granulocyte macrophage-colony stimulating factor), a second linker (Linker2) comprising a second protease-cleavable release segment (RS2), and a second mask polypeptide (Mask2), wherein each release segment (RS1 and RS2) (i) is not capable of being cleaved by legumain in human plasma, (ii) is cleaved by legumain in human plasma at a rate that is less than 25% of the rate that RSR-2295 (EAGRSANHTPAGLTGP; SEQ ID NO:7048) is cleaved by legumain in human plasma, or (iii) comprises an amino acid sequence comprising the sequence EAGRSAXHTPAGLTGP, wherein X is any amino acid other than N (SEQ ID NO:7627).

[0223] Cytokines that may be incorporated in the fusion proteins of the disclosure may be composed of more than one subunit polypeptide. For example, IL-12 is composed of an IL-12 p35 polypeptide subunit and an IL-12 p40 polypeptide subunit that come together to form IL-12. Similarly, IL-23 is composed of an IL-23 p19 polypeptide subunit and an IL-12 p40 polypeptide subunit that come

together to form IL-23. Thus, in some embodiments, the fusion protein of the disclosure comprises a first subunit polypeptide comprising a first subunit of a cytokine, and a second subunit polypeptide comprising a second subunit of the cytokine.

[0224] In one aspect, the disclosure provides a fusion protein comprising: a. a first polypeptide subunit comprising the following elements in an N-to-C or C-to-N terminal orientation: i. a first mask polypeptide (Mask1), ii. a first linker (Linker1) comprising a first protease-cleavable release segment (RS1), iii. a first cytokine subunit polypeptide, iv. a second linker (Linker2) comprising a second protease-cleavable release segment (RS2), and v. a second mask polypeptide (Mask2); and b. a second polypeptide subunit comprising the following elements in an N-to-C or C-to-N terminal orientation: i. a third mask polypeptide (Mask3), ii. a third linker (Linker3) comprising a third protease-cleavable release segment (RS3), iii. a second cytokine subunit polypeptide, iv. a fourth linker (Linker4) comprising a fourth protease-cleavable release segment (RS4), and v. a fourth mask polypeptide (Mask4); wherein the first cytokine subunit polypeptide of the first polypeptide subunit and the second cytokine subunit polypeptide of the second polypeptide subunit are linked thereby forming a linked heterodimer; wherein each release segment (RS1, RS2, RS3, and RS4) is capable of being cleaved by at least one protease that is present in a tumor; and wherein each release segment (RS1, RS2, RS3, and RS4) (i) is not capable of being cleaved by legumain in human plasma, (ii) is cleaved by legumain in human plasma at a rate that is less than 25% of the rate that RSR-2295 (EAGRSANHTPAGLTGP; SEQ ID NO:7048) is cleaved by legumain in human plasma, or (iii) comprises an amino acid sequence comprising the sequence EAGRSAXHTPAGLTGP, wherein X is any amino acid other than N (SEQ ID NO:7627).

[0225] In another aspect, the disclosure provides a fusion protein comprising:

- a. a first polypeptide subunit comprising the following elements in an N-to-C or C-to-N terminal orientation:
 - i. a first mask polypeptide (Mask1),
 - ii. a first linker (Linker1) comprising a first protease-cleavable release segment (RS1),
 - iii. an IL-12 p35 polypeptide,
 - iv. a second linker (Linker2) comprising a second protease-cleavable release segment (RS2), and
 - v. a second mask polypeptide (Mask2); and
- b. a second polypeptide subunit comprising the following elements in an N-to-C or C-to-N terminal orientation:
 - i. a third mask polypeptide (Mask3),
 - ii. a third linker (Linker3) comprising a third protease-cleavable release segment (RS3),

- iii. an IL-12 p40 polypeptide,
- iv. a fourth linker (Linker4) comprising a fourth protease-cleavable release segment (RS4), and
- v. a fourth mask polypeptide (Mask4);

wherein the IL-12 p35 polypeptide of the first polypeptide subunit and the IL-12 p40 polypeptide of the second polypeptide subunit are disulfide-linked thereby forming a disulfide-linked heterodimer;

wherein each release segment (RS1, RS2, RS3, and RS4) is capable of being cleaved by at least one protease that is present in a tumor;

and wherein each release segment (RS1, RS2, RS3, and RS4) (i) is not capable of being cleaved by legumain in human plasma, (ii) is cleaved by legumain in human plasma at a rate that is less than 25% of the rate that RSR-2295 (EAGRSANHTPAGLTGP; SEQ ID NO:7048) is cleaved by legumain in human plasma, or (iii) comprises an amino acid sequence comprising the sequence EAGRSAXHTPAGLTGP, wherein X is any amino acid other than N (SEQ ID NO:7627).

[0226] In some embodiments, the first mask polypeptide (Mask1) comprises a first barcode fragment (Barcode1), the second mask polypeptide (Mask2) comprises a second barcode fragment (Barcode2), the third mask polypeptide (Mask3) comprises a third barcode fragment (Barcode3), and the fourth mask polypeptide (Mask4) comprises a fourth barcode fragment (Barcode4); and wherein each barcode fragment (Barcode1, Barcode2, Barcode3, and Barcode4) is releasable from the fusion protein upon digestion with a non-mammalian protease, and each barcode fragment differs in both sequence and molecular weight from all other barcode fragments or peptide fragments that are releasable from the fusion protein upon complete digestion of the fusion protein by the non-mammalian protease.

[0227] In another aspect, the disclosure provides a fusion protein comprising:

- a. a first polypeptide subunit comprising the following elements in an N-to-C or C-to-N terminal orientation:
 - i. a first mask polypeptide (Mask1) comprising a first barcode fragment (Barcode1),
 - ii. a first linker (Linker1) comprising a first protease-cleavable release segment (RS1),
 - iii. an IL-12 p35 polypeptide,
 - iv. a second linker (Linker2) comprising a second protease-cleavable release segment (RS2), and
 - v. a second mask polypeptide (Mask2) comprising a second barcode fragment (Barcode2); and

- b. a second polypeptide subunit comprising the following elements in an N-to-C or C-to-N terminal orientation:
 - i. a third mask polypeptide (Mask3) comprising a third barcode fragment (Barcode3),
 - ii. a third linker (Linker3) comprising a third protease-cleavable release segment (RS3),
 - iii. an IL-12 p40 polypeptide,
 - iv. a fourth linker (Linker4) comprising a fourth protease-cleavable release segment (RS4), and
 - v. a fourth mask polypeptide (Mask4) comprising a fourth barcode fragment (Barcode4);

wherein the IL-12 p35 polypeptide of the first subunit and the IL-12 p40 polypeptide of the second unit are disulfide-linked thereby forming a disulfide-linked heterodimer;

wherein each release segment (RS1, RS2, RS3, and RS4) is capable of being cleaved by at least one protease that is present in a tumor;

and wherein each barcode fragment (Barcode1, Barcode2, Barcode3, and Barcode4) is releasable from the fusion protein upon digestion with a non-mammalian protease, and each barcode fragment differs in both sequence and molecular weight from all other barcode fragments or peptide fragments that are releasable from the fusion protein upon complete digestion of the fusion protein by the non-mammalian protease.

[0228] In some embodiments, each release segment comprises an amino acid sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity, to a sequence of Table 6a.

[0229] In some embodiments, each release segment (RS1, RS2, RS3, and RS4) (i) is not capable of being cleaved by legumain in human plasma, (ii) is cleaved by legumain in human plasma at a rate that is less than 25% of the rate that RSR-2295 (EAGRSANHTPAGLTGP; SEQ ID NO:7048) is cleaved by legumain in human plasma, or (iii) comprises an amino acid sequence comprising the sequence EAGRSAXHTPAGLTGP, wherein X is any amino acid other than N (SEQ ID NO:7627).

[0230] It will be appreciated that various amino acid substitutions (especially conservative amino acid substitutions) can be made in a fusion protein sequence to create variants without departing from the spirit of the present disclosure with respect to the biological activity or pharmacologic properties of, e.g., a protease-activable cytokine fusion protein. Examples of conservative substitutions for amino acids in polypeptide sequences are shown in Table 4a. In addition, variants can also include, for instance, polypeptides wherein one or more amino acid residues are added or deleted at the N- or C-terminus of the full-length native amino acid sequence of a fusion protein that retains at least a portion of the biological activity of the native peptide.

Table 4a: Exemplary conservative amino acid substitutions

Original Residue	Exemplary Substitutions
Ala (A)	Val; Leu; Ile
Arg (R)	Lys; Gln; Asn
Asn (N)	Gln; His; Lys; Arg
Asp (D)	Glu
Cys (C)	Ser
Gln (Q)	Asn
Glu (E)	Asp
Gly (G)	Pro
His (H)	Asn; Gln; Lys; Arg
Ile (I)	Leu; Val; Met; Ala; Phe; Norleucine
Leu (L)	Norleucine; Ile; Val; Met; Ala; Phe
Lys (K)	Arg; Gln; Asn
Met (M)	Leu; Phe; Ile
Phe (F)	Leu; Val; Ile; Ala
Pro (P)	Gly
Ser (S)	Thr
Thr (T)	Ser
Trp (W)	Tyr
Tyr (Y)	Trp; Phe; Thr; Ser
Val (V)	Ile; Leu; Met; Phe; Ala; Norleucine

SPACERS & CYTOKINE RELEASE SEGMENTS

[0231] Included herein are fusion proteins comprising a cytokine that either becomes biologically active or has an increase in biological activity upon release from an ELNN by cleavage of an optional cleavage sequence incorporated within optional spacer sequences into the fusion protein, e.g., as described herein.

[0232] In some embodiments, the spacer may be provided to enhance expression of the fusion protein from a host cell and/or to decrease steric hindrance such that the cytokine may assume its desired tertiary structure and/or interact appropriately with its target molecule. For spacers and methods of identifying desirable spacers, see, for example, George, *et al.* (2003) Protein Engineering 15:871-879, specifically incorporated by reference herein. In some embodiments, the spacer comprises one or more peptide sequences that are between 1 to 50 amino acid residues in length, or about 1 to 25 residues, or about 1 to 10 residues in length. Spacer sequences, exclusive of cleavage sites, can comprise any of the 20 natural L amino acids, and will preferably comprise hydrophilic amino acids that are sterically unhindered that can include, but not be limited to, glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) or proline (P). In some embodiments, the spacer can be a polyglycine or polyalanine, or predominately a mixture of combinations of glycine and alanine residues. In some embodiments, the spacer polypeptide exclusive of a cleavage sequence is substantially devoid of secondary structure. In some embodiments, one or both spacer sequences in a fusion protein composition may each further contain a cleavage sequence, which may be identical or may be different, wherein the cleavage sequence may be acted on by a protease to release the cytokine from the fusion protein.

Table 4b: Exemplary Spacers between a Release Segment and a Cytokine

Amino Acid Sequence	SEQ ID NO:
STEPS	89
SATPESGPGT	90
ATSGSETPGT	91
GTAEAAASASG	92
STEPSEGSAPGTS	93
SGPGTS	94
GTSTEPS	95
GTSESATPES	96
GTATPESGPG	97
TSESATPES	98
GTPESATPES	99
GTATPESAP	100

[0233] In some embodiments of the polypeptides of this disclosure, a release segment (RS) (e.g., a first release segment (RS1), a second release segment (RS2), a third release segment (RS3), a fourth release segment (RS4), etc.) can be fused to a cytokine by a spacer. In some embodiments, a spacer can (each independently) comprise at least 4 types of amino acids that are glycine (G), alanine (A), serine (S), threonine (T), glutamate (E) or proline (P). In some embodiments, the fusion protein of this disclosure can comprise a first release segment fused to the cytokine via a first spacer and a second release segment fused to the cytokine via a second spacer. In some embodiments, a spacer (e.g., a first spacer, a second spacer, etc.) can (each independently) comprise an amino acid sequence having at least (about) 80%, at least (about) 90%, or 100% sequence identity to a sequence set forth in Table C. In some embodiments, the spacer (e.g., the first spacer, the second spacer, etc.) can (each independently) comprise an amino acid sequence identical to a sequence set forth in Table C.

[0234] In some embodiments, the incorporation of the cleavage sequence into a fusion protein is designed to permit release of a cytokine that becomes active or more active upon its release from one or more ELNNs. In some embodiments, the cleavage sequences are located sufficiently close to the cytokine sequences, generally within 18, or within 12, or within 6, or within 2 amino acids of the cytokine sequence terminus, such that any remaining residues attached to the cytokine after cleavage do not appreciably interfere with the activity (e.g., such as binding to a receptor) of the cytokine yet provide sufficient access to the protease to be able to effect cleavage of the cleavage sequence. In some embodiments, the cleavage site is a sequence that can be cleaved by a protease endogenous to the mammalian subject such that a fusion protein can be cleaved after administration to a subject. In such cases, the fusion protein can serve as a circulating depot for the cytokine. Examples of cleavage sites contemplated herein include, but are not limited to, a polypeptide sequence cleavable by a mammalian endogenous protease listed in Table 5.

[0235] In some embodiments, a fusion protein comprises spacer sequences that comprise one or more cleavage sequences configured to release the cytokine from the fusion protein when acted on by a protease. In some embodiments, a spacer sequence does not comprise a cleavage sequence. In some embodiments, the one or more cleavage sequences can be a sequence having at least about 80% (e.g., at least about 85%, at least about 90%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or 100%) sequence identity to a sequence from Table 6a or 6b.

[0236] In some embodiments, the disclosure provides cytokine release segment polypeptides (or release segments (RSs)) that are substrates for one or more mammalian proteases associated with or produced by disease tissues or cells found in proximity to disease tissues. The one or more mammalian proteases are present in a tumor. The one or more mammalian proteases may be expressed by cancer cells within the tumor or by non-cancerous cells within the tumor. Such proteases can include, but not be limited to the classes of proteases such as metalloproteinases, cysteine proteases, aspartate proteases, and serine proteases, including, but not limited to, the proteases of Table 5. The RSs are useful for, amongst other things, incorporation into the subject recombinant polypeptides, conferring an inactive format that can be activated by the cleavage of the RSs by mammalian proteases. As described herein, the RSs are incorporated into the subject recombinant polypeptide compositions, linking the incorporated cytokines to the ELNN (exemplary configurations of which are described herein) such that upon cleavage of the RSs by action of the one or more proteases for which the RSs are substrates, the cytokines and ELNN are released from the composition and the cytokines, no longer shielded by the ELNN, regain their full potential to bind their ligands.

Table 5: Proteases of Target Tissues

Class of Proteases	Protease
Metalloproteinases	Meprin
	Neprilysin (CD10)
	PSMA
	BMP-1
	A disintegrin and metalloproteinases (ADAMs)
	ADAM8
	ADAM9
	ADAM10
	ADAM12
	ADAM15
	ADAM17 (TACE)
	ADAM19
	ADAM28 (MDC-L)
	ADAM with thrombospondin motifs (ADAMTS)
	ADAMTS1
	ADAMTS4
	ADAMTS5
	Matrix Metalloproteinases (MMPs)
	MMP-1 (Collagenase 1)
	MMP-2 (Gelatinase A)
	MMP-3 (m1)

Class of Proteases	Protease
	MMP-7 (Matrilysin 1)
	MMP-8 (Collagenase 2)
	MMP-9 (Gelatinase B)
	MMP-10 (Stromelysin 2)
	MMP-11 (Stromelysin 3)
	MMP-12 (Macrophage elastase)
	MMP-13 (Collagenase 3)
	MMP-14 (MT1-MMP)
	MMP-15 (MT2-MMP)
	MMP-19
	MMP-23 (CA-MMP)
	MMP-24 (MT5-MMP)
	MMP-26 (Matrilysin 2)
	MMP-27 (CMMP)
Cysteine Proteases	Legumain
	Cysteine cathepsins
	Cathepsin B
	Cathepsin C
	Cathepsin K
	Cathepsin L
	Cathepsin S
Aspartate Proteases	Cathepsin X
	Cathepsin D
	Cathepsin E
	Secretase
Serine Proteases	Urokinase (uPA)
	Tissue-type plasminogen activator (tPA)
	Plasmin
	Thrombin
	Prostate-specific antigen (PSA, KLK3)
	Human neutrophil elastase (HNE)
	Elastase
	Tryptase
	Type II transmembrane serine proteases (TTSPs)
	DESC1
	Hepsin (HPN)
	Matriptase
	Matriptase-2
	TMPRSS2
	TMPRSS3
	TMPRSS4 (CAP2)
	Fibroblast Activation Protein (FAP)
	kallikrein-related peptidase (KLK family)
	KLK4
	KLK5
	KLK6
	KLK7
	KLK8
	KLK10
	KLK11
	KLK13
	KLK14

[0237] In some embodiments, the disclosure provides activatable recombinant polypeptides (e.g., fusion proteins) comprising a first release segment (RS1) sequence having at least 88%, or at least

94%, or 100% sequence identity, when optimally aligned, to a sequence identified in Table 6a, wherein the RS1 is a substrate for one or more mammalian proteases. In some embodiments, the RS is further engineered to remove a legumain cleavage site. In some embodiments, the disclosure provides activatable recombinant polypeptides comprising a RS1 and a second release segment (RS2) sequence, each having at least 88%, or at least 94%, or 100% sequence identity, when optimally aligned, to a sequence identified herein by the sequences set forth in Table 6a, wherein the RS1 and the RS2 each are a substrate for one or more mammalian proteases. In some embodiments, the disclosure provides activatable recombinant polypeptides comprising a RS1, a RS2, a third release segment (RS3) sequence, and a fourth release segment sequence (RS4), each having at least 88%, or at least 94%, or 100% sequence identity, when optimally aligned, to a sequence identified herein by the sequences set forth in Table 6a, wherein the RS1, the RS2, the RS3, and the RS4 each are a substrate for one or more mammalian proteases. In some embodiments, the RS1, the RS2, the RS3, and the RS4 each do not serve as substrates for legumain.

[0238] In some embodiments, disclosure provides activatable recombinant polypeptides comprising a first RS (RS1) sequence having at least 90%, at least 93%, at least 97%, or 100% identity, when optimally aligned, to a sequence identified in Table 6b, wherein the RS1 is a substrate for one or more mammalian proteases. In some embodiments, the disclosure provides activatable recombinant polypeptides comprising a RS1 and a second release segment (RS2) sequence, each having at least 88%, or at least 94%, or 100% sequence identity, when optimally aligned, to a sequence identified herein by the sequences set forth in Table 6b, wherein the RS1 and the RS2 are each a substrate for one or more mammalian proteases (e.g., at one, two, or three cleavage sites within each release segment sequence). In some embodiments, the disclosure provides activatable recombinant polypeptides comprising a RS1, a RS2, a third release segment (RS3) sequence, and a fourth release segment sequence (RS4), each having at least 88%, or at least 94%, or 100% sequence identity, when optimally aligned, to a sequence identified herein by the sequences set forth in Table 6b, wherein the RS1, the RS2, the RS3, and the RS4 are each a substrate for one or more mammalian proteases (e.g., at one, two, or three cleavage sites within each release segment sequence). In some embodiments of activatable recombinant polypeptides comprising the RS1, the RS2, the RS3, and the RS4, the four release segments can be identical. In some embodiments of activatable recombinant polypeptides comprising the RS1, the RS2, the RS3, and the RS4, the four release segments can be different.

[0239] The present disclosure contemplates release segments that are substrates for one, two or three different classes of proteases that are metalloproteinases, cysteine proteases, aspartate proteases, or serine proteases, including the proteases of Table 5. In some embodiments, a fusion protein comprises RSs (e.g., RS1, RS2, RS3, and RS4) that serve as substrates for one or more proteases found in close association with or are co-localized with tumors or cancer cells, and upon cleavage of the RSs, the cytokines that are otherwise shielded by ELNNs of the fusion protein (and thus have a lower binding

affinity for their respective ligands) are released from the ELNNs and regain their full potential to bind target and effector cell ligands. In some embodiments, a fusion protein comprises RSs (e.g., RS1 and RS2), that each comprise an amino acid sequence that is a substrate for one or more cellular proteases located within a targeted cell, including but not limited to a protease of Table 5. In some embodiments, RSs are substrates for two or three classes of proteases that cleave different portions of each RS. In some embodiments, each RS that is a substrate for two, three, or more classes of proteases has two, three, or more distinct cleavage sites, but cleavage by a single protease nevertheless results in the release of the cytokines from an ELNN.

[0240] In some embodiments, an RS of the disclosure for incorporation into a fusion protein is a substrate for one or more proteases including but not limited to meprin, neprilysin (CD10), PSMA, BMP-1, A disintegrin and metalloproteinases (ADAMs), ADAM8, ADAM9, ADAM10, ADAM12, ADAM15, ADAM17 (TACE), ADAM19, ADAM28 (MDC-L), ADAM with thrombospondin motifs (ADAMTS), ADAMTS1, ADAMTS4, ADAMTS5, MMP-1 (collagenase 1), matrix metalloproteinase-1 (MMP-1), matrix metalloproteinase-2 (MMP-2, gelatinase A), matrix metalloproteinase-3 (MMP-3, stromelysin 1), matrix metalloproteinase-7 (MMP-7, Matrilysin 1), matrix metalloproteinase-8 (MMP-8, collagenase 2), matrix metalloproteinase-9 (MMP-9, gelatinase B), matrix metalloproteinase-10 (MMP-10, stromelysin 2), matrix metalloproteinase-11 (MMP-11, stromelysin 3), matrix metalloproteinase-12 (MMP-12, macrophage elastase), matrix metalloproteinase-13 (MMP-13, collagenase 3), matrix metalloproteinase-14 (MMP-14, MT1-MMP), matrix metalloproteinase-15 (MMP-15, MT2-MMP), matrix metalloproteinase-19 (MMP-19), matrix metalloproteinase-23 (MMP-23, CA-MMP), matrix metalloproteinase-24 (MMP-24, MT5-MMP), matrix metalloproteinase-26 (MMP-26, matrilysin 2), matrix metalloproteinase-27 (MMP-27, CMMP), legumain, cathepsin B, cathepsin C, cathepsin K, cathepsin L, cathepsin S, cathepsin X, cathepsin D, cathepsin E, secretase, urokinase (uPA), tissue-type plasminogen activator (tPA), plasmin, thrombin, prostate-specific antigen (PSA, KLK3), human neutrophil elastase (HNE), elastase, tryptase, Type II transmembrane serine proteases (TTSPs), DESC1, hepsin (HPN), matriptase, matriptase-2, TMPRSS2, TMPRSS3, TMPRSS4 (CAP2), fibroblast activation protein (FAP), kallikrein-related peptidase (KLK family), KLK4, KLK5, KLK6, KLK7, KLK8, KLK10, KLK11, KLK13, and KLK14. In some embodiments, the RS is a substrate for ADAM17. In some embodiments, the RS is a substrate for BMP-1. In some embodiments, the RS is a substrate for cathepsin. In some embodiments, the RS is a substrate for HtrA1. In some embodiments, the RS is a substrate for legumain. In some embodiments, the RS is a substrate for MMP-1. In some embodiments, the RS is a substrate for MMP-2. In some embodiments, the RS is a substrate for MMP-7. In some embodiments, the RS is a substrate for MMP-9. In some embodiments, the RS is a substrate for MMP-11. In some embodiments, the RS is a substrate for MMP-14. In some embodiments, the RS is a substrate for uPA. In some embodiments, the RS is a substrate for

matriptase. In some embodiments, the RS is a substrate for MT-SP1. In some embodiments, the RS is a substrate for neutrophil elastase. In some embodiments, the RS is a substrate for thrombin. In some embodiments RS is a substrate for TMPRSS3. In some embodiments, the RS is a substrate for TMPRSS4. In some embodiments, the RS of the subject recombinant polypeptide compositions is a substrate for at least two proteases including but not limited to legumain, MMP-1, MMP-2, MMP-7, MMP-9, MMP-11, MMP-14, uPA, and matriptase. In some embodiments, the RS of the subject recombinant polypeptide compositions is a substrate for legumain, MMP-1, MMP-2, MMP-7, MMP-9, MMP-11, MMP-14, uPA, and matriptase. In specific embodiments, the RS of the subject recombinant polypeptide compositions is not a substrate for legumain. In some embodiments, the RS of the subject recombinant polypeptide compositions is a substrate for uPA, matriptase (also known as MT-SP1 and ST14), MMP2, MMP7, MMP9, and MMP14. In some embodiments, the RS of the subject recombinant polypeptide compositions is substrate for uPA, matriptase, MMP2, MMP7, MMP9, and MMP14 but not legumain.

Table 6a: Fusion Protein Release Segment Sequences.

Name	Amino Acid Sequence	SEQ ID NO:
RSR-1517	EAGRSANHEPLGLVAT	7001
BSRS-A1-1	ASGRSTNAGPSGLAGP	7002
BSRS-A2-1	ASGRSTNAGPQGLAGQ	7003
BSRS-A3-1	ASGRSTNAGPPGLTGP	7004
VP-1	ASSRGTNAGPAGLTGP	7005
RSR-1752	ASSRTTNTGPSTLTGP	7006
RSR-1512	AAGRSDNGTPLELVAP	7007
RSR-1517	EAGRSANHEPLGLVAT	7008
VP-2	ASGRGTNAGPAGLTGP	7009
RSR-1018	LFGRNDNHEPLELGGG	7010
RSR-1053	TAGRSDNLEPLGLVFG	7011
RSR-1059	LDGRSDNFHPPELVAG	7012
RSR-1065	LEGRSDNEEPENLVAG	7013
RSR-1167	LKGRSDNNAPLALVAG	7014
RSR-1201	VYSRGTNAGPHGLTGR	7015
RSR-1218	ANSRGTNKGFAGLIGP	7016
RSR-1226	ASSRLTNEAPAGLTIP	7017
RSR-1254	DQSRGTNAGPEGLTDP	7018
RSR-1256	ESSRGTNIGQGGLTGP	7019
RSR-1261	SSSRGTNQDPAGLTIP	7020
RSR-1293	ASSRGQNHSPMGLTGP	7021
RSR-1309	AYSRGPNAGPAGLEGR	7022
RSR-1326	ASERGNNAGPANLTGF	7023
RSR-1345	ASHRGTNPKPAILTGP	7024
RSR-1354	MSSRRTNANPAQLTGP	7025
RSR-1426	GAGRTDNHEPLELGAA	7026
RSR-1478	LAGRSENTAPLELTAG	7027
RSR-1479	LEGRPDNHEPLALVAS	7028
RSR-1496	LSGRSDNEEPLALPAG	7029
RSR-1508	EAGRTDNHEPLELSAP	7030
RSR-1513	EGGRSDNHGPLELVSG	7031
RSR-1516	LSGRSDNEAPLELEAG	7032
RSR-1524	LGGRADNHEPPELGAG	7033
RSR-1622	PFSRGTNAEPAGLTGE	7034

Name	Amino Acid Sequence	SEQ ID NO:
RSR-1629	ASTRGENAGPAGLEAP	7035
RSR-1664	ESSRGTNGAPEGLTGP	7036
RSR-1667	ASSRATNESPAGLTGE	7037
RSR-1709	ASSRGENPPPGGLTGP	7038
RSR-1712	AASRGTTNTGPAELTGS	7039
RSR-1727	AGSRTTNAGPGGLEGP	7040
RSR-1754	APSRGENAGPATLTGA	7041
RSR-1819	ESGRAANTGPPTLTAP	7042
RSR-1832	NPGRAANEGPPGLPGS	7043
RSR-1855	ESSRAANLTPPELTGP	7044
RSR-1911	ASGRAANETPPGLTGA	7045
RSR-1929	NSGRGENLGAPGLTGT	7046
RSR-1951	TTGRAANLTPAGLTGP	7047
RSR-2295	EAGRSANHTPAGLTGP	7048
RSR-2298	ESGRAANTPAGLTGP	7049
RSR-2038	TTGRATEAANLTPAGLTGP	7050
RSR-2072	TTGRAEEAANLTPAGLTGP	7051
RSR-2089	TTGRAGEAANLTPAGLTGP	7052
RSR-2302	TTGRATEAANATPAGLTGP	7053
RSR-3047	TTGRAGEAEGATSAGATGP	7054
RSR-3052	TTGEAGEAANATSAGATGP	7055
RSR-3043	TTGEAGEAAGLTPAGLTGP	7056
RSR-3041	TTGAAGEAANATPAGLTGP	7057
RSR-3044	TTGRAGEAAGLTPAGLTGP	7058
RSR-3057	TTGRAGEAANATSAGATGP	7059
RSR-3058	TTGEAGEAAGATSAGATGP	7060
RSR-2485	ESGRAANTEPPPELGAG	7061
RSR-2486	ESGRAANTAPEGLTGP	7062
RSR-2488	EPGRAANHEPSSLTEG	7063
RSR-2599	ESGRAANHTGAPPGGLTGP	7064
RSR-2706	TTGRTGEGANATPGGLTGP	7065
RSR-2707	RTGRSGEAANETPEGLEGP	7066
RSR-2708	RTGRTGESANETPAGLGGP	7067
RSR-2709	STGRTGEPANETPAGLSGP	7068
RSR-2710	TTGRAGEPANATPTGLSGP	7069
RSR-2711	RTGRPGEANATPTGLPGP	7070
RSR-2712	RTGRGGEAANATPSGLGGP	7071
RSR-2713	STGRSGESANATPGGLGGP	7072
RSR-2714	RTGRTGEEANATPAGLPGP	7073
RSR-2715	ATGRPGEFANTTPEGLEGP	7074
RSR-2716	STGRSGEPANATPGGLTGP	7075
RSR-2717	PTGRGGEANTTPTGLPGP	7076
RSR-2718	PTGRSGEGANATPSGLTGP	7077
RSR-2719	TTGRASEGANSTPAPLTGP	7078
RSR-2720	TYGRAEEAANTTPAGLTAP	7079
RSR-2721	TTGRATEGANATPAELTGP	7080
RSR-2722	TVGRASEEANTTPASLTGP	7081
RSR-2723	TTGRAPEAANATPAPLTGP	7082
RSR-2724	TWGRATEPANATPAPLTSP	7083
RSR-2725	TVGRASESANATPAELTSP	7084
RSR-2726	TVGRAPEGANSTPAGLTGP	7085
RSR-2727	TWGRATEAPNLEPATLTTP	7086
RSR-2728	TTGRATEAPNLTPAPLTGP	7087
RSR-2729	TQGRATEAPNLSPAALTSP	7088
RSR-2730	TQGRAAEAPNLTPATLTAP	7089
RSR-2731	TSGRAPEATNLAPAPLTGP	7090
RSR-2732	TQGRAAEANLTPAGLTGP	7091

Name	Amino Acid Sequence	SEQ ID NO:
RSR-2733	TTGRAGSAPNLPPTGLTTP	7092
RSR-2734	TTGRAGGAENLPPEGLTAP	7093
RSR-2735	TTSRAGTATNLTPPEGLTAP	7094
RSR-2736	TTGRAGTATNLPPSGLTTP	7095
RSR-2737	TTARAGEAENLSPSGLTAP	7096
RSR-2738	TTGRAGGAGNLAPGGLTEP	7097
RSR-2739	TTGRAGTATNLPPPEGLTGP	7098
RSR-2740	TTGRAGGAANLAPTGLTEP	7099
RSR-2741	TTGRAGTAENLAPSGLTTP	7100
RSR-2742	TTGRAGSATNLGPGGLTGP	7101
RSR-2743	TTARAGGAENLTPAGLTEP	7102
RSR-2744	TTARAGSAENLSPSGLTGP	7103
RSR-2745	TTARAGGAGNLAPEGLTTP	7104
RSR-2746	TTSRAGAAENLTPTGLTGP	7105
RSR-2747	TYGRITTPGNEPPASLEAE	7106
RSR-2748	TYSRGESGPNEPPPGLTGP	7107
RSR-2749	AWGRTGASENETFAPLGGE	7108
RSR-2750	RWGRAETTPNTPEEGLETE	7109
RSR-2751	ESGRAANHTGAEPPELGAG	7110
RSR-2754	TTGRAGEAANLTPAGLTES	7111
RSR-2755	TTGRAGEAANLTPAALTES	7112
RSR-2756	TTGRAGEAANLTPAPLTES	7113
RSR-2757	TTGRAGEAANLTPEPLTES	7114
RSR-2758	TTGRAGEAANLTPAGLTGA	7115
RSR-2759	TTGRAGEAANLTPEGLTGA	7116
RSR-2760	TTGRAGEAANLTPEPLTGA	7117
RSR-2761	TTGRAGEAANLTPAGLTEA	7118
RSR-2762	TTGRAGEAANLTPEGLTEA	7119
RSR-2763	TTGRAGEAANLTPAPLTEA	7120
RSR-2764	TTGRAGEAANLTPEPLTEA	7121
RSR-2765	TTGRAGEAANLTPEPLTGP	7122
RSR-2766	TTGRAGEAANLTPAGLTGG	7123
RSR-2767	TTGRAGEAANLTPEGLTGG	7124
RSR-2768	TTGRAGEAANLTPEALTGG	7125
RSR-2769	TTGRAGEAANLTPEPLTGG	7126
RSR-2770	TTGRAGEAANLTPAGLTEG	7127
RSR-2771	TTGRAGEAANLTPEGLTEG	7128
RSR-2772	TTGRAGEAANLTPAPLTEG	7129
RSR-2773	TTGRAGEAANLTPEPLTEG	7130
RSR-3213	EAGRSASHTPAGLTGP	7628

Table 6b: Release Segment Sequences

Name	Amino Acid Sequence	SEQ ID NO:	Name	Amino Acid Sequence	SEQ ID NO:
RSN-0001	GSAPGSAGGYAELRMGG AIATSGSETPGT	7131	RSC-0001	GTAEAAASASGGSAGGYA ELRMGGAIPGSP	7379
RSN-0002	GSAPGTGGGYAPLRMGG GAATSGSETPGT	7132	RSC-0002	GTAEAAASASGGTGGGYA PLRMGGGAPGSP	7380
RSN-0003	GSAPGAEGGYAALRMGG EIATSGSETPGT	7133	RSC-0003	GTAEAAASASGGAEGGYA ALRMGGIIPGSP	7381
RSN-0004	GSAPGGPGGYALLRMGG PAATSGSETPGT	7134	RSC-0004	GTAEAAASASGGPGGYA LLRMGGPAPGSP	7382
RSN-0005	GSAPGEAGGYAFLRMGG SIATSGSETPGT	7135	RSC-0005	GTAEAAASASGGEAGGYA FLRMGGSIPGSP	7383
RSN-0006	GSAPGPGGGYASLRMGG TAATSGSETPGT	7136	RSC-0006	GTAEAAASASGGPGGGYA SLRMGGTAPGSP	7384

Name	Amino Acid Sequence	SEQ ID NO:	Name	Amino Acid Sequence	SEQ ID NO:
RSN-0007	GSAPGSEGGYATLRMGG AIATSGSETPGT	7137	RSC-0007	GTAEAAASASGGSEGGYA TLRMGGAIPGSP	7385
RSN-0008	GSAPGTPGGYANLRMGG GAATSGSETPGT	7138	RSC-0008	GTAEAAASASGGTPGGYA NLRMGGGAPGSP	7386
RSN-0009	GSAPGASGGYAHLRMGG EIATSGSETPGT	7139	RSC-0009	GTAEAAASASGGASGGYA HLRMGGEIPGSP	7387
RSN-0010	GSAPGGTGGYGELRMGG PAATSGSETPGT	7140	RSC-0010	GTAEAAASASGGTGGYG ELRMGGPAPGSP	7388
RSN-0011	GSAPGEAGGYPELRMGG SIATSGSETPGT	7141	RSC-0011	GTAEAAASASGGEAGGY ELRMGGSIIPGSP	7389
RSN-0012	GSAPGPGGGYVELRMGG TAATSGSETPGT	7142	RSC-0012	GTAEAAASASGPGGGYV ELRMGGTAPGSP	7390
RSN-0013	GSAPGSEGGYLELRMGG AIATSGSETPGT	7143	RSC-0013	GTAEAAASASGGSEGGYL ELRMGGAIPGSP	7391
RSN-0014	GSAPGTPGGYSELRMGG GAATSGSETPGT	7144	RSC-0014	GTAEAAASASGGTPGGYS ELRMGGGAPGSP	7392
RSN-0015	GSAPGASGGYTELRMGG EIATSGSETPGT	7145	RSC-0015	GTAEAAASASGGASGGYT ELRMGGEIPGSP	7393
RSN-0016	GSAPGGTGGYQELRMGG PAATSGSETPGT	7146	RSC-0016	GTAEAAASASGGTGGYQ ELRMGGPAPGSP	7394
RSN-0017	GSAPGEAGGYEELRMGG SIATSGSETPGT	7147	RSC-0017	GTAEAAASASGGEAGGYE ELRMGGSIIPGSP	7395
RSN-0018	GSAPGPGIGPAELRMGG TAATSGSETPGT	7148	RSC-0018	GTAEAAASASGPGIGPA ELRMGGTAPGSP	7396
RSN-0019	GSAPGSEIGAAELRMGG AIATSGSETPGT	7149	RSC-0019	GTAEAAASASGGSEIGAA ELRMGGAIPGSP	7397
RSN-0020	GSAPGTPIGSAELRMGG GAATSGSETPGT	7150	RSC-0020	GTAEAAASASGGTPIGSA ELRMGGGAPGSP	7398
RSN-0021	GSAPGASIGTAEELRMGG EIATSGSETPGT	7151	RSC-0021	GTAEAAASASGGASIGTA ELRMGGEIPGSP	7399
RSN-0022	GSAPGGTIGNAELRMGG PAATSGSETPGT	7152	RSC-0022	GTAEAAASASGGTIGNA ELRMGGPAPGSP	7400
RSN-0023	GSAPGEAIGQAEELRMGG SIATSGSETPGT	7153	RSC-0023	GTAEAAASASGGEAIGQA ELRMGGSIIPGSP	7401
RSN-0024	GSAPGPGGPYAEELRMGG TAATSGSETPGT	7154	RSC-0024	GTAEAAASASGGPGGPYA ELRMGGTAPGSP	7402
RSN-0025	GSAPGSEGAYAEELRMGG AIATSGSETPGT	7155	RSC-0025	GTAEAAASASGGSEGAYA ELRMGGAIPGSP	7403
RSN-0026	GSAPGTPGVYAEELRMGG GAATSGSETPGT	7156	RSC-0026	GTAEAAASASGGTPGVYA ELRMGGGAPGSP	7404
RSN-0027	GSAPGASGLYAEELRMGG EIATSGSETPGT	7157	RSC-0027	GTAEAAASASGGASGLYA ELRMGGEIPGSP	7405
RSN-0028	GSAPGGTGIYAEELRMGG PAATSGSETPGT	7158	RSC-0028	GTAEAAASASGGGTGIYA ELRMGGPAPGSP	7406
RSN-0029	GSAPGEAGFYAEELRMGG SIATSGSETPGT	7159	RSC-0029	GTAEAAASASGGEAGFYA ELRMGGSIIPGSP	7407
RSN-0030	GSAPGPGGYYAEELRMGG TAATSGSETPGT	7160	RSC-0030	GTAEAAASASGGPGGYYA ELRMGGTAPGSP	7408
RSN-0031	GSAPGSEGSYAEELRMGG AIATSGSETPGT	7161	RSC-0031	GTAEAAASASGGSEGSYA ELRMGGAIPGSP	7409
RSN-0032	GSAPGTPGNYAEELRMGG GAATSGSETPGT	7162	RSC-0032	GTAEAAASASGGTPGNYA ELRMGGGAPGSP	7410
RSN-0033	GSAPGASGEYAEELRMGG EIATSGSETPGT	7163	RSC-0033	GTAEAAASASGGASGEYA ELRMGGEIPGSP	7411
RSN-0034	GSAPGGTGHYAEELRMGG PAATSGSETPGT	7164	RSC-0034	GTAEAAASASGGTGHYA ELRMGGPAPGSP	7412
RSN-0035	GSAPGEAGGYAEARMGG SIATSGSETPGT	7165	RSC-0035	GTAEAAASASGGEAGGYA EARMGGSIIPGSP	7413

Name	Amino Acid Sequence	SEQ ID NO:	Name	Amino Acid Sequence	SEQ ID NO:
RSN-0036	GSAPGPGGGYAEVRMGG TAATSGSETPGT	7166	RSC-0036	GTAEAAASASGGPGGGYA EVRMGGTAPGSP	7414
RSN-0037	GSAPGSEGGYAEIRMGG AIATSGSETPGT	7167	RSC-0037	GTAEAAASASGGSEGGYA EIRMGGAI PGSP	7415
RSN-0038	GSAPGTPGGYAEFRMGG GAATSGSETPGT	7168	RSC-0038	GTAEAAASASGGTPGGYA EFRMGGGAPGSP	7416
RSN-0039	GSAPGASGGYAEYRMGG EIATSGSETPGT	7169	RSC-0039	GTAEAAASASGGASGGYA EYRMGGEI PGSP	7417
RSN-0040	GSAPGGTGGYAESRMGG PAATSGSETPGT	7170	RSC-0040	GTAEAAASASGGGTGGYA ESRMGGPAPGSP	7418
RSN-0041	GSAPGEAGGYAETRMGG SIATSGSETPGT	7171	RSC-0041	GTAEAAASASGGEAGGYA ETRMGGSI PGSP	7419
RSN-0042	GSAPGPGGGYAEELMGG TRATSGSETPGT	7172	RSC-0042	GTAEAAASASGGPGGGYA ELAMGGTRPGSP	7420
RSN-0043	GSAPGSEGGYAEELVMGG ARATSGSETPGT	7173	RSC-0043	GTAEAAASASGGSEGGYA ELVMGGARPGSP	7421
RSN-0044	GSAPGTPGGYAEELMGG GRATSGSETPGT	7174	RSC-0044	GTAEAAASASGGTPGGYA ELLMGGGRPGSP	7422
RSN-0045	GSAPGASGGYAEELIMG ERATSGSETPGT	7175	RSC-0045	GTAEAAASASGGASGGYA ELIMGGERPGSP	7423
RSN-0046	GSAPGGTGGYAEELWMGG PRATSGSETPGT	7176	RSC-0046	GTAEAAASASGGGTGGYA ELWMGGPRPGSP	7424
RSN-0047	GSAPGEAGGYAELSMGG SRATSGSETPGT	7177	RSC-0047	GTAEAAASASGGEAGGYA ELSMGGSRPGSP	7425
RSN-0048	GSAPGPGGGYAEELTMGG TRATSGSETPGT	7178	RSC-0048	GTAEAAASASGGPGGGYA ELTMGGTRPGSP	7426
RSN-0049	GSAPGSEGGYAEELQMG ARATSGSETPGT	7179	RSC-0049	GTAEAAASASGGSEGGYA ELQMGGARPGSP	7427
RSN-0050	GSAPGTPGGYAEELNMGG GRATSGSETPGT	7180	RSC-0050	GTAEAAASASGGTPGGYA ELNMGGGRPGSP	7428
RSN-0051	GSAPGASGGYAEELMGG ERATSGSETPGT	7181	RSC-0051	GTAEAAASASGGASGGYA ELEMGGGERPGSP	7429
RSN-0052	GSAPGGTGGYAEELRPGG PIATSGSETPGT	7182	RSC-0052	GTAEAAASASGGGTGGYA ELRPGGPI PGSP	7430
RSN-0053	GSAPGEAGGYAELRAGG SAATSGSETPGT	7183	RSC-0053	GTAEAAASASGGEAGGYA ELRAGGSAPGSP	7431
RSN-0054	GSAPGPGGGYAEELRLGG TIATSGSETPGT	7184	RSC-0054	GTAEAAASASGGPGGGYA ELRLGGTI PGSP	7432
RSN-0055	GSAPGSEGGYAEELRIGG AAATSGSETPGT	7185	RSC-0055	GTAEAAASASGGSEGGYA ELRIGGAAPGSP	7433
RSN-0056	GSAPGTPGGYAEELRSGG GIATSGSETPGT	7186	RSC-0056	GTAEAAASASGGTPGGYA ELRSGGGI PGSP	7434
RSN-0057	GSAPGASGGYAEELRNGG EAATSGSETPGT	7187	RSC-0057	GTAEAAASASGGASGGYA ELRNGGEAPGSP	7435
RSN-0058	GSAPGGTGGYAEELRQGG PIATSGSETPGT	7188	RSC-0058	GTAEAAASASGGGTGGYA ELRQGGPI PGSP	7436
RSN-0059	GSAPGEAGGYAELRDGG SAATSGSETPGT	7189	RSC-0059	GTAEAAASASGGEAGGYA ELRDGGSA PGSP	7437
RSN-0060	GSAPGPGGGYAEELREGG TIATSGSETPGT	7190	RSC-0060	GTAEAAASASGGPGGGYA ELREGGTI PGSP	7438
RSN-0061	GSAPGSEGGYAEELRHGG AAATSGSETPGT	7191	RSC-0061	GTAEAAASASGGSEGGYA ELRHGGAAPGSP	7439
RSN-0062	GSAPGTPGGYAEELRMFG GIATSGSETPGT	7192	RSC-0062	GTAEAAASASGGTPGGYA ELRMFGGI PGSP	7440
RSN-0063	GSAPGASGGYAEELRMAG EAATSGSETPGT	7193	RSC-0063	GTAEAAASASGGASGGYA ELRMAGEAPGSP	7441
RSN-0064	GSAPGGTGGYAEELRMVG PIATSGSETPGT	7194	RSC-0064	GTAEAAASASGGGTGGYA ELRMVGPI PGSP	7442

Name	Amino Acid Sequence	SEQ ID NO:	Name	Amino Acid Sequence	SEQ ID NO:
RSN-0065	GSAPGEAGGYAELRMLG SAATSGSETPGT	7195	RSC-0065	GTAEAASASGGEAGGYA ELRMLGSAPGSP	7443
RSN-0066	GSAPGPGGGYAE LRMI G TIATSGSETPGT	7196	RSC-0066	GTAEAASASGGPGGGYA ELRMIGTIPGSP	7444
RSN-0067	GSAPGSEGGYAE LRMY G AIATSGSETPGT	7197	RSC-0067	GTAEAASASGGSEGGYA ELRMYGAIPGSP	7445
RSN-0068	GSAPGTPGGYAE LRMS G GAATSGSETPGT	7198	RSC-0068	GTAEAASASGGTPGGYA ELRMSGGAPGSP	7446
RSN-0069	GSAPGASGGYAE LRNM G EIATSGSETPGT	7199	RSC-0069	GTAEAASASGGASGGYA ELRMNGEIPGSP	7447
RSN-0070	GSAPGGTGGYAE LRMQ G PAATSGSETPGT	7200	RSC-0070	GTAEAASASGGTGGYA ELRMQGPAPGSP	7448
RSN-0071	GSAPGANHTPAGLTGP G ARATSGSETPGT	7201	RSC-0071	GTAEAASASGGANHTPA GLTGPGARPGSP	7449
RSN-0072	GSAPGANTAPEGLTG PS TRATSGSETPGT	7202	RSC-0072	GTAEAASASGGANTAPE GLTGPTSTRPGSP	7450
RSN-0073	GSAPGTGAPPGGLTG PG TRATSGSETPGT	7203	RSC-0073	GTAEAASASGGTGAPPG GLTGPTTRPGSP	7451
RSN-0074	GSAPGANHEPSGLTE GS PRATSGSETPGT	7204	RSC-0074	GTAEAASASGGANHEPS GLTEGSPRPGSP	7452
RSN-0075	GSAPGANTEPPELGAG T ERATSGSETPGT	7205	RSC-0075	GTAEAASASGGANTEPP ELGAGTERPGSP	7453
RSN-0076	GSAPGASGPPPGLTGP P GRATSGSETPGT	7206	RSC-0076	GTAEAASASGGASGPPP GLTGPPGRPGSP	7454
RSN-0077	GSAPGASGTPAPLGGE P GRATSGSETPGT	7207	RSC-0077	GTAEAASASGGASGTPA PLGGEPRPGSP	7455
RSN-0078	GSAPGPAGPPEGLETE A GRATSGSETPGT	7208	RSC-0078	GTAEAASASGGPAGPPE GLETEAGRPGSP	7456
RSN-0079	GSAPGPTSGQGGLTG PE SRATSGSETPGT	7209	RSC-0079	GTAEAASASGGPTSGQG GLTGPESTRPGSP	7457
RSN-0080	GSAPGSAGGAANLVR GG AIATSGSETPGT	7210	RSC-0080	GTAEAASASGGSAGGAA NLVRGGAIPGSP	7458
RSN-0081	GSAPGTGGGAAPLVR GG GAATSGSETPGT	7211	RSC-0081	GTAEAASASGGTGGGAA PLVRGGGAPGSP	7459
RSN-0082	GSAPGAEGGAALVR GG EIATSGSETPGT	7212	RSC-0082	GTAEAASASGGAEAGGAA ALVRGGEIPGSP	7460
RSN-0083	GSAPGGPGGAALLVR GG PAATSGSETPGT	7213	RSC-0083	GTAEAASASGGPGGGAA LLVRGGPAPGSP	7461
RSN-0084	GSAPGEAGGAFLVR GG SIATSGSETPGT	7214	RSC-0084	GTAEAASASGGEAGGAA FLVRGG SIPGSP	7462
RSN-0085	GSAPGPGGAASLVR GG TAATSGSETPGT	7215	RSC-0085	GTAEAASASGGPGGGAA SLVRGGTAPGSP	7463
RSN-0086	GSAPGSEGGAA TLVR GG AIATSGSETPGT	7216	RSC-0086	GTAEAASASGGSEGGAA TLVRGGAIPGSP	7464
RSN-0087	GSAPGTPGGAAGLVR GG GAATSGSETPGT	7217	RSC-0087	GTAEAASASGGTPGGAA GLVRGGGAPGSP	7465
RSN-0088	GSAPGASGGAADLVR GG EIATSGSETPGT	7218	RSC-0088	GTAEAASASGGASGGAA DLVRGGEIPGSP	7466
RSN-0089	GSAPGGTGGAGNLVR GG PAATSGSETPGT	7219	RSC-0089	GTAEAASASGGGTGGAG NLVRGGPAPGSP	7467
RSN-0090	GSAPGEAGGAPNLVR GG SIATSGSETPGT	7220	RSC-0090	GTAEAASASGGEAGGAP NLVRGGSIPGSP	7468
RSN-0091	GSAPGPGGGAVNLVR GG TAATSGSETPGT	7221	RSC-0091	GTAEAASASGGPGGGAV NLVRGGTAPGSP	7469
RSN-0092	GSAPGSEGGALNLVR GG AIATSGSETPGT	7222	RSC-0092	GTAEAASASGGSEGGAL NLVRGGAIPGSP	7470
RSN-0093	GSAPGTPGGASNLVR GG GAATSGSETPGT	7223	RSC-0093	GTAEAASASGGTPGGAS NLVRGGGAPGSP	7471

Name	Amino Acid Sequence	SEQ ID NO:	Name	Amino Acid Sequence	SEQ ID NO:
RSN-0094	GSAPGASGGATNLVRGG EIATSGSETPGT	7224	RSC-0094	GTAEASASGGASGGAT NLVRGGEIPGSP	7472
RSN-0095	GSAPGGTGGQNLVRGG PAATSGSETPGT	7225	RSC-0095	GTAEASASGGGTGGAQ NLVRGGPAPGSP	7473
RSN-0096	GSAPGEAGGAENLVRGG SIATSGSETPGT	7226	RSC-0096	GTAEASASGGGEAGGAE NLVRGGSIPGSP	7474
RSN-1517	GSAPAEAGRSANHEPLGL VATATSGSETPGT	7227	RSC-1517	GTAEASASGEAGRSAN HEPLGLVATPGSP	7475
BSRS-A1-2	GSAPASGRSTNAGPSGL AGPATSGSETPGT	7228	BSRS-A1-3	GTAEASASGASGRSTN AGPSGLAGPPGSP	7476
BSRS-A2-2	GSAPASGRSTNAGPQGL AGQATSGSETPGT	7229	BSRS-A2-3	GTAEASASGASGRSTN AGPQGLAGQPGSP	7477
BSRS-A3-2	GSAPASGRSTNAGPPGL TGPATSGSETPGT	7230	BSRS-A3-3	GTAEASASGASGRSTN AGPPGLTGPPGSP	7478
VP-1	GSAPASSRGTNAGPAGL TGPATSGSETPGT	7231	VP-1	GTAEASASGASSRGTN AGPAGLTGPPGSP	7479
RSN-1752	GSAPASSRTTNTGPSTL TGPATSGSETPGT	7232	RSC-1752	GTAEASASGASSRRTTN TGPSTLTGPPGSP	7480
RSN-1512	GSAPAAGRSNNGTLEL VAPATSGSETPGT	7233	RSC-1512	GTAEASASGAAGRSND GTLELVAPPGSP	7481
RSN-1517	GSAPAEAGRSANHEPLGL VATATSGSETPGT	7234	RSC-1517	GTAEASASGEAGRSAN HEPLGLVATPGSP	7482
VP-2	GSAPASGRGTNAGPAGL TGPATSGSETPGT	7235	VP-2	GTAEASASGASGRGTN AGPAGLTGPPGSP	7483
RSN-1018	GSAPLFGRNDNHELEL GGATSGSETPGT	7236	RSC-1018	GTAEASASGLFGRNDN HEPLELGGGPGSP	7484
RSN-1053	GSAPTAGRSNDNLEPLGL VFGATSGSETPGT	7237	RSC-1053	GTAEASASGTAGRSND LEPLGLVFGPGSP	7485
RSN-1059	GSAPLDGRSDNFHPEL VAGATSGSETPGT	7238	RSC-1059	GTAEASASGLDGRSDN FHPPELVAGPGSP	7486
RSN-1065	GSAPLEGRSDNEEPENL VAGATSGSETPGT	7239	RSC-1065	GTAEASASGLEGRSDN EPEPNLVAGPGSP	7487
RSN-1167	GSAPLKGRSDNNAPLAL VAGATSGSETPGT	7240	RSC-1167	GTAEASASGLKGRSDN NAPLALVAGPGSP	7488
RSN-1201	GSAPVYSRGTNAGPHGL TGRATSGSETPGT	7241	RSC-1201	GTAEASASGVYSRGTN AGPHGLTGRPGSP	7489
RSN-1218	GSAPANSRGTNKGFAGL IGPATSGSETPGT	7242	RSC-1218	GTAEASASGANSRGTN KGFAGLIGPPGSP	7490
RSN-1226	GSAPASSRLTNEAPAGL TIPATSGSETPGT	7243	RSC-1226	GTAEASASGASSRLTN EAPAGLTIPPSP	7491
RSN-1254	GSAPDQSRGTNAGPEGL TDPATSGSETPGT	7244	RSC-1254	GTAEASASGDQSRGTN AGPEGLTDPGSP	7492
RSN-1256	GSAPESSRGTNIGQGGL TGPATSGSETPGT	7245	RSC-1256	GTAEASASGESRGTN IGQGGLTGPPGSP	7493
RSN-1261	GSAPSSSRGTNQDPAGL TIPATSGSETPGT	7246	RSC-1261	GTAEASASGSSSRGTN QDPAGLTIPPSP	7494
RSN-1293	GSAPASSRGQNHPMGL TGPATSGSETPGT	7247	RSC-1293	GTAEASASGASSRGQN HSPMGLTGPPGSP	7495
RSN-1309	GSAPAYSRGPNAGPAGL EGRATSGSETPGT	7248	RSC-1309	GTAEASASGAYSRGPN AGPAGLEGRPGSP	7496
RSN-1326	GSAPASERGNNAHPANL TGFATSGSETPGT	7249	RSC-1326	GTAEASASGASERGN AGPANLTGPPGSP	7497
RSN-1345	GSAPASHRGTNPKPAIL TGPATSGSETPGT	7250	RSC-1345	GTAEASASGASHRGTN PKPAILTGPPGSP	7498
RSN-1354	GSAPMSSRRTNANPAQL TGPATSGSETPGT	7251	RSC-1354	GTAEASASGMSSRRTN ANPAQLTGPPGSP	7499
RSN-1426	GSAPGAGRTDNHELEL GAAATSGSETPGT	7252	RSC-1426	GTAEASASGGAGRTDN HEPLELGAAPGSP	7500

Name	Amino Acid Sequence	SEQ ID NO:	Name	Amino Acid Sequence	SEQ ID NO:
RSN-1478	GSAPLAGRSENTAPLEL TAGATSGSETPGT	7253	RSC-1478	GTAEAAASASGLAGRSEN TAPLELTAGPGSP	7501
RSN-1479	GSAPLEGRPDNHEPLAL VASATSGSETPGT	7254	RSC-1479	GTAEAAASASGLEGRPDN HEPLALVASPGSP	7502
RSN-1496	GSAPLSGRSDNEEPLAL PAGATSGSETPGT	7255	RSC-1496	GTAEAAASASGLSGRSDN EEPLALPAGPGSP	7503
RSN-1508	GSAPAEAGRDTNHEPLEL SAPATSGSETPGT	7256	RSC-1508	GTAEAAASASGEAGRDTN HEPLELSAPPGSP	7504
RSN-1513	GSAPEGGRSDNHGPLEL VSGATSGSETPGT	7257	RSC-1513	GTAEAAASASGEGGRSDN HGPLELVSGPGSP	7505
RSN-1516	GSAPLSGRSDNEAPLEL EAGATSGSETPGT	7258	RSC-1516	GTAEAAASASGLSGRSDN EAPLELEAGPGSP	7506
RSN-1524	GSAPLGGRADNHEPPEL GAGATSGSETPGT	7259	RSC-1524	GTAEAAASASGLGGRADN HEPPELGAGPGSP	7507
RSN-1622	GSAPPPSRGTNAEPAGL TGEATSGSETPGT	7260	RSC-1622	GTAEAAASASGPPSRGTN AEPAGLTGEPGSP	7508
RSN-1629	GSAPASTRGENAGPAGL EAPATSGSETPGT	7261	RSC-1629	GTAEAAASASGASTRGEN AGPAGLEAPPGSP	7509
RSN-1664	GSAPESSRGTNGAPEGL TGPATSGSETPGT	7262	RSC-1664	GTAEAAASASGESRGTN GAPEGLTGPPGSP	7510
RSN-1667	GSAPASSRATNESPAGL TGEATSGSETPGT	7263	RSC-1667	GTAEAAASASGASSRATN ESPAGLTGEPGSP	7511
RSN-1709	GSAPASSRGENPPPGL TGPATSGSETPGT	7264	RSC-1709	GTAEAAASASGASSRGEN PPPGLTGPPGSP	7512
RSN-1712	GSAPAASRGTTNTPAEL TGSATSGSETPGT	7265	RSC-1712	GTAEAAASASGAASRGTN TGPAELTGSPPGSP	7513
RSN-1727	GSAPAGSRTTNAGPGL EGPATSGSETPGT	7266	RSC-1727	GTAEAAASASGAGSRTTN AGPGGLEGP PGSP	7514
RSN-1754	GSAPAPSRGENAGPATL TGAATSGSETPGT	7267	RSC-1754	GTAEAAASASGAPSRGEN AGPATLTGAPGSP	7515
RSN-1819	GSAPESGRAANTGPPTL TAPATSGSETPGT	7268	RSC-1819	GTAEAAASASGESGRAAN TGPPTLTAPPGSP	7516
RSN-1832	GSAPNPGRAANEGPPGL PGSATSGSETPGT	7269	RSC-1832	GTAEAAASASGNPGRAAN EGPPGLPGSPGSP	7517
RSN-1855	GSAPESSRAANLTPEL TGPATSGSETPGT	7270	RSC-1855	GTAEAAASASGESGRAAN LTPELTGPPGSP	7518
RSN-1911	GSAPASGRAANETPPGL TGAATSGSETPGT	7271	RSC-1911	GTAEAAASASGASGRAAN ETPPGLTGAPGSP	7519
RSN-1929	GSAPNSGRGENLGAPGL TGTATSGSETPGT	7272	RSC-1929	GTAEAAASASGNSGRGEN LGAPGLTGTPGSP	7520
RSN-1951	GSAPTTGRAANLTPAGL TGPATSGSETPGT	7273	RSC-1951	GTAEAAASASGTTGRAAN LTPAGLTGPPGSP	7521
RSN-2295	GSAPAEAGRSANHTPAGL TGPATSGSETPGT	7274	RSC-2295	GTAEAAASASGEAGRSAN HTPAGLTGPPGSP	7522
RSN-2298	GSAPESGRAANTTPAGL TGPATSGSETPGT	7275	RSC-2298	GTAEAAASASGESGRAAN TTPAGLTGPPGSP	7523
RSN-2038	GSAPTTGRATEAANLTP AGLTGPATSGSETPGT	7276	RSC-2038	GTAEAAASASGTTGRATE AANLTPAGLTGPPGSP	7524
RSN-2072	GSAPTTGRAEEAANLTP AGLTGPATSGSETPGT	7277	RSC-2072	GTAEAAASASGTTGRAEE AANLTPAGLTGPPGSP	7525
RSN-2089	GSAPTTGRAGEAANLTP AGLTGPATSGSETPGT	7278	RSC-2089	GTAEAAASASGTTGRAGE AANLTPAGLTGPPGSP	7526
RSN-2302	GSAPTTGRATEAANATP AGLTGPATSGSETPGT	7279	RSC-2302	GTAEAAASASGTTGRATE AANATPAGLTGPPGSP	7527
RSN-3047	GSAPTTGRAGEAEGATS AGATGPATSGSETPGT	7280	RSC-3047	GTAEAAASASGTTGRAGE AEGATSAGATGPPGSP	7528
RSN-3052	GSAPTTGEAGEAANATS AGATGPATSGSETPGT	7281	RSC-3052	GTAEAAASASGTTGEAGE AANATSAGATGPPGSP	7529

Name	Amino Acid Sequence	SEQ ID NO:	Name	Amino Acid Sequence	SEQ ID NO:
RSN-3043	GSAPTTGEAGEAAGLTP AGLTGPATSGSETPGT	7282	RSC-3043	GTAEAAASASGTTGEAGE AAGLTPAGLTGPPGSP	7530
RSN-3041	GSAPTTGAAGEAANATP AGLTGPATSGSETPGT	7283	RSC-3041	GTAEAAASASGTTGAAGE AANATPAGLTGPPGSP	7531
RSN-3044	GSAPTTGRAGEAAGLTP AGLTGPATSGSETPGT	7284	RSC-3044	GTAEAAASASGTTGRAGE AAGLTPAGLTGPPGSP	7532
RSN-3057	GSAPTTGRAGEAANATS AGATGPATSGSETPGT	7285	RSC-3057	GTAEAAASASGTTGRAGE AANATSAGATGPPGSP	7533
RSN-3058	GSAPTTGEAGEAAGATS AGATGPATSGSETPGT	7286	RSC-3058	GTAEAAASASGTTGEAGE AAGATSAGATGPPGSP	7534
RSN-2485	GSAPESGRAANTEPEL GAGATSGSETPGT	7287	RSC-2485	GTAEAAASASGESGRAAN TEPPELGAGPGSP	7535
RSN-2486	GSAPESGRAANTAPEGL TGPATSGSETPGT	7288	RSC-2486	GTAEAAASASGESGRAAN TAPEGLTGPPGSP	7536
RSN-2488	GSAPEPGRAANHEPSGL TEGATSGSETPGT	7289	RSC-2488	GTAEAAASASGEPGRAAN HEPSGLTEGPGSP	7537
RSN-2599	GSAPESGRAANHTGAPP GGLTGPATSGSETPGT	7290	RSC-2599	GTAEAAASASGESGRAAN HTGAPPGGLTGPPGSP	7538
RSN-2706	GSAPTTGRTGEGANATP GGLTGPATSGSETPGT	7291	RSC-2706	GTAEAAASASGTTGRTGE GANATPGGLTGPPGSP	7539
RSN-2707	GSAPRTGRSGEAAANETP EGLEGPATSGSETPGT	7292	RSC-2707	GTAEAAASASGRTGRSGE AANETPEGLEGPPGSP	7540
RSN-2708	GSAPRTGRTGESANETP AGLGGPATSGSETPGT	7293	RSC-2708	GTAEAAASASGRTGRTGE SANETPAGLGGPPGSP	7541
RSN-2709	GSAPSTGRTGEPANETP AGLSGPATSGSETPGT	7294	RSC-2709	GTAEAAASASGSTGRTGE PANETPAGLSGPPGSP	7542
RSN-2710	GSAPTTGRAGEPANATP TGLSGPATSGSETPGT	7295	RSC-2710	GTAEAAASASGTTGRAGE PANATPTGLSGPPGSP	7543
RSN-2711	GSAPRTGRPGEGANATP TGLPGPATSGSETPGT	7296	RSC-2711	GTAEAAASASGRTGRPGE GANATPTGLPGPPGSP	7544
RSN-2712	GSAPRTGRGGEAANATP SGLGGPATSGSETPGT	7297	RSC-2712	GTAEAAASASGRTGRGGE AANATPSGLGGPPGSP	7545
RSN-2713	GSAPSTGRSGESANATP GGLGGPATSGSETPGT	7298	RSC-2713	GTAEAAASASGSTGRSGE SANATPGGLGGPPGSP	7546
RSN-2714	GSAPRTGRTGEEANATP AGLPGPATSGSETPGT	7299	RSC-2714	GTAEAAASASGRTGRTGE EANATPAGLPGPPGSP	7547
RSN-2715	GSAPATGRPGEPANTTP EGLEGPATSGSETPGT	7300	RSC-2715	GTAEAAASASGATGRPGE PANTTPEGLEGPPGSP	7548
RSN-2716	GSAPSTGRSGEPANATP GGLTGPATSGSETPGT	7301	RSC-2716	GTAEAAASASGSTGRSGE PANATPGGLTGPPGSP	7549
RSN-2717	GSAPPTGRGGEGANTTP TGLPGPATSGSETPGT	7302	RSC-2717	GTAEAAASASGPTGRGGE GANTTPTGLPGPPGSP	7550
RSN-2718	GSAPPTGRSGEGANATP SGLTGPATSGSETPGT	7303	RSC-2718	GTAEAAASASGPTGRSGE GANATPSGLTGPPGSP	7551
RSN-2719	GSAPTTGRASEGANSTP APLTEPATSGSETPGT	7304	RSC-2719	GTAEAAASASGTTGRASE GANSTPAPLTEPPGSP	7552
RSN-2720	GSAPTYGRAEEAANTTP AGLTAPATSGSETPGT	7305	RSC-2720	GTAEAAASASGTYGRAAE AANTTPAGLTAPPGSP	7553
RSN-2721	GSAPTTGRATEGANATP AELTEPATSGSETPGT	7306	RSC-2721	GTAEAAASASGTTGRATE GANATPAELTEPPGSP	7554
RSN-2722	GSAPTVGRASEEANTTP ASLTGPATSGSETPGT	7307	RSC-2722	GTAEAAASASGTVGRASE EANTTPASLTGPPGSP	7555
RSN-2723	GSAPTTGRAPEAANATP APLTGPATSGSETPGT	7308	RSC-2723	GTAEAAASASGTTGRAPE AANATPAPLTGPPGSP	7556
RSN-2724	GSAPTWGRATEPANATP APLTPSPATSGSETPGT	7309	RSC-2724	GTAEAAASASGTWGRATE PANATPAPLTPPPGSP	7557
RSN-2725	GSAPTVGRASESANATP AELTSPATSGSETPGT	7310	RSC-2725	GTAEAAASASGTVGRASE SANATPAELTSPPPGSP	7558

Name	Amino Acid Sequence	SEQ ID NO:	Name	Amino Acid Sequence	SEQ ID NO:
RSN-2726	GSAPTVGRAPEGANSTP AGLTGPATSGSETPGT	7311	RSC-2726	GTAEAAASASGTVGRAPE GANSTPAGLTGPPGSP	7559
RSN-2727	GSAPTWGRATEAPNLEP ATLTTPATSGSETPGT	7312	RSC-2727	GTAEAAASASGTWGRATE APNLEPATLTTPPGSP	7560
RSN-2728	GSAPTTGRATEAPNLTP APLTGPATSGSETPGT	7313	RSC-2728	GTAEAAASASGTTGRATE APNLTPAPLTTPPGSP	7561
RSN-2729	GSAPTQGRATEAPNLS P AALTSPATSGSETPGT	7314	RSC-2729	GTAEAAASASGTQGRATE APNLS PAALTSPPGSP	7562
RSN-2730	GSAPTQGRAEAPNLTP ATLTAPATSGSETPGT	7315	RSC-2730	GTAEAAASASGTQGRAE APNLTPATLTAPPGSP	7563
RSN-2731	GSAPTSGRAPEATNLAP APLTGPATSGSETPGT	7316	RSC-2731	GTAEAAASASGTSGRAPE ATNLAPAPLTGPPGSP	7564
RSN-2732	GSAPTQGRAEAAANLTP AGLTGPATSGSETPGT	7317	RSC-2732	GTAEAAASASGTQGRAE AANLTPAGLTTPPGSP	7565
RSN-2733	GSAPTTGRAGSAPNLFP TGLTTPATSGSETPGT	7318	RSC-2733	GTAEAAASASGTTGRAGS APNLPTGLTTPPGSP	7566
RSN-2734	GSAPTTGRAGGAENLFP EGLTAPATSGSETPGT	7319	RSC-2734	GTAEAAASASGTTGRAGG AENLPP EGLTAPPGSP	7567
RSN-2735	GSAPTTSRAGTATNLTP EGLTAPATSGSETPGT	7320	RSC-2735	GTAEAAASASGTTSRAGT ATNLTP EGLTAPPGSP	7568
RSN-2736	GSAPTTGRAGTATNLFP SGLTTPATSGSETPGT	7321	RSC-2736	GTAEAAASASGTTGRAGT ATNLPPSGLTTPPGSP	7569
RSN-2737	GSAPTTARAGEAENLSP SGLTAPATSGSETPGT	7322	RSC-2737	GTAEAAASASGTTARAGE AENLPSGLTAPPGSP	7570
RSN-2738	GSAPTTGRAGGAGNLAP GGLTEPATSGSETPGT	7323	RSC-2738	GTAEAAASASGTTGRAGG AGNLAPGGLTEPPGSP	7571
RSN-2739	GSAPTTGRAGTATNLFP EGLTGPATSGSETPGT	7324	RSC-2739	GTAEAAASASGTTGRAGT ATNLPP EGLTGPPGSP	7572
RSN-2740	GSAPTTGRAGGAANLAP TGLTEPATSGSETPGT	7325	RSC-2740	GTAEAAASASGTTGRAGG AANLAPTGLTEPPGSP	7573
RSN-2741	GSAPTTGRAGTAENLAP SGLTTPATSGSETPGT	7326	RSC-2741	GTAEAAASASGTTGRAGT AENLAPSGLTTPPGSP	7574
RSN-2742	GSAPTTGRAGSATNLGP GGLTGPATSGSETPGT	7327	RSC-2742	GTAEAAASASGTTGRAGS ATNLGPGGLTGPPGSP	7575
RSN-2743	GSAPTTARAGGAENLTP AGLTGPATSGSETPGT	7328	RSC-2743	GTAEAAASASGTTARAGG AENLTPAGLTTPPGSP	7576
RSN-2744	GSAPTTARAGSAENLSP SGLTGPATSGSETPGT	7329	RSC-2744	GTAEAAASASGTTARAGS AENLPSGLTGPPGSP	7577
RSN-2745	GSAPTTARAGGAGNLAP EGLTTPATSGSETPGT	7330	RSC-2745	GTAEAAASASGTTARAGG AGNLAPEGLTTPPGSP	7578
RSN-2746	GSAPTTSRAGAAENLTP TGLTGPATSGSETPGT	7331	RSC-2746	GTAEAAASASGTTSRAGA AENLTP TGLTGPPGSP	7579
RSN-2747	GSAPTYGRTTTPGNEFP ASLEAEATSGSETPGT	7332	RSC-2747	GTAEAAASASGTYGRTTT PGNEPPASLEAEPGSP	7580
RSN-2748	GSAPTYSRGESGPNEFP PGLTGPATSGSETPGT	7333	RSC-2748	GTAEAAASASGTYSRGES GPNEPPPGLTGPPGSP	7581
RSN-2749	GSAPAWGRTGASENETP APLGGEATSGSETPGT	7334	RSC-2749	GTAEAAASASGAWGRTGA SENETPAPLGGEPPGSP	7582
RSN-2750	GSAPRWGRAETTPNTPP EGLETEATSGSETPGT	7335	RSC-2750	GTAEAAASASGRWGRAET TPNTPPEGLETEPPGSP	7583
RSN-2751	GSAPESGRAANHTGAEP PELGAGATSGSETPGT	7336	RSC-2751	GTAEAAASASGESGRAAN HTGAEPPELGAGPGSP	7584
RSN-2754	GSAPTTGRAGEAANLTP AGLTESATSGSETPGT	7337	RSC-2754	GTAEAAASASGTTGRAGE AANLTPAGLTESPGSP	7585
RSN-2755	GSAPTTGRAGEAANLTP AALTESATSGSETPGT	7338	RSC-2755	GTAEAAASASGTTGRAGE AANLTPAALTESPGSP	7586
RSN-2756	GSAPTTGRAGEAANLTP APLTESATSGSETPGT	7339	RSC-2756	GTAEAAASASGTTGRAGE AANLTPAPLTESPGSP	7587

Name	Amino Acid Sequence	SEQ ID NO:	Name	Amino Acid Sequence	SEQ ID NO:
RSN-2757	GSAPTTGRAGEAANLTP EPLTESATSGSETPGT	7340	RSC-2757	GTAEAASASGTTGRAGE AANLTPEPLTESPGSP	7588
RSN-2758	GSAPTTGRAGEAANLTP AGLTGAATSGSETPGT	7341	RSC-2758	GTAEAASASGTTGRAGE AANLT PAGLTGAPGSP	7589
RSN-2759	GSAPTTGRAGEAANLTP EGLTGAATSGSETPGT	7342	RSC-2759	GTAEAASASGTTGRAGE AANLTPEGLTGAPGSP	7590
RSN-2760	GSAPTTGRAGEAANLTP EPLTGAATSGSETPGT	7343	RSC-2760	GTAEAASASGTTGRAGE AANLTPEPLTGAPGSP	7591
RSN-2761	GSAPTTGRAGEAANLTP AGLTEAATSGSETPGT	7344	RSC-2761	GTAEAASASGTTGRAGE AANLT PAGLTEAPGSP	7592
RSN-2762	GSAPTTGRAGEAANLTP EGLTEAATSGSETPGT	7345	RSC-2762	GTAEAASASGTTGRAGE AANLTPEGLTEAPGSP	7593
RSN-2763	GSAPTTGRAGEAANLTP APLTEAATSGSETPGT	7346	RSC-2763	GTAEAASASGTTGRAGE AANLT PAPLTEAPGSP	7594
RSN-2764	GSAPTTGRAGEAANLTP EPLTEAATSGSETPGT	7347	RSC-2764	GTAEAASASGTTGRAGE AANLTPEPLTEAPGSP	7595
RSN-2765	GSAPTTGRAGEAANLTP EPLTGPATSGSETPGT	7348	RSC-2765	GTAEAASASGTTGRAGE AANLTPEPLTGPPGSP	7596
RSN-2766	GSAPTTGRAGEAANLTP AGLTGGATSGSETPGT	7349	RSC-2766	GTAEAASASGTTGRAGE AANLT PAGLTGGPGSP	7597
RSN-2767	GSAPTTGRAGEAANLTP EGLTGGATSGSETPGT	7350	RSC-2767	GTAEAASASGTTGRAGE AANLTPEGLTGGPGSP	7598
RSN-2768	GSAPTTGRAGEAANLTP EALTGGATSGSETPGT	7351	RSC-2768	GTAEAASASGTTGRAGE AANLTPEALTGGPGSP	7599
RSN-2769	GSAPTTGRAGEAANLTP EPLTGGATSGSETPGT	7352	RSC-2769	GTAEAASASGTTGRAGE AANLTPEPLTGGPGSP	7600
RSN-2770	GSAPTTGRAGEAANLTP AGLTGATSGSETPGT	7353	RSC-2770	GTAEAASASGTTGRAGE AANLT PAGLTGPGSP	7601
RSN-2771	GSAPTTGRAGEAANLTP EGLTEGATSGSETPGT	7354	RSC-2771	GTAEAASASGTTGRAGE AANLTPEGLTEGPGSP	7602
RSN-2772	GSAPTTGRAGEAANLTP APLTEGATSGSETPGT	7355	RSC-2772	GTAEAASASGTTGRAGE AANLT PAPLTEGPGSP	7603
RSN-2773	GSAPTTGRAGEAANLTP EPLTEGATSGSETPGT	7356	RSC-2773	GTAEAASASGTTGRAGE AANLTPEPLTEGPGSP	7604
RSN-3047	GSAPTTGRAGEAAGATS AGATGPATSGSETPGT	7357	RSC-3047	GTAEAASASGTTGRAGE AEGATSAGATGPPGSP	7605
RSN-2783	GSAPTEAGRSAEATSAGA TGPATSGSETPGT	7358	RSC-2783	GTAEAASASGEAGRSAE ATSAGATGPPGSP	7606
RSN-3107	GSAPSASGTYSRGESGP GSPATSGSETPGT	7359	RSC-3107	GTAEAASASGSASGTY RGESGPGSPPGSP	7607
RSN-3103	GSAPSASGEAGRTDTHP GSPATSGSETPGT	7360	RSC-3103	GTAEAASASGSASGEAG RTDTHPGSPPGSP	7608
RSN-3102	GSAPSASGEPGRAAEHP GSPATSGSETPGT	7361	RSC-3102	GTAEAASASGSASGEP RAAEHPGSPPGSP	7609
RSN-3119	GSAPSPAGESSRGTTIA GSPATSGSETPGT	7362	RSC-3119	GTAEAASASGSPAGESS RGTTIAGSPPGSP	7610
RSN-3043	GSAPTTGEAGEAAGLTP AGLTGPATSGSETPGT	7363	RSC-3043	GTAEAASASGTTGEAGE AAGLT PAGLTGPPGSP	7611
RSN-2789	GSAPTEAGESAGAT PAGL TGPATSGSETPGT	7364	RSC-2789	GTAEAASASGEAGESAG AT PAGLTGPPGSP	7612
RSN-3109	GSAPSASGAPLELEAGP GSPATSGSETPGT	7365	RSC-3109	GTAEAASASGSASGAPL ELEAGPGSPPGSP	7613
RSN-3110	GSAPSASGEPPELGAGP GSPATSGSETPGT	7366	RSC-3110	GTAEAASASGSASGEP ELGAGPGSPPGSP	7614
RSN-3111	GSAPSASGEP SGLTEGP GSPATSGSETPGT	7367	RSC-3111	GTAEAASASGSASGEP SGLTEGPSPPGSP	7615
RSN-3112	GSAPSASGTAPLTEPP GSPATSGSETPGT	7368	RSC-3112	GTAEAASASGSASGTAP LTEPPGSPPGSP	7616

Name	Amino Acid Sequence	SEQ ID NO:	Name	Amino Acid Sequence	SEQ ID NO:
RSN-3113	GSAPSASGTPAELTEPP GSPATSGSETPGT	7369	RSC-3113	GTAEASASGSASGTPA ELTEPPGSPPGSP	7617
RSN-3114	GSAPSASGPPPGLTGFP GSPATSGSETPGT	7370	RSC-3114	GTAEASASGSASGPPP GLTGPPGSPPGSP	7618
RSN-3115	GSAPSASGTPAPLGGEPP GSPATSGSETPGT	7371	RSC-3115	GTAEASASGSASGTPA PLGGEPPGSPPGSP	7619
RSN-3125	GSAPSPAGAPEGLTGPA GSPATSGSETPGT	7372	RSC-3125	GTAEASASGSAPAGAPE GLTGPPGSPPGSP	7620
RSN-3126	GSAPSPAGPPEGLETEA GSPATSGSETPGT	7373	RSC-3126	GTAEASASGSAPGPPE GLETEAGSPPGSP	7621
RSN-3127	GSAPSPTSGQGGLTGPG SEPATSGSETPGT	7374	RSC-3127	GTAEASASGSPTSGQG GLTGPGSEPPGSP	7622
RSN-3131	GSAPSESAPPEGLETES TEPATSGSETPGT	7375	RSC-3131	GTAEASASGSSESAPPE GLETESTEPPGSP	7623
RSN-3132	GSAPSEGSEPLELGAAS ETPATSGSETPGT	7376	RSC-3132	GTAEASASGSSEGSEPL ELGAASETPPGSP	7624
RSN-3133	GSAPSEGSGPAGLEAPS ETPATSGSETPGT	7377	RSC-3133	GTAEASASGSSEGSGPA GLEAPSETPPGSP	7625
RSN-3138	GSAPSEPTPPASLEAEP GSPATSGSETPGT	7378	RSC-3138	GTAEASASGSSEPTPPA SLEAEPGSPPGSP	7626

[0241] In some embodiments, a fusion protein comprises an RS1 and an RS2 that have different rates of cleavage and different cleavage efficiencies to multiple proteases for which they are substrates. As a given protease may be found in different concentrations in a tumor, compared to healthy tissues or in circulation, the disclosure provides RSs that have a higher or lower cleavage efficiency for a given protease in order to ensure that a fusion protein is preferentially converted from the inactive form to the active form (i.e., by the separation and release of the cytokines and ELNNs from the fusion protein after cleavage of the RSs) when in proximity to the cancer cell or tissue and its co-localized proteases compared to the rate of cleavage of the RSs in healthy tissue or the circulation such that the released cytokine of the fusion proteins have a greater ability to bind to ligands in the tumor compared to the inactive form that remains in circulation. By such selective designs, the therapeutic index of the resulting compositions can be improved, resulting in reduced side effects relative to convention therapeutics that do not incorporate such site-specific activation.

[0242] In some embodiments, cleavage efficiency is the $\log_2$ value of the ratio of the percentage of the test substrate comprising the RS cleaved to the percentage of the control substrate AC1611 cleaved when each is subjected to the protease enzyme in biochemical assays in which reaction is conducted wherein the initial substrate concentration is 6 μ M, the reactions are incubated at 37°C for 2 hours before being stopped by adding EDTA, with the amount of digestion products and uncleaved substrate analyzed by non-reducing SDS-PAGE to establish the ratio of the percentage cleaved. The cleavage efficiency may be calculated as follows: $\text{Log}_2 \left(\frac{\% \text{ Cleaved for substrate of interest}}{\% \text{ cleaved for AC1611 in the same experiment}} \right)$. Thus, a cleavage efficiency of -1 means that the amount of test substrate cleaved was 50% compared to that of the control substrate, while a cleavage efficiency of +1 means that the amount of test

substrate cleaved was 200% compared to that of the control substrate. A higher rate of cleavage by the test protease relative to the control would result in a higher cleavage efficiency, and a slower rate of cleavage by the test protease relative to the control would result in a lower cleavage efficiency. A control RS sequence AC1611 (RSR-1517), having the amino acid sequence EAGRSANHEPLGLVAT (SEQ ID NO: 7001), was established as having an appropriate baseline cleavage efficiency by the proteases legumain, MMP-2, MMP-7, MMP-9, MMP-14, uPA, and matriptase, when tested in *in vitro* biochemical assays for rates of cleavage by the individual proteases. By selective substitution of amino acids at individual locations in the RS peptides, libraries of RS were created and evaluated against the panel of the 7 proteases, resulting in profiles that were used to establish guidelines for appropriate amino acid substitutions in order to achieve RS with desired cleavage efficiencies. In some embodiments, in making RSs with desired cleavage efficiencies, substitutions using the hydrophilic amino acids A, E, G, P, S, and T are preferred. however other L-amino acids can be substituted at given positions in order to adjust the cleavage efficiency so long as the RSs retain at least some susceptibility to cleavage by a given protease. Conservative substitutions of amino acids in a peptide to retain or effect activity is well within the knowledge and capabilities of a person within skill in the art. In some embodiments, the disclosure provides an RS in which the RS is cleaved by a protease including but not limited to MMP-2, MMP-7, MMP-9, MMP-14, uPA, or matriptase (also known as MT-SP1) with at least a 0.2 log₂, or 0.4 log₂, or 0.8 log₂, or 1.0 log₂ higher cleavage efficiency in an *in vitro* biochemical competitive assay compared to the cleavage by the same protease of a control sequence RSR-1517 having the sequence EAGRSANHEPLGLVAT (SEQ ID NO: 7001). In some embodiments, the disclosure provides an RS in which the RS is cleaved by a protease including but not limited to MMP-2, MMP-7, MMP-9, MMP-11, uPA, or matriptase with at least a 0.2 log₂, or 0.4 log₂, or 0.8 log₂, or 1.0 log₂ lower cleavage efficiency in an *in vitro* biochemical competitive assay compared to the cleavage by the same protease of a control sequence RSR-1517 having the sequence EAGRSANHEPLGLVAT (SEQ ID NO: 7001). In some embodiments, the disclosure provides an RS in which the rate of cleavage of the RS by a protease including but not limited to MMP-2, MMP-7, MMP-9, MMP-14, uPA, or matriptase is at least 2-fold, or at least 4-fold, or at least 8 fold, or at least 16-fold faster compared to the control sequence RSR-1517 having the sequence EAGRSANHEPLGLVAT (SEQ ID NO: 7001). In some embodiments, the disclosure provides an RS in which the rate of cleavage of the RS by a protease including but not limited to MMP-2, MMP-7, MMP-9, MMP-14, uPA, or matriptase is at least 2-fold, or at least 4-fold, or at least 8-fold, or at least 16-fold slower compared to the control sequence RSR-1517 having the sequence EAGRSANHEPLGLVAT (SEQ ID NO: 7001).

[0243] In some embodiments, the RS comprises the amino acid sequence EAGRSAXHTPAGLTGP (SEQ ID NO: 7627), wherein X is any amino acid other than N. In some embodiments, X is S. In some embodiments, X is T. In some embodiments, X is Y. In some embodiments, X is Q. In some

embodiments, X is G. In some embodiments, X is A. In some embodiments, X is V. In some embodiments, X is C. In some embodiments, X is P. In some embodiments, X is L. In some embodiments, X is I. In some embodiments, X is M. In some embodiments, X is F. In some embodiments, X is K. In some embodiments, X is R. In some embodiments, X is H. In some embodiments, X is D. In some embodiments, X is E. In some embodiments, the RS is not cleaved by legumain. In some embodiments, the RS is not cleavable by legumain in human blood, plasma, or serum. Advantageously, the RS amino acid sequence EAGRSAXHTPAGLTGP (SEQ ID NO: 7627), wherein X is any amino acid other than N, eliminates a glycosylation site on the sequence. Removal of said glycosylation site may improve accessibility of the release site to be acted upon by the tumor associate protease. Removal of said glycosylation site may improve expression of the fusion protein harboring said RS sequence.

[0244] In some embodiments, the RS is not cleavable upon incubation with about 1nM or less legumain for about 20 hours. In some embodiments, the RS is cleaved by legumain less quickly or efficiently than RSR-2295 (EAGRSANHTPAGLTGP; SEQ ID NO: 7048) is cleaved by legumain. In some embodiments, the RS is cleaved by legumain at a rate that is less than about 50% of the rate that legumain cleaves RSR-2295 (EAGRSANHTPAGLTGP; SEQ ID NO: 7048). In some embodiments, the RS is cleaved by legumain at a rate that is less than about 25% of the rate that legumain cleaves RSR-2295. In some embodiments, the RS is cleaved by legumain at a rate that is less than about 10% of the rate that legumain cleaves RSR-2295. In some embodiments, the RS is cleaved by legumain at a rate that is less than about 5% of the rate that legumain cleaves RSR-2295. In some embodiments, the RS is cleaved by legumain at a rate that is less than about 2.5% of the rate that legumain cleaves RSR-2295.

[0245] In some embodiments, the RS is cleaved by legumain at a rate that is less than about 50% of the rate that legumain cleaves RSR-2295 (EAGRSANHTPAGLTGP; SEQ ID NO: 7048) in human plasma. In some embodiments, the RS is cleaved by legumain at a rate that is less than about 25% of the rate that legumain cleaves RSR-2295 in human plasma. In some embodiments, the RS is cleaved by legumain at a rate that is less than about 10% of the rate that legumain cleaves RSR-2295 in human plasma. In some embodiments, the RS is cleaved by legumain at a rate that is less than about 5% of the rate that legumain cleaves RSR-2295 in human plasma. In some embodiments, the RS is cleaved by legumain at a rate that is less than about 2.5% of the rate that legumain cleaves RSR-2295 in human plasma.

[0246] In some embodiments, the disclosure provides fusion proteins comprising multiple RSs wherein each RS sequence is identified herein by the group of sequences set forth in Table 6a and the RSs are linked to each other by 1 to 6 amino acids that are glycine, serine, alanine, and threonine. In some embodiments, a fusion protein comprises a first RS and a second RS different from the first RS wherein each RS sequence is identified herein by a sequence set forth in Table 6a and the RSs are

linked to each other by 1 to 6 amino acids that are glycine, serine, alanine, and threonine. In some embodiments, the fusion protein comprises a first RS, a second RS different from the first RS, and a third RS different from the first and the second RS wherein each sequence is identified herein by a sequence set forth in Table 6a and the first and the second and the third RS are linked to each other by 1 to 6 amino acids that are glycine, serine, alanine, and threonine. In some embodiments, multiple RS of the fusion protein can be concatenated to form a sequence that can be cleaved by multiple proteases at different rates or efficiency of cleavage. In some embodiments, the disclosure provides a fusion protein comprising an RS1 and an RS2, wherein each has a sequence set forth in Table 6a or 6b and ELNNs (e.g., an ELNN1 and ELNN2), such as those described herein, wherein the RS1 is fused between the ELNN1 and the cytokines and the RS2 is fused between the ELNN2 and the cytokines. In some embodiments, a fusion protein is more readily cleaved in target tissues that express multiple proteases (e.g., tumor tissues), compared with healthy tissues or when in the normal circulation, with the result that the resulting fragments bearing the cytokines would more readily penetrate the target tissue; e.g., a tumor, and have an enhanced ability to bind and link the cancer cell and the effector cell.

[0247] In some embodiments, a fusion protein comprises a first release segment (RS1) positioned between a first ELNN and a cytokine. In some embodiments, the polypeptide further comprises a second release segment (RS2) positioned between the cytokine and a second ELNN. In some embodiments, RS1 and RS2 are identical in sequence. In some embodiments, RS1 and RS2 are not identical in sequence. In some embodiments, the RS1 comprises an amino acid sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to a sequence identified herein in Table 6a or 6b or a subset thereof. In some embodiments, the RS2 comprises an amino acid sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% sequence identity to a sequence identified herein in Table 6a or 6b or a subset thereof. In some embodiments, the RS1 and RS2 are each a substrate for cleavage by multiple proteases at one, two, or three cleavage sites within each release segment sequence.

[0248] In some embodiments, the fusion protein further comprises one or more reference fragments (e.g., barcode fragments) releasable from the fusion protein upon digestion by the protease. In some embodiments, the one or more reference fragments is a single reference fragment that differs in sequence and molecular weight from all other peptide fragments that are releasable from the polypeptide upon digestion of the polypeptide by the protease.

Exemplary Fusion Proteins

[0249] In some embodiments, a fusion protein comprises an amino acid sequence having at least (about) 80% sequence identity to a sequence set forth in Table A1 and Table A2 (SEQ ID NOs: 1000-1042) or a subset thereof. In some embodiments, the fusion protein comprises an amino acid sequence having at least (about) 81%, at least (about) 82%, at least (about) 83%, at least (about) 84%, at least (about) 85%, at least (about) 86%, at least (about) 87%, at least (about) 88%, at least (about) 89%, at

least (about) 90%, at least (about) 91%, at least (about) 92%, at least (about) 93%, at least (about) 94%, at least (about) 95%, at least (about) 96%, at least (about) 97%, at least (about) 98%, at least (about) 99%, or (about) 100% sequence identity to a sequence set forth in SEQ ID NOs: 1000-1042 or a subset thereof. In some embodiments, the fusion protein comprises an amino acid sequence having at least (about) 90%, at least (about) 91%, at least (about) 92%, at least (about) 93%, at least (about) 94%, at least (about) 95%, at least (about) 96%, at least (about) 97%, at least (about) 98%, at least (about) 99%, or (about) 100% sequence identity to a sequence set forth in SEQ ID NOs: 1000-1042 or a subset thereof. In some embodiments, the fusion protein comprises an amino acid sequence identical to a sequence set forth in SEQ ID NOs: 1000-1042.

[0250] In some embodiments, a fusion protein comprises a first polypeptide subunit comprising an amino acid sequence having at least (about) 81%, at least (about) 82%, at least (about) 83%, at least (about) 84%, at least (about) 85%, at least (about) 86%, at least (about) 87%, at least (about) 88%, at least (about) 89%, at least (about) 90%, at least (about) 91%, at least (about) 92%, at least (about) 93%, at least (about) 94%, at least (about) 95%, at least (about) 96%, at least (about) 97%, at least (about) 98%, at least (about) 99%, or (about) 100% sequence identity to a sequence set forth in SEQ ID NO: 1000, and a second polypeptide subunit comprising an amino acid sequence having at least (about) 81%, at least (about) 82%, at least (about) 83%, at least (about) 84%, at least (about) 85%, at least (about) 86%, at least (about) 87%, at least (about) 88%, at least (about) 89%, at least (about) 90%, at least (about) 91%, at least (about) 92%, at least (about) 93%, at least (about) 94%, at least (about) 95%, at least (about) 96%, at least (about) 97%, at least (about) 98%, at least (about) 99%, or (about) 100% sequence identity to a sequence set forth in SEQ ID NO: 1033.

[0251] In some embodiments, a fusion protein comprises a first polypeptide subunit comprising an amino acid sequence having at least (about) 81%, at least (about) 82%, at least (about) 83%, at least (about) 84%, at least (about) 85%, at least (about) 86%, at least (about) 87%, at least (about) 88%, at least (about) 89%, at least (about) 90%, at least (about) 91%, at least (about) 92%, at least (about) 93%, at least (about) 94%, at least (about) 95%, at least (about) 96%, at least (about) 97%, at least (about) 98%, at least (about) 99%, or (about) 100% sequence identity to a sequence set forth in SEQ ID NO: 1001, and a second polypeptide subunit comprising an amino acid sequence having at least (about) 81%, at least (about) 82%, at least (about) 83%, at least (about) 84%, at least (about) 85%, at least (about) 86%, at least (about) 87%, at least (about) 88%, at least (about) 89%, at least (about) 90%, at least (about) 91%, at least (about) 92%, at least (about) 93%, at least (about) 94%, at least (about) 95%, at least (about) 96%, at least (about) 97%, at least (about) 98%, at least (about) 99%, or (about) 100% sequence identity to a sequence set forth in SEQ ID NO: 1033.

[0252] In some embodiments, a fusion protein comprises a first polypeptide subunit comprising an amino acid sequence having at least (about) 81%, at least (about) 82%, at least (about) 83%, at least (about) 84%, at least (about) 85%, at least (about) 86%, at least (about) 87%, at least (about) 88%, at

least (about) 89%, at least (about) 90%, at least (about) 91%, at least (about) 92%, at least (about) 93%, at least (about) 94%, at least (about) 95%, at least (about) 96%, at least (about) 97%, at least (about) 98%, at least (about) 99%, or (about) 100% sequence identity to a sequence set forth in SEQ ID NO: 1005, and a second polypeptide subunit comprising an amino acid sequence having at least (about) 81%, at least (about) 82%, at least (about) 83%, at least (about) 84%, at least (about) 85%, at least (about) 86%, at least (about) 87%, at least (about) 88%, at least (about) 89%, at least (about) 90%, at least (about) 91%, at least (about) 92%, at least (about) 93%, at least (about) 94%, at least (about) 95%, at least (about) 96%, at least (about) 97%, at least (about) 98%, at least (about) 99%, or (about) 100% sequence identity to a sequence set forth in SEQ ID NO: 1036.

[0256] In some embodiments, a fusion protein comprises a first polypeptide subunit comprising an amino acid sequence having at least (about) 81%, at least (about) 82%, at least (about) 83%, at least (about) 84%, at least (about) 85%, at least (about) 86%, at least (about) 87%, at least (about) 88%, at least (about) 89%, at least (about) 90%, at least (about) 91%, at least (about) 92%, at least (about) 93%, at least (about) 94%, at least (about) 95%, at least (about) 96%, at least (about) 97%, at least (about) 98%, at least (about) 99%, or (about) 100% sequence identity to a sequence set forth in SEQ ID NO: 1000, and a second polypeptide subunit comprising an amino acid sequence having at least (about) 81%, at least (about) 82%, at least (about) 83%, at least (about) 84%, at least (about) 85%, at least (about) 86%, at least (about) 87%, at least (about) 88%, at least (about) 89%, at least (about) 90%, at least (about) 91%, at least (about) 92%, at least (about) 93%, at least (about) 94%, at least (about) 95%, at least (about) 96%, at least (about) 97%, at least (about) 98%, at least (about) 99%, or (about) 100% sequence identity to any one sequence set forth in SEQ ID NOs: 1033-1036.

[0257] In some embodiments, a fusion protein comprises a first polypeptide subunit comprising an amino acid sequence having at least (about) 81%, at least (about) 82%, at least (about) 83%, at least (about) 84%, at least (about) 85%, at least (about) 86%, at least (about) 87%, at least (about) 88%, at least (about) 89%, at least (about) 90%, at least (about) 91%, at least (about) 92%, at least (about) 93%, at least (about) 94%, at least (about) 95%, at least (about) 96%, at least (about) 97%, at least (about) 98%, at least (about) 99%, or (about) 100% sequence identity to a sequence set forth in SEQ ID NO: 1001, and a second polypeptide subunit comprising an amino acid sequence having at least (about) 81%, at least (about) 82%, at least (about) 83%, at least (about) 84%, at least (about) 85%, at least (about) 86%, at least (about) 87%, at least (about) 88%, at least (about) 89%, at least (about) 90%, at least (about) 91%, at least (about) 92%, at least (about) 93%, at least (about) 94%, at least (about) 95%, at least (about) 96%, at least (about) 97%, at least (about) 98%, at least (about) 99%, or (about) 100% sequence identity to any one sequence set forth in SEQ ID NOs: 1033-1036.

[0258] In some embodiments, a fusion protein comprises a first polypeptide subunit comprising an amino acid sequence having at least (about) 81%, at least (about) 82%, at least (about) 83%, at least (about) 84%, at least (about) 85%, at least (about) 86%, at least (about) 87%, at least (about) 88%, at

least (about) 89%, at least (about) 90%, at least (about) 91%, at least (about) 92%, at least (about) 93%, at least (about) 94%, at least (about) 95%, at least (about) 96%, at least (about) 97%, at least (about) 98%, at least (about) 99%, or (about) 100% sequence identity to a sequence set forth in SEQ ID NO: 1002, and a second polypeptide subunit comprising an amino acid sequence having at least (about) 81%, at least (about) 82%, at least (about) 83%, at least (about) 84%, at least (about) 85%, at least (about) 86%, at least (about) 87%, at least (about) 88%, at least (about) 89%, at least (about) 90%, at least (about) 91%, at least (about) 92%, at least (about) 93%, at least (about) 94%, at least (about) 95%, at least (about) 96%, at least (about) 97%, at least (about) 98%, at least (about) 99%, or (about) 100% sequence identity to any one sequence set forth in SEQ ID NOs: 1033-1036.

[0259] In some embodiments, a fusion protein comprises a first polypeptide subunit comprising an amino acid sequence having at least (about) 81%, at least (about) 82%, at least (about) 83%, at least (about) 84%, at least (about) 85%, at least (about) 86%, at least (about) 87%, at least (about) 88%, at least (about) 89%, at least (about) 90%, at least (about) 91%, at least (about) 92%, at least (about) 93%, at least (about) 94%, at least (about) 95%, at least (about) 96%, at least (about) 97%, at least (about) 98%, at least (about) 99%, or (about) 100% sequence identity to a sequence set forth in SEQ ID NO: 1003, and a second polypeptide subunit comprising an amino acid sequence having at least (about) 81%, at least (about) 82%, at least (about) 83%, at least (about) 84%, at least (about) 85%, at least (about) 86%, at least (about) 87%, at least (about) 88%, at least (about) 89%, at least (about) 90%, at least (about) 91%, at least (about) 92%, at least (about) 93%, at least (about) 94%, at least (about) 95%, at least (about) 96%, at least (about) 97%, at least (about) 98%, at least (about) 99%, or (about) 100% sequence identity to any one sequence set forth in SEQ ID NOs: 1033-1036.

[0260] In some embodiments, a fusion protein comprises a first polypeptide subunit comprising an amino acid sequence having at least (about) 81%, at least (about) 82%, at least (about) 83%, at least (about) 84%, at least (about) 85%, at least (about) 86%, at least (about) 87%, at least (about) 88%, at least (about) 89%, at least (about) 90%, at least (about) 91%, at least (about) 92%, at least (about) 93%, at least (about) 94%, at least (about) 95%, at least (about) 96%, at least (about) 97%, at least (about) 98%, at least (about) 99%, or (about) 100% sequence identity to a sequence set forth in SEQ ID NO: 1004, and a second polypeptide subunit comprising an amino acid sequence having at least (about) 81%, at least (about) 82%, at least (about) 83%, at least (about) 84%, at least (about) 85%, at least (about) 86%, at least (about) 87%, at least (about) 88%, at least (about) 89%, at least (about) 90%, at least (about) 91%, at least (about) 92%, at least (about) 93%, at least (about) 94%, at least (about) 95%, at least (about) 96%, at least (about) 97%, at least (about) 98%, at least (about) 99%, or (about) 100% sequence identity to any one sequence set forth in SEQ ID NOs: 1033-1036.

[0261] In some embodiments, a fusion protein comprises a first polypeptide subunit comprising an amino acid sequence having at least (about) 81%, at least (about) 82%, at least (about) 83%, at least (about) 84%, at least (about) 85%, at least (about) 86%, at least (about) 87%, at least (about) 88%, at

least (about) 89%, at least (about) 90%, at least (about) 91%, at least (about) 92%, at least (about) 93%, at least (about) 94%, at least (about) 95%, at least (about) 96%, at least (about) 97%, at least (about) 98%, at least (about) 99%, or (about) 100% sequence identity to a sequence set forth in SEQ ID NO: 1005, and a second polypeptide subunit comprising an amino acid sequence having at least (about) 81%, at least (about) 82%, at least (about) 83%, at least (about) 84%, at least (about) 85%, at least (about) 86%, at least (about) 87%, at least (about) 88%, at least (about) 89%, at least (about) 90%, at least (about) 91%, at least (about) 92%, at least (about) 93%, at least (about) 94%, at least (about) 95%, at least (about) 96%, at least (about) 97%, at least (about) 98%, at least (about) 99%, or (about) 100% sequence identity to any one sequence set forth in SEQ ID NOs: 1033-1036.

[0262] It is specifically contemplated that the compositions of this disclosure can comprise sequence variants of the amino acid sequences set forth in Table A, such as with linker sequence(s) substituted or inserted or with purification tag sequence(s) attached thereto, so long as the variants exhibit substantially similar or same bioactivity/bioactivities and/or activation mechanism(s).

Table A1: Exemplary amino acid sequences of IL-12p35 polypeptides

SEQ ID NO:	AMINO ACID SEQUENCE
1000 (P72H Sub35)	ASGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSEFATSGS ETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEFATSGSE TPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESG PGTSESATPESGPGTSESATPESGPGSEFATSGSETPGSEFATSGSETPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGSEFATSGSETPGTSESATEAGRSASHTPAGLTGPGT PESATPESRNLPVATPDPMFPCLLHHSQNLRLRAVSNMLQKARQTLEFYPTCTSEEDHEDIT KDKTSTVEACLPLELTKNESCLNSRETSFITNGSCLASRKTSFMMALCLSSIIYEDLKMYQV EFKTMNAKLLMDPKRQIFLDQNMLAVIDELMQALNFNSETVPPQKSLEEDPFYKTKIKLCI LLHAFRIRAVTIDRVMSYLNASGTATPESAPEAGRSASHTPAGLTGPATPESGPGTSTEPS EGSAPGTSESATPESGPGSEFATSGSETPGTSESATPESGPGSEFATSGSETPGTSESATP ESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEFATSGSETPGTSESATPE SGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPES GPGTSASATPESGPGSEFATSGSETPGSEFATSGSETPGSPAGSPTSTEEGTSTEPSEGS PGTSTEPSEGSAPGSEFATSGSETPGPEA
1001 (P73H Sub35)	ASGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSEFATSGS ETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEFATSGSE TPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESG PGTSESATPESGPGTSESATPESGPGSEFATSGSETPGSEFATSGSETPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGSEFATSGSETPGTSESATEAGRSASHTPAGLTGPGT PESATPESRNLPVATPDPMFPCLLHHSQNLRLRAVSNMLQKARQTLEFYPTCTSEEDHEDIT KDKTSTVEACLPLELTKNESCLNSRETSFITNGSCLASRKTSFMMALCLSSIIYEDLKMYQV EFKTMNAKLLMDPKRQIFLDQNMLAVIDELMQALNFNSETVPPQKSLEEDPFYKTKIKLCI LLHAFRIRAVTIDRVMSYLSAGTATPESAPEAGRSASHTPAGLTGPATPESGPGTSTEPS EGSAPGTSESATPESGPGSEFATSGSETPGTSESATPESGPGSEFATSGSETPGTSESATP ESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEFATSGSETPGTSESATPE SGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPES GPGTSASATPESGPGSEFATSGSETPGSEFATSGSETPGSPAGSPTSTEEGTSTEPSEGS PGTSTEPSEGSAPGSEFATSGSETPGPEA
1002 (P74H Sub35)	ASGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSEFATSGS ETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEFATSGSE TPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESG PGTSESATPESGPGTSESATPESGPGSEFATSGSETPGSEFATSGSETPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGSEFATSGSETPGTSESATEAGRSASHTPAGLTGPGT PESATPESRNLPVATPDPMFPCLLHHSQNLRLRAVSNMLQKARQTLEFYPTCTSEEDHEDIT

SEQ ID NO:	AMINO ACID SEQUENCE
	KDKTSTVEACLPLELTKNESCLNSRETSTFITNGSCLASRKTSFMMALCLSSIIYEDLKMYQV EFKTMNAKLMLDPKRQIFLDQNMLAVIDELMQALNFNSETVTPQKSSLEEDPFYKTKIKLCI LLHAFRIRAVTIDRVMSYLNASGTATPESSPGEAGRSASHTPAGLTGPATPESGPGTSTEPS EGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATP ESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP SGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPES GPGTSASATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSA PGTSTEPSEGSAPGSEPATSGSETPGPEA
1003 (P31H Sub35)	ASSATPESGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSE PATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSE ATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSES ATPESGPGTSESATPESGPGTSESATPEAGRSASHTPAGLTGPATPESATPESRNLPVATP DPGMFPCLLHHSQNLRLRAVSNMLQKARQTLEFYPTCTSEEDHEDITKDKTSTVEACLPLELT KNESCLNSRETSTFITNGSCLASRKTSFMMALCLSSIIYEDLKMYQVEFKTMNAKLMLDPKRQ IFLDQNMLAVIDELMQALNFNSETVTPQKSSLEEDPFYKTKIKLCILLHAFRIRAVTIDRV SYLNASGTATPESSPGEAGRSASHTPAGLTGPATPESGPGTSTEPSEGSAPGSPAGSPTST EEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTST EGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSASATPESGPGSEPATSGSET GSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATGEPEA
1004 (P27H Sub35)	ASSATPESGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSE PATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSE ATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSES ATPESGPGTSESATPESGPGTSESATPEAGRSASHTPAGLTGPATPESATPESRNLPVATP DPGMFPCLLHHSQNLRLRAVSNMLQKARQTLEFYPTCTSEEDHEDITKDKTSTVEACLPLELT KNESCLNSRETSTFITNGSCLASRKTSFMMALCLSSIIYEDLKMYQVEFKTMNAKLMLDPKRQ IFLDQNMLAVIDELMQALNFNSETVTPQKSSLEEDPFYKTKIKLCILLHAFRIRAVTIDRV SYLNASGTATPESSPGEAGRSASHTPAGLTGPATPESGPGTSTEPSEGSAPGSPAGSPTST EEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTST EGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSASATPESGPGSEPATSGSET GSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATGEPEA
1005 (P25H Sub35)	ASGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESG PGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPEAGRSASHTPAGLTGP TPESATPESRNLPVATPDGMFPCLLHHSQNLRLRAVSNMLQKARQTLEFYPTCTSEEDHEDIT TKDKTSTVEACLPLELTKNESCLNSRETSTFITNGSCLASRKTSFMMALCLSSIIYEDLKMYQ VEFKTMNAKLMLDPKRQIFLDQNMLAVIDELMQALNFNSETVTPQKSSLEEDPFYKTKIKLC ILLHAFRIRAVTIDRVMSYLNASGTATPESSPGEAGRSASHTPAGLTGPATPESGPGTST PSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESA TPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESA PESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATP ESGPGTSASATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSE SAPGTSTEPSEGSAPGSEPATSGSETPGPEA
1006 (P01H Sub35)	HHHHHHHGSAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTST PSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGS PTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPS EGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSG SETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPEAGRSANHT PAGLTGPGTAEASASGRNLPVATPDGMFPCLLHHSQNLRLRAVSNMLQKARQTLEFYPTCT SEEDHEDITKDKTSTVEACLPLELTKNESCLNSRETSTFITNGSCLASRKTSFMMALCLSSI YEDLKMYQVEFKTMNAKLMLDPKRQIFLDQNMLAVIDELMQALNFNSETVTPQKSSLEEDPF YKTKIKLCILLHAFRIRAVTIDRVMSYLNASGTAEASASG
1007 (P71M Sub35)	ASGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESG PGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPEAGRSASHTPAGLTGP PESATPESRVIPVSGPARCLSQSRNLLKTTDDMVKTAREKLKHYSCTAEDIDHEDITRDQT

SEQ ID NO:	AMINO ACID SEQUENCE
	STLKTCLPLELHKNESCLATRETSSTTRGSCLPQKTSLMMTLCGLSIYEDLKMYQTEFQA INAALQNNHHQIILDKGMLVAIDELMQSLNHNGETLRQKPPVGEADPYRVKMKLCILLHA FSTRVVTINRVMGYLNASGTATPESAPEAGRSASHTPAGLTGPATPESGPGTSTEPSEGS PGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESG GTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPG SPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT SASATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTS TEPSEGSAPGSEPATSGSETPGEPEA
1008 (P70M Sub35)	ASGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESG PGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATEAGRSASHTPAGLTGPGT PESATPESRVI PVSGPARCLSQSRNLLKTDDMVKTAREKLKHYSCTAEDIDHEDITRDQT STLKTCLPLELHKNESCLATRETSSTTRGSCLPQKTSLMMTLCGLSIYEDLKMYQTEFQA INAALQNNHHQIILDKGMLVAIDELMQSLNHNGETLRQKPPVGEADPYRVKMKLCILLHA FSTRVVTINRVMGYLSSAGTATPESAPEAGRSASHTPAGLTGPATPESGPGTSTEPSEGS PGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGP GTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPG SPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT SASATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTS TEPSEGSAPGSEPATSGSETPGEPEA
1009 (P30M Sub35)	ASSATPESGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSE PATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSE ATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSES ATPESGPGTSESATPESGPGTSESATPEAGRSASHTPAGLTGPGT PESATPESRVI PVSGP ARCLSQSRNLLKTDDMVKTAREKLKHYSCTAEDIDHEDITRDQTSTLKTCLPLELHKNES CLATRETSSTTRGSCLPQKTSLMMTLCGLSIYEDLKMYQTEFQA INAALQNNHHQIILD KGMLVAIDELMQSLNHNGETLRQKPPVGEADPYRVKMKLCILLHAFSTRVVTINRVMGYLS SAGTATPESSPGEAGRSASHTPAGLTGPATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT SESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTS TEPSEGSAPGTSESATPESGPGTSESATPESGPGTSASATPESGPGSEPATSGSETPGSEP ATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATGPEA
1030 (P26M Sub35)	ASSATPESGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSE PATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSE ATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSES ATPESGPGTSESATPESGPGTSESATPEAGRSASHTPAGLTGPGT PESATPESRVI PVSGP ARCLSQSRNLLKTDDMVKTAREKLKHYSCTAEDIDHEDITRDQTSTLKTCLPLELHKNES CLATRETSSTTRGSCLPQKTSLMMTLCGLSIYEDLKMYQTEFQA INAALQNNHHQIILD KGMLVAIDELMQSLNHNGETLRQKPPVGEADPYRVKMKLCILLHAFSTRVVTINRVMGYLS SAGTATPESSPGEAGRSASHTPAGLTGPATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGT SESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTS TEPSEGSAPGTSESATPESGPGTSESATPESGPGTSASATPESGPGSEPATSGSETPGSEP ATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATGPEA
1031 (P24M Sub35)	ASGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESG PGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPEAGRSASHTPAGLTGPG TPESATPESRVI PVSGPARCLSQSRNLLKTDDMVKTAREKLKHYSCTAEDIDHEDITRDQ TSTLKTCLPLELHKNESCLATRETSSTTRGSCLPQKTSLMMTLCGLSIYEDLKMYQTEFQ A INAALQNNHHQIILDKGMLVAIDELMQSLNHNGETLRQKPPVGEADPYRVKMKLCILLH AFSTRVVTINRVMGYLSSAGTATPESSPGEAGRSASHTPAGLTGPATPESGPGTSTEPSE SAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPES GPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESG PGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGP GTSASATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGSEPATSGSETPGEPEA

SEQ ID NO:	AMINO ACID SEQUENCE
1032 (P01M Sub35)	GSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPG TSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGT SESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGSPAGSPTSTEEGTSTEPSEGSAPGTS TEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTST EPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPEAGRSANHTPAGLTGPG TAAEASASGMWELEKDVYVVEVDWTPDAPGETVNLTCDTPEEDDITWTSDQRHGVIGSGKT LITITVKEFLDAGQYTCHKGGETLSHSHLLHKKENGIWSTEILKNFKNKTFLKCEAPNYS G RFTCSWLVRNMDLKFNKSSSSSPDSRAVTCGMASLSAEKVTLDQRDYEKYSVSCQEDVT CPTAEETLPIELALEARQQNKYENYSTSF FIRDIIKPDPPKNLQMKPLKNSQVEVSWEY PD SWSTPHSYFSLKFFVRIQRKKEKMKETEEGCNQKGAFLEKTSSTEVQCKGNGVCQAQDRY YNSSCSKWACVPCRVRSGTAAEASASGEAGRSANHTPAGLTGPGTSESATPESGPGSE PAT SGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGS P TSSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGS PT STEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSE PATSGS ET PGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSE PATSGSE TPGTSESATPESGPGTSTEPSEGSAP

Table A2: Exemplary amino acid sequences of IL-12p40 polypeptides

SEQ ID NO:	AMINO ACID SEQUENCE
1033 (P72H/P73 H/P74H Sub40)	ASGPGTSTEPSEGSAPGTSESATPESGPGSGPATSESATPGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESG PGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATEAGRSASHTPAGLTGPTS ESATPESIWEKKDVYVVELDWYPDAPGEMVVLTCDTPEEDGITWTLDQSSEVLGSGKTLT IQVKEFGDAGQYTCHKGGEVLSHSLLLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYS GRFTCWLLTTISTDLTFSVKSSRGSSDPQGVTCGAATLSAERVGRDNKEYEYSVEQCEDSA CPAAEESLPIEVMVDAVHKLKYENYTSFF FIRDIIKPDPPKNLQLKPLKNSRQVEVSWEY PDTWSTPHSYFSLTFCVQVGKSKREKKDRVFTDKTSATVICRKNASISVRAQDRYSSSSWS EWASVPCSGTATPESGPGSEAGRSASHTPAGLTGPATPESGPGSPAGSPTSTEEGSPAGS PT STEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTST EEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTST EGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSPSATPESGPGSEPATSGSET PGSEPATSGSETPGSPAG
1034 (P31H Sub40)	ASSATPESGPGTSTEPSEGSAPGTSESATPESGPGSGPATSESATPGTSESATPESGPGSE PATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSE PATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSES ATPESGPGTSESATPESGPGTSESATPEAGRSASHTPAGLTGPTSSESATPESIWEKKDVY VVELDWYPDAPGEMVVLTCDTPEEDGITWTLDQSSEVLGSGKTLTIQVKEFGDAGQYTCHK GGEVLSHSLLLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRFTCWLLTTISTDLT FSVKSSRGSSDPQGVTCGAATLSAERVGRDNKEYEYSVEQCEDSACPAAEESLPIEVMVDA VHKLKYENYTSFF FIRDIIKPDPPKNLQLKPLKNSRQVEVSWEYPDWSTPHSYFSLTFCV QVGKSKREKKDRVFTDKTSATVICRKNASISVRAQDRYSSSSWSEWASVPCSGTATPES GTPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAPGTSTEPS EGSAPGTSESATPESGPGPSESATPESGPGSPAGSPTSTEEGTSESATPESGPGSEPATSG SETPGTSESATPESGPGT
1035 (P27H Sub40)	ASSATPESGPGTSTEPSEGSAPGTSESATPESGPGSGPATSESATPGTSESATPESGPGSE PATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSE PATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSES ATPESGPGTSESATPESGPGTSESATPEAGRSASHTPAGLTGPTSSESATPESIWEKKDVY VVELDWYPDAPGEMVVLTCDTPEEDGITWTLDQSSEVLGSGKTLTIQVKEFGDAGQYTCHK GGEVLSHSLLLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRFTCWLLTTISTDLT FSVKSSRGSSDPQGVTCGAATLSAERVGRDNKEYEYSVEQCEDSACPAAEESLPIEVMVDA VHKLKYENYTSFF FIRDIIKPDPPKNLQLKPLKNSRQVEVSWEYPDWSTPHSYFSLTFCV QVGKSKREKKDRVFTDKTSATVICRKNASISVRAQDRYSSSSWSEWASVPCSGTATPESG

[illegible]

SEQ ID NO:	AMINO ACID SEQUENCE
	QNKYENYSTSFFIRDIIKPDPPKNLQMKPLKNSQVEVSWEYPDSWSTPHSYFSLKFFVRIQ RKKEKMKETEEGCNQKGAFLVEKTSTEVQCKGGNVCVQAQDRYNNSSCSKWACVPCRVRSA PGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSTEPSEGSAP GTSTEPSEGSAPGTSESATPESGPGPSESATPESGPGSPAGSPTSTEEGTSESATPESGPG SEPATSGSETPGTSESATPESGPGT
1040 (P26M Sub40)	ASSATPESGPGTSTEPSEGSAPGTSESATPESGPGSGPATSESATPGTSESATPESGPGSE PATSGSETPGTSESATPESGPGTSTEPSEGSAPGS PAGESPTSTEEGTSESATPESGPGSE ATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSES ATPESGPGTSESATPESGPGTSESATPEAGRSASHTPAGLTGPTSESATPESMWELEKDVY VVEVDWTPDAPGETVNLTCDDTPEEDDITWTSQQRHGVIGSGKTLTITVKEFLDAGQYTCHK GGETLSHSHLLHKKKENGWSTEILKNFKNKTFLKCEAPNYSGRFTCSWLVRNMDLKFN KSSSSSPDSRAVTCGMASLSAEKVTLDQRDYEKYSVSCQEDVTCPTAEETLP IELALEARQ QNKYENYSTSFFIRDIIKPDPPKNLQMKPLKNSQVEVSWEYPDSWSTPHSYFSLKFFVRIQ RKKEKMKETEEGCNQKGAFLVEKTSTEVQCKGGNVCVQAQDRYNNSSCSKWACVPCRVRSG TATPESGPGSEAGRSASHTPAGLTGPATPESGPGSEPATSGSETPGTSESATPESGPGSEPA TSGSETPGTSESATPESGPGTSTEPSEGSAPGS PAGESPTSTEEGTSESATPESGPGSEPA SGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESAT PESGPGTSESATPESGPGTSPSATPESGPGSEPATSGSETPGSEPATSGSET
1041 (P24M Sub40)	ASGPGTSTEPSEGSAPGTSESATPESGPGSGPATSESATPGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGTSTEPSEGSAPGS PAGESPTSTEEGTSESATPESGPGSEPATSGSE TPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESG PGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEE GTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESATPEAGRSASHTPAGLTGPT SESATPESMWELEKDVYVVEVDWTPDAPGETVNLTCDDTPEEDDITWTSQQRHGVIGSGKTL TITVKEFLDAGQYTCHKGGETLSHSHLLHKKKENGWSTEILKNFKNKTFLKCEAPNYSGR FTCSWLVRNMDLKFNKSSSSSPDSRAVTCGMASLSAEKVTLDQRDYEKYSVSCQEDVTC PTAEETLP IELALEARQQNKYENYSTSFFIRDIIKPDPPKNLQMKPLKNSQVEVSWEYPDS WSTPHSYFSLKFFVRIQRKKEKMKETEEGCNQKGAFLVEKTSTEVQCKGGNVCVQAQDRY NNSSCSKWACVPCRVRSGTATPESGPGSEAGRSASHTPAGLTGPATPESGPGSPAGSPTSTEE GSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPG SEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGS PAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSP AGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSPSATPESGPGSE ATSGSETPGSEPATSGSETPGSPAG
1042 (P01M Sub40)	HHHHHHHHGSPAGSPTSTEEGTSESATPESGPGTSTEPSEGSAPGS PAGESPTSTEEGTSTE PSEGSAPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGS PTSTEEGTSESATPESGPGTSTEPSEGSAPGTSTEPSEGSAPGS PAGESPTSTEEGTSTEPS EGSAPGTSTEPSEGSAPGTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSG SETPGTSTEPSEGSAPGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGSEAGRSANHT PAGLTGPGTAAEASASGRVIVPVGPARCLSQSRNLKTTDDMVKTAREKLKHYSCTAEDID HEDITRDQSTLKTCLPLELHKNESCLATRETSSTTRGSCLPPQKTSLMMTLCGLSIYEDL KMYQTEFQAINAALQNNHQQIILDKGMLVAIDELMQSLNHNGETLRQKPPVGEADPYRVK MKLCILHAFSTRVVTINRVMGYLSSAGTAAEASASGEAGRSANHTPAGLTGPGTSESATP ESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSE SAPGS PAGESPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTST EEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESG PGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAP GSEPATSGSETPGTSESATPESGPGTSTEPSEGSAP

RECOMBINANT PRODUCTION

[0263] Also provided are polynucleotides that encode any polypeptide disclosed herein and/or the reverse complements of such polynucleotides.

[0264] The disclosure herein includes an expression vector that comprises a polynucleotide sequence, such as any described in the preceding paragraph, and a regulatory sequence operably linked to the polynucleotide sequence.

[0265] The disclosure herein includes a host cell comprising an expression vector, such as described any in the preceding paragraph. In some embodiments, the host cell is a prokaryote. In some embodiments, the host cell is *E. coli*. In some embodiments, the host cell is a mammalian cell. In some embodiments, the mammalian cell is a mammalian cell line, such as a HEK 293 cell line or a CHO cell line.

[0266] In some embodiments, the disclosure provides methods of manufacturing the subject compositions. In some embodiments, such a method comprises culturing a host cell comprising a nucleic acid construct that encodes a fusion protein described herein under conditions that promote the expression of the fusion protein, followed by recovery of the fusion protein using standard purification methods (e.g., column chromatography, HPLC, and the like) wherein the composition is recovered wherein at least 70%, or at least 80%, or at least 90%, or at least 95%, or at least 97%, or at least 99% of the fusion proteins are correctly folded. In some embodiments of the method of making, the expressed fusion protein is recovered in which at least or at least 90%, or at least 95%, or at least 97%, or at least 99% of the fusion protein is recovered in monomeric, soluble form.

[0267] In some embodiments, the disclosure relates to methods of making a fusion protein at high fermentation expression levels of functional protein using an *E. coli* or mammalian host cell, as well as providing expression vectors encoding the fusion protein useful in methods to produce the cytotoxically active fusion protein at high expression levels. In some embodiments, the method comprises the steps of 1) preparing a polynucleotide encoding a fusion protein disclosed herein, 2) cloning the polynucleotide into an expression vector, which can be a plasmid or other vector under the control of appropriate transcription and translation sequences for high level protein expression in a biological system, 3) transforming an appropriate host cell with the expression vector, and 4) culturing the host cell in conventional nutrient media under conditions suitable for the expression of the fusion protein. Where desired, the host cell is *E. coli*. As used herein, the term “correctly folded” means that the cytokine component of the composition has the ability to specifically bind the cytokine target ligand (e.g., a cytokine receptor and upon activation). In some embodiments, the disclosure provides a method for producing a fusion protein, the method comprising culturing in a fermentation reaction a host cell that comprises a vector encoding the fusion protein under conditions effective to express the fusion protein product.

PHARMACEUTICAL COMPOSITION

[0268] Disclosed herein includes a pharmaceutical composition comprising a polypeptide (such as a fusion protein disclosed herein), and one or more pharmaceutically acceptable excipients. In some embodiments, the pharmaceutical composition is formulated for intradermal, subcutaneous, intravenous, intra-arterial, intraabdominal, intraperitoneal, intravitreal, intrathecal, or intramuscular administration. In some embodiments, the pharmaceutical composition is formulated for intravenous injection. In some embodiments, the pharmaceutical composition is in a liquid form or frozen. In some embodiments, the pharmaceutical composition is formulated as a lyophilized powder to be reconstituted prior to administration.

[0269] The pharmaceutical compositions can be administered for therapy by any suitable route. In some embodiments, the dose is administered intradermally, subcutaneously, intravenously (e.g., by intravenous bolus or intravenous infusion), intra-arterially, intra-abdominally, intraperitoneally, intrathecally, or intramuscularly. In some embodiments, the subject is a mouse, rat, monkey, or human. In preferred embodiments, the subject is a human.

[0270] In some embodiments, the pharmaceutical composition can be administered subcutaneously, intramuscularly, or intravenously. In some embodiments, the pharmaceutical composition is administered at a therapeutically effective amount. In some embodiments, the therapeutically effective amount results in a gain in time spent within a therapeutic window for the fusion protein compared to the corresponding cytokine of the fusion protein not linked to the ELNN and administered at a comparable dose to a subject.

[0271] In some embodiments, the pharmaceutical composition is administered subcutaneously. In some embodiments, the pharmaceutical composition is administered intravenously.

[0272] In some embodiments, the composition may be supplied as a lyophilized powder or cake to be reconstituted prior to administration. In some embodiments, the composition may also be supplied in a liquid form or frozen, which can be administered directly to a subject.

PHARMACEUTICAL KITS

[0273] In some embodiments, the present disclosure provides kits to facilitate the use of the fusion proteins disclosed herein. In some embodiments, a kit comprises (a) a first container comprising pharmaceutically effective amount of the fusion proteins disclosed herein in a lyophilized composition; and (b) a second container comprising a diluent for reconstituting the lyophilized formulation. In some embodiments, the kit further comprises instructions for storage of the kit, information regarding a cancer that is treatable with the fusion proteins disclosed herein, instructions for the reconstitution of the lyophilized formulation, and/or administration instructions.

METHODS OF TREATMENT

[0274] Disclosed herein are uses of a polypeptide, such as any described herein, in the preparation of a medicament for the treatment of a disease in a subject. In some embodiments, the particular disease to be treated will depend on the choice of the biologically active proteins. In some embodiments, the disease is cancer. Included herein are polypeptides (e.g., a fusion protein comprising a cytokine, a linker comprising a protease-cleavable release segment, and a mask polypeptide) for use in the treatment of cancer. In some embodiments, the cancer or tumor is a solid tumor.

[0275] The present disclosure includes a method of treating a disease in a subject, the method comprising administering to the subject in need thereof a therapeutically effective amount of the pharmaceutical composition, such as any described herein. In some embodiments, the disease is cancer. In some embodiments, the subject is a mouse, rat, monkey, or human. In some embodiments, the subject is a human.

[0276] In some embodiments, the disclosure provides a method of treating cancer in a subject in need thereof, the method comprising administering to the subject an amount of the fusion protein described herein to the subject, wherein the cancer comprises a solid tumor, and treating the cancer comprises reducing the volume of the solid tumor.

[0277] The following are examples of compositions and evaluations of compositions of the disclosure. It is understood that various some embodiments may be practiced, given the general description provided above.

INCORPORATION BY REFERENCE

[0278] All publications, patents, and patent applications mentioned in this specification are incorporated herein 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.

EXAMPLES

Example 1. Design of protease-activatable cytokines

[0279] This example demonstrates the design of protease-activatable cytokine fusion proteins including one to four masks, where each mask is linked to the cytokine by a linker having a protease-cleavable release segment. Fusion proteins including mouse IL-12 (having a p35 subunit and a p40 subunit) were designed having between one and four ELNN masks, with various lengths (as shown in amino acid number (AA)), with the following formats:

Table 7. Protease activable cytokine descriptions

Number of Masks	Identifier	p40 mask(s)	p35 masks(s)
Single	AP2414 (P40~X864/ P35)	864 AA on C-terminus	None

	AP2447 (X864~P40/ P35)	864 AA on N-terminus	None
Double	AP2410 (X864~P40/ P35~X288)	864 AA on N-terminus	288 AA on C-terminus
	AP2411 (X864~P40/ X288~P35)	864 AA on N-terminus	288 AA on N-terminus
	AP2415 (P40~X864/ P35~X288)	864 AA on C-terminus	288 AA on C-terminus
	AP2416 (P40~X864/ X288~288)	864 AA on C-terminus	288 AA on N-terminus
	AP2404 (P40/ X432~P35~X432)	None	432 AA on N-terminus; 432 AA on C-terminus
Triple	AP2405 (P40~X288/ X288~P35~X288)	288 AA on C-terminus	288 AA on N-terminus; 288 AA on C-terminus
	AP2406 (P40~X288/ X432~P35~X432)	288 AA on C-terminus	432 AA on N-terminus; 432 AA on C-terminus
	AP2407 (X288~P40/ X288~P35~288)	288 AA on N-terminus	288 AA on N-terminus; 288 AA on C-terminus
	AP2408 (X288~P40/ X432~P35~X432)	288 AA on N-terminus	432 AA on N-terminus; 432 AA on C-terminus
	AP2409 (P40~X432/ X288~P35~X288)	432 AA on C-terminus	288 AA on N-terminus; 288 AA on C-terminus
	AP2412 (X432~P40~X432/ P35~X288)	432 AA on N-terminus; 432 AA on C-terminus	288 AA on C-terminus
	AP2413 (X432~P40~X432/ X288~P35)	432 AA on N-terminus; 432 AA on C-terminus	288 AA on N-terminus
	AP2450 (X288~P40~X288/ P35~X288)	288 AA on N-terminus; 288 AA on C-terminus	288 AA on C-terminus
Quadruple	AP2446 (X288~P40~X288/ X288~P35~X288)	288 AA on N-terminus; 288 AA on C-terminus	288 AA on N-terminus; 288 AA on C-terminus

[0280] The fusion proteins were expressed in Expi293 cells (ThermoFisher) and purified by a one-step histidine tag-based purification method. Briefly, the subunit constructs were co-transfected into Expi293 cells in a transient manner. Supernatants expressing the protein of interest were harvested 6-7 days post transfection. Protein was purified from the supernatants using a single IMAC step to bind via a polyhistidine tag (HisTrap™ HP 5ml; Cytiva, 17524801). Unmasked versions of were generated proteolytically using MMP9, which cleaves the inbuilt protease activation sites and effects mask release. For MMP9 digestion, the MMP9 (R&D systems) was used at 20 mM, in 20 mM histidine, 154 mM NaCl, 10 mM CaCl₂, pH 6.5 buffer with 150nM each protein. Incubation was done at 37 C for four hours. Some of the MMP9 digests were noted to be partial (less than 100% unmasking),

suggesting that the actual difference in fold masking may be greater than noted based on this experiment.

[0281] The fusion proteins were assayed for their ability to bind to activate the mouse IL-12 receptor in both their masked and unmasked forms by HEK Blue reporter assay. HEK-Blue IL-12 reporter cells (InvivoGen) were cultured at 37°C, 5% CO₂ in a culture media consisting of DMEM, 4.5 g/l glucose, 2 mM L-glutamine, 10% (v/v) heat-inactivated fetal bovine serum, 100 U/ml penicillin, 100 µg/ml streptomycin, 100 µg/ml Normocin, 1X HEK-Blue Selection. For the IL-12 activity assay, a test medium was prepared as described in the immediately preceding sentence but without Normocin and selection antibiotics. The test medium and 1X PBS were warmed to 37°C in a water bath. Cells were dislodged from the flask by washing the flask with the pre-warmed PBS, followed by a centrifugation at 300xg (1200 rpm) for 5 mins at room temperature, determination of cell viability, and a resuspension of the cell pellet in the test medium to 0.833x10⁶ cells/mL. Ninety microliters (90 µL) of the cells were aliquoted into each well of a 96-well flat-clear-bottom plate (Costar, cat#3595). IL-12 test articles were prepared at 10X concentration in the test medium with 17 nM being the highest concentration, followed by a serial 10-fold dilution to 1.7 pM. Then, 10 µL of the 10X solution were added to the 90 µL of cells, and the plate was incubated for 24h. The next day, a QuantiBlue solution, the detection reagent for secreted embryonic alkaline phosphatase (SEAP), was prepared by diluting QB reagent and QB buffer in room temperature MilliQ water to 1% (v/v) concentration each. The mixture was incubated at room temperature for 10 minutes. Subsequently, 180 µL were aliquoted to each well of a 96-well flat bottom tissue culture plate, and to each well was added 20 µL of the supernatant. The plate was incubated at 37°C, 5% CO₂ for 6h. At different incubation time intervals (15min, 30min, 1h, 2h, 3h), a microplate reader was used to measure the optical density (O.D.) at 650nm. The results were analyzed by Excel software and presented here from the 3hr timepoint. The results of this assay are shown in FIGs. 2-3.

[0282] A single-masked IL-12 (AP2447; X864~P40), a triple-masked IL-12 (AP2450; X288~P40~X288/ P35~X288), and quadruple-masked IL-12 (AP2446; X288~P40~X288/ X288~P35~X288) were further transiently expressed in Expi293 cells and purified as detailed above.

[0283] The fusion proteins, in their masked and unmasked forms, were assayed for their IL-12 activity according to the HEK Blue reporter assay described above and compared to the activity of WT mouse IL-12. The masked versions of these fusion proteins provided 8-fold (single mask; AP2447), 11-fold (triple-mask, AP2447), or 78-fold (quadruple-mask, AP2446) decrease in activity as compared to the unmasked equivalent. The results of this assay are provided in FIG. 4.

[0284] A further experiment was performed to determine the masking effect on IL-12 activity of a quadruple-masked protease-activatable IL-12 (AP2551, X288~P40~X288/ X288~P35~X288), where the fusion protein was fully purified and fully (100%) unmasked. AP2551 was purified by harvesting

and clarification of the conditioned medium; size-exclusion column chromatography (Capto Core 400, Cytiva); anion-exchange (AEX) capture (Q Sepharose Fast Flow (QFF), Cytiva); and AEX polishing (Capto Q ImpRes (Q-IR), Cytiva).

[0285] Quantitative fold masking difference between unmasked and fully masked protein was determined according to the HEK blue assay as detailed above. The quadruple-masked IL-12 demonstrated 430-fold masking of IL-12 activity relative to its unmasked equivalent (FIG. 5).

[0286] The results of this example demonstrate that fusion of ELNN masks can provide significant masking of IL-12 activity. Additionally, ELNN-dependent masking of IL-12 activity was highest when four ELNN masks were used. Additionally, three or four shorter (288 AA) masks provided better masking as compared to a single, longer (864 AA) mask.

Example 2. Selection of ELNN lengths

[0287] Protease-activatable IL-12 molecules having different mask lengths were compared for masking efficacy as determined by induction of IFN γ and for tumor growth inhibition in a mouse tumor model. Anti-tumor activity and mechanism of action were assessed in syngeneic mouse models using murine surrogates (including WT murine IL-12) as the human IL-12 does not bind to the mouse IL-12 receptors. The masking activity of an IL-12 molecule having four 288 amino acid length ELNN masks (linked by protease-cleavable linkers at C- and N-termini of p35 and p40 subunits) was compared to an equivalent molecule having four 210 amino acid ELNN masks. The molecules were expressed and fully purified as detailed above.

[0288] The mouse surrogate molecules P24M (4 x 288aa ELNNs) and P26M (4 x 210aa ELNNs) were assessed for induction of IFN γ by mouse splenocyte assay. The purpose of this assay is to assess the ability of the masked and proteolytic-activated cytokines to engage the IL-12R and induce the subsequent secretion of IFN γ . This activity is measured by quantifying IFN γ released into the supernatant by the cells. Recombinant mouse IL-12 is included as reference. In brief, spleens were collected from multiple female C57Bl/6 mice and mechanically dissociated into a single-cell suspension of pooled splenocytes. After elimination of red blood cells, aliquots of 5E5 splenocytes were treated with titrated therapeutic P24M, MMP-activated (unmasked) P24M, therapeutic P26M, MMP-activated (unmasked) P26M, or recombinant murine IL-12. The splenocytes and test articles were incubated for 48 hours following standard tissue culture methods. Supernatants were collected for quantitative analysis of mouse IFN γ by ELISA. The results of this assay are provided in FIG. 6. The results demonstrate superior masking by P24M, with the 288aa ELNN masks, as compared to P26M with the shorter 210aa ELNN masks.

[0289] The mouse surrogate molecules P24M (4 x 288aa ELNNs) and P26M (4 x 210aa ELNNs) were also assessed for differential tumor growth inhibition in the B16F10 melanoma C57BL/6 mouse model. Briefly, for the tumor growth inhibition experiment, 130 mice were subcutaneously implanted

with 100k B16F10 cells in the right flank. After 11 days, 24 mice bearing tumors with an average tumor volume of 111 mm³ were randomized into 3 groups of 8 mice each. Mice were treated intravenously with a weekly dose of diluent control, 3 µg of P24M, or 3 µg of P26M. Mice were measured twice per week for tumor volume and body weight. At day 21, the end of study, mice treated with 3 µg of P24M or P26M had an average tumor growth inhibition of 80% and 61%, respectively, compared to the diluent control group. Treatment with P24M, with the 288aa ELNN masks, had a trend of greater tumor growth inhibition compared to P26M with the shorter 210aa ELNN masks (FIG. 7). Mice did not exhibit signs of toxicity from the treatment.

[0290] Quadruple masks of 288 amino acids each provided better masking in in vitro mouse splenocyte activity assays and in vivo efficacy assays done in mouse tumor models.

Example 3. Design of barcoded ELNNs

[0291] ELNN polypeptide sequences can optionally contain a barcode fragment releasable from the polypeptide upon digestion by a protease. A barcode fragment may be, e.g., (1) a portion of the ELNN that includes at least part of a (non-recurring, non-overlapping) sequence motif that occurs only once within the ELNN; and (2) differs in sequence and molecular weight from all other peptide fragments that are releasable from the polypeptide containing them upon complete digestion of the polypeptide by a protease. The term “barcode fragment” (“barcode,” or “barcode sequence”) can refer to either the portion of the ELNN cleavably fused within the polypeptide, or the resulting peptide fragment released from the polypeptide. Previous barcode sequences (see, e.g., PCT International Patent Publication No. WO2021/263058, the entire content of which is incorporated herein by reference) were designed with the intention of creating unique barcode polypeptide sequences with as minimal mutations in the original ELNN sequence as possible. However, such barcode sequences required 1000µg/mL of Glu-C and an overnight digest to release them from peptides containing them. The barcode polypeptide sequences described in this Example were designed and tested to perform against a second criteria: That the barcode polypeptide is releasable from the ELNN polypeptide rapidly (in approximately two hours vs an overnight digest) by a low concentration of protease (less than 30µg/mL protease); in addition to the criteria of introducing the fewest mutations to the original ELNN sequence as possible.

[0292] In order to determine which peptide sequences were most favorably cleaved by Glu-C protease in a two-hour protease digest, a library of approximately 1000 peptides was constructed with each peptide containing a different cleavage sequence for the protease Glu-C. Equimolar concentrations of these Glu-C site-containing peptides were tested in a 2-hour digest against a range of Glu-C protease concentrations from 0.05 µg/mL to 1000 µg/mL of protease. After digestion the peptides were analyzed by liquid chromatography mass spectrometry. The Glu-C cleavage site sequences that were cleaved by the lowest concentrations of protease were cataloged. From this list of the fastest sequences, a select few were selected that were most compatible with ELNN

polypeptides. These sequences were then implemented to flank new “Generation 2” barcode sequences.

[0293] A selection of Generation 2 barcode sequences was cloned into ELNN sequences and their performance as barcode peptides was tested by Glu-C digestion and subsequent liquid chromatography mass spectrometry (LC-MS) analyses. Successful barcode sequences from this experiment had 3 criteria: 1.) The barcode peptide was fully releasable from the ELNN polypeptide in a 2-hour digest by a concentration of 40 µg/mL of protease. 2.) The barcode peptide was not cleaved or otherwise degraded by much higher concentrations of protease, and 3.) The barcode peptide that met conditions 1 and 2 contained the fewest mutations from the original ELNN polypeptide sequence. Table 8 provides examples of successful Generation 2 barcode sequences according to the criteria of the aforementioned selection process.

Table 8. Exemplary Generation 2 Barcode Sequences

Name	Sequence	SEQ ID NO:
Gen 2 Barcode 01	SGPE . SGPGTGTSATPE . SGPG	3001
Gen 2 Barcode 02	ATPE . SGPGSGPGTSE . SATP	3002
Gen 2 Barcode 03	ATPE . SGPGTTPGTTPPE . SGPG	3003
Gen 2 Barcode 04	ATPE . SGPGTTPPTSTPE . SGPG	3004
Gen 2 Barcode 05	ATPE . SGPGTSPSATPE . SGPG	3005
Gen 2 Barcode 06	ATPE . SGPGTGSAGTPE . SGPG	3006
Gen 2 Barcode 07	ATPE . SGPGTGGAGTPE . SGPG	3007
Gen 2 Barcode 08	ATPE . SGPGTSPGATPE . SGPG	3008
Gen 2 Barcode 09	GTPE . SGPGTSGSGTPE . SGPG	3009
Gen 2 Barcode 10	GTPE . SGPGTSSASTPE . SGPG	3010
Gen 2 Barcode 11	GTPE . SGPGTGAGTTPPE . SGPG	3011
Gen 2 Barcode 12	GTPE . SGPGTGSTSTPE . SGPG	3012
Gen 2 Barcode 13	GTPE . TPGSEPATSGSE . TGTP	3013
Gen 2 Barcode 14	GTPE . GSAPGTSTEPSE . SATP	3014
Gen 2 Barcode 15	ATPE . SGPGTAGSGTPE . SGPG	3015
Gen 2 Barcode 16	ATPE . SGPGTSSGGTPE . SGPG	3016
Gen 2 Barcode 17	ATPE . SGPGTAGPATPE . SGPG	3017
Gen 2 Barcode 18	ATPE . SGPGTPGTGTPE . SGPG	3018
Gen 2 Barcode 19	TTPE . SGPGTGGPTTPE . SGPG	3019
Gen 2 Barcode 20	STPE . SGPGTSGGSTPE . SGPG	3020

[0294] Table 9 provides Generation 3 barcode sequences selected based on rate of Glu-C cleavage and based on separation and quantification by LC-MS. FIG. 8 shows the incorporation of barcode fragments into the ELNNs of an IL-12 fusion protein.

Table 9. Exemplary Generation 3 Barcode Sequences

Name	Sequence	SEQ ID NO:
Gen 3 Barcode 01	ATPE . SGPGSGPGTSE . SATP	3021

Gen 3 Barcode 02	ATPE . SGPGTSPSATPE . SGPG	3022
Gen 3 Barcode 03	ATPE . SGPGSGPATSE . SATP	3023
Gen 3 Barcode 04	ATPE . SGPGTSASATPE . SGPG	3024

Example 4. Release Site Engineering

[0295] Release site engineering was performed to (1) remove a glycosylation site, and (2) reduce the rate of cleavage in plasma which was surprisingly attributed to legumain.

Glycosylation of release site

[0296] To assess glycosylation of the IL-12 fusion protein, a modified peptide-mapping protocol was performed. Briefly, IL-12 fusion protein samples were digested with Trypsin/LysC and run via LC/MS. Data was processed with ProteinMetrics software searching for N-glycosylated peptides. Observed glycosylation events were manually validated in ProteinMetrics software and a report was generated. The report contains the site of glycosylation and the site occupancy of glycosylation (intensity of glycosylated peptide/intensity of glycosylated peptide + intensity of un-glycosylated wild-type peptide).

[0297] An IL-12 fusion protein including RSR-2295 (EAGRSAN~~N~~HTPAGLTGP; SEQ ID NO:7048) was expressed in a HEK 293 cell line. RSR-2295 was found to be 50% N-glycosylated. N-glycosylation and the resulting heterogeneity of the release site is undesirable as it may affect protease cleavage rate and specificity. RSR-2295 was mutated to remove the N-glycosylation site. RSR-3213 (EAGRSAS~~N~~HTPAGLTGP; SEQ ID NO:7628) lacks an N-glycosylation site and therefore is not expected to be glycosylated.

Legumain cleavage in plasma

[0298] Incubation of a fusion protein including RSR-2295 in human plasma showed some cleavage that, though not high, was unexpected. Further investigation revealed that the cleavage in human plasma was surprisingly due to legumain, which has previously believed to be specifically present in tumor tissues. Additionally, it was previously believed that legumain cleavage provided meaningful levels of unmasking and fusion protein activation in tumor tissues.

[0299] RSR-3213 was designed to avoid cleavage by legumain. Surprisingly, a protease-cleavable linker including the RSR-3213 release sequence was cleaved less in plasma but at comparable amounts to a corresponding protease-cleavable linker including the RSR-2295 release sequences in multiple tumor types (including gastric carcinoma (NCI-N87), colorectal adenocarcinoma (HT-29), colon carcinoma (HT-55) tumors). Thus, fusion proteins comprising RSR-3213 have enhanced specificity for tumor tissues without a significant loss of activation in tumor tissues.

In vitro digest:

[0300] *In vitro* digest assays were performed to demonstrate that RSR-3213 is cleaved by MMP and ST14/matriptase, but not legumain. Protease activatable T cell engager (paTCE) fusion proteins were designed, each including a bispecific TCE core that binds to EpCAM and CD3, and two protease-cleavable release sites where each release site is bound to a polypeptide mask such that cleavage of the release site release the polypeptide mask. A first paTCE fusion protein included RSR-2295 in both release sites flanking the TCE core. A second paTCE fusion protein included RSR-3213 in both release sites flanking the TCE core. The paTCEs were expressed in *E. coli*. The paTCEs (RSR-2295-containing and RSR-3213 containing) were digested with 5-fold dilutions of MMP9, legumain, or ST14/matriptase. Similar banding patterns were observed for both MMP9 and matriptase, suggesting the mutation of the legumain cleavage site did not affect cleavability of the MMP and serine protease cleavage sites. Unmasked TCE (uTCE) was observed for the paTCE containing RSR-2295 after digestion with legumain, indicating cleavage at the protease cleavable linker by legumain. uTCE was not observed for the paTCE containing RSR-3213 after digestion with legumain, indicating the mutation successfully prevented cleavage at the protease cleavable linker by legumain (FIG. 9A and FIG. 9B).

Plasma Stability – *In Vivo* Cleavability

[0301] Fluorescently labeled variants of a paTCE containing either RSR-2295 or RSR-3213 were labeled with Sulfo-Cy5.5 or Sulfo-Cy7.5. Opposite colors were co-injected into mice containing NCI-N87, HT-29, or HT-55 xenograft tumors. 48 hours after injection, tumors were harvested, homogenized, and protein extracts were analyzed by SDS-PAGE and LI-COR. Relative abundances for paTCE, a metabolite including a single C-terminal mask (1x-C), a metabolite including a single N-terminal mask (1x-N), and uTCE were quantified. No significant differences were observed in uTCE and 1x-C between the two protease cleavable linkers. paTCE containing RSR-2295 showed a small but statistically significant increase (average 2.19% more) in 1x-N than the corresponding paTCE containing RSR-3213. (FIG. 10A and FIG.10B).

[0302] The observed cleavability *in vivo* from tumor homogenates was also determined from 3 different mouse tumor models. The % abundance for metabolites 1x-C, 1x-N, and uTCE was measured with results depicted in FIG. 10C. Finally, FIG. 10D depicts the % of total for paTCE plus the 3 metabolites (1x-N, 1x-C, and uTCE) when employing RSR-2295 or RSR-3213.

[0303] Overall, these data suggest that differences between *in vivo* cleavability of RSR-2295 and RSR-3213 are minor across 3 different tumor models.

Tumor Uptake:

[0304] Tumor uptake between paTCEs containing either RSR-2295 or RSR-3213 were compared using the ratio of calculated concentrations of total drug (paTCE, 1x-C, 1x-N, and uTCE). While differences in tumor uptake were observed across 3 different tumor models, no significant differences were observed between RSR-2295 and RSR-3213 within each model. This indicates that the changes to the protease cleavable linkers between RSR-2295 and RSR-3213 do not affect tumor uptake of paTCE (FIG. 11).

Efficacy of protease-activatable cytokine including RSR-3213

[0305] Protease-activatable IL-12 molecules having either RSR-2295 or RSR-3213 release segments were compared for masking efficacy as determined by induction of IFN γ and for tumor growth inhibition in a mouse tumor model. Anti-tumor activity and mechanism of action were assessed in syngeneic mouse models using murine surrogates (including WT murine IL-12) as the human IL-12 does not bind to the mouse IL-12 receptors. A quadruple ELNN-masked IL-12 molecule having an RSR-3213 release site linking IL-12 to each of ELNN masks (P24M) was compared to a quadruple ELNN-masked IL-12 molecule having an RSR-2295 release site linking IL-12 to each of ELNN masks. The molecules were expressed and fully purified as detailed above.

[0306] The mouse surrogate molecules P24M (4 x 288aa ELNNs; 4 x RSR-3213) and P01M (4 x 288aa ELNNs; 4 x RSR-2295) were assessed for induction of IFN γ by mouse splenocyte assay. The purpose of this assay is to assess the ability of the masked and proteolytic-activated cytokines to engage the IL-12R and induce the subsequent secretion of IFN γ . This activity is measured by quantifying IFN γ released into the supernatant by the cells. Recombinant mouse IL-12 is included as reference. In brief, spleens were collected from multiple female C57Bl/6 mice and mechanically dissociated into a single-cell suspension of pooled splenocytes. After elimination of red blood cells, aliquots of 5E5 splenocytes were treated with titrated therapeutic P24M, MMP-activated (unmasked) P24M, therapeutic P01M, MMP-activated (unmasked) P01M, or recombinant murine IL-12. The splenocytes and test articles were incubated for 48 hours following standard tissue culture methods. Supernatants were collected for quantitative analysis of mouse IFN γ by ELISA. The results of this assay are provided in FIG. 12.

[0307] The mouse surrogate molecules P24M (4 x 288aa ELNNs; 4 x RSR-3213) and P01M (4 x 288aa ELNNs; 4 x RSR-2295) were also assessed for differential tumor growth inhibition in the B16F10 melanoma C57BL/6 mouse model. Briefly, for the tumor growth inhibition experiment, 130 mice were subcutaneously implanted with 100k B16F10 cells in the right flank. After 11 days, 24 mice bearing tumors with an average tumor volume of 111 mm³ were randomized into 3 groups of 8 mice each. Mice were treated intravenously with a weekly dose of diluent control, 3 μ g of P24M, or 3 μ g of P01M. Mice were measured twice per week for tumor volume and body weight. At day 21, the

end of study, mice treated with 3 µg of P24M or P01M had an average tumor growth inhibition of 80% and 44%, respectively, compared to the diluent control group. Treatment with P24M, with RSR-3213, had a trend of greater tumor growth inhibition compared to P01M with RSR-2295 (FIG. 13). Mice did not exhibit signs of toxicity from the treatment.

Example 5. Glycosylation of human IL-12

[0308] To assess glycosylation of human IL-12 p35 and p40 subunits in the context of a quadruple-masked fusion protein, a modified peptide-mapping protocol was performed. Briefly, fusion protein samples were digested with Trypsin/LysC and run via LC/MS. Data was processed with ProteinMetrics software searching for N-glycosylated peptides. Observed glycosylation events were manually validated in ProteinMetrics software and a report was generated. The report contains the site of glycosylation and the site occupancy of glycosylation (intensity of glycosylated peptide/intensity of glycosylated peptide + intensity of un-glycosylated wild-type peptide).

[0309] Human and mouse protease-activatable IL-12 fusion proteins and human IL-12 (not in a fusion protein) were each produced in HEK 293 cells and were assayed for glycosylation. The results are provided in Table 10.

Table 10. N-Glycosylation of human IL-12

IL-12 subunit	Glycosylation site (AA)	hIL-12 (no fusion)	hIL-12 fusion protein (with four ELNN masks)	mIL-12 fusion protein (with four ELNN masks)
P35	N93	100%	0	0%
	N107	100%	0	n/a
	N195	0%	99%	n/a
P40	N125	Partial	0	0%
	N135	n.d.	0	0%
	N222	100%	99%	99%
	N303	Partial	0-0.05%	n/a

n/a = no glycosylation site at this position in mouse IL-12

n.d. = no data

[0310] N195 of the p35 subunit of huIL-12 in protease-activatable fusion protein with four ELNN masks was 99% glycosylated but glycosylation of N195 was not observed in recombinant huIL-12 (not masked). It is hypothesized that N-glycosylation of N195 was created in the IL-12 fusion proteins by the addition of a polypeptide (e.g., a linker and ELNN mask polypeptide) to the C-terminus of the p35 subunit. In endogenous or recombinantly produced IL-12, N195 is not glycosylated as it is 3 amino acids from the C-terminus of the polypeptide chain and is unable to be glycosylated. Protease-activatable IL-12 fusion proteins were designed to remove the p35 N195 glycosylation site. When mutating N195, amino acid substitutions that did not increase potential immunogenicity of the fusion protein were selected, for example N195Q.

Example 6. P72H and P74H, exemplary protease-activatable IL-12 fusion proteins

[0311] This example provides data relating to exemplary protease-activatable IL-12 fusion proteins, P72H and P74H.

[0312] P72H comprises a first polypeptide subunit comprising IL-12p35 with the amino acid sequence set forth as SEQ ID NO: 1000 and a second polypeptide subunit comprising IL-12p40 with the amino acid sequence set forth as SEQ ID NO: 1033. The annotated amino acid sequence for P72H is provided below in Table 11:

Table 11. P72H Sequences

Subunit	Component	SEQ ID NO:	Sequence
P35	N-ELNN	8026	ASGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESAT* (barcode underlined)
	RSR-3213	7628	EAGRSASHTPAGLTGP
	Spacer	99	GPESATPES
	hIL-12-p35	8050	RNLPVATPDGMFPC LHHSQNLLRAVSNMLQKARQTLEFYPC TSEEIDHEDITKDKTSTVEACLPLELTKNESCLNSRETSFITNGSCLASRKTSFMMALCLSS IYEDLKMYQVEFKTMNAKLMDPKRQIFLDQNMLAVIDELMQALNFNSETVPQKSSLEEPDFYKTKIKLCILLHAFRIRAVTIDRVMSYLNAS
	Spacer	100	GTATPESAP
	RSR-3213	7628	EAGRSASHTPAGLTGP
	C-ELNN	8028	ATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSASATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGEPEA (barcode underlined)
P40	N-ELNN	8023	ASGPGTSTEPSEGSAPGTSESATPESGPGSGPATSESATPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESAT* (barcode underlined)
	RSR-3213	7628	EAGRSASHTPAGLTGP
	Spacer	98	TSESATPES
	hIL-12-p40	8051	IWELKKDVYVVELDWYPDAPGEMVVLTCDTPEEDGITWTLDQSSEVLGSGKTLTIQVKEFGDAGQYTC HKGGEVLSHSLLLLHKKE DGIWSTDILKDQKEPKNKTF LRCEAKNYSGRFTCWWLTTISTDLTFSVKSSRGSSDPQGVTCGAATLSAERVVRGDNKEYEYSVECEDSACPAAEESLP I EVMVDAVHKLKYENYTSFFIRDIKPDPPKNLQLKPLKNSRQVEVSWEYPD TWSTPHSYFSLTFCVQVQGK

			SKREKKDRVFTDKTSATVICRKNASISVRAQDRYSSSWSEWASVPCS
	Spacer	97	GTATPESGPG
	RSR-3213	7628	EAGRSASHTPAGLTGP
	C-FLNN	8025	ATPESGPGSPAGSPPTSTEEGSPAGSPPTSTEEGSPAGSPPTSTEEGTSATPESGPGSTSTEPESGAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSTSTEPESGAPGSPAGSPPTSTEEGTSATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPPTSTEEGSPAGSPPTSTEEGTSTEPESGAPGTSESATPESGPGTSESATPESGPGTSPSATPESGPGSEPATSGSETPGSEPATSGSETPGSPAG

*may optionally include a C-terminal “proline” residue on N-ELNN of each subunit

[0313] P74H comprises a first polypeptide subunit comprising IL-12p35 with the amino acid sequence set forth as SEQ ID NO: 1002 and a second polypeptide subunit comprising IL-12p40 with the amino acid sequence set forth as SEQ ID NO: 1033. P74H includes a p35 variant subunit that is aglycosylated at N195 (N195Q). The annotated amino acid sequence for P74H is provided below in Table 12:

Table 12. P74H Sequences

Subunit	Component	SEQ ID NO:	Sequence
P35	N-ELNN	8026	ASGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPPTSTEEGSPAGSPPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESAT* (barcode underlined)
	RSR-3213	7628	EAGRSASHTPAGLTGP
	Spacer	99	GTPESATPES
	hIL-12-p35 (N195Q)	8053	RNLPVATPDPMFFCLHHSQNLLRAVSNMLQKARQTLEFYFCTSEEDHEDITKDKTSTVEACIPLLELTKNESCLNSRETSFITNGSCLASRKSTFMMALCLSSIYEDLKMYQVEFKTMNAKLMDPKRQIFLDQNLMAVIDELMQALNFNSETVPQKSSLEEDPFYKTIKLCILLHAFRIRAVTIDRVMSYL QAS (N195Q bold)
	Spacer	100	GTATPESAP
	RSR-3213	7628	EAGRSASHTPAGLTGP
	C-ELNN	8028	ATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPPTSTEEGSPAGSPPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSASATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGEFEA (barcode underlined)
P40	N-ELNN	8023	ASGPGTSTEPSEGSAPGTSESATPESGPGSGPATSESATPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGSPPTSTEEGSPAGSPPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPPTSTEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGTSESAT* (barcode underlined)

	RSR-3213	7628	EAGRSASHTPAGLTGP
	Spacer	98	TSESATPES
	hIL-12-p40	8051	IWELKKDVYVVELDWYPDAPGEMVLTCDTPEEDGITWTLDQS SEVLGSGKTLTIQVKEFGDAGQYCHKGGEVLSHSLLLHKKKE DGIWSTDILKDQKEPKNKTFLRCEAKNYSGRFTCWWTITSTD LTFSVKSSRGSSDPQGVTCGAATLSAERVRGDNKEYEYSVECO EDSACPAAEESLPFIEVMVDVHKLKYENYTSSFFIRDIKPD PKNLQKPLKNSRQVEVSWEYPDWTSTPHSYFSLTFCVQVQGK SKREKKDRVFTDKTSATVICRKNASISVRAQDRYSSSWSEWA SVPCS
	Spacer	97	GTATPESGPG
	RSR-3213	7628	EAGRSASHTPAGLTGP
	C-ELNN	8025	ATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEE GTSESATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPAT GSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGT STEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGS ETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTST EPSEGSAPGTSESATPESGPGTSESATPESGPGTSPSATPESG PGSEPATSGSETPGSEPATSGSETPGSPAG

*may optionally include a C-terminal "proline" residue on N-ELNN of each subunit

[0314] Methods for producing protease-activatable cytokine fusion proteins are known in the art, e.g., as described in PCT International Patent Publication No. WO2021/262985.

[0315] In some of the experiments below, mouse surrogates for P72H and P74H are used. The anti-tumor activity and mechanism of action were assessed in syngeneic mouse models using murine surrogates as the human IL-12 does not bind to the mouse IL-12 receptors.

[0316] P71M is the mouse surrogate for P72H. P71M is identical to P72H except that human IL-12 has been replaced with mouse IL-12 in the P71M mouse surrogate. Additionally, mouse IL-12 of P71M has been mutated to introduce an N195 glycosylation site (not present in mouse, introduced to better mimic glycosylation pattern of human IL-12 in P72H).

[0317] P70M is the mouse surrogate for P74H except that human IL-12 has been replaced with mouse IL-12 in the P70M mouse surrogate. P74H contains an N-195 aglycosylated version of human IL-12 and therefore WT mouse IL-12, which naturally lacks N195, can be used as a surrogate.

In vitro characterization of P71M and P70M surrogate mouse fusion proteins

[0318] P71M is the mouse surrogate for P72H, as described above. P70M is the mouse surrogate for P74H, as described above.

[0319] Fusion of an ELNN to the C-terminal end of human IL-12 P35 subunit results in additional glycosylation of the N195 site as described above. The C-terminal end of the mouse IL-12 P35 is differs in sequence from the corresponding human sequence and does not exhibit this extra glycosylation. Equivalent mouse surrogate variant P71M was designed to have a similar glycosylation pattern to human P72H to investigate if there are any in vitro or in vivo effects of N195 glycosylation in human P72H on IL-12 activity and/or the pharmacokinetic profile.

[0320] The mouse surrogate molecules P71M (glycosylated at N195 of P35; surrogate for P72H) and P70M (lacking glycosylation position at N195; surrogate for P74H which contains a P35 N195Q mutation) were assessed for binding to the mouse IL-12 receptor (mIL-12R) by SPR. The SPR assay was run on Biacore T200 at 25 degrees Celsius. The chip type that was used was Series S Sensor Chip CM5. For each antigen, a separate chip was made. For hIL-12RB2, the anti-His antibody was immobilized on the CM5 chip surface. For mIL-12RB2 and cIL-12RB2, the anti-mouse Fc antibody was immobilized on the CM5 chip surface. The running buffer that was used was PBS-P+ from Cytiva. Chips were regenerated by 10 mM Glycine HCl, pH 1.5. Sample flow was 30µl/min, while capture flow was 10µl/min. The association time was 100 seconds. The dissociation time was 360 seconds. Each sample was assayed as a titration ranging from 500nM to 0.68nM. Sersogram was fitted globally as a 1:1 kinetic interaction model using Biacore T200 Evaluation software.

[0321] Binding affinity data is shown in Table 13. A biphasic binding curve observed with mIL-12Rβ1. Affinity trends were similar to that determined for mIL-12Rβ2.

Table 13. Affinity of mouse surrogate molecules to mIL-12Rβ2

Test Item	mIL-12Rβ2 KD (nM)
mIL-12	6.8
P70 unmasked	26
P70 masked	370
P71 unmasked	23
P71 masked	350

[0322] P71M and P70M were also assessed for induction of IFNγ by mouse splenocyte assay. The purpose of this assay is to assess the ability of the masked and proteolytic-activated cytokines to engage the IL-12R and induce the subsequent secretion of IFNγ. This activity is measured by quantifying IFNγ released into the supernatant by the cells. Recombinant mouse IL-12 is included as reference.

[0323] Spleens were collected from multiple female C57Bl/6 mice and mechanically dissociated into a single-cell suspension of pooled splenocytes. After elimination of red blood cells, aliquots of 5E5 splenocytes were treated with titrated therapeutic P70M, MMP-activated (unmasked) P70M, therapeutic P71M, MMP-activated (unmasked) P71M, or recombinant murine IL-12. The splenocytes and test articles were incubated for 48 hours following standard tissue culture methods. Supernatants were collected for quantitative analysis of mouse IFNγ by ELISA. The results are provided in FIG. 14.

[0324] The mouse surrogates, P70M (N195 aglycosylated) and P71M (N195 glycosylated), exhibited similar EC₅₀ values of 2.6 nM and 2.8 nM, respectively. Their proteolytic-activated counterparts were more likewise active at 0.0031 nM and 0.0013 nM, respectively. Both were comparable to recombinant mouse IL-12 (EC₅₀ 0.0017 pM). The fold protection conferred by the masked constructs relative to the MMP-activated counterparts demonstrates a reduced potency ranging from 100-fold to 1000-fold.

Pharmacokinetics of P71M and P70M surrogate mouse fusion proteins

[0325] P71M is the mouse surrogate for P72H, as described above. P70M is the mouse surrogate for P74H, as described above.

[0326] The impact of glycosylation of N195 in the P35 domain on the pharmacokinetics of masked and unmasked murine IL-12 was evaluated in non-tumor-bearing C57BL/6 mice. Twelve mice each were administered a single bolus intravenous dose of 70 µg unmasked P70M or P71M (24 total mice). Plasma samples were collected from 3 mice in each dose group prior to IV administration and at 0.083, 0.25, 0.5, 1, 2, 4, 8, 24, and 48 hours post-dose. Additionally, 24 mice each were administered a single bolus intravenous dose of 30 µg masked P70M and P71M. Plasma samples were collected from 3 mice in each dose group prior to administration and at 0.083, 0.25, 0.5, 1, 2, 4, 8, 24 (n=6), 48 (n=6), 72, 96, 168, and 240 hours post-dose. The concentration of masked and unmasked P70M and P71M was quantified using qualified ligand binding assays. Pharmacokinetic parameters were determined for each test item by non-compartment analysis using composite profiles.

[0327] As shown in FIG. 15, the unmasked murine IL-12 was cleared approximately 100-fold faster than the masked murine IL-12, supporting the use of the masks to prolong half-life. The addition of the N195 glycosylation site in P71M resulted in a modestly shorter elimination half-life of the unmasked murine IL-12, suggesting that glycosylation of N195 has an impact on the pharmacokinetics of unmasked murine IL-12. In contrast, the addition of the N195 glycosylation site had no impact on the pharmacokinetics of masked murine IL-12, suggesting that the presence of the masks may interfere with the glycosylation-specific elimination of murine IL-12. Clearance values (CL) and terminal half-life (T_{1/2}) are provided in Table 14.

Table 14. Clearance values (CL) and terminal half-life (T_{1/2}) of mouse surrogate molecules

Test Article	Dose (µg)	CL (mL/h/kg)	T _{1/2} (h)
P70 masked	30	1.44	21.4
P71 masked	30	1.77	21.5
P70 unmasked	70	266	1.94

P71 unmasked	70	202	1.43
-------------------------	----	-----	------

Efficacy and toxicity of P71M and P70M surrogate mouse fusion proteins

[0328] P71M is the mouse surrogate for P72H, as described above. P70M is the mouse surrogate for P74H, as described above.

[0329] P70M (native mouse IL-12 sequence) and P71M (added N195 glycosylation site on the P35 subunit) were assessed for differential tumor growth inhibition and toxicity in the MC38 colon carcinoma C57BL/6 mouse model. The results are provided in FIG. 16.

[0330] Briefly, for the tumor growth inhibition experiment, 120 mice were subcutaneously implanted with 250k MC38 cells in the right flank. After 14 days, 40 mice bearing tumors with an average tumor volume of 101 mm³ were randomized into 5 groups of 8 mice each. Mice were treated intravenously with a weekly dose of diluent control, 3 µg of P70M, 10 µg of P70M, 3 µg of P71M, or 10 µg of P71M. Mice were measured twice per week for tumor volume and body weight. At day 31, the end of study, mice treated with 3 µg of P70M or P71M had an average tumor growth inhibition of 56% and 45%, respectively, compared to the diluent control group. Mice treated with 10 µg of P70M or P71M had an average tumor growth inhibition of 91% and 86%, respectively. These data indicate that P70M and P71M exhibit similar dose-dependent anti-tumor efficacy when using P70M or P71M. Mice did not exhibit signs of toxicity from the treatment.

[0331] For the toxicity experiment, 130 mice were subcutaneously implanted with 250k MC38 cells in the right flank. After 14 days, 54 mice bearing tumors with an average tumor volume of 80 mm³ were randomized into 3 groups of 18 mice each. Mice were treated intravenously twice weekly with diluent control, 100 µg of P70M, or 100 µg P71M. Mice were assessed daily for signs of overt toxicity, which was defined by death, morbidity, or body weight percentage loss greater than 15%. For the last 5 days of the study (day 11- and 15-days post-dosing start), P70M and P71M had similar overt toxicity of 44% and 50% of animals, respectively.

In vitro characterization of P72H and P74H fusion proteins

[0332] In vitro assays were performed to characterize the IL-12 receptor (IL-12R) binding and activity of human P72H and P74H. Binding affinity of P72H and P74H were measured against human and cynomolgus monkey IL-12Rβ, by SPR. The assay was run on Biacore T200 at 25 degrees Celsius. The chip type that was used was Series S Sensor Chip CM5. For each antigen, a separate chip was made. For hIL-12RB2, the anti-His antibody was immobilized on the CM5 chip surface. For mIL-12RB2 and cIL-12RB2, the anti-mouse Fc antibody was immobilized on the CM5 chip surface. The running buffer that was used was PBS-P+ from Cytiva. Chips were regenerated by 10 mM Glycine HCl, pH 1.5. Sample flow was 30µl/min, while capture flow was 10µl/min. The association

time was 100 seconds. The dissociation time was 360 seconds. Each sample was assayed as a titration ranging from 500nM to 0.68nM. Sersogram was fitted globally as a 1:1 kinetic interaction model using Biacore T200 Evaluation software.

[0333] Affinity data is shown in Table 15. A biphasic binding curve was observed with IL-12R β 1. Affinity trends were similar to that determined for IL-12R β 2.

Table 15. Affinity of P72H and P74H to human and cynomolgus monkey IL-12R β

Test Item	hIL-12R β 2 KD (nM)	cIL-12R β 2 KD (nM)
hIL-12	52.6	16.7
cIL-12	105.7	52.1
P72 unmasked	50.7	11.9
P72 masked	88.2	77.9
P74 unmasked	53.9	10.7
P74 masked	78.1	45.4

[0334] Activity was determined by their respective ability to induce the downstream secretion of IFN γ into the supernatant by cells. Frozen PBMCs from a single human donor were quickly thawed, washed, and resuspended. Aliquots of 5E6 PBMCs are activated with 5 μ g/ml anti-CD3 monoclonal antibody (clone OKT3) in 3 ml of media for 48 hours following standard tissue culture methods. Activated PBMCs were aliquoted in 1E5 cells and combined with titrated therapeutic P72H, MMP-activated (unmasked) P72H, therapeutic P74H, MMP-activated (unmasked) P74H, or recombinant human IL-12. The cells and therapeutic reagents are incubated for 48 hours. Supernatants were collected for quantitative analysis of human IFN γ by ELISA. This experiment was repeated in triplicate, each utilizing a unique human PBMC donor. The results are provided in FIG. 17.

[0335] The human constructs, P72H and P74H, elicited a median EC₅₀ values of 1 nM and 1.4 nM, respectively, across the three donors making. Similarly, the proteolytically-activated counterparts were more active, 0.0055 nM and 0.0093 nM, respectively. This is comparable to the activity exhibited by recombinant human IL-12 (EC₅₀ 0.0065 nM). However, the masks conferred a reduced potency of 100-fold.

[0336] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. It is not intended that the invention be limited by the specific examples provided within the specification. While the invention has been described with reference to the aforementioned specification, the descriptions and illustrations of the embodiments herein are not meant to be construed in a limiting sense. Numerous variations, changes, and substitutions will now occur to

those skilled in the art without departing from the invention. Furthermore, it shall be understood that all aspects of the invention are not limited to the specific depictions, configurations or relative proportions set forth herein which depend upon a variety of conditions and variables. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is therefore contemplated that the invention shall also cover any such alternatives, modifications, variations or equivalents. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

CLAIMS

WHAT IS CLAIMED IS:

1. A fusion protein comprising:
 - a cytokine,
 - a linker (Linker1) comprising a protease-cleavable release segment (RS1), and
 - a mask polypeptide (Mask1),wherein the linker (Linker1) comprising the release segment (RS1) is positioned between the cytokine and the mask polypeptide (Mask1);
wherein the release segment is capable of being cleaved by at least one protease that is present in a tumor;
and wherein the release segment (RS1) (i) is not capable of being cleaved by legumain in human plasma, (ii) is cleaved by legumain in human plasma at a rate that is less than 25% of the rate that EAGRSANHTPAGLTGP (RSR-2295; SEQ ID NO:7048) is cleaved by legumain in human plasma, or (iii) comprises an amino acid sequence comprising the sequence EAGRSAXHTPAGLTGP, wherein X is any amino acid other than N (SEQ ID NO:7627).
2. The fusion protein of claim 1, wherein the cytokine is an interleukin, a transforming growth factor, an interferon, a tumor necrosis factor, a chemokine, or granulocyte macrophage-colony stimulating factor.
3. The fusion protein of claim 2, wherein the cytokine is an interleukin.
4. The fusion protein of claim 3, wherein the interleukin is selected from IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, IL-21, and IL-25.
5. The fusion protein of claim 4, wherein the interleukin is IL-12.
6. The fusion protein of claim 5, wherein the IL-12 is a disulfide-linked heterodimer comprising a p35 polypeptide and a p40 polypeptide, and wherein the linker and mask are covalently joined to (i) the N-terminus or C-terminus of the p35 polypeptide, or (ii) the N-terminus or C-terminus of the p40 polypeptide.
7. A fusion protein comprising:
 - (a) a first polypeptide subunit comprising the following elements in an N-to-C or C-to-N terminal orientation:

- i. a first mask polypeptide (Mask1),
 - ii. a first linker (Linker1) comprising a first protease-cleavable release segment (RS1),
 - iii. an IL-12 p35 polypeptide,
 - iv. a second linker (Linker2) comprising a second protease-cleavable release segment (RS2), and
 - v. a second mask polypeptide (Mask2); and
- (b) a second polypeptide subunit comprising the following elements in an N-to-C or C-to-N terminal orientation:
- i. a third mask polypeptide (Mask3),
 - ii. a third linker (Linker3) comprising a third protease-cleavable release segment (RS3),
 - iii. an IL-12 p40 polypeptide,
 - iv. a fourth linker (Linker4) comprising a fourth protease-cleavable release segment (RS4), and
 - v. a fourth mask polypeptide (Mask4);

wherein the IL-12 p35 polypeptide of the first polypeptide subunit and the IL-12 p40 polypeptide of the second polypeptide subunit are disulfide-linked thereby forming a disulfide-linked heterodimer;

wherein each release segment (RS1, RS2, RS3, and RS4) is capable of being cleaved by at least one protease that is present in a tumor;

and wherein each release segment (RS1, RS2, RS3, and RS4) (i) is not capable of being cleaved by legumain in human plasma, (ii) is cleaved by legumain in human plasma at a rate that is less than 25% of the rate that RSR-2295 (EAGRSANHTPAGLTGP; SEQ ID NO:7048) is cleaved by legumain in human plasma, or (iii) comprises an amino acid sequence comprising the sequence EAGRSAXHTPAGLTGP, wherein X is any amino acid other than N (SEQ ID NO:7627).

8. The fusion protein of claim 7, wherein the first mask polypeptide (Mask1) comprises a first a first barcode fragment (Barcode1), the second mask polypeptide (Mask2) comprises a second barcode fragment (Barcode2), the third mask polypeptide (Mask3) comprises a third barcode fragment (Barcode3), and the fourth mask polypeptide (Mask4) comprises a fourth barcode fragment (Barcode4);

and wherein each barcode fragment (Barcode1, Barcode2, Barcode3, and Barcode4) is releasable from the fusion protein upon digestion with a non-mammalian protease, and each barcode fragment differs in both sequence and molecular weight from all other barcode fragments or peptide fragments that are releasable from the fusion protein upon complete digestion of the fusion protein by the non-mammalian protease.

9. A fusion protein comprising:
- (a) a first polypeptide subunit comprising the following elements in an N-to-C or C-to-N terminal orientation:
 - i. a first mask polypeptide (Mask1) comprising a first barcode fragment (Barcode1),
 - ii. a first linker (Linker1) comprising a first protease-cleavable release segment (RS1),
 - iii. an IL-12 p35 polypeptide,
 - iv. a second linker (Linker2) comprising a second protease-cleavable release segment (RS2), and
 - v. a second mask polypeptide (Mask2) comprising a second barcode fragment (Barcode2); and
 - (b) a second polypeptide subunit comprising the following elements in an N-to-C or C-to-N terminal orientation:
 - i. a third mask polypeptide (Mask3) comprising a third barcode fragment (Barcode3),
 - ii. a third linker (Linker3) comprising a third protease-cleavable release segment (RS3),
 - iii. an IL-12 p40 polypeptide,
 - iv. a fourth linker (Linker4) comprising a fourth protease-cleavable release segment (RS4), and
 - v. a fourth mask polypeptide (Mask4) comprising a fourth barcode fragment (Barcode4);
- wherein the IL-12 p35 polypeptide of the first subunit and the IL-12 p40 polypeptide of the second unit are disulfide-linked thereby forming a disulfide-linked heterodimer;
- wherein each release segment (RS1, RS2, RS3, and RS4) is capable of being cleaved by at least one protease that is present in a tumor;
- and wherein each barcode fragment (Barcode1, Barcode2, Barcode3, and Barcode4) is releasable from the fusion protein upon digestion with a non-mammalian protease, and each barcode fragment differs in both sequence and molecular weight from all other barcode fragments or peptide fragments that are releasable from the fusion protein upon complete digestion of the fusion protein by the non-mammalian protease.
10. The fusion protein of claim 9, wherein each release segment comprises an amino acid sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity, to a sequence of Table 6a.

11. The fusion protein of claim 9, wherein each release segment (RS1, RS2, RS3, and RS4) (i) is not capable of being cleaved by legumain in human plasma, (ii) is cleaved by legumain in human plasma at a rate that is less than 25% of the rate that RSR-2295 (EAGRSANHTPAGLTGP; SEQ ID NO:7048) is cleaved by legumain in human plasma, or (iii) comprises an amino acid sequence comprising the sequence EAGRSAXHTPAGLTGP, wherein X is any amino acid other than N (SEQ ID NO:7627).
12. The fusion protein of any one of claims 1-11, wherein each release segment is a substrate for at least one protease of Table 5.
13. The fusion protein of any one of claims 1-12, wherein each release segment is capable of being cleaved by uPA, ST14, MMP2, MMP7, MMP9, and MMP14.
14. The fusion protein of any one of claims 1-13, wherein each release segment is not capable of being cleaved by legumain in human plasma.
15. The fusion protein of any one of claims 1-14, wherein each release segment is not capable of being cleaved by legumain in human blood, plasma, or serum.
16. The fusion protein of any one of claims 1-15, wherein each release segment is not capable of being cleaved upon incubation with about 1nM or less legumain for about 20 hours.
17. The fusion protein of any one of claims 1-16, wherein each release segment is cleaved by legumain in human plasma at a rate that is less than 25%, less than 15%, less than 10%, less than 5%, or less than 2.5% of the rate that RSR-2295 (EAGRSANHTPAGLTGP; SEQ ID NO:7048) is cleaved by legumain in human plasma.
18. The fusion protein of any one of claims 1-17, wherein each release segment comprises the amino acid sequence of RSR-3213 (EAGRSASHTPAGLTGP; SEQ ID NO:7628).
19. The fusion protein of any one of claims 1-18, wherein each mask polypeptide is an extended length non-natural polypeptide (ELNN).
20. The fusion protein of claim 19, wherein each ELNN is, independently, characterized in that: at least 90% of the amino acid residues of each ELNN are selected from glycine (G), alanine (A), serine (S), threonine (T), glutamate (E), proline (P), or any combination thereof; and

each ELNN comprises at least 3 types of amino acid residues selected from the group consisting of G, A, S, T, E, and P.

21. The fusion protein of claim 20, wherein each ELNN, independently, comprises a plurality of non-overlapping sequence motifs that are each from 9 to 14 amino acid residues in length, wherein the plurality of non-overlapping sequence motifs comprises a set of non-overlapping sequence motifs, wherein each non-overlapping sequence motif of the set of non-overlapping sequence motifs is repeated at least two times in the ELNN.

22. The fusion protein of claim 21, wherein the plurality of non-overlapping sequence motifs comprises at least one non-overlapping sequence motif that occurs only once within the ELNN.

23. The fusion protein of claim 21 or 22, wherein the non-overlapping sequence motifs comprise one of or any combination of the sequence motifs listed in Table 1b.

24. The fusion protein of any one of claims 21-23, wherein the non-overlapping sequence motifs comprise at least 2, 3, or 4 of the sequence motifs listed in Table 1b.

25. The fusion protein of any one of claims 21-24, wherein the non-overlapping sequence motifs comprise any one of or any combination of GTSTEPSEGSAP (SEQ ID NO:189), GTSESATPESGP (SEQ ID NO:188), GSGPGTSESATP (SEQ ID NO:201), GSEPATSGSETP (SEQ ID NO:187), GSPAGSPTSTEE (SEQ ID NO:186), and GTSPSATPESGP (SEQ ID NO:202).

26. The fusion protein of any one of claims 19-25, wherein each ELNN comprises at least 4 types of amino acid residues selected from the group consisting of G, A, S, T, E, and P.

27. The fusion protein of any one of claims 19-26, wherein the amino acid residues of each ELNN consists of A, E, G, S, P, and/or T.

28. The fusion protein of any one of claims 19-27, wherein each ELNN, independently, has a length of least 100 amino acid residues, at least 200 amino acid residues, or at least 250 amino acid residues.

29. The fusion protein of any one of claims 19-28, wherein each ELNN, independently, has a length of 100 to 1000 amino acid residues, 100 to 500 amino acid residues, 200 to 100 amino acid residues, 200 to 500 amino acid residues, 200 to 300 amino acid residues, 250 to 1000 amino acid residues, 250 to 500 amino acid residues, or 250 to 300 amino acid residues.

30. The fusion protein of claim 23, wherein each ELNN has a length of between 250 and 300 amino acid residues.

31. The fusion protein of claim 30, wherein each ELNN has a length of 287 or 288 amino acid residues.

32. The fusion protein of any one of claims 19-31, wherein each ELNN comprises an amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to an amino acid sequence listed in Table 3a or 3b.

33. The fusion protein of any one of claims 7-32, wherein the first polypeptide mask is a first ELNN (ELNN1), and wherein ELNN1 comprises an amino sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity to:

ASGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSEPATSG
SETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGT
SESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATP
ESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGT
STEPSEGSAPGSEPATSGSETPGTSESAT (SEQ ID NO: 8026).

34. The fusion protein of any one of claims 7-33, wherein the second polypeptide mask is a second ELNN (ELNN2), and wherein ELNN2 comprises an amino sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity to:

ATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPAT
SGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETP
GTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESA
TPESGPGTSASATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAP
GTSTEPSEGSAPGSEPATSGSETPGEPEA (SEQ ID NO: 8028).

35. The fusion protein of any one of claims 7-34, wherein the third polypeptide mask is a third ELNN (ELNN3), and wherein ELNN3 comprises an amino sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity to:

ASGPGTSTEPSEGSAPGTSESATPESGPGSGPATSESATPGTSESATPESGPGSEPATSG
SETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGT
SESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATP
ESGPGTSESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGT
STEPSEGSAPGSEPATSGSETPGTSESAT (SEQ ID NO: 8023).

36. The fusion protein of any one of claims 7-35, wherein the fourth polypeptide mask is a fourth ELNN (ELNN4), and wherein ELNN4 comprises an amino sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity to:

ATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGTSTEP
SEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATPESGP
GTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSPAGS
PTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSPSATPESGP
GSEPATSGSETPGSEPATSGSETPGSPAG (SEQ ID NO: 8025).

37. The fusion protein of any one of claims 6-36 wherein the IL-12 p35 polypeptide is a human IL-12 p35 polypeptide and the IL12 p40 polypeptide is a human IL-12 p40 polypeptide.

38. The fusion protein of claim 37, wherein the human IL-12 p35 polypeptide comprises an amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity to:

RNLPVATPDPGMFPCLHHSQNLLRAVSNMLQKARQTLEFYPTSEEIDHEDITKDKTS
TVEACLPLELTKNESCLNSRETSTFITNGSCLASRKTSFMMALCLSSIEDLKMYQVEFKTMNA
KLLMDPKRQIFLDQNMLAVIDELMQALNFNSETVPQKSSLEEPDFYKTKIKLCILLHAFRIRA
VTIDRVMSYLNAS (SEQ ID NO: 8050).

39. The fusion protein of claim 38, wherein the human IL-12 p35 polypeptide comprises the amino acid sequence of:

RNLPVATPDPGMFPCLHHSQNLLRAVSNMLQKARQTLEFYPTSEEIDHEDITKDKTS
TVEACLPLELTKNESCLNSRETSTFITNGSCLASRKTSFMMALCLSSIEDLKMYQVEFKTMNA
KLLMDPKRQIFLDQNMLAVIDELMQALNFNSETVPQKSSLEEPDFYKTKIKLCILLHAFRIRA
VTIDRVMSYLNAS (SEQ ID NO: 8050).

40. The fusion protein of claim 37 or 38, wherein the human IL-12 p35 polypeptide is an N195 aglycosylated variant.

41. The fusion protein of claim 40, wherein the human IL-12 p35 polypeptide comprises the amino acid sequence of:

RNLPVATPDPGMFPCLHHSQNLLRAVSNMLQKARQTLEFYPTSEEIDHEDITKDKTS
TVEACLPLELTKNESCLNSRETSTFITNGSCLASRKTSFMMALCLSSIEDLKMYQVEFKTMNA
KLLMDPKRQIFLDQNMLAVIDELMQALNFNSETVPQKSSLEEPDFYKTKIKLCILLHAFRIRA
VTIDRVMSYLNAS (SEQ ID NO: 8055).

wherein X is any amino acid residue other than asparagine (N).

42. The fusion protein of claim 40 or 41, wherein the human IL-12 p35 polypeptide comprises an amino acid sequence selected from:

RNLPVATPDPMFPCLLHHSQNLLRAVSNNMLQKARQTLEFYPCTSEEIDHEDITKDKTS
TVEACLPLELTKNESCLNSRETSFITNGSCLASRKTSFMMALCLSSIEDLKMYQVEFKTMNA
KLLMDPKRQIFLDQNMLAVIDELMQALNFNSETVPQKSSLEEPDFYKTKIKLCILLHAFRIRA
VTIDRVMSYLSSA (SEQ ID NO: 8052);

RNLPVATPDPMFPCLLHHSQNLLRAVSNNMLQKARQTLEFYPCTSEEIDHEDITKDKTS
TVEACLPLELTKNESCLNSRETSFITNGSCLASRKTSFMMALCLSSIEDLKMYQVEFKTMNA
KLLMDPKRQIFLDQNMLAVIDELMQALNFNSETVPQKSSLEEPDFYKTKIKLCILLHAFRIRA
VTIDRVMSYQLAS (SEQ ID NO: 8053); and

RNLPVATPDPMFPCLLHHSQNLLRAVSNNMLQKARQTLEFYPCTSEEIDHEDITKDKTS
TVEACLPLELTKNESCLNSRETSFITNGSCLASRKTSFMMALCLSSIEDLKMYQVEFKTMNA
KLLMDPKRQIFLDQNMLAVIDELMQALNFNSETVPQKSSLEEPDFYKTKIKLCILLHAFRIRA
VTIDRVMSYLDAS (SEQ ID NO: 8054).

43. The fusion protein of any one of claims 37-42, wherein the human IL-12 p40 polypeptide comprises an amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity to:

IWELKKDVYVVELDWYPDAPGEMVVLTCDTPEEDGITWTLDQSSEVLGSGKTLTIQ
VKEFGDAGQYTCHKGGVLSHSLLLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRF
TCWWLTITISTDLTFSVKSSRGSSDPQGVTCGAATLSAERVGRDNKEYEYSVEQCQEDSACPAA
EESLPIEVMVDAVHKLKYENYTSSFFIRDIIKPDPPKNLQLKPLKNSRQVEVSWEYPDTWSTP
HSYFSLTFCVQVQGKSKREKKDRVFTDKTSATVICRKNASISVRAQDRYYSSSWSEWASVPC
S (SEQ ID NO: 8051).

44. The fusion protein of claim 43, wherein the human IL-12 p40 polypeptide comprises the amino acid sequence of:

IWELKKDVYVVELDWYPDAPGEMVVLTCDTPEEDGITWTLDQSSEVLGSGKTLTIQ
VKEFGDAGQYTCHKGGVLSHSLLLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRF
TCWWLTITISTDLTFSVKSSRGSSDPQGVTCGAATLSAERVGRDNKEYEYSVEQCQEDSACPAA
EESLPIEVMVDAVHKLKYENYTSSFFIRDIIKPDPPKNLQLKPLKNSRQVEVSWEYPDTWSTP
HSYFSLTFCVQVQGKSKREKKDRVFTDKTSATVICRKNASISVRAQDRYYSSSWSEWASVPC
S (SEQ ID NO: 8051).

45. The fusion protein of any one of claims 6-36, wherein the IL-12 p35 polypeptide is a mouse IL-12 p35 polypeptide and the IL-12 p40 polypeptide is a mouse IL-12 p40 polypeptide.
46. The fusion protein of any one of claims 1-45, wherein each linker further comprises a spacer.
47. The fusion protein of claim 46, wherein each spacer is positioned between a release segment and a cytokine, an IL-12 p35 polypeptide, or an IL-12 p40 polypeptide.
48. The fusion protein of claim 46 or 47, wherein each spacer is, independently, characterized in that:
at least 90% of the amino acid residues of each spacer are selected from glycine (G), alanine (A), serine (S), threonine (T), glutamate (E), proline (P), or any combination thereof; and
each spacer comprises at least 3 types of amino acid residues selected from the group consisting of G, A, S, T, E, and P.
49. The fusion protein of any one of claims 46-48, wherein each spacer is, independently, 5 to 14 amino acid residues in length.
50. The fusion protein of any one of claims 46-49, wherein each spacer comprises at least 4 types of amino acid residues selected from the group consisting of G, A, S, T, E, and P.
51. The fusion protein of any one of claims 46-50, wherein the amino acid residues of each spacer comprise only amino acid residues selected from G, A, S, T, E, and/or P.
52. The fusion protein of any one of claims 46-51, wherein each spacer is cleavable by a non-mammalian protease.
53. The fusion protein of claim 52, wherein the non-mammalian protease is Glu-C.
54. The fusion protein of any one of claims 46-53, wherein each spacer, independently, comprises an amino acid sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity, to a sequence listed in Table C.
55. The fusion protein of any one of claims 46-54, wherein each spacer, independently, comprises an amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity to TSESATPES (SEQ ID NO:98), GTATPESGPG (SEQ ID NO:97), GTPESATPES (SEQ ID NO:99), or GTATPESAP (SEQ ID NO:100).

56. The fusion protein of any one of claims 46-55, wherein no spacer shares 100% sequence identity with any other spacer.
57. The fusion protein of any one of claims 1-56, wherein the fusion protein does not comprise a His-tag, wherein a His-tag is a polypeptide sequence comprising an uninterrupted chain of between 5 and 12 histidine residues.
58. The fusion protein of any one of claims 8-57, wherein each barcode fragment is releasable from the fusion protein upon digestion with a non-mammalian protease and the non-mammalian protease is Glu-C.
59. The fusion protein of any one of claims 8-58, wherein none of the barcode fragments include any of the following: the N-terminal amino acid of the first polypeptide subunit, the C-terminal acid of the first polypeptide subunit, the N-terminal amino acid of the second polypeptide subunit, and the C-terminal amino acid of the polypeptide subunit.
60. The fusion protein of any one of claims 8-59, wherein the entirety of each barcode fragment is 5 to 100, 5 to 50, 10 to 100, or 10 to 50 amino acid residues from any of the following: the N-terminal amino acid of the first polypeptide subunit, the C-terminal acid of the first polypeptide subunit, the N-terminal amino acid of the second polypeptide subunit, and the C-terminal amino acid of the polypeptide subunit.
61. The fusion protein of any one of claims 8-60, wherein each barcode fragment is, independently, at least 4 amino acid residues in length.
62. The fusion protein of any one of claims 8-61, wherein each barcode fragment is, independently, 4 to 20, 5 to 15, 6 to 12, 7 to 10, 8 to 13, 10 to 13, or 11 to 12 amino acid residues in length.
63. The fusion protein of any one of claims 8-62, wherein each barcode fragment comprises a glutamate at the C-terminus thereof.
64. The fusion protein of claim 63, wherein each barcode fragment does not include a second glutamate at a position other than the C-terminus of the barcode fragment unless the second glutamate is immediately followed by a proline.

65. The fusion protein of any one of claims 8-641, wherein the amino acid residue that is directly N-terminal to each barcode fragment in the primary amino acid sequence of the fusion protein is a glutamate.
66. The fusion protein of claim 65, wherein the glutamate that is directly N-terminal to each barcode fragment is not immediately adjacent to another glutamate.
67. The fusion protein of any one of claims 8-66, wherein none of the barcode fragments includes a glutamate that is immediately adjacent to another glutamate, if present, in the mask polypeptide that contains the barcode fragment.
68. The fusion protein of any one of claims 8-67, wherein each barcode fragment is independently, selected from Table 2 or Table 3a.
69. The fusion protein of any one of claims 8-68, wherein each barcode fragment comprises only amino acid residues selected from A, E, G, S, P, and T.
70. The fusion protein of any one of claims 8-69, wherein each barcode fragment comprises an amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity, to SGPGSGPGTSE (SEQ ID NO:78), SGPGTSASATPE (SEQ ID NO:1029), SGPGSGPATSE (SEQ ID NO:1028), or SGPGTSPSATPE (SEQ ID NO:79).
71. The fusion protein of any one of claims 8-70, wherein each barcode fragment is selected from the following four amino acid sequences: SGPGSGPGTSE (SEQ ID NO:78), SGPGTSASATPE (SEQ ID NO:1029), SGPGSGPATSE (SEQ ID NO:1028), or SGPGTSPSATPE (SEQ ID NO:79).
72. The fusion protein of any one of claims 8-71, wherein the first barcode fragment (Barcode1) comprises the amino acid sequence of SGPGSGPGTSE (SEQ ID NO:78).
73. The fusion protein of any one of claims 8-72, wherein the second barcode fragment (Barcode2) comprises the amino acid sequence of SGPGTSASATPE (SEQ ID NO:1029).
74. The fusion protein of any one of claims 8-73, wherein the third barcode fragment (Barcode3) comprises the amino acid sequence of SGPGSGPATSE (SEQ ID NO:1028).
75. The fusion protein of any one of claims 8-74, wherein the fourth barcode fragment (Barcode4) comprises the amino acid sequence of SGPGTSPSATPE (SEQ ID NO:79).

76. The fusion protein of any one of claims 1-75, comprising a Glu-C cleavage site comprising one of the following amino acid sequences: ATPESGPG (SEQ ID NO:2223), SGSETPGT (SEQ ID NO:2242), and GTSESATP (SEQ ID NO:204).

77. The fusion protein of any one of claims 1-76, comprising at least one of the following amino acid sequences: SGPE.SGPGXnSGPE.SGPG (SEQ ID NO:2001), SGPE.SGPGXnATPE.SGPG (SEQ ID NO:2002), SGPE.SGPGXnGTSE.SATP (SEQ ID NO:2003), SGPE.SGPGXnTTPE.SGPG (SEQ ID NO:2004), SGPE.SGPGXnSTPE.SGPG (SEQ ID NO:2005), SGPE.SGPGXnGTPE.SGPG (SEQ ID NO:2006), SGPE.SGPGXnGTPE.TPGS (SEQ ID NO:2007), SGPE.SGPGXnSGSE.TGTP (SEQ ID NO:2008), SGPE.SGPGXnGTPE.GSAP (SEQ ID NO:2009), SGPE.SGPGXnEPSE.SATP (SEQ ID NO:2010), ATPE.SGPGXnSGPE.SGPG (SEQ ID NO:2011), ATPE.SGPGXnATPE.SGPG (SEQ ID NO:2012), ATPE.SGPGXnGTSE.SATP (SEQ ID NO:2013), ATPE.SGPGXnATSE.SATP (SEQ ID NO:2014), ATPE.SGPGXnTTPE.SGPG (SEQ ID NO:2015), ATPE.SGPGXnSTPE.SGPG (SEQ ID NO:2016), ATPE.SGPGXnGTPE.SGPG (SEQ ID NO:2017), ATPE.SGPGXnGTPE.TPGS (SEQ ID NO:2018), ATPE.SGPGXnSGSE.TGTP (SEQ ID NO:2019), ATPE.SGPGXnGTPE.GSAP (SEQ ID NO:2020), ATPE.SGPGXnEPSE.SATP (SEQ ID NO:2021), GTSE.SATPXnSGPE.SGPG (SEQ ID NO:2022), GTSE.SATPXnATPE.SGPG (SEQ ID NO:2023), GTSE.SATPXnGTSE.SATP (SEQ ID NO:2024), GTSE.SATPXnTTPE.SGPG (SEQ ID NO:2025), GTSE.SATPXnSTPE.SGPG (SEQ ID NO:2026), GTSE.SATPXnGTPE.SGPG (SEQ ID NO:2027), GTSE.SATPXnGTPE.TPGS (SEQ ID NO:2028), GTSE.SATPXnSGSE.TGTP (SEQ ID NO:2029), GTSE.SATPXnGTPE.GSAP (SEQ ID NO:2030), GTSE.SATPXnEPSE.SATP (SEQ ID NO:2031), TTPE.SGPGXnSGPE.SGPG (SEQ ID NO:2032), TTPE.SGPGXnATPE.SGPG (SEQ ID NO:2033), TTPE.SGPGXnGTSE.SATP (SEQ ID NO:2034), TTPE.SGPGXnTTPE.SGPG (SEQ ID NO:2035), TTPE.SGPGXnSTPE.SGPG (SEQ ID NO:2036), TTPE.SGPGXnGTPE.SGPG (SEQ ID NO:2037), TTPE.SGPGXnGTPE.TPGS (SEQ ID NO:2038), TTPE.SGPGXnSGSE.TGTP (SEQ ID NO:2039), TTPE.SGPGXnGTPE.GSAP (SEQ ID NO:2040), TTPE.SGPGXnEPSE.SATP (SEQ ID NO:2041), STPE.SGPGXnSGPE.SGPG (SEQ ID NO:2042), STPE.SGPGXnATPE.SGPG (SEQ ID NO:2043), STPE.SGPGXnGTSE.SATP (SEQ ID NO:2044), STPE.SGPGXnTTPE.SGPG (SEQ ID NO:2045), STPE.SGPGXnSTPE.SGPG (SEQ ID NO:2046), STPE.SGPGXnGTPE.SGPG (SEQ ID NO:2047), STPE.SGPGXnGTPE.TPGS (SEQ ID NO:2048), STPE.SGPGXnSGSE.TGTP (SEQ ID NO:2049), STPE.SGPGXnGTPE.GSAP (SEQ ID NO:2050), STPE.SGPGXnEPSE.SATP (SEQ ID NO:2051), GTPE.SGPGXnSGPE.SGPG (SEQ ID NO:2052), GTPE.SGPGXnATPE.SGPG (SEQ ID NO:2053), GTPE.SGPGXnGTSE.SATP (SEQ ID NO:2054), GTPE.SGPGXnTTPE.SGPG (SEQ ID NO:2055), GTPE.SGPGXnSTPE.SGPG (SEQ ID NO:2056), GTPE.SGPGXnGTPE.SGPG (SEQ ID NO:2057), GTPE.SGPGXnGTPE.TPGS (SEQ ID NO:2058), GTPE.SGPGXnSGSE.TGTP (SEQ ID NO:2059), GTPE.SGPGXnGTPE.GSAP (SEQ ID NO:2060), GTPE.SGPGXnEPSE.SATP (SEQ ID NO:2061), GTPE.TPGSXnSGPE.SGPG

(SEQ ID NO:2062), GTPE.TPGSX_nATPE.SGPG (SEQ ID NO:2063), GTPE.TPGSX_nGTSE.SATP (SEQ ID NO:2064), GTPE.TPGSX_nTTPE.SGPG (SEQ ID NO:2065), GTPE.TPGSX_nSTPE.SGPG (SEQ ID NO:2066), GTPE.TPGSX_nGTPE.SGPG (SEQ ID NO:2067), GTPE.TPGSX_nGTPE.TPGS (SEQ ID NO:2068), GTPE.TPGSX_nSGSE.TGTP (SEQ ID NO:2069), GTPE.TPGSX_nGTPE.GSAP (SEQ ID NO:2070), GTPE.TPGSX_nEPSE.SATP (SEQ ID NO:2071), SGSE.TGTPX_nSGPE.SGPG (SEQ ID NO:2072), SGSE.TGTPX_nATPE.SGPG (SEQ ID NO:2073), SGSE.TGTPX_nGTSE.SATP (SEQ ID NO:2074), SGSE.TGTPX_nTTPE.SGPG (SEQ ID NO:2075), SGSE.TGTPX_nSTPE.SGPG (SEQ ID NO:2076), SGSE.TGTPX_nGTPE.SGPG (SEQ ID NO:2077), SGSE.TGTPX_nGTPE.TPGS (SEQ ID NO:2078), SGSE.TGTPX_nSGSE.TGTP (SEQ ID NO:2079), SGSE.TGTPX_nGTPE.GSAP (SEQ ID NO:2080), SGSE.TGTPX_nEPSE.SATP (SEQ ID NO:2081), GTPE.GSAPX_nSGPE.SGPG (SEQ ID NO:2082), GTPE.GSAPX_nATPE.SGPG (SEQ ID NO:2083), GTPE.GSAPX_nGTSE.SATP (SEQ ID NO:2084), GTPE.GSAPX_nTTPE.SGPG (SEQ ID NO:2085), GTPE.GSAPX_nSTPE.SGPG (SEQ ID NO:2086), GTPE.GSAPX_nGTPE.SGPG (SEQ ID NO:2087), GTPE.GSAPX_nGTPE.TPGS (SEQ ID NO:2088), GTPE.GSAPX_nSGSE.TGTP (SEQ ID NO:2089), GTPE.GSAPX_nGTPE.GSAP (SEQ ID NO:2090), GTPE.GSAPX_nEPSE.SATP (SEQ ID NO:2091), EPSE.SATPX_nSGPE.SGPG (SEQ ID NO:2092), EPSE.SATPX_nATPE.SGPG (SEQ ID NO:2093), EPSE.SATPX_nGTSE.SATP (SEQ ID NO:2094), EPSE.SATPX_nTTPE.SGPG (SEQ ID NO:2095), EPSE.SATPX_nSTPE.SGPG (SEQ ID NO:2096), EPSE.SATPX_nGTPE.SGPG (SEQ ID NO:2097), EPSE.SATPX_nGTPE.TPGS (SEQ ID NO:2098), EPSE.SATPX_nSGSE.TGTP (SEQ ID NO:2099), EPSE.SATPX_nGTPE.GSAP (SEQ ID NO:2100), or EPSE.SATPX_nEPSE.SATP (SEQ ID NO:2101), wherein each “.” is a Glu-C cleavage site and n is any integer from 0 to 50.

78. The fusion protein of claim 77, comprising at least one of the following amino acid sequences: ATPE.SGPGX_nATPE.SGPG (SEQ ID NO:2102), ATPE.SGPGX_nGTSE.SATP (SEQ ID NO:2103), ATPE.SGPGX_nATSE.SATP (SEQ ID NO:2104), GPE.SGPGX_nATPE.SGPG (SEQ ID NO:2116), ATPE.SGPGX_nTTPE.SGPG (SEQ ID NO:2106), ATPE.SGPGX_nSTPE.SGPG (SEQ ID NO:2107), ATPE.SGPGX_nGTPE.SGPG (SEQ ID NO:2108), GTPE.SGPGX_nGTPE.SGPG (SEQ ID NO:2109), GTPE.SGPGX_nSTPE.SGPG (SEQ ID NO:2110), GTPE.SGPGX_nTTPE.SGPG (SEQ ID NO:2111), GTPE.TPGSX_nSGSE.TGTP (SEQ ID NO:2112), GTPE.GSAPX_nEPSE.SATP (SEQ ID NO:2113), TTPE.SGPGX_nTTPE.SGPG (SEQ ID NO:2114), or STPE.SGPGX_nSTPE.SGPG (SEQ ID NO:2115), wherein each “.” is a Glu-C cleavage site and n is any integer from 0 to 30.

79. The fusion protein of claim 78, comprising at least one of the following amino acid sequence ATPE.SGPGX_nATSE.SATP, wherein “.” is a Glu-C cleavage site and n is any integer from 0 to 30 (SEQ ID NO:2104).

80. The fusion protein of any one of claims 77-79, wherein n is any integer from 1 to 20, 5 to 15, 3 to 7, 2 to 5, or 5 to 10.

81. The fusion protein of any one of claims 77-79, wherein n is 9.

82. The fusion protein of any one of claims 77-79, wherein n is 4.

83. The fusion protein of any one of claims 77-79 wherein n is 3.

84. The fusion protein of any one of claims 77-79, wherein X_n is PGTGTSAT (SEQ ID NO:2233), PGSGPGT (SEQ ID NO:2221), PGTPGTT (SEQ ID NO:2241), PGTPPTST (SEQ ID NO:2235), PGTSPSAT (SEQ ID NO:2238), PGTGSAGT (SEQ ID NO:2230), PGTGGAGT (SEQ ID NO:2228), PGTSPGAT (SEQ ID NO:2237), PGTSGSGT (SEQ ID NO:2236), PGTSSAST (SEQ ID NO:2239), PGTGAGTT (SEQ ID NO:2227), PGTGSTST (SEQ ID NO:2232), GSEPATSG (SEQ ID NO:2224), APGTSTEP (SEQ ID NO:2222), PGTAGSGT (SEQ ID NO:2226), PGTSSGGT (SEQ ID NO:2240), PGTAGPAT (SEQ ID NO:2225), PGTPGTGT (SEQ ID NO:2234), PGTGGPTT (SEQ ID NO:2229), or PGTGSGST (SEQ ID NO:2231).

85. The fusion protein of any one of claims 77-79, wherein X_n is TSAS (SEQ ID NO:2214), TGTS (SEQ ID NO:2211), SGP, TTPG (SEQ ID NO:2220), TPPT (SEQ ID NO:2213), TSPS (SEQ ID NO:2217), TGSA (SEQ ID NO:2208), TGGA (SEQ ID NO:2206), TSPG (SEQ ID NO:2216), TSGS (SEQ ID NO:2215), TSSA (SEQ ID NO:2218), TGAG (SEQ ID NO:2205), TGST (SEQ ID NO:2210), EPAT (SEQ ID NO:2201), GTST (SEQ ID NO:2202), TAGS (SEQ ID NO:2204), TSSG (SEQ ID NO:2219), TAGP (SEQ ID NO:2203), TPGT (SEQ ID NO:2212), TGGP (SEQ ID NO:2207), or TGSG (SEQ ID NO:2209).

86. The fusion protein of any one of claim 77-79, wherein X_n is SGP,

87. The fusion protein of any one of claims 7-86, wherein the first polypeptide subunit comprises an amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity, to a sequence listed in Table A1 (SEQ ID NOs: 1000-1005).

88. The fusion protein of any one of claims 7-86, wherein the first polypeptide subunit comprises an amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity, to:

ASGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSEPATSGSETPGT
SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP

ESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT
SESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE
GSAPGSEPATSGSETPGTSESAT**EAGRSASHTPAGLTGPGT**PESATPESRNLLPVATPDPMFP
CLHHSQNLLRAVSNMLQKARQTLEFYPTCTSEEIDHEDITKDKTSTVEACLPLELTKNESCLNS
RETSFITNGSCLASRKTSFMMALCLSSIEDLKMYQVEFKTMNAKLLMDPKRQIFLDQNMMLA
VIDELMQALNFNSETVPQKSSLEEPDFYKTKIKLCILLHAFRIRAVTIDRVMSYLNASGTATPE
SAPEAGRSASHTPAGLTGPATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGT
SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATP
ESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGT
SESATPESGPGTSESATPESGPGTSASATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPT
STEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGEPEA (SEQ ID NO: 1000).

89. The fusion protein of any one of claims 7-88, wherein the second polypeptide subunit comprises an amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity, to a sequence listed in Table A2 (SEQ ID NOs: 1033-1036).

90. The fusion protein of any one of claims 7-89, wherein the second polypeptide subunit comprises an amino acid sequence that has at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity, to:

ASGPGTSTEPSEGSAPGTSESATPESGPGSPATSESATPGTSESATPESGPGSEPATSGSETPGT
SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP
ESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT
SESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE
GSAPGSEPATSGSETPGTSESAT**EAGRSASHTPAGLTGPT**SESATPESIWELKKDVYVVELD
WYPDAPGEMVVLTCDTPEEDGITWLDQSSEVLGSGKTLTIQVKEFGDAGQYTCHKGGEVL
SHSLLLLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRFTCWWLTTISTDLTFSVKSSR
GSSDPQGVTCGAATLSAERVGRDNKEYEYSVEQCQEDSACPAAEESLPIEVMVDVAVHKLKYE
NYTSSFFIRDIIKPDPPKNLQKPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCVQVQGKSKRE
KKDRVFTDKTSATVICRKNASISVRAQDRYYSSSWSEWASVPCSGTATPESGPG**EAGRSASH**
TPAGLTGPATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGT
STEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATP
ESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGS
PAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSPSATP
ESGPGSEPATSGSETPGSEPATSGSETPGSPAG (SEQ ID NO: 1033).

91. The fusion protein of any one of claims 7-87, wherein the first polypeptide subunit comprises the amino acid sequence of:

ASGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSEPATSGSETPGT
SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP
ESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT
SESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE
GSAPGSEPATSGSETPGTSESAT**EAGRSASHTPAGLTG**PGTPESATPESRNLPVATPDPMFP
CLHHSQNLLRAVSNMLQKARQTLEFYPCTSEEIDHEDITKDKTSTVEACLPLELTKNESCLNS
RETSFITNGSCLASRKTSFMMALCLSSIIYEDLKMYQVEFKTMNAKLLMDPKRQIFLDQNMLA
VIDELMQALNFNSETVPQKSSLEEPDFYKTKIKLCILLHAFRIRAVTIDRVMSYLNASGTATPE
SA**EAGRSASHTPAGLTG**PATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGT
SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATP
ESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGT
SESATPESGPGTSESATPESGPGTSASATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPT
STEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGEPEA (SEQ ID NO: 1000).

and wherein the second polypeptide subunit comprises the amino acid sequence of:

ASGPGTSTEPSEGSAPGTSESATPESGPGSGPATSESATPGTSESATPESGPGSEPATSGSETPGT
SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP
ESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT
SESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE
GSAPGSEPATSGSETPGTSESAT**EAGRSASHTPAGLTG**PTSESATPESIWELKKDVYVVELD
WYPDAPGEMVVLTCDTPEEDGITWTLDQSSEVLGSGKTLTIQVKEFGDAGQYTCHKGGEVL
SHSLLLLHKKEDGIWSTDILKDQKEPKNKTFLRCEAKNYSGRFTCWWTITISTDLTFSVKSSR
GSSDPQGVTCGAATLSAERVRGDNKEYEYSVEQEDSACPAAEESLPIEVMVDVAVHKLKYE
NYTSSFFIRDIHKDPPPKNLQLKPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCVQVQGKSKRE
KKDRVFTDKTSATVICRKNASISVRAQDRYYSSSWSEWASVPCSGTATPESGPG**EAGRSASH**
TPAGLTGPATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGT
STEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATP
ESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGS
PAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSPSATP
ESGPGSEPATSGSETPGSEPATSGSETPGSPAG (SEQ ID NO: 1033).

92. A fusion protein comprising a first polypeptide subunit and a second polypeptide subunit, wherein the first polypeptide subunit and the second polypeptide subunit are disulfide-linked thereby forming a disulfide-linked heterodimer, wherein the first polypeptide subunit comprises an amino acid sequence having 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity, to:

ASGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSEPATSGSETPGT
 SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP
 ESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT
 SESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE
 GSAPGSEPATSGSETPGTSESATEAGRSASHTPAGLTGPGTPESATPESRNLPVATPDPGMFPC
 LHHSQNLLRAVSNMLQKARQTLEFYPTSEEIDHEDITKDKTSTVEACLPLELTKNESCLNSR
 ETSFITNGSCLASRKTSFMMALCLSSIIYEDLKMYQVEFKTMNAKLLMDPKRQIFLDQNMALAV
 IDELMQALNFNSETVPQKSSLEEPDFYKTKIKLCILLHAFRIRAVTIDRVMSYLNASGTATPES
 APEAGRSASHTPAGLTGPATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSE
 SATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPES
 GPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSE
 SATPESGPGTSESATPESGPGTSASATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTST
 EEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGPEA (SEQ ID NO: 1000).

and wherein the second polypeptide subunit comprises an amino acid sequence having 85%,

90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity, or 100% identity, to:

ASGPGTSTEPSEGSAPGTSESATPESGPGSGPATSESATPGTSESATPESGPGSEPATSGSETPGT
 SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP
 ESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT
 SESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE
 GSAPGSEPATSGSETPGTSESATE**AGRSASHTPAGLTGPT**SESATPES**IWELKKDVVVELD**
WYPDAPGEMVVLTCDTPEEDGITW**LDQSS**EV**LGSGKTLTIQVKEFGDAGQY**CHK**GGEVL**
SHSLLLL**HKKEDGIW**STD**ILKDQKEPKNK**TFLRCEAKNYSGRFTCW**WLT**TTISTDLTFSVK**SSR**
GSSDPQGVTCGAATLSAERV**R**GD**NKEYEYS**VE**CQEDS**ACPA**AEESLPI**EV**MVD**AV**HKLKYE**
NYTSSFF**IRDI**IK**PDPPKNLQ**L**KPLKNSRQ**VEVS**WEY**PDTW**STPHSY**FS**L**TFCV**QVQ**GK**SKRE**
KKDRVFTDKTSATVICRKNASIS**VRAQDRYSSSWSEWASVPC**SGTATPESGPG**EAGRSASH**
TPAGLTGPATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGT
 STEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATP
 ESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGS
 PAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSPSATP
 ESGPGSEPATSGSETPGSEPATSGSETPGSPAG (SEQ ID NO: 1033).

93. The fusion protein of claim 92, wherein the first polypeptide subunit comprises the amino acid sequence of:

ASGPGTSTEPSEGSAPGTSESATPESGPGSGPGTSESATPGTSESATPESGPGSEPATSGSETPGT
 SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP
 ESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT
 SESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE

GSAPGSEPATSGSETPGTSESATE**EAGRSASHTPAGLTGPGT**PESATPESRNLVPATPDPMFP
CLHHSQNLLRAVSNMLQKARQTLEFYPTCTSEEIDHEDITKDKTSTVEACLPLELTKNESCLNS
RETSFITNGSCLASRKTSFMMALCLSSIEDLKMYQVEFKTMNAKLLMDPKRQIFLDQNMLA
VIDELMQALNFNSETVPQSSLEEDFYKTKIKLCILLHAFRIRAVTIDRVMSYLNASGTATPE
 SAPE**EAGRSASHTPAGLTGP**ATPESGPGTSTEPSEGSAPGTSESATPESGPGSEPATSGSETPGT
 SESATPESGPGSEPATSGSETPGTSESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATP
 ESGPGSEPATSGSETPGTSESATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGT
 SESATPESGPGTSESATPESGPGTSASATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPT
 STEEGTSTEPSEGSAPGTSTEPSEGSAPGSEPATSGSETPGEPEA (SEQ ID NO: 1000),

and wherein the second polypeptide subunit comprises the amino acid sequence of:

ASGPGTSTEPSEGSAPGTSESATPESGPGSPATSESATPGTSESATPESGPGSEPATSGSETPGT
 SESATPESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATP
 ESGPGSPAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGT
 SESATPESGPGSEPATSGSETPGSEPATSGSETPGSPAGSPTSTEEGTSTEPSEGSAPGTSTEPSE
 GSAPGSEPATSGSETPGTSESATE**EAGRSASHTPAGLTGPT**SESATPESIWELKKDVYVVELD
WYPDAPGEMVVLTCDTPEEDGITWTLDOSSSEVLGSGKTLTIQVKEFGDAGQYTCHKGGEVL
SHSLLLLHKKEDGIWSTDILKDQKEPKNTFLRCEAKNYSGRFTCWWTITISTDLTFSVKSSR
GSSDPQGVTCGAATLSAERVGRDNKEYEYSVECOEDSACPAAEESLPIEVMVDAVHKLKYE
NYTSSFFIRDIIKPDPPKNLQLKPLKNSRQVEVSWEYPDTWSTPHSYFSLTFCVQVQGKSKRE
KKDRVFTDKTSATVICRKNASISVRAQDRYYSSSWSEWASVPCSGTATPESGPG**EAGRSASH**
TPAGLTGPATPESGPGSPAGSPTSTEEGSPAGSPTSTEEGSPAGSPTSTEEGTSESATPESGPGT
 STEPSEGSAPGTSESATPESGPGSEPATSGSETPGTSESATPESGPGSEPATSGSETPGTSESATP
 ESGPGTSTEPSEGSAPGSPAGSPTSTEEGTSESATPESGPGSEPATSGSETPGTSESATPESGPGS
 PAGSPTSTEEGSPAGSPTSTEEGTSTEPSEGSAPGTSESATPESGPGTSESATPESGPGTSPSATP
 ESGPGSEPATSGSETPGSEPATSGSETPGSPAG (SEQ ID NO: 1033).

94. A pharmaceutical composition comprising the fusion of any one of claims 1-93 and at least one pharmaceutically acceptable excipient.

95. A polynucleotide sequence encoding the fusion protein of any one of claims 1-93.

96. An expression vector comprising the polynucleotide sequence of claim 95.

97. A host cell comprising the expression vector of claim 96.

98. A method of producing the fusion protein of any one of claims 1-93.

99. The method of claim 98, further comprising isolating the fusion protein from a host cell.
100. A method of treating cancer in a subject in need thereof, the method comprising administering an effective amount of the fusion protein of any one of claims 1-93 or the pharmaceutical composition of claim 94 to the subject.
101. The method of claim 100, wherein the cancer comprises a solid tumor.