



US 20250258180A1

(19) **United States**

(12) **Patent Application Publication**

KABAYAMA

(10) **Pub. No.: US 2025/0258180 A1**

(43) **Pub. Date: Aug. 14, 2025**

(54) **PEPTIDE TAG AND NUCLEIC ACID
ENCODING SAME**

(71) Applicant: **STAND Therapeutics Co., Ltd.**,
Minato-ku, Tokyo (JP)

(72) Inventor: **Hiroyuki KABAYAMA**, Tokyo (JP)

(73) Assignee: **STAND Therapeutics Co., Ltd.**,
Minato-ku, Tokyo (JP)

(21) Appl. No.: **18/292,562**

(22) PCT Filed: **Jul. 26, 2022**

(86) PCT No.: **PCT/JP2022/028746**

§ 371 (c)(1),
(2) Date: **Jan. 26, 2024**

(30) **Foreign Application Priority Data**

Jul. 27, 2021 (JP) 2021-122510

Publication Classification

(51) **Int. Cl.**

G01N 33/68 (2006.01)

C07K 16/18 (2006.01)

(52) **U.S. Cl.**

CPC **G01N 33/6845** (2013.01); **C07K 16/18**
(2013.01); **C07K 2317/622** (2013.01); **C07K
2317/80** (2013.01); **C07K 2319/00** (2013.01)

(57)

ABSTRACT

The present disclosure provides a peptide tag, and a nucleic acid encoding the peptide tag. The peptide tag of the present disclosure can reduce an aggregation property of a protein in a cell. Specifically, the peptide tag of the present disclosure can be a peptide tag in which 5% or more and less than 45% of amino acids contained in an amino acid sequence thereof are acidic amino acids, and (b) 20% or more of the amino acids contained in the amino acid sequence are amino acids selected from the group consisting of F, P, Y, G, S, Q, N, and A.

Specification includes a Sequence Listing.

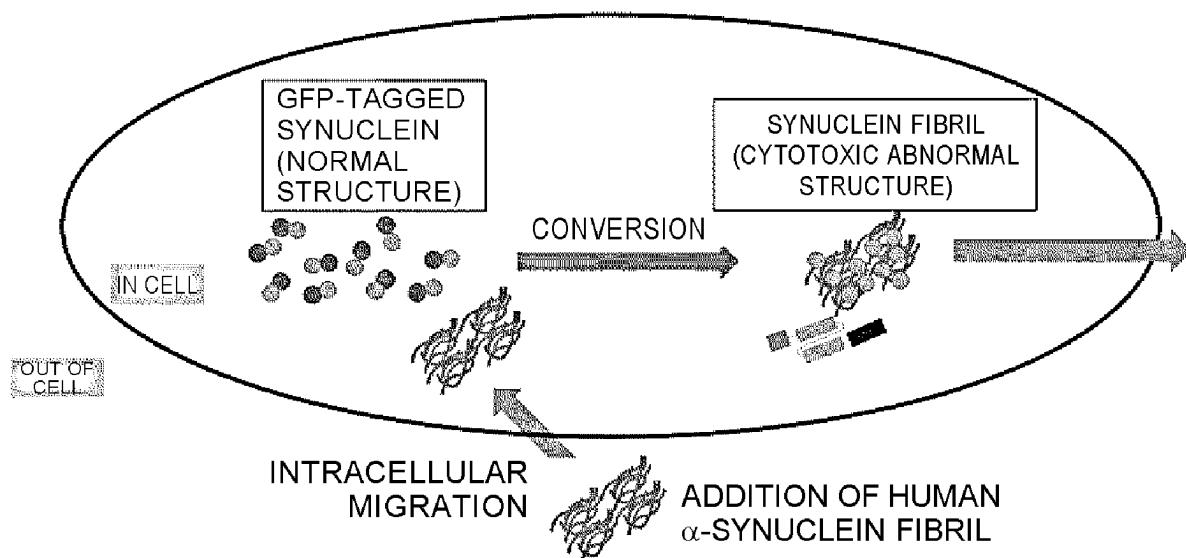


FIG. 1

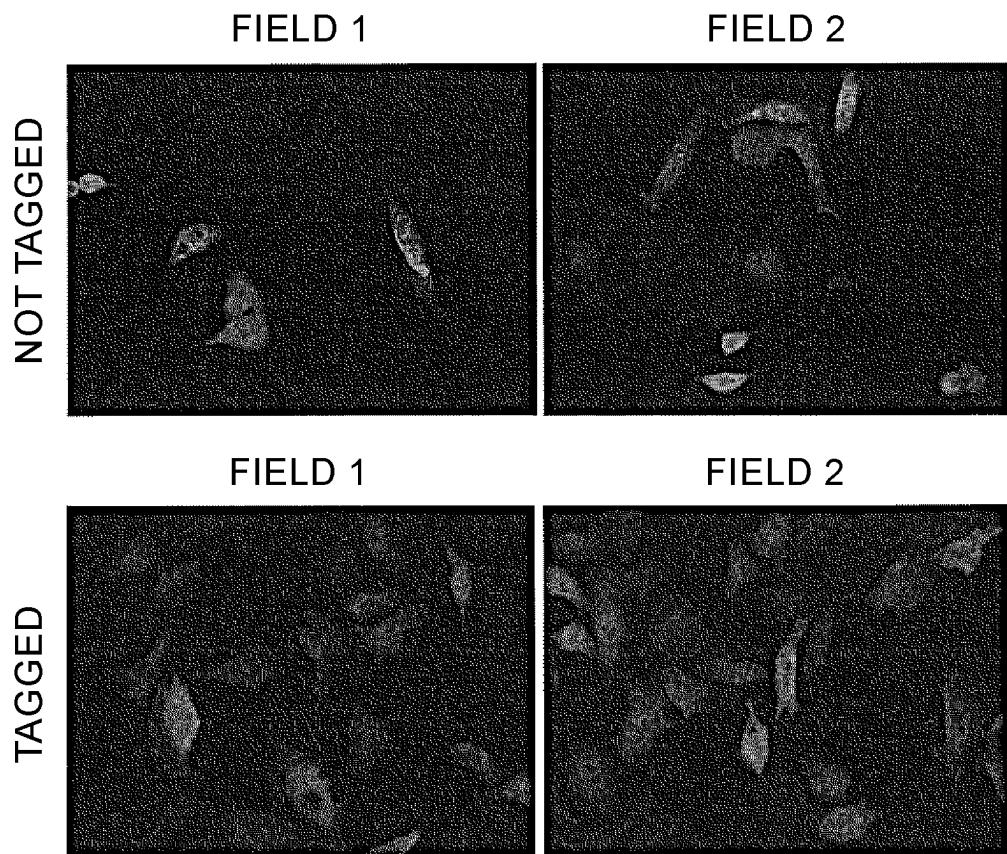


FIG. 2A

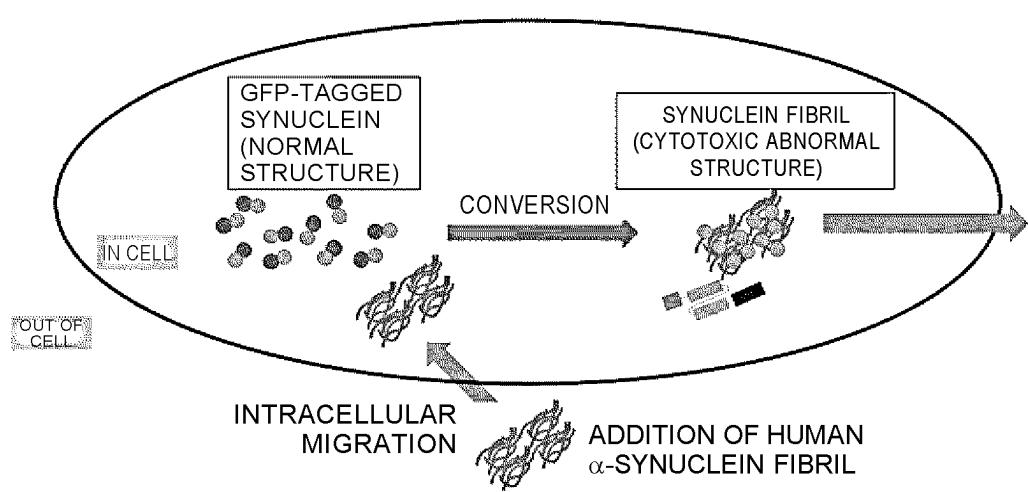


FIG. 2B

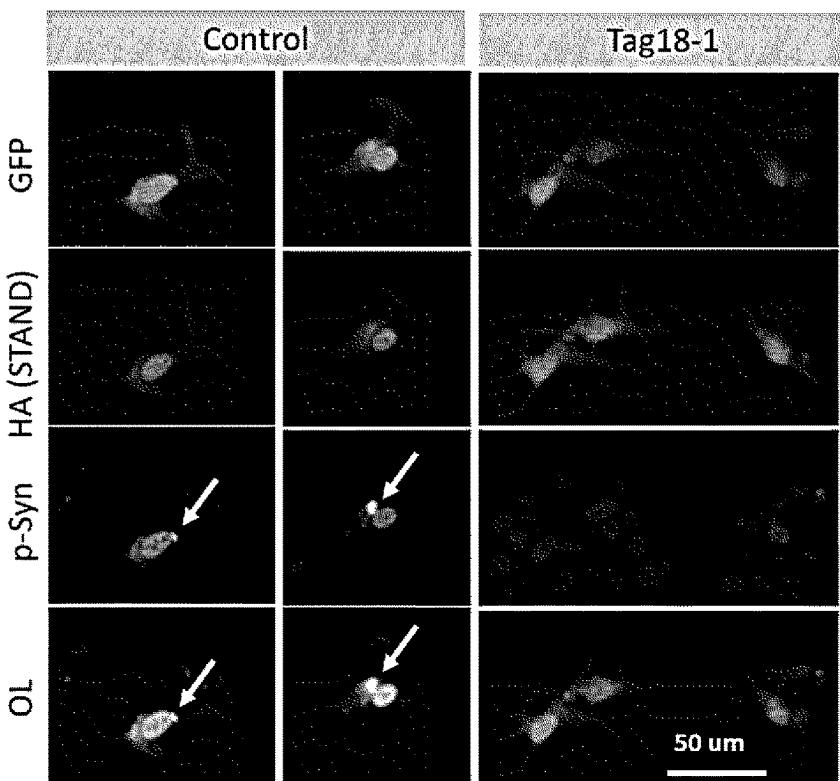


FIG. 2C

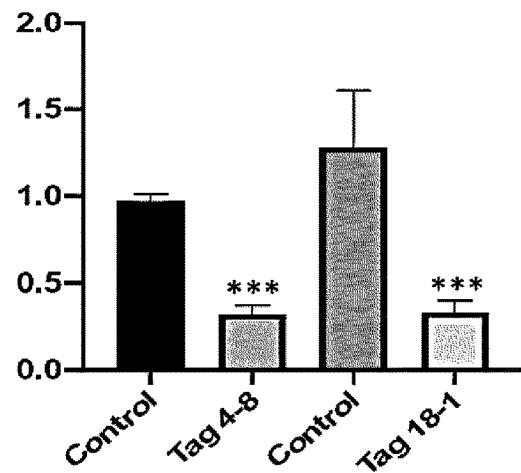


FIG. 3A

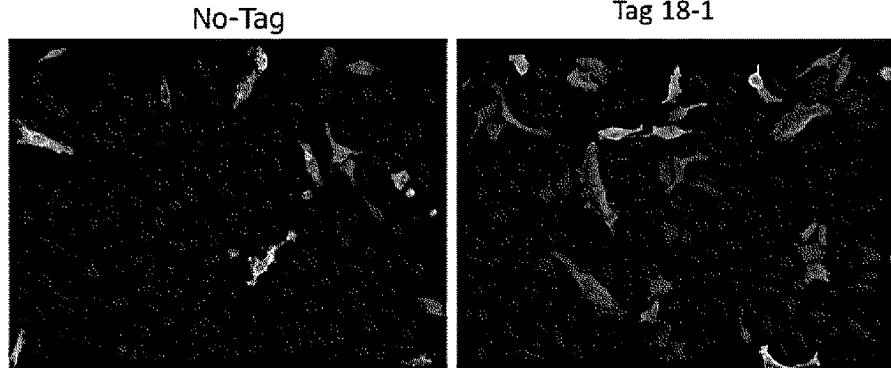
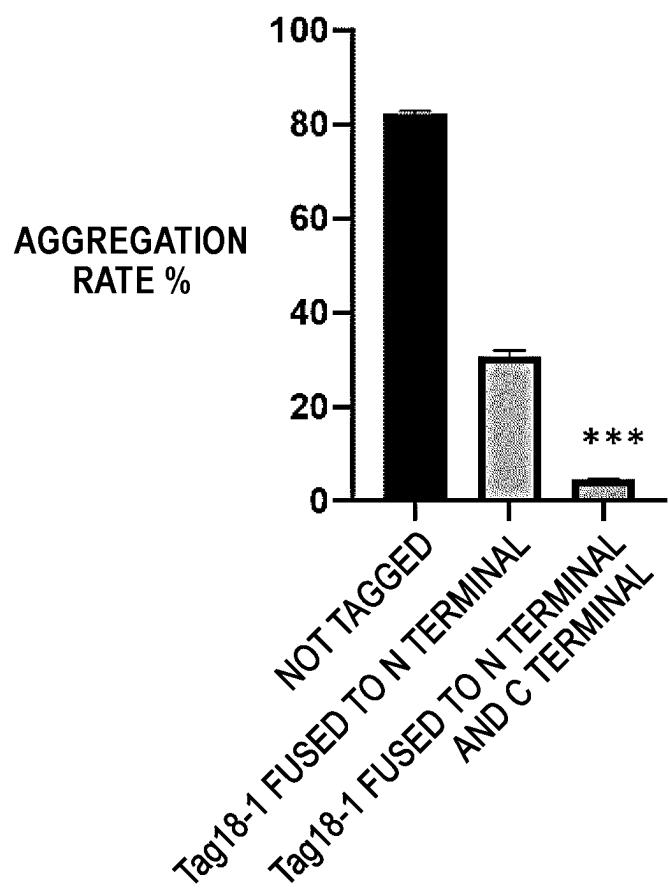


FIG. 3B



**PEPTIDE TAG AND NUCLEIC ACID
ENCODING SAME****TECHNICAL FIELD**

[0001] The present disclosure relates to a peptide tag, and a nucleic acid encoding the peptide tag.

BACKGROUND ART

[0002] An antibody functioning in a cell, namely, an intrabody (intracellular antibody) can affect the function of the cell by recognizing and binding to an antigen (target molecule) in the cell of a higher organism. Such an antigen can be a significant intracellular therapeutic target that can be inactivated by binding to an intracellular antibody. As a research method, use of an intracellular antibody attracts attention as means for specifically inhibiting the function of a protein directly by binding to the antibody in the cell.

[0003] In case of an intracellular antibody, a hybridoma producing a monoclonal antibody recognizing an antigen is first produced by a standard method, and from the cDNA thereof, an intracellular expression vector containing a DNA encoding a single chain antibody (single chain Fv: scFv) is constructed to obtain a complex of a heavy chain (VH) and a light chain (VL) as an intracellular antibody. Recently, a phage library for presenting a human scFv from an antibody isolated from a human B cell is produced, and is used for isolation of a scFv binding to an intracellular antigen in some cases.

[0004] An antibody usually moves around in an extracellular space such as blood in a body, and recognizes an extracellular antigen to function, and hence works in the extracellular space as a premise. Accordingly, if an antibody is expressed in the cytoplasm, there arise problems of reduction of the expression level, folding causing limitation of a half-life of an antibody domain, and stability. The problem of stability of an intracellular antibody in the cytoplasm can lead to formation of an aggregation of the intracellular antibody in the cytoplasm. The formation of the aggregation can lead to reduction of a production amount of the intracellular antibody, and inhibition of expression of normal function. The same applies to a protein except for the intracellular antibody. The intracellular antibody has a characteristic of easily aggregating in particular, but a protein except for the intracellular antibody also can form an aggregation in the cytoplasm when produced in the cytoplasm.

[0005] By contrast, it has been shown that a peptide tag having an amino acid sequence containing 45% or more of acidic amino acids improves stability of an intracellular antibody (Patent Literature 1, and Non Patent Literature 1). In proposing the effectiveness of a peptide tag having an amino acid sequence containing 45% or more of acidic amino acids, Patent Literature 1 and Non Patent Literature 1 point out, as a design guideline for the peptide tag, significance of designing the peptide tag in such a manner that a charge value and a pI value are sufficiently low based not on the pH environment of the cytoplasm but on the pH environment on the surface of an endosome on the side of the cytoplasm. In Non Patent Literature 3, a membrane localization signal of HRAS is added to a heavy chain variable region of an antibody.

CITATION LIST**Patent Literature**

[0006] Patent Literature 1: WO2019/004213

Non Patent Literature

[0007] Non Patent Literature 1: Kabayama et al., 2020, Nature Communication, 11, 336

[0008] Non Patent Literature 2: Shubhada et al., 2012, Biochemical genetics, Vol. 50, No. 7-8, pp. 625-41

[0009] Non Patent Literature 3: Tanaka et al., 2007, EMBO Journal, 26:3250-3259

SUMMARY OF INVENTION

[0010] The present disclosure provides a peptide tag, and a nucleic acid encoding the peptide tag. The peptide tag of the present disclosure can reduce the aggregation property of a protein in a cell.

[0011] The present inventors made earnest studies on peptide tags having various amino acid sequences, resulting in finding a peptide tag having an effect of reducing an aggregation property of a protein in a cell.

[0012] The present disclosure provides the following inventions:

[0013] [1] A peptide having an amino acid sequence with a length of, for example, 600 amino acids or less, for example, 10 to 200 amino acids (for example, 10 to 90 amino acids),

[0014] wherein (a) 5% or more and less than 45% of amino acids contained in the amino acid sequence are acidic amino acids, and

[0015] (b) 20% or more, and preferably 30% or more of the amino acids contained in the amino acid sequence are amino acids selected from the group consisting of F, P, Y, G, S, Q, N, and A, and

[0016] the peptide is preferably capable of reducing an aggregation property in a cell of a protein linked to the peptide, wherein 10% or 15% or more of the amino acids contained in the amino acid sequence are N or P.

[0017] [2] The peptide according to [1] above, wherein 30% or less, preferably 20% or less, more preferably 15% or less, and further preferably 10% or less of the amino acids contained in the amino acid sequence are amino acids selected from the group consisting of M, T, W, C, I, V, and L.

[0018] [3] The peptide according to [1] or [2] above, wherein each of A and G constitutes less than 10% of the amino acids contained in the amino acid sequence thereof.

[0019] [4] The peptide according to any one of [1] to [3] above, wherein

[0020] (a) 20% or more and less than 45% of the amino acids contained in the amino acid sequence are acidic amino acids,

[0021] (b) 30% or more and less than 70% of the amino acids contained in the amino acid sequence are amino acids selected from the group consisting of F, P, Y, G, S, Q, N, and A,

[0022] (c) 20% or less of the amino acids contained in the amino acid sequence are amino acids selected from the group consisting of M, T, W, C, I, V, and L, and

[0023] (d) each of A and G constitutes less than 10% of the amino acids contained in the amino acid sequence.

- [0024] [5] A peptide having an amino acid sequence set forth in any one of SEQ ID NOs: 2 to 11.
- [0025] [6] A nucleic acid encoding the peptide according to any one of [1] to [5] above.
- [0026] [7] A protein expression vector comprising: the nucleic acid according to [6] above operably linked to a regulatory sequence; and a nucleic acid encoding a protein of interest in-frame to the nucleic acid according to [6] above.
- [0027] [8] The protein expression vector according to [7] above, wherein the protein of interest is an antibody, or an antigen-binding fragment of an antibody.
- [0028] [9] The protein expression vector according to [8] above, wherein the antigen-binding fragment of the antibody is a single chain Fv (scFv).
- [0029] [10] A fusion protein of the peptide according to any one of [1] to [5] above and a protein of interest.
- [0030] [11] The fusion protein according to [10] above, wherein the protein of interest is an antibody, or an antigen-binding fragment of an antibody.
- [0031] [12] The fusion protein according to [11] above, wherein the antigen-binding fragment of the antibody is a single chain Fv (scFv).
- [0032] [13] A protein-producing cell comprising: the nucleic acid according to [6] above operably linked to a regulatory sequence; and a nucleic acid encoding a protein of interest in-frame to the nucleic acid according to [6] above.
- [0033] [14] A method for selecting or identifying an amino acid sequence having a length of 600 amino acids or less, for example, 10 to 200 amino acids (for example, 10 to 90 amino acids), the method comprising:
- [0034] acquiring, from an amino acid sequence (group) (that can include an amino acid sequence (group) having a length of 10 to 200 amino acids (for example, 10 to 90 amino acids)), an amino acid sequence (group) in which (a) 5% or more and less than 45% of amino acids contained in the amino acid sequence are acidic amino acids, and (b) 20% or more of the amino acids contained in the amino acid sequence are amino acids selected from the group consisting of F, P, Y, G, S, Q, N, and A;
- [0035] selecting or identifying an amino acid sequence of a peptide tag that, when the fusion protein of a peptide tag having the selected or identified amino acid sequence and a reference protein is expressed in a mammal cell (preferably in a human cell), provides reduction of a proportion of cells in which the fusion protein forms an aggregation (for example, the proportion which is not more than a predetermined value); and
- [0036] obtaining a peptide tag having the amino acid sequence, or a nucleic acid encoding the peptide tag.
- [0037] [15] The method according to claim 14, wherein the amino acid sequence group to be acquired is the peptide according to any one of [1] to [5] above.
- [0038] [16] The method according to [14] or [15] above, wherein the amino acid sequence group to be acquired is a group of amino acid sequences encoded by coding regions of human genome.
- [0039] [17] The method according to any one of [14] to [16] above, wherein the amino acid sequence to be acquired contains a neo-antigen.
- [0040] [18] A peptide satisfying one or more selected from the group consisting of (a) to (h) described below, and capable of reducing an aggregation property in a cell of a protein linked to the peptide.
- [0041] [19] A peptide selected from the group consisting of (A) to (AE) and (AF) to (AU) described below, and capable of reducing an aggregation property in a cell of a protein linked to the peptide.
- [0042] [20] A nucleic acid encoding the peptide according to [18] above.
- [0043] [21] A protein expression vector comprising: the nucleic acid according to [20] above operably linked to a regulatory sequence; and a nucleic acid encoding a protein of interest in-frame to the nucleic acid according to [20] above.
- [0044] [22] A fusion protein of the peptide according to [18] above and a protein of interest.
- [0045] [23] The fusion protein according to [22] above, wherein the protein of interest is an antibody, or an antigen-binding fragment of an antibody.
- [0046] [24] The fusion protein according to [23] above, wherein the antigen-binding fragment of the antibody is a single chain Fv (scFv).
- [0047] [25] A nucleic acid encoding the peptide according to [19] above.
- [0048] [26] A protein expression vector comprising: the nucleic acid according to [25] above operably linked to a regulatory sequence; and a nucleic acid encoding a protein of interest in-frame to the nucleic acid according to [25] above.
- [0049] [27] A fusion protein of the peptide according to [25] above, and a protein of interest.
- [0050] [28] The fusion protein according to [26] above, wherein the protein of interest is an antibody, or an antigen-binding fragment of an antibody.
- [0051] [29] The fusion protein according to [27] above, wherein the antigen-binding fragment of the antibody is a single chain Fv (scFv).
- [0052] [30] The method according to [14] above, wherein the amino acid sequence (group) includes the peptide according to [18] above.
- [0053] [31] The method according to [14] above, wherein the amino acid sequence (group) includes the peptide according to [19] above.
- [0054] [32] The method according to [14] above, wherein the reference protein is a scFv, and the predetermined value is a value of 30% or less.
- [0055] [33] The method according to [14] above, wherein the reference protein is a scFv, and the predetermined value is a value of 20% or less.
- [0056] [34] The method according to [14] above, wherein the reference protein is a scFv, and the predetermined value is a value of 15% or less.
- [0057] [35] The method according to [14] above, wherein the reference protein is a scFv, and the predetermined value is a value of 10% or less.
- [0058] [36] The method according to [14] above, wherein the reference protein is a scFv, and the predetermined value is a value of 5% or less.
- [0059] [37] The method according to [14] above, wherein the reference protein is a scFv, the amino acid sequence group includes the peptide according to (A) described below, and the amino acid sequence to be acquired preferably further satisfies a condition of (A) described below.

[0108] [70] The method according to any one of [37] to [67] and [67A] to [67N] above, wherein the reference protein is a scFv, and the predetermined value is a value of 5% or less.

[0109] [71] The method according to any one of [37] to [70] above, wherein a proportion of cells in which the reference protein forms an aggregation is a value more than 30%.

[0110] [72] The method according to any one of [37] to [70] above, wherein a proportion of cells in which the reference protein forms an aggregation is a value in a range of 30 to 40%.

[0111] [73] The method according to any one of [37] to [70] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 40 to 50%.

[0112] [74] The method according to any one of [37] to [70] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 50 to 60%.

[0113] [75] The method according to any one of [37] to [70] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 60 to 70%.

[0114] [76] The method according to any one of [37] to [70] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 70 to 80%.

[0115] [77] The method according to any one of [37] to [70] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 80 to 90%.

[0116] [78] The method according to any one of [37] to [70] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 90 to 95%.

[0117] [79] The method according to any one of [37] to [70] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 95 to 99%.

[0118] [80] The method according to any one of [37] to [70] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 99 to 99.9%.

[0119] [81] The method according to any one of [37] to [70] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 99.9 to 100%.

[0120] [82] The method according to [69] above, wherein a proportion of cells in which the reference protein forms an aggregation is a value more than 30%.

[0121] [83] The method according to [69] above, wherein a proportion of cells in which the reference protein forms an aggregation is a value in a range of 30 to 40%.

[0122] [84] The method according to [69] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 40 to 50%.

[0123] [85] The method according to [69] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 50 to 60%.

[0124] [86] The method according to [69] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 60 to 70%.

[0125] [87] The method according to [69] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 70 to 80%.

[0126] [88] The method according to [69] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 80 to 90%.

[0127] [89] The method according to [69] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 90 to 95%.

[0128] [91] The method according to [69] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 99 to 99.9%.

[0129] [92] The method according to [69] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 99.9 to 100%.

[0130] [93] The method according to [70] above, wherein a proportion of cells in which the reference protein forms an aggregation is a value more than 30%.

[0131] [94] The method according to [70] above, wherein a proportion of cells in which the reference protein forms an aggregation is a value in a range of 30 to 40%.

[0132] [95] The method according to [70] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 40 to 50%.

[0133] [96] The method according to [70] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 50 to 60%.

[0134] [97] The method according to [70] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 60 to 70%.

[0135] [98] The method according to [70] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 70 to 80%.

[0136] [99] The method according to [70] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 80 to 90%.

[0137] [100] The method according to [70] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 90 to 95%.

[0138] [101] The method according to [70] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 95 to 99%.

[0139] [102] The method according to [70] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 99 to 99.9%.

[0140] [103] The method according to [70] above, wherein a rate of cells in which the reference protein forms an aggregation is a value in a range of 99.9 to 100%.

[0141] [104] The peptide according to any one of [1] to [5] above, wherein a peptide tag is capable of reducing an aggregation property of a scFv having at least an amino acid sequence set forth in SEQ ID NO: 1.

[0142] [105] The nucleic acid according to [6] above, wherein a peptide tag is capable of reducing an aggregation property of a scFv having at least an amino acid sequence set forth in SEQ ID NO: 1.

[0143] [106] The protein expression vector according to any one of [7] to [9] above, wherein a peptide tag is capable of reducing an aggregation property of a scFv having at least an amino acid sequence set forth in SEQ ID NO: 1.

[0144] [107] The protein expression vector according to any one of [7] to [9] above, wherein the protein expression vector is a virus vector.

[0145] [108] The protein expression vector according to [107] above, wherein the virus vector is selected from the group consisting of a retrovirus vector, a lentivirus vector, an adenovirus vector, an adeno-associated virus vector, a herpes simplex virus vector, a vaccinia virus vector, a Sendai virus vector, and a vesicular stomatitis virus vector.

[0146] [109] The nucleic acid according to [6] above, wherein the nucleic acid is an mRNA.

[0147] [110] The nucleic acid according to [109] above, wherein the nucleic acid has a cap structure at the 5' end, and a poly A chain at the 3' UTR.

[0148] [111] The nucleic acid according to [109] or [110] above, wherein the nucleic acid contains pseudouridine as U.

[0149] [112] A nanoparticle, comprising the nucleic acid according to any one of [109] to [111] above.

[0150] [113] The nanoparticle according to [112], wherein the nanoparticle is a lipid nanoparticle.

[0151] [114] The method according to any one of [14] to [17] and [30] to [103] above, wherein the reference protein has an amino acid sequence set forth in SEQ ID NO: 1.

[0152] [115] The method, the peptide, the fusion protein, the nucleic acid, or the vector according to any one of those described above, wherein the cell is a eukaryotic cell.

[0153] [116] The method, the peptide, the fusion protein, the nucleic acid, or the vector according to any one of those described above, wherein the cell is a human cell.

[0154] [117] The method, the peptide, the fusion protein, the nucleic acid, or the vector according to any one of those described above, wherein a peptide tag does not prevent free localization of a protein of interest.

[0155] A peptide tag of the present disclosure can cause an intracellular stability of a tagged protein. Accordingly, the peptide tag of the present disclosure can be a more highly biocompatible peptide tag.

BRIEF DESCRIPTION OF DRAWINGS

[0156] FIG. 1 illustrates an effect of a peptide tag Tag4-1 on an aggregation property of a single chain Fv (scFv) in a cell.

[0157] FIG. 2A illustrates a scheme for constructing a model of an intracellular accumulation of α -synuclein, that is, an amyloid.

[0158] FIG. 2B illustrates fluorescence microscope images showing influence on intracellular synuclein fibril caused by intracellular expression of scFv-E6-CMA peptide fusion protein having Tag18-1, that is, one of peptide tags of the present disclosure.

[0159] FIG. 2C illustrates an effect of removing synuclein fibril by intracellular expression of scFv-E6-CMA peptide fusion protein having Tag4-8 or Tag18-1, that is, one of peptide tags of the present disclosure.

[0160] FIG. 3A illustrates fluorescence microscope images showing intracellular localization of a scFv-C2 itself having Tag18-1, that is, one of peptide tags of the present disclosure expressed in the cell.

[0161] FIG. 3B illustrates a stabilizing action of the scFv-C2 having Tag18-1, that is, one of peptide tags of the present disclosure.

DESCRIPTION OF EMBODIMENTS

[0162] In the present invention, the term "subject" is a vertebrate, examples include birds and mammals, and spe-

cific examples include mammals such as a mouse, a rat, a hamster, a guinea pig, a horse, a cow, a pig, a goat, sheep, a donkey, a dog, and a cat, and primates such as a monkey, a chimpanzee, a gorilla, an orangutan, a bonobo, and a human, and particularly a human. Herein, the term "subject" is used in the meaning including a human as described above, and when a human is excluded, the term "non-human" is used.

[0163] Herein, the term "antibody" means an immunoglobulin, and refers to a protein having a structure in which two heavy chains (H chains) and two light chains (L chains) stabilized through a disulfide bond are associated with each other. The heavy chain contains a heavy chain variable region VH, heavy chain constant regions CH1, CH2, and CH3, and a hinge region positioned between the CH1 and the CH2, and the light chain contains a light chain variable region VL (wherein VL can be V κ or V λ), and a light chain constant region CL. Among these regions, a variable region fragment (Fv) consisting of the VH and the VL is a region directly involved in an antigen bond, and imparting variety to the antibody. An antigen binding region consisting of the VL, the CL, the VH, and the CH1 is designated as a Fab region, and a region consisting of the hinge region, the CH2 and the CH3 is designated as a Fc region.

[0164] Among the variable regions, a region directly contacting an antigen is particularly largely changed, and is designated as a complementarity-determining region (CDR). A portion except for the CDRs that is comparatively less mutated is designated as a framework region (FR). There are three CDRs in each variable region of the heavy chain and the light chain, and these are designated, successively from the N terminal side, heavy chain CDR1 to CDR3, and light chain CDR1 to CDR3, respectively. Each CDR is incorporated into the framework regions. The heavy chain variable region of the antibody includes, from the N terminal side to the C terminal side, a heavy chain framework region 1, the heavy chain CDR1, a heavy chain framework region 2, the heavy chain CDR2, a heavy chain framework region 3, the heavy chain CDR3, and a heavy chain framework region 4 in the stated order. The light chain variable region of the antibody includes, from the N terminal side to the C terminal side, a light chain framework region 1, the light chain CDR1, a light chain framework region 2, the light chain CDR2, a light chain framework region 3, the light chain CDR3, and a light chain framework region 4 in the stated order. The antibody may be a recombinant protein (recombinant antibody), and can be produced in an animal cell such as a Chinese hamster ovarian cell (CHO cell). The derivation of the antibody is not especially limited, and examples include an antibody of a non-human animal, an antibody of a non-human mammal (such as a mouse antibody, a rat antibody, or a camel antibody), and a human antibody. The antibody may be a chimeric antibody, a humanized antibody, or a fully humanized antibody. The antibody may be a polyclonal antibody or a monoclonal antibody, and is preferably a monoclonal antibody. A "chimeric antibody" refers to an antibody in which a heavy chain variable region and a light chain variable region are respectively linked to a heavy chain constant region and a light chain constant region of different species. A humanized antibody means an antibody in which an amino acid sequence characteristic to a non-human-derived antibody is substituted in the corresponding position of a human antibody, and an example includes an antibody having heavy

chain CDR1 to CDR3 and light chain CDR1 to CDR3 of an antibody produced by immunizing a mouse or a rat, and having the other regions including four framework regions (FR) each of the heavy chain and the light chain all derived from a human antibody. Such an antibody is designated as a CDR-grafted antibody in some cases. A “humanized antibody” encompasses a human chimeric antibody in some cases. A “human chimeric antibody” refers to a non-human-derived antibody in which a constant region of the non-human-derived antibody is substituted with a constant region of a human antibody. The antibody can be an isolated antibody, or a purified antibody. The antibody can be, for example, an IgG.

[0165] A variable region of an immunoglobulin chain generally has the same entire structure including relatively preserved framework regions (FR) linked through three hypervariable regions (more frequently designated as “complementarity-determining regions” or CDRs). The CDRs obtained from the two chains of each heavy chain/light chain pair are typically arranged parallel by the framework region for forming a structure specifically binding to a specific epitope on a protein of interest (such as PCSK9). Light chain and heavy chain variable regions present in nature all usually have these elements in the following order from the N terminal to the C terminal: FR1, CDR1, FR2, CDR2, FR3, CDR3 and FR4. In order to assign numbers to amino acids positioned in these respective domains, a numbering system has been devised. This numbering system is defined in “Kabat Sequences of Proteins of Immunological Interest (1987 and 1991, NIH, Bethesda, MD)”, or “Chothia & Lesk, 1987, J. Mol. Biol. 196: 901-917; Chothia et al., 1989, Nature, 342: 878-883”.

[0166] Herein, the antibody encompasses an antigen-binding fragment of an antibody. Herein, an antibody not fragmented may be referred to as a full length antibody. A full length antibody can contain the full length of the antibody excluding a signal sequence.

[0167] Herein, the term “antigen-binding fragment” means a part of an antibody maintaining a binding property to an antigen. The antigen-binding fragment can contain either or both of a heavy chain variable region and a light chain variable region of the antibody of the present disclosure. The antigen-binding fragment may be chimerized or humanized. Examples of the antigen-binding fragment include Fab, Fab', F(ab')₂, and Fv. The antigen-binding fragment may contain a bonded product or functional equivalent produced by recombination (for example, a part of another antibody in the form of a scFv (single chain Fv), a diabody, a scDb, a tandem scFv, a leucine zipper type, or a sc(Fv)₂ (single chain (Fv)₂)). Such an antigen-binding fragment of an antibody can be obtained, for example, by treating the antibody with an enzyme, although not especially limited. For example, when an antibody is digested with papain, a Fab can be obtained. Alternatively, when an antibody is digested with pepsin, a F(ab')₂ can be obtained, and when this is further reduced, a Fab' can be obtained. Herein, such an antigen-binding fragment of the antibody can be used. In an scFv, the VL and the VH are linked via an artificial polypeptide linker, and thus, the same antigen specificity as that of the original antibody can be maintained. The VL and the VH can be linked in the order of the VH and the VL, or the VL and the VH from the N terminal side. The linker can have a length of about 10 to 25 amino acids. The linker may contain glycine in a large amount, and may

contain an amino acid such as serine or threonine for purpose of increasing water solubility.

[0168] Herein, the term “intracellular antibody” (intra-body) refers to an antibody expressed in a cell (for example, in the cytoplasm or in the nucleus). Although an antibody is extracellularly secreted to function, an intracellular antibody is different in that it is designed to be expressed in a cell to function. The intracellular antibody can affect the function of an intracellular protein, and can inhibit the function thereof in the cytoplasm, the nucleus, or the secretory pathway. A cancer gene product can be a target of the intracellular antibody (Biocca, S., Pierandrei-Amaldi, P., and Cattaneo, A. (1993), Biochem Biophys Res Commun, Vol. 197, p. 422 to 427; Biocca, S., Pierandrei-Amaldi, P., Campioni, N., and Cattaneo, A. (1994), Biotechnology (NY), Vol. 12, p. 396 to 399; Cochet, O. et al., (1998), Cancer Res, Vol. 58, p. 1170 to 1176). The intracellular antibody directly binds to a protein for purpose of inhibiting the protein function. The bond may directly inhibit the function of the protein in some cases, and may inhibit the protein from binding to another protein in other cases.

[0169] Examples of the intracellular antibody include various antibodies and antigen-binding fragments thereof, and although not especially limited, a scFv, a tandem scFv, a VHH antibody (nanobody), a minibody, and a diabody can be preferably used. A scFv is typically an antibody fragment having a heavy chain variable region and a light chain variable region of an antibody, and the heavy chain variable region and the light chain variable region are linked via a linker. A tandem scFv is typically an antibody fragment having two scFvs having different antigen specificities, and these are linked via a linker. A diabody is typically a dimer of a scFv. Diabodies are roughly divided into bivalent monospecific diabodies and bispecific diabodies. A minibody is typically dimerized two fusion proteins each of a dimerized domain and a scFv via the dimerized domain. A VHH antibody is an antibody fragment containing a heavy chain variable domain of a heavy chain antibody. The VHH antibody is typically a heavy chain variable domain of a heavy chain antibody derived from a camelid (such as a camel, a llama, or an alpaca). Although a general antibody is extracellularly expressed, and hence can be caused to function only extracellularly, the intracellular antibody is superior because it can be caused to exhibit the antibody function in a cell. The intracellular antibody can be used in various applications in a cell such as activation and inactivation of a target protein, and neutralization and block of protein-protein interaction. A scFv tends to exhibit an aggregation property when expressed in a cell. Accordingly, in such a case, it is useful to reduce the aggregation property by obtaining a fusion protein by linking a peptide tag of the present disclosure to the intracellular antibody. When the aggregation property of a protein is reduced, the protein can be caused to exhibit functions inherent to the protein in the cell.

[0170] Herein, the term “peptide tag” refers to one that labels a protein of interest, or changes a biochemical property of the protein of interest when fused with the protein of interest. Examples of the peptide tag include various tags such as a FLAG tag, a 3xFLAG tag, a Myc tag, an HA tag, T7, a 6xHis tag, a PA tag, an S tag, an E tag, VSV-G, Glu-Glu, Strep-tag II, a HSV tag, a Chitin Binding Domain (CBD) tag, a Calmodulin Binding Peptide (CBP) tag, a V5 tag, a GST tag, a maltose binding protein (MBP) tag, a

thioredoxin (Trx) tag, and a mini-AID tag. These can be used for affinity purification of a protein of interest by utilizing affinity for the tag, or for detection of the protein of interest with an antibody to the tag produced. An antibody recognizing a tag is generally designated as a tag antibody, and a tag sequence corresponding to an epitope of the tag antibody is designated as an epitope tag. A tag can have a polypeptide chain generally with a length of several amino acids to several tens amino acids.

[0171] Herein, the term "protein of interest" refers to a protein to be expressed in a cell. The protein of interest may be an aggregating protein or a non-aggregating protein. In either case, when the peptide tag of the present disclosure is added thereto, the stability is further increased, and robustness against formation of aggregation can be obtained. Even when added to an aggregating protein, however, the peptide tag of the present disclosure can reduce the aggregation property thereof in a cell, and therefore, the protein of interest can be preferably an aggregating protein. Even when the protein of interest is a secretory protein, aggregation may be formed in a cell before the secretory protein is secreted extracellularly in some cases. The peptide tag of the present disclosure can be advantageously used also for a secretory protein, preferably a secretory protein having an aggregation property.

[0172] Herein, the term "aggregating protein" refers to a protein that forms aggregation (particularly, an insoluble aggregation) in a cell. Herein, the term "aggregation property" means a property of forming aggregation, and the term "non-aggregation property" means a property of not forming aggregation. Attenuation of the aggregation property can be promotion of the non-aggregation property, and promotion of the aggregation property can be attenuation of the non-aggregation property. Herein, the term "non-aggregation property" is used interchangeably with the term "stability". Aggregation can be observed, for example, as a bright point under a microscope by immunocytochemistry (IC). An aggregation rate can be calculated, for example, as a proportion of cells exhibiting aggregation in cells forcedly expressing a protein. Reduction of the aggregation rate thus calculated means increase of cells that forcedly express a protein and are not affected by the aggregation, and therefore can be an index of physiological favorability. Reduction of the aggregation property (for example, reduction of the aggregation rate) and increase of solubility are different indexes. The increase of solubility means increase of a concentration in an aqueous solution of available protein, and does not directly lead to the number of aggregations, or a proportion of cells having the aggregations. Accordingly, the increase of solubility does not always mean the reduction of the aggregation property (for example, the reduction of the aggregation rate).

[0173] Herein, an amino acid sequence is described by one letter amino acid code. Specifically, A denotes alanine, R denotes arginine, N denotes asparagine, D denotes aspartic acid, C denotes cysteine, Q denotes glutamine, E denotes glutamic acid, G denotes glycine, H denotes histidine, I denotes isoleucine, L denotes leucine, K denotes lysine, M denotes methionine, F denotes phenylalanine, P denotes proline, S denotes serine, T denotes threonine, W denotes tryptophan, Y denotes tyrosine, and V denotes valine. Amino acids are usually 20 types of L-amino acids mentioned above.

[0174] Herein, the term "regulatory sequence" refers to a sequence having activity of driving a gene operably linked thereto to transcribe RNA from the gene. The regulatory sequence is, for example, a promoter. Examples of the promoter include a class I promoter (usable for transcription of an rRNA precursor), a class II promoter (containing a core promoter and an upstream promoter element, and usable for transcription of an mRNA), and a class III promoter (further roughly divided into type I, type II, and type III).

[0175] The present invention provides a peptide tag that reduces aggregation tendency of an aggregating protein. The present invention provides a protein expression vector operably linked to a regulatory sequence, and containing a gene encoding the peptide tag. The present invention provides a protein of interest fused with the peptide tag. The present invention provides a protein expression vector operably linked to a regulatory sequence, and containing a gene encoding a protein of interest fused with the peptide tag. The protein of interest can be an intracellular protein in one embodiment. The protein of interest can be an intracellular antibody in one embodiment. The protein of interest can be an scFv in one embodiment.

[0176] Hereinafter, the peptide tag of the present disclosure that reduces aggregation tendency of an aggregating protein will be described in detail. The peptide tag of the present disclosure can reduce aggregation tendency of a protein of interest in a eukaryotic cell, particularly, in a human cell. In examination of pharmaceutical application and the like, it can be useful to reduce the aggregation tendency of a protein of interest in a human cell.

[0177] The length of the peptide tag of the present disclosure is not especially limited, and can be, for example, 600 amino acids or less, 500 amino acids or less, 400 amino acids or less, 300 amino acids or less, or 200 amino acids or less, and for example, 5 amino acids to 100 amino acids, such as 10 amino acids to 90 amino acids, 20 amino acids to 80 amino acids, 30 amino acids to 70 amino acids, 40 amino acids to 60 amino acids, 10 amino acids to 50 amino acids, 10 amino acids to 40 amino acids, or 10 amino acids to 30 amino acids. In this embodiment, the lower limit of the length of the peptide tag of the present disclosure can be 5 amino acids or more, 10 amino acids or more, 15 amino acids or more, 20 amino acids or more, 30 amino acids or more, 40 amino acids or more, 50 amino acids or more, 60 amino acids or more, 70 amino acids or more, or 80 amino acids or more, and/or the upper limit can be 100 amino acids or less, 90 amino acids or less, 80 amino acids or less, 70 amino acids or less, 60 amino acids or less, 50 amino acids or less, 40 amino acids or less, 30 amino acids or less, or 20 amino acids or less.

[0178] (a) The peptide tag of the present disclosure can contain acidic amino acids (amino acids belonging to Element 1) in the following ratio.

[0179] In the peptide tag of the present disclosure, 45% or more of amino acids contained in the amino acid sequence thereof can be acidic amino acids.

[0180] In the peptide tag of the present disclosure, less than 45% of amino acids contained in the amino acid sequence thereof can be acidic amino acids. In a preferable embodiment, in the peptide tag of the present disclosure, 5% or more and less than 45% of amino acids contained in the amino acid sequence thereof can be acidic amino acids, more preferably, 10% or more and less than 45% of amino acids contained in the amino acid sequence thereof can be

acidic amino acids, further preferably, 20% or more and less than 45% of amino acids contained in the amino acid sequence thereof can be acidic amino acids, further preferably, 30% or more and less than 45% of amino acids contained in the amino acid sequence thereof can be acidic amino acids, still further preferably, 35% or more and less than 45% of amino acids contained in the amino acid sequence thereof can be acidic amino acids, and particularly preferably, 40% or more and less than 45% of amino acids contained in the amino acid sequence thereof can be acidic amino acids. The acidic amino acids are D or E. For example, an acidic amino acid content in the peptide tag of the present disclosure can be 44% or less, 43.5% or less, 43% or less, 42.5% or less, 42% or less, 41.5% or less, 41% or less, 40% or less, 35% or less, 30% or less, 25% or less, or 20% or less. Thus, in one embodiment, a risk of occurrence of unexpected interaction with an intracellular molecule or the like having a positive charge based on a high acidic amino acid ratio in the peptide tag can be reduced.

[0181] (b) The peptide tag of the present disclosure can contain basic amino acids (amino acids belonging to Element 2) in the following ratio.

[0182] The peptide tag of the present disclosure can contain basic amino acids in a rate of preferably 25% or less or 20% or less, and more preferably can contain basic amino acids in a rate of 15% or less of amino acids, can contain basic amino acids further preferably in a rate of 10% or less, can contain basic amino acids further preferably in a rate of 5% or less, and can contain basic amino acids particularly preferably in a rate less than 3%, less than 2%, or less than 1%. In a most preferable embodiment, the peptide tag of the present disclosure does not contain a basic amino acid in the amino acid sequence thereof. The basic amino acids are K, R, or H.

[0183] (c) The peptide tag of the present disclosure can contain amino acids belonging to Element 3 in the following ratio.

[0184] The amino acids belonging to the Element 3 can be F, P, Y, G, S, Q, N, and A.

[0185] In the peptide tag of the present disclosure, 10% or more, preferably 20% or more, more preferably 30% or more, or 40% or more of amino acids contained in the amino acid sequence thereof can be preferably the amino acids of the Element 3. In the peptide tag of the present disclosure, 50% or more, 60% or more, or 70% or more of amino acids contained in the amino acid sequence thereof can be amino acids of the Element 3. In the peptide tag of the present disclosure, preferably 80% or less, more preferably 70% or less, and further preferably 60% or less of amino acids contained in the amino acid sequence thereof can be the amino acids of the Element 3. 50% or less, 40% or less, 30% or less, or 20% or less of amino acids contained in the amino acid sequence thereof can be the amino acids of the Element 3. In a preferable embodiment, in the peptide tag of the present disclosure, 20% or more and 80% or less, 30% or more and 70% or less, 30% or more and 60% or less, 30% or more and 50% or less, 30% or more and 40% or less, 40% or more and 70% or less, 40% or more and 60% or less, 40% or more and 50% or less, 50% or more and 70% or less, 50% or more and 60% or less, 60% or more and 80% or less, or 60% or more and 70% or less of amino acids contained in the amino acid sequence thereof can be the amino acids of the Element 3.

[0186] In a particularly preferable embodiment, in the peptide tag of the present disclosure, 5% or more, 10% or more, 15% or more, or 20% or more (preferably 21% or more, 25% or more, or 30% or more) of amino acids contained in the amino acid sequence thereof are either N or P. In the peptide tag of the present disclosure, for example, 90% or less, 80% or less, 70% or less, 60% or less, 50% or less, 45% or less, 40% or less, 35% or less, 30% or less, or 25% or less of amino acids contained in the amino acid sequence thereof can be either N or P. In one embodiment, in the peptide tag of the present disclosure, 10% or more and 20% or less of amino acids contained in the amino acid sequence thereof can be either N or P. In one embodiment, in the peptide tag of the present disclosure, 55% or more and 90% or less of amino acids contained in the amino acid sequence thereof can be either N or P. In one embodiment, in the peptide tag of the present disclosure, more than 10% and 20% or less of amino acids contained in the amino acid sequence thereof can be either N or P. In one embodiment, in the peptide tag of the present disclosure, more than 20% and 30% or less of amino acids contained in the amino acid sequence thereof can be either N or P. In one embodiment, in the peptide tag of the present disclosure, more than 30% and 40% or less of amino acids contained in the amino acid sequence thereof can be either N or P. In one embodiment, in the peptide tag of the present disclosure, more than 40% and 50% or less of amino acids contained in the amino acid sequence thereof can be either N or P. In one embodiment, in the peptide tag of the present disclosure, more than 50% and 60% or less of amino acids contained in the amino acid sequence thereof can be either N or P.

[0187] For example, in the peptide tag of the present disclosure, 5% or more, 10% or more, 15% or more, or 20% or more (preferably 21% or more, 25% or more, or 30% or more) of amino acids contained in the amino acid sequence thereof are N. In the peptide tag of the present disclosure, for example, 90% or less, 80% or less, 70% or less, 60% or less, 50% or less, 45% or less, 40% or less, 35% or less, 30% or less, or 25% or less of amino acids contained in the amino acid sequence thereof can be N. In one embodiment, in the peptide tag of the present disclosure, 10% or more and 20% or less of amino acids contained in the amino acid sequence thereof can be N. In one embodiment, in the peptide tag of the present disclosure, 55% or more and 90% or less of amino acids contained in the amino acid sequence thereof can be N. In one embodiment, in the peptide tag of the present disclosure, more than 10% and 20% or less of amino acids contained in the amino acid sequence thereof can be N. In one embodiment, in the peptide tag of the present disclosure, more than 20% and 30% or less of amino acids contained in the amino acid sequence thereof can be N. In one embodiment, in the peptide tag of the present disclosure, more than 30% and 40% or less of amino acids contained in the amino acid sequence thereof can be N. In one embodiment, in the peptide tag of the present disclosure, more than 40% and 50% or less of amino acids contained in the amino acid sequence thereof can be N. In one embodiment, in the peptide tag of the present disclosure, more than 50% and 60% or less of amino acids contained in the amino acid sequence thereof can be N.

[0188] For example, in the peptide tag of the present disclosure, 5% or more, 6% or more, 7% or more, 8% or more, 9% or more, 10% or more, 15% or more, or 20% or more (preferably 21% or more, 25% or more, or 30% or

(more) of amino acids contained in the amino acid sequence thereof are P. In the peptide tag of the present disclosure, for example, 90% or less, 80% or less, 70% or less, 60% or less, 50% or less, 45% or less, 40% or less, 35% or less, 30% or less, or 25% or less of amino acids contained in the amino acid sequence thereof can be P. In one embodiment, in the peptide tag of the present disclosure, 55% or more and 90% or less of amino acids contained in the amino acid sequence thereof can be P. In one embodiment, in the peptide tag of the present disclosure, 10% or more and 20% or less of amino acids contained in the amino acid sequence thereof can be P. In one embodiment, in the peptide tag of the present disclosure, more than 10% and 20% or less of amino acids contained in the amino acid sequence thereof can be P. In one embodiment, in the peptide tag of the present disclosure, more than 20% and 30% or less of amino acids contained in the amino acid sequence thereof can be P. In one embodiment, in the peptide tag of the present disclosure, more than 30% and 40% or less of amino acids contained in the amino acid sequence thereof can be P. In one embodiment, in the peptide tag of the present disclosure, more than 40% and 50% or less of amino acids contained in the amino acid sequence thereof can be P. In one embodiment, in the peptide tag of the present disclosure, more than 50% and 60% or less of amino acids contained in the amino acid sequence thereof can be P.

[0189] In a particularly preferable embodiment, in the peptide tag of the present disclosure, 5% or less, 10% or less, 15% or less, or 20% or less of amino acids contained in the amino acid sequence thereof are F or Y. In a particularly preferable embodiment, in the peptide tag of the present disclosure, 5% or less, 10% or less, 15% or less, or 20% or less of amino acids contained in the amino acid sequence thereof are F and/or Y. In a particularly preferable embodiment, in the peptide tag of the present disclosure, 5% or less, 10% or less, 15% or less, or 20% or less of amino acids contained in the amino acid sequence thereof are F and Y. In a particularly preferable embodiment, 5% or more, 10% or more, 15% or more, or 20% or more of amino acids contained in the amino acid sequence thereof are either N or P, and 5% or less, 10% or less, 15% or less, or 20% or less thereof are F and/or Y.

[0190] (d) The peptide tag of the present disclosure can contain amino acids belonging to Element 4 in the following ratio.

[0191] The amino acids belonging to the Element 4 can be amino acids that are none of an acidic amino acid, a basic amino acid, and the amino acids of the Element 3. The amino acids of the Element 4 can be, for example, M, T, W, C, I, V, and L.

[0192] In the peptide tag of the present disclosure, 80% or less, 70% or less, 60% or less, 50% or less, 40% or less, 30% or less, 20% or less, 15% or less, 10% or less, or 5% or less of amino acids contained in the amino acid sequence thereof can be preferably the amino acids of the Element 4. In a preferable embodiment, the peptide tag of the present disclosure does not contain the amino acids of the Element 4.

[0193] (e) In the peptide tag of the present disclosure, 40% or less, 30% or less, 20% or less, 15% or less, 10% or less, or 5% or less of amino acids contained in the amino acid sequence thereof can be preferably G. In one embodiment, the peptide tag of the present disclosure does not contain G.

[0194] (f) In the peptide tag of the present disclosure, 40% or less, 30% or less, 20% or less, 15% or less, 10% or less,

or 5% or less of amino acids contained in the amino acid sequence thereof can be preferably A. In one embodiment, the peptide tag of the present disclosure does not contain A. [0195] (g) In the peptide tag of the present disclosure, 40% or less, 30% or less, 20% or less, 15% or less, 10% or less, or 5% or less of amino acids contained in the amino acid sequence thereof can be preferably G, and 40% or less, 30% or less, 20% or less, 15% or less, 10% or less, or 5% or less of amino acids contained in the amino acid sequence thereof can be A.

[0196] (h) The peptide tag of the present disclosure can preferably contain S. The peptide tag of the present disclosure preferably does not contain S. The peptide tag of the present disclosure can contain S in a rate of 10% or more, 20% or more, 30% or more, 40% or more, or 50% or more of amino acids contained in the amino acid sequence thereof. The peptide tag of the present disclosure can contain S in a rate of 60% or less, 50% or less, 40% or less, 30% or less, 20% or less, 15% or less, 10% or less, or 5% or less of amino acids contained in the amino acid sequence thereof.

[0197] (A) In the peptide tag of the present disclosure, less than 45% of amino acids contained in the amino acid sequence thereof can be acidic amino acids, and 20% or more and 80% or less (preferably 30% or more and 70% or less) of the amino acids contained in the amino acid sequence thereof can be the amino acids of the Element 3.

[0198] (B) In the peptide tag of the present disclosure, less than 45% of amino acids contained in the amino acid sequence thereof can be acidic amino acids, 20% or more and 80% or less (preferably 30% or more and 70% or less) of the amino acids contained in the amino acid sequence thereof can be the amino acids of the Element 3, and 10% or less of the amino acids contained in the amino acid sequence thereof can be A.

[0199] (C) In the peptide tag of the present disclosure, less than 45% of amino acids contained in the amino acid sequence thereof can be acidic amino acids, 20% or more and 80% or less (preferably 30% or more and 70% or less) of the amino acids contained in the amino acid sequence thereof can be the amino acids of the Element 3, and 10% or less of the amino acids contained in the amino acid sequence thereof can be G.

[0200] (D) In the peptide tag of the present disclosure, less than 45% of amino acids contained in the amino acid sequence thereof can be acidic amino acids, 20% or more and 80% or less (preferably 30% or more and 70% or less) of the amino acids contained in the amino acid sequence thereof can be the amino acids of the Element 3, 10% or less of the amino acids contained in the amino acid sequence thereof can be A, and 10% or less of the amino acids contained in the amino acid sequence thereof can be G.

[0201] (E) In the peptide tag of the present disclosure, less than 45% of amino acids contained in the amino acid sequence thereof can be acidic amino acids, 20% or more and 80% or less (preferably 30% or more and 70% or less) of the amino acids contained in the amino acid sequence thereof can be the amino acids of the Element 3, and 30% or less of the amino acids contained in the amino acid sequence thereof can be the amino acids of the Element 4.

[0202] (F) In the peptide tag of the present disclosure, less than 45% of amino acids contained in the amino acid sequence thereof can be acidic amino acids, 20% or more and 80% or less (preferably 30% or more and 70% or less) of the amino acids contained in the amino acid sequence

the amino acid sequence thereof can be acidic amino acids, 10% or less of the amino acids contained in the amino acid sequence thereof can be basic amino acids, 30% or more and 70% or less of the amino acids contained in the amino acid sequence thereof can be the amino acids of the Element 3, 30% or less of the amino acids contained in the amino acid sequence thereof can be the amino acids of the Element 4, 10% or less of the amino acids contained in the amino acid sequence thereof can be G, and 10% or less of the amino acids contained in the amino acid sequence thereof can be A. In this embodiment, preferably 40% or more and 60% or less of the amino acids contained in the amino acid sequence thereof are the amino acids of the Element 3. In this embodiment, more preferably 35% or more and less than 45% of the amino acids contained in the amino acid sequence thereof can be acidic amino acids, and further preferably 40% or more and less than 45% of the amino acids contained in the amino acid sequence thereof can be acidic amino acids. In this embodiment, preferably 40% or more and 60% or less of the amino acids contained in the amino acid sequence thereof are the amino acids of the Element 3, and 35% or more and less than 45% of the amino acids contained in the amino acid sequence thereof are acidic amino acids, and further preferably 40% or more and 60% or less of the amino acids contained in the amino acid sequence thereof can be the amino acids of the Element 3, and 40% or more and less than 45% of the amino acids contained in the amino acid sequence thereof can be acidic amino acids.

[0226] (AC) In the peptide tag of the present disclosure, 30% or more and less than 45% of amino acids contained in the amino acid sequence thereof can be acidic amino acids, 10% or less of the amino acids contained in the amino acid sequence thereof can be basic amino acids, 30% or more and 70% or less of the amino acids contained in the amino acid sequence thereof can be the amino acids of the Element 3, 20% or less of the amino acids contained in the amino acid sequence thereof can be the amino acids of the Element 4, 10% or less of the amino acids contained in the amino acid sequence thereof can be G, and 10% or less of the amino acids contained in the amino acid sequence thereof can be A. In this embodiment, preferably 15% or less, 10% or less, or 5% or less of the amino acids contained in the amino acid sequence thereof can be G. In one embodiment, the peptide tag of the present disclosure does not contain G.

[0227] (AD) In the peptide tag of the present disclosure, 35% or more and less than 45% of amino acids contained in the amino acid sequence thereof can be acidic amino acids, 10% or less of the amino acids contained in the amino acid sequence thereof can be basic amino acids, 30% or more and 70% or less of the amino acids contained in the amino acid sequence thereof can be the amino acids of the Element 3, 10% or less of the amino acids contained in the amino acid sequence thereof can be the amino acids of the Element 4, 10% or less of the amino acids contained in the amino acid sequence thereof can be G, and 10% or less of the amino acids contained in the amino acid sequence thereof can be A.

[0228] (AE) The peptide tag of the present disclosure can have, for example, an amino acid sequence set forth in any of SEQ ID NOs: 2 to 11. The peptide tag of the present disclosure can have preferably an amino acid sequence of SEQ ID NO: 5.

[0229] (AF) The peptide tag of the present disclosure can have, for example, any one of amino acid sequences shown

in Tables 1 to 11, Table 12-1, Table 12-2, Table 13-1, Table 13-2, Table 14-1, and Table 14-2.

[0230] (AG) The peptide tag of the present disclosure may have one or more selected from the group consisting of addition and insertion of one or more amino acids selected from the group consisting of N and P, and substitution with the amino acids (for example, substitution of one to about 30% of amino acids, such as substitution with 1 to 20, 10, or several amino acids) in any amino acids of any amino acid sequences of (AE) and (AF) described above (for example, the amino acids of the Element 1, the Element 2, the Element 3, or the Element 4, or for example, the amino acids of the Element 2, the Element 3, or the Element 4) The peptide tag of the present disclosure may have addition and insertion of S in any amino acids of any amino acid sequences of (AF) and (AG) described above (for example, the amino acids of the Element 1, the Element 2, the Element 3, or the Element 4, or for example, amino acids of the Element 2, the Element 3, or the Element 4), and substitution of S with the amino acids (for example, substitution of one to about 30% of amino acids, such as substitution with 1 to 20, 10, or several amino acids). The peptide tag of the present disclosure may have deletion of an arbitrary amino acid of any one of the amino acid sequences of (AF) and (AG) described above (for example, the amino acids of the Element 1, the Element 2, the Element 3, or the Element 4, for example, the amino acids of the Element 2, the Element 3 (particularly, F and/or Y), and the Element 4 (A or G).

[0231] (AH) In the peptide tag of the present disclosure, preferably 5% or more, 6% or more, 7% or more, 8% or more, 9% or more, 10% or more, 11% or more, 12% or more, 13% or more, 14% or more, 15% or more, 16% or more, 17% or more, 18% or more, 19% or more, or 20% or more of amino acids contained in the amino acid sequence thereof can be either N or P, or N and P. In the peptide tag of the present disclosure, preferably 5% or more, 6% or more, 7% or more, 8% or more, 9% or more, 10% or more, 11% or more, 12% or more, 13% or more, 14% or more, 15% or more, 16% or more, 17% or more, 18% or more, 190% or more, 20% or more, 21% or more, 25% or more, or 30% or more of the amino acids contained in the amino acid sequence thereof can be P. In the peptide tag of the present disclosure, preferably 5% or more, 6% or more, 7% or more, 8% or more, 9% or more, 10% or more, 11% or more, 12% or more, 13% or more, 14% or more, 15% or more, 16% or more, 17% or more, 18% or more, 19% or more, 20% or more, 21% or more, 25% or more, or 30% or more of the amino acids contained in the amino acid sequence thereof can be N.

[0232] (AI) In the peptide tag of the present disclosure, 45% or more of amino acids contained in the amino acid sequence thereof can be acidic amino acids, and 5% or more, 6% or more, 7% or more, 8% or more, 9% or more, 10% or more, 11% or more, 12% or more, 13% or more, 14% or more, 15% or more, 16% or more, 17% or more, 18% or more, 19% or more, 20% or more, 21% or more, 25% or more, or 30% or more thereof can be N or P. In this embodiment, less than 10% (preferably less than 5%, and more preferably 0%) of the amino acids contained in the amino acid sequence thereof can be G, less than 10% (preferably less than 5%, and more preferably 0%) thereof can be A, and/or less than 10% (preferably less than 5%, and more preferably 0%) thereof can be F and Y.

more of the amino acids contained in the amino acid sequence can be N or P, and 10% or less of the amino acids contained in the amino acid sequence can be F and/or Y,

[0265] (c) 10% or less of the amino acids contained in the amino acid sequence can be amino acids selected from the group consisting of M, T, W, C, I, V, and L, and

[0266] (d) each of A and G can constitute less than 10% of the amino acids contained in the amino acid sequence. In this embodiment, the amino acid sequence can have a length of 10 to 200 amino acids (such as 10 to 90 amino acids).

[0267] (AP) In the peptide tag of the present disclosure,

[0268] (a) 5% or more and less than 45% of amino acids contained in the amino acid sequence can be acidic amino acids, (b1) 55% or more and less than 90% of the amino acids contained in the amino acid sequence can be either of N and P, and

[0269] the rest of the amino acids contained in the amino acid sequence can be other amino acids. In this embodiment, 20% or less (preferably 15% or less, 10% or less, 5% or less, 4% or less, 3% or less, 2% or less, 1% or less, or 0%) of the amino acids contained in the amino acid sequence are neither an acidic amino acid nor N and P.

[0270] (AQ) In the peptide tag of the present disclosure,

[0271] (a) 45% or more of amino acids contained in the amino acid sequence can be acidic amino acids, (b1) 21% or more of the amino acids contained in the amino acid sequence can be N, and/or 7% or more thereof can be P, and

[0272] 20% or less (preferably 15% or less, 10% or less, 5% or less, 4% or less, 3% or less, 2% or less, 1% or less, or 0%) of the amino acids contained in the amino acid sequence can be other amino acids.

[0273] (AR) In the (AK) to (AQ) above, a rate of the Element 4 can be preferably 0%.

[0274] (AS) In the (AK) to (AQ) above, a rate of A or G can be 0%. In the (AK) to (AO) above, a rate of A and G is preferably 0%.

[0275] (AT) In the (AK) to (AQ) above, preferably, a rate of the Element 4 is 0%, and a rate of A and G is 0%.

[0276] (AU) In the (AR) to (AT) above, a rate of the Element 2 is preferably 0%.

[0277] In the (a) above, preferably 10% or more and less than 45% of the amino acids contained in the amino acid sequence are acidic amino acids, more preferably 15% or more and less than 45% thereof are acidic amino acids, further preferably 20% or more and less than 45% thereof are acidic amino acids, still further preferably 25% or more and less than 45% thereof are acidic amino acids, and particularly preferably 30% or more and less than 45% thereof are acidic amino acids. In these examples, the upper limit of the acidic amino acid content can be, for example, less than 45%, 40% or less, or 35% or less.

[0278] The peptide tag of the present disclosure can be added to a protein (such as an intracellular aggregating protein). Accordingly, the present disclosure provides a fusion protein of the peptide tag of the present disclosure and an intracellular aggregating protein. The peptide tag of the present disclosure may be added to an intracellular non-aggregating protein. When the tag is added thereto, toughness of the non-aggregating protein against a non-aggrega-

tion property can be increased. The protein can be, for example, an intracellular antibody. The intracellular antibody can be an antigen-binding fragment of an antibody. The peptide tag of the present disclosure may be added to a fusion protein of an intracellular antibody and a degradation-inducing sequence. Thus, selective degradation of a target to which the intracellular antibody binds can be induced. Examples of the intracellular antibody include the above-described antibody fragments. Other examples of the intracellular antibody include antibodies that bind to α-synuclein, LRRK2, Tau, β-amyloid, amyloid precursor protein (APP), C9orf72, superoxide dismutase 1 (SOD1), TAR DNA-binding protein 43 (TDP43), Fused in Sarcoma (FUS), and a prion protein, and pathological forms thereof. Another example of the intracellular antibody includes an antibody inhibiting protein-protein interaction (PPI). Still another example of the intracellular antibody includes one in the form of a fusion protein with a degradation-inducing sequence that binds to a target. Other examples of the intracellular antibody include intracellular antibodies described in Molecular Therapy, 29(2): 859-872, 2021 (such as CP13 iB, PHF1 iB, and Tau5 iB), and intracellular antibodies each having all CDR sequences of these intracellular antibodies. "iB" is an abbreviation of "intrabody", and specifically means an intracellular antibody. The present disclosure provides a fusion protein of, for example, such an intracellular antibody and the peptide tag of the present disclosure. The intracellular antibody may preferably further include a degradation-inducing sequence. Other examples of the intracellular antibody include intracellular antibodies described in Molecular Therapy, 30(4): 1484-1499, 2022 (such as VHH E4-1, and VHHZ70), and intracellular antibodies each having all CDR sequences of these intracellular antibodies. The intracellular antibody may preferably further include a degradation-inducing sequence. Still other examples of the intracellular antibody include an intracellular antibody described in J. Biol. Chem., 295(31): 10662-10676, 2020 (such as M204-scFv), and an intracellular antibody having all CDR sequences of this intracellular antibody. The intracellular antibody may preferably further include a degradation-inducing sequence. Still other examples of the intracellular antibody include an intracellular antibody described in WO2018/231254 (such as BIIB092 antibody), an intracellular antibody described in WO2016/207245, an antibody described in WO2018/011073 (such as C10-2), intracellular antibodies described in WO2015/114538 (such as VHH tau A2, VHH tau A2-SH, and VHH tau A2var-SH), intracellular antibodies described in WO2014/059442 (such as F9T, D11C, D4G, G12C, H2A and H7T), and JP2020/515233 (such as IE4, 9B11, 3A9, 10F10, 11F11, AC8, AE8, AA9, DG5, AD2, AD7, DG11, DG8, and DA9) and intracellular antibodies each having all CDR sequences of these intracellular antibodies. The intracellular antibody may preferably further include a degradation-inducing sequence. Examples of the intracellular antibody further include an intracellular antibodies capable of degrading and removing abnormal TDP-43 (such as SEQ ID NOs: 21 to 24) described in WO2019177138, and an intracellular antibody having all CDR sequences of this intracellular antibody. The intracellular antibody may preferably further include a degradation-inducing sequence. Thus, an intracellular antibody (such as a scFv or VHH) binding to tau, an intracellular antibody (such as a scFv or VHH) binding to α-synuclein, and other intracellular antibodies

(such as a scFv or VHH) against amyloid causing cytotoxicity in a cell are preferred, and can be linked to the tag to form a fusion protein with the tag.

[0279] In one embodiment, the peptide tag of the present disclosure does not have a CAA motif (such as SEQ ID NO: 58: KLNPPDESGPGCMSCKCVLS). In one embodiment, the peptide of the present disclosure does not have a membrane localization signal. In one embodiment, the peptide tag of the present disclosure does not have a signal peptide sequence for extracellular secretion from the viewpoint of expressing a protein of interest in a cell. In one embodiment, the peptide tag of the present disclosure may have a signal sequence for extracellular secretion from the viewpoint of promoting extracellular secretion of a protein of interest. When a secretory protein has an aggregation property, the peptide tag of the present disclosure containing a signal sequence in the sequence thereof, or the peptide tag of the present disclosure linked tandem to the signal sequence can be advantageous. In one embodiment, the peptide tag of the present disclosure can contain a nuclear localization signal. In one embodiment, the peptide tag of the present disclosure does not have a sequence preventing protein localization in the cytoplasm. In one embodiment, the peptide tag of the present disclosure promotes free distribution in a cell of the protein of interest. In one embodiment, the peptide tag of the present disclosure can promote intracellular bond of the protein of interest to an original binding partner, and co-localization with the binding partner. In one embodiment, the peptide tag of the present disclosure can have a sequence that imposes unique constraints on the distribution in a cell (or a sequence that prevents free distribution) of the protein of interest, but is possible not to have such a sequence.

[0280] In any embodiment, the peptide tag of the present disclosure does not have the following sequence (Enzymol. 326, 362-267 (2000)): S-tag: KETAAAKFERQHMDS (SEQ ID NO: 14). In one embodiment, the peptide tag of the present disclosure can have a sequence in which a rate of the Element 2 is 10% or less, and/or a rate of A is 10% or less.

[0281] In any embodiment, the peptide tag of the present disclosure does not have KLNPPDESGPGCMSCKCVLS (SEQ ID NO: 15) (Tanaka et al., 2007, EMBO Journal, 26: 3250-3259). In one embodiment, the peptide tag of the present disclosure does not have a sequence having 90% or more sequence identity to this sequence.

[0282] In any embodiment, the peptide tag of the present disclosure does not have EFGGAPEFPKPSTPPGSSGL (SEQ ID NO: 16), and a sequence having 90% or more sequence identity to this sequence (Paolo et al., 2003, Clinical Cancer Research, 9: 2837-2848). In one embodiment, the peptide tag of the present disclosure does not have a sequence having 90% or more sequence identity to this sequence.

[0283] In any embodiment, the peptide tag of the present disclosure does not have any one of the following sequences (Arimori et al., 2017, Structures, 25: 1611-1622):

[0284] hMst1: DYEFLLKSWTVEDLQKRLALDPM-MEQEIEEIRQKYQSKRQPILDIAEK (SEQ ID NO: 17);

[0285] hMST2: DFDFLKNLSLEELQMRLKALDPM-MEREIEELRQRYTAKRQPILDAMDAK (SEQ ID NO: 18);

[0286] hRaf1: GEVNWDAFSMPELHNFL-RILQREEEHLRQILQKYYSRQKIQEALHAS (SEQ ID NO: 19);

[0287] hRaf5: GEVEWDAFSIPELQNFLTILEKE-EQDKIQQQVKKYDKFRQKLEALRES (SEQ ID NO: 20);

[0288] hSAV1: HILKWELFQLADLDTYQGMLKLL-FMKELEQIVKMYEAYRQALLTELENR (SEQ ID NO: 21). In one embodiment, the peptide tag of the present disclosure does not have a sequence having 90% or more sequence identity to any one of these sequences.

[0289] In any embodiment, the peptide tag of the present disclosure have none of the following (Zhang et al., 2004, Protein Expression and Purification, 36(2): 207-216):

T7C:

[0290] LEDPFQSGVMLGVASTVAASPEEASVT-STEETLTPAQEAARTRAANKARKEAELAAA-TAEQ (SEQ ID NO: 22);

[0291] T7B: LEDPSEEASVTSTEETLTPAQEAARTRAANKARKEAELAAATAEQ (SEQ ID NO: 23);

[0292] T7B1: LEDPSEEASVTSTEETLTPAQEAARTRPPNKARKEAELAAATAEQ (SEQ ID NO: 24);

[0293] T7B2: LEDPSEEASVTSTEETLTPAQEAARTRGGNKARKEAELAAATAEQ (SEQ ID NO: 25);

[0294] T7B3: LEDPSEEASVTSTEETLTPAQEAARTRAANKARKEAELTAEQ (SEQ ID NO: 26);

[0295] T7B4: LEDPSEEASVTSTEETLTPAQEAARTRAANKARKEAELEAATAEQ (SEQ ID NO: 27);

[0296] T7B5: LEDPSEEASVTSTEETLTPAQEAARTRAAKARKEAELAAATAEQ (SEQ ID NO: 28);

[0297] T7B6: LEDPSEEASVTSTEETLTPAQEAARTRKARKEAELAAATAEQ (SEQ ID NO: 29);

[0298] T7B7: LEDPSEEASVTSTEETLTPAQEAARTRAANKARKEAELAA (SEQ ID NO: 30);

[0299] T7B8: LEDPSEEASVTSTEETLTPAQEAARTRAANKARKEAELAAA (SEQ ID NO: 31);

[0300] T7B9: LEDPSEEASVTSTEETLTTPAQEAARTRAEEAETEAANKARKEAELEAATAEQ (SEQ ID NO: 32);

[0301] T7B10: LEDPTPAQEAARTRAANKARKEAELAAATAEQ (SEQ ID NO: 33);

[0302] T7A: LEDPAANKARKEAELAAATAEQ (SEQ ID NO: 34);

[0303] T7A1: LEDPERNKERKEAELAAATAEQ (SEQ ID NO: 35);

[0304] T7A2: LEDPERNKERKEAELEAATAEQ (SEQ ID NO: 36);

[0305] T7A3: LEDPERNKERKEAELEAETAEQ (SEQ ID NO: 37);

[0306] T3: LEDPAVWEAGKVAKGVGTADIT-ATTSNGLIASCKVIVNAATS (SEQ ID NO: 38);

[0307] T3A: LEDPAVWEAGKVAKGVGTADIT-ATTSNGLIASSEEADNAATS (SEQ ID NO: 39). In one embodiment, the peptide tag of the present disclosure does not have a sequence having 90% or more sequence identity to any one of these sequences.

[0308] In any embodiment, the peptide tag of the present disclosure have none of the following (Japanese Patent Laid-Open No. 2015-97519):

-continued

Zif628:
 (SEQ ID NO: 40)
 ERPYACPVESCDRRFSRSDELTRHIRIHTGQKPFQCRICMR
 NFSRSDHLLTTHIRHTGEKPFACDICGRKFARSDERKRHTK
 IHLRQKD;
 HinR:
 (SEQ ID NO: 41)
 GRPRAITKHEQEIQISRLLEKGHPROQLAIIFGIGVSTLYRY
 FPASSIKKRMN;
 and
 TrpR:
 (SEQ ID NO: 42)
 MAQQSPYSAAMAEQRHXXQEWLRFDLKNAYQNXDLHLP
 LLNLMLTPDERXXEALGTRVRIVEELLRGEMSQRELKNELG
 AGIATITRGSNSLKAAPVELROWLEEVLLKSD.

[0309] In one embodiment, the peptide tag of the present disclosure can be a natural sequence found in a non-human living thing. In a preferable embodiment, the peptide tag of the present disclosure can be a non-natural sequence or a part thereof. In either embodiment, the peptide tag of the present disclosure is none of the following (WO2010/034183):

NE-1:
 (SEQ ID NO: 43)
 TKENPRSNQEESYDDNES;
 NE-8:
 (SEQ ID NO: 44)
 TKENPRTNQEESYDDNES;
 NE-9:
 (SEQ ID NO: 45)
 TKENPRSNQDESYDDNES;
 NE-10:
 (SEQ ID NO: 46)
 TKENPRSNOPPSYDDNES.

[0310] In one embodiment, the peptide tag of the present disclosure is none of the following (WO2011/034605):
 ACID.P1: GGSQALEKELQALEKE-
 NAQLEWELQALEKELAQGAT (SEQ ID NO: 50).

[0311] In one embodiment, the peptide tag of the present disclosure is none of the following (WO2009/023270):

rPEG_K288-GFP:
 (SEQ ID NO: 51)
 (GEGEGEGEG) 32

[0312] In one embodiment, the peptide tag of the present disclosure is none of the following (WO2020/059228):

Hero7:
 (SEQ ID NO: 1038)
 MTRGNQRELARQKNMKQSDSVKGKRRDDGLSAAARKQRDS
 EIMQQOKKANEKKEEPK;

Hero9:
 (SEQ ID NO: 1039)
 MSGPNGDLGMPVEAGAEGEEDGFGEAEYAAINSMLDQINSC
 LDHLEEKNDHLHARLQELLESNRQTRLEFQQQLGEAPSDAS
 P;
 Herol11:
 (SEQ ID NO: 1040)
 MAQGQRKFOAHKPAKSCTAAAASEKNRGPRKGRVIAPKKA
 RVVQQQKLKKNLEVGIRKKIEHDVVMKASSSLPKKLALLKA
 PAKKKGAAAATSSKTPS.

[0313] In one embodiment, the peptide tag of the present disclosure is none of the following (Protein Engineering, Design & Selection, 26(8): 490-501, 2013):

PAS#1:
 (SEQ ID NO: 52)
 ASPAAPAPASPAAPAPSAPAA;
 1P2:
 (SEQ ID NO: 53)
 ASAAAAPAAAASAAAAPSAAAAA;
 PAS#5:
 (SEQ ID NO: 54)
 AASPAAPSAPPAAASPAAPSAPPAA;

and repeated sequences of these (the number of repetition being, for example, 200±20 times, 400±40 times, or 600±60 times).

[0314] In one embodiment, the peptide tag of the present disclosure is none of the following (Protein Engineering, Design & Selection, 17(11): 779-786, 2004):

Z (W):
 (SEQ ID NO: 55)
 VDNKFNKEQQNAFYIELHLPNLNEEQRNAFIQ
 SLKDDPSQSANLLAEAKKLNDQAPK;
 Z (a1):
 (SEQ ID NO: 56)
 VDNKFNKEQQNAEYEIEHLPNLNEEQNAFIQ
 SLEDDPSQSANLLAEAKKLNDQAPK;
 Z (a2):
 (SEQ ID NO: 57)
 VDNKFNKEEEEAEIIIHLPNLNEEQEEAFIE
 SLEDDPSQSANLLAEAKKLNDQAPK

[0315] In one embodiment, the peptide tag of the present disclosure may have a mutation selected from the group consisting of substitution, insertion, deletion, addition, and elimination of one or more, preferably two or more amino acids in any one of the amino acid sequences of SEQ ID NOS: 43 to 46 and 47 to 58. In one embodiment, the peptide tag of the present disclosure can have less than 90%, 85% or less, 80% or less, 75% or less, 70% or less, 60% or less, 50% or less, 40% or less, 30% or less, 20% or less, or 10% or less sequence identity to any one of the amino acid sequences of SEQ ID NOS: 44 to 47 and 47 to 58.

[0316] In any embodiment, the peptide tag of the present disclosure does not contain the following sequences (Protein Science (2019) 28, 823-836):

PA12-tag:	(SEQ ID NO: 47)
GVAMPGAEDDVV;	
PA14-tag:	(SEQ ID NO: 48)
EGGVAMPGAEDDVV.	

In any embodiment, the peptide tag of the present disclosure does not contain the following sequences:

(SEQ ID NO: 49)	DYKDDDDVEAEEESDNVDSADAEE-DDSDVWWGGADTDY
ADGSEDKVVEVAEEEEVAEVEEEEADDDEDDEDGDEV	
EEEAEEPYEEATERTTSIATTTTTESVEEVYPGQV	
GYPGQVGYPGQV.	

[0317] In one embodiment, the peptide tag of the present disclosure has less than 90%, 80% or less, 70% or less, 60% or less, 50% or less, 40% or less, 30% or less, 20% or less, or 10% or less sequence identity to any one of SEQ ID NOS: 14 to 48 and 49 to 58.

[0318] In one embodiment, the peptide tag of the present disclosure can have a sequence satisfying one, two or all of the following (i) to (iii): (i) a rate of the Element 2 is 10% or less, (ii) a rate of A is 10% or less, and (iii) a rate of G is 10% or less.

[0319] In a preferable embodiment, the peptide tag of the present disclosure does not have a sequence consecutively containing 5 or more As. In a preferable embodiment, the peptide tag of the present disclosure does not have a sequence consecutively containing 5 or more Qs. In a preferable embodiment, the peptide tag of the present disclosure does not have a sequence consecutively containing 5 or more Ss. In a preferable embodiment, the peptide tag of the present disclosure does not have a sequence consecutively containing 5 or more Ns. In a preferable embodiment, the content of a specific single amino acid in the amino acid sequence of the peptide tag does not exceed 50%, 40%, 35%, 30%, 25%, or 20%. In a preferable embodiment, the peptide tag of the present disclosure does not contain an amino acid sequence having a length of 3 to 8 amino acids described in Table 1 in WO2002/092132, or does not contain a consecutive repeat (for example, consecutive repeat of 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, or 10 or more times) of the amino acid sequence.

[0320] For example, the peptide tag of the present disclosure can inhibit an intracellular antibody from forming aggregation to promote uniform distribution in a cell of the antibody, and/or inhibit aggregation formation to promote bond of the intracellular antibody to an antigen in a cell, and co-localization with the antigen. In one embodiment, the peptide tag of the present disclosure is possible not to have a sequence that imposes unique constraints on the distribution in a cell (or a sequence that prevents free distribution) of the intracellular antibody, and/or is possible not to have bond in a cell of the intracellular antibody to an antigen, and co-localization with the antigen.

[0321] All the peptide tags of the present disclosure can mitigate, inhibit, or improve aggregation tendency; increase, promote, or improve a non-aggregation property; or increase, promote, or improve stability of a tagged protein. The peptide tag of the present disclosure is possible not to have a sequence that imposes unique constraints on the distribution in a cell (or a sequence that prevents free distribution) of the intracellular antibody, and/or is possible not to have bond in a cell of the intracellular antibody to an antigen and co-localization with the antigen. When the protein of interest is an antigen-binding fragment of an antibody, the peptide tag of the present disclosure can promote co-localization with an antigen through bond of the protein to the antigen.

[0322] The peptide tag of the present disclosure can be, for example, a gene product encoded by a gene of a living thing, or a fragment thereof, and here, can be a gene product encoded by a gene of a non-human living thing (for example, a microorganism such as a bacteria, an alga, or a fungus, an animal such as a mammal, a bird, or fish, or a plant), or a fragment thereof. Alternatively, in a preferable embodiment, the peptide tag of the present disclosure can be a gene product encoded by a gene of a human, or a fragment thereof.

[0323] When fused with an aggregating protein, the peptide tag of the present disclosure can mitigate, inhibit or improve the aggregation tendency of the aggregating protein, or can increase, promote, or improve the non-aggregation property of the aggregating protein. Aggregation of a protein can adversely affect a cell in which the protein is expressed, and in addition, can adversely affect the protein production amount and functionality by the aggregation. Accordingly, the mitigation, inhibition, or improvement of the aggregation tendency reduces the influence of the aggregation on the cell, and can lead to reduction of the influence on the protein production amount and the functionality. In this manner, the peptide tag of the present disclosure can be beneficial in improvement of the expression level of an aggregating protein expressed in a cell and/or improvement of the functionality, and accordingly, can be useful for forced expression of the aggregating protein *in vivo*.

[0324] Accordingly, the peptide tag of the present disclosure may be fused with an aggregating protein. The aggregating protein can be a protein that forms aggregation in 30% or more, 40% or more, 50% or more, 60% or more, 70% or more, 80% or more, or 90% or more of cells when intracellularly expressed. The aggregating protein may be a protein that forms aggregation in 90% or less, 80% or less, 70% or less, 60% or more, or 50% or less of cells when intracellularly expressed. The fusion can be performed, for example, on the N terminal and/or the C terminal (preferably, both the N terminal and the C terminal) of the aggregating protein. The fusion can be conducted by, for example, linking a nucleic acid encoding the peptide tag of the present disclosure and a nucleic acid encoding the aggregating protein, in-frame (in such a manner as to match the reading frames of codons). The peptide tag of the present disclosure may be fused with a non-aggregating protein. Thus, the non-aggregation property of the non-aggregating protein can be further increased. The non-aggregating protein may be a protein that forms aggregation in 20% or less, 15% or less, 10% or less, 9% or less, 8% or less, 7% or less, 6% or less,

5% or less, 4% or less, 3% or less, 2% or less, or 1% or less of cells or a protein that does not form aggregation, when intracellularly expressed.

[0325] When fused with a protein, the peptide tag of the present disclosure can increase, promote, or improve the stability under an intracellular environment of the protein. The stability under an intracellular environment of a protein is beneficial in both an aggregating protein and a non-aggregating protein.

[0326] In one embodiment, the protein of interest may be linked to a second peptide tag. The second peptide tag can be added to the protein of interest for purpose of, for example, detection or purification. In this case, the peptide tag of the present disclosure can be used for reducing, inhibiting, or improving aggregation tendency; increasing, promoting, or improving a non-aggregation property; or increasing, promoting, or improving stability of a fusion protein of the protein of interest and the second peptide tag. As the second peptide tag, a usual peptide tag, such as an HA tag, can be used.

[0327] The aggregating protein is not especially limited, and can be, for example, an antigen-binding fragment of an antibody. The aggregating protein can be preferably a single chain Fv (scFv) or a VH_H antibody. A scFv is a fusion protein containing a heavy chain variable region and a light chain variable region of an antibody in which the heavy chain variable region and the light chain variable region are linked via a linker (preferably, a flexible linker). For example, there is an undruggable therapeutic target in a cell. This is conspicuous, for example, when a site for binding to a low molecular weight compound cannot be found in a therapeutic target. An antibody can bind to the target with strong binding affinity with different principles from the low molecular weight compound, and hence can effectively work on the therapeutic target regarded as undruggable with the low molecular weight compound. An antibody is, however, usually extracellularly secreted, and extracellularly functions. Therefore, in order to express, in a cell, a secretory protein (protein extracellularly secreted), a gene can be designed to express a secretory protein (intracellular antibody) in a cell. For example, in order to express a secretory protein (intracellular antibody), a signal sequence of the protein can be disrupted, preferably removed or the like. In particular, a scFv can exhibit an aggregation property in a cell. Accordingly, the tag of the present disclosure can be fused with an aggregating protein, particularly a secretory protein, particularly an antigen-binding fragment of an antibody, and with preferably a scFv. A secretory protein exhibits an aggregation property in a cell in some cases. For stabilizing such a secretory protein before secretion in a cell, the tag of the present disclosure can be effective.

[0328] The antibody, or the antigen-binding fragment of the antibody can have binding affinity (KD) to an antigen thereof of, for example, 10⁻⁵ M or less, 10⁻⁶ M or less, 10⁻⁷ M or less, 10⁻⁸ M or less, 10⁻⁹ M or less, 10⁻¹⁰ M or less, 10⁻¹¹ M or less, or 10⁻¹² M or less. A test of the binding property and a test of the binding affinity can be performed, for example, in a buffered saline.

[0329] An example of the antigen includes an intracellular antigen such as an intracellular protein. Examples of the intracellular protein include an intracytoplasmic protein (such as an intracellular extravesicular cytoplasmic protein), a nuclear protein (in this case, the peptide tag or the fusion protein may contain a nuclear localization signal), a nuclear

transcription factor, a protein binding to a transcription factor, a protein binding to a genomic DNA, a protein binding to a protein binding to a genomic DNA, a constituent protein of chromatin, a protein binding to chromatin, an intracellular cell skeleton, and a protein binding to an intracellular cell skeleton. The intracellular protein is not especially limited, and other examples include a gene product of a cancer driver gene, a protein in an activated signal cascade (particularly in an activated immune cell of a patient having a cancer cell or an immune-related disease), and a gene product of a tumor suppressor gene (particularly under regulation of a binding partner for negative regulation thereof). In one embodiment, the antigen can be Kras.

[0330] Introduction of the protein into a cell can be conducted by introducing, into the cell, a protein expression vector containing a nucleic acid encoding a fusion protein (a fusion protein of the peptide tag and the protein of interest) operably linked to a regulatory sequence. The protein expression vector is introduced into a protein-producing cell or a mammal cell, and can express the fusion protein in the cell.

[0331] The regulatory sequence can be a promoter capable of transcribing an mRNA, and for example, various types of pol II promoters can be used. The pol II promoters are not especially limited, and examples include a CMV promoter, an EF1 promoter (EF1 α promoter), an SV40 promoter, an MSCV promoter, an hTERT promoter, a β actin promoter, a CAG promoter, and a CBh promoter. Further, a promoter driving bacteriophage-derived RNA polymerase, such as a T7 promoter, a T3 promoter, or an SP6 promoter, and a pol III promoter such as a U6 promoter can be used as the promoter. For a cyclic DNA, the T7 promoter can be preferably used, and for a linear DNA, the SP6 promoter can be preferably used. The promoter may be a promoter of a virus. Alternatively, the promoter may be an inducible promoter. The inducible promoter is a promoter capable of inducing expression of a polynucleotide functionally linked to the promoter only in the presence of an inducer driving the promoter. An example of the inducible promoter includes a promoter inducing gene expression by heat such as a heat shock promoter. Another example of the inducible promoter includes a promoter using a drug as the inducer driving the promoter. Examples of such a drug inducible promoter include a Cumate operator sequence, a λ operator sequence (such as 12 \times λ Op), and a tetracycline-based inducible promoter. An example of the tetracycline-based inducible promoter includes a promoter driving gene expression in the presence of tetracycline or a derivative thereof (such as doxycycline), or a reverse tetracycline controlled trans-activator (rtTA). An example of the tetracycline-based inducible promoter includes a TRE3G promoter.

[0332] The protein expression vector is not especially limited, and can be a virus vector or a plasmid vector. The virus vector is not especially limited, and examples include a retrovirus vector, a lentivirus vector, an adenovirus vector, an adeno-associated virus vector, a herpes simplex virus vector, a vaccinia virus vector, a Sendai virus vector, and a vesicular stomatitis virus vector. From the viewpoint of changing infectiveness to a cell, these vectors may be of pseudo type. These vectors may be derived from attenuated strains. Such a vector can be appropriately prepared by known technique.

[0333] From the viewpoint of convenience in production of the protein expression vector, the protein expression

vector may contain a nucleic acid encoding the regulatory sequence and the peptide tag of the present disclosure operably linked to the regulatory sequence, and have, on the downstream of the nucleic acid, a cloning site of a nucleic acid encoding the protein of interest. The cloning site has a restriction enzyme cleavage site uniquely present in the vector, and is suitable for introducing a fragment of a gene encoding the protein of interest. A gene encoding the fusion protein of the peptide tag and the protein of interest is obtained by linking a gene encoding the protein of interest in-frame to a gene encoding the peptide tag. Accordingly, the present disclosure provides a protein expression vector containing a nucleic acid encoding the regulatory sequence and the peptide tag of the present disclosure operably linked to the regulatory sequence. In one preferable embodiment, this vector has, on the downstream of the nucleic acid, a cloning site of a nucleic acid encoding the protein of interest. The present disclosure provides a protein expression vector containing: a nucleic acid encoding the regulatory sequence and the peptide tag of the present disclosure operably linked to the regulatory sequence; and a nucleic acid encoding the protein of interest linked in-frame to the former nucleic acid. In this manner, the fusion protein of the peptide tag and the protein of interest can be expressed in a cell.

[0334] The present disclosure provides a messenger RNA (mRNA) containing a nucleic acid encoding the peptide tag of the present disclosure. The mRNA further contains a nucleic acid encoding the protein of interest. The nucleic acid encoding the protein of interest is linked in-frame to the nucleic acid encoding the peptide tag. In one embodiment, at least one or more uridines may be changed to pseudouridines in the mRNA. The pseudouridine can be 1-methyl-pseudouridine. The mRNA can be one transcribed from a cDNA, namely, may not have an intron. The mRNA may have a cap structure at the 5' end (Furuichi Y. & Miura K., Nature, 1975; 253 (5490): 374-5). As the cap structure, a Cap0 structure can be added to the mRNA by Anti-Reverse Cap Analogues (ARCA) method using a cap analogue (Stepinski J. et al., RNA, 2001 Oct; 7(10): 1486-95). When 2'-O methyltransferase treatment is further performed, the Cap0 structure of the mRNA can be converted to a Cap1 structure. Such an operation can be performed by an ordinary method, and can be practiced using, for example, a commercially available kit, such as ScriptCap m7G Capping System, ScriptCap 2'-O-Methyltransferase Kit, or T7 mScript Standard mRNA Production System (AR Brown Co., Ltd.). The mRNA may have a poly A chain. The addition of a poly A chain can be performed by an ordinary method, and can be performed, for example, with A-Plus Poly(A) Polymerase Tailing Kit (AR Brown Co., Ltd.). Accordingly, in one embodiment, the mRNA can be an mRNA that has a cap structure at the 5' end, has a poly A chain at the 3' end, and preferably has pseudouridine (preferably 1-methyl-pseudouridine) as at least a part of uridines. The mRNA can be an isolated mRNA or a synthesized mRNA.

[0335] The mRNA can be encapsulated in a nanoparticle, such as a lipid nanoparticle (LNP). Thus, degradation of the mRNA in a living body is prevented, and efficiency of delivering the mRNA into a cell is improved. Accordingly, in one embodiment, the mRNA can be an mRNA that has a cap structure at the 5' end, has a poly A chain at the 3' end, and preferably has pseudouridine (preferably 1-methyl-pseudouridine) as at least a part of uridines. Such a lipid

nanoparticle encapsulating the mRNA is also provided. The lipid nanoparticle is not especially limited, and lipid nanoparticles described in, for example, U.S. Pat. Nos. 9,364,435B, 8,822,668B, 8,802,644B, and 8,058,069B2 can be used. Alternatively, the mRNA may be encapsulated in a polyion complex micelle, or a polyion complex polymersome (Miyata et al., Chem. Soc. Rev., 2012, 41, 2562-2574). A nanoparticle refers to a particle having a diameter (for example, a hydrodynamic diameter) less than 1 μm.

[0336] Accordingly, the present disclosure provides a nanovesicle (such as a lipid nanovesicle, or a polyion complex polymersome) containing an mRNA at least containing a nucleic acid encoding the peptide tag of the present disclosure. The mRNA can contain a nucleic acid encoding the fusion protein of the peptide tag and the protein of interest.

[0337] The present disclosure provides a method for reducing, inhibiting, or improving aggregation tendency of a protein, including fusing, to the protein, the peptide tag of the present disclosure that reduces, inhibits, or improves the aggregation tendency. In this embodiment, the protein can be an aggregating protein. The aggregation tendency can be aggregation tendency under an intracellular environment. The fusion is conducted usually on the N terminal and/or the C terminal of the protein. The method can be an in vitro method.

[0338] The present disclosure provides a method for increasing, promoting, or improving a non-aggregation property a protein, including fusing, to the protein, the peptide tag of the present disclosure that reduces, inhibits, or improves the aggregation tendency. In this embodiment, the protein can be an aggregating protein. The non-aggregation property can be a non-aggregation property under an intracellular environment. The method can be an in vitro method.

[0339] The present disclosure provides a method for increasing, promoting, or improving stability of a protein, including fusing, to the protein, the peptide tag of the present disclosure that reduces, inhibits, or improves aggregation tendency. In this embodiment, the protein can be an aggregating protein. In this embodiment, the protein can be a non-aggregation property. The stability can be stability under an intracellular environment. The method can be an in vitro method.

[0340] The present disclosure provides use of the peptide tag of the present disclosure for reducing, inhibiting, or improving aggregation tendency of a protein. The present disclosure also provides use of the peptide tag of the present disclosure for increasing, promoting, or improving a non-aggregation property of a protein. The present disclosure also provides use of the peptide tag of the present disclosure for increasing, promoting, or improving stability of a protein. The use can be use in vitro.

[0341] It can be tested by in vitro assay how strongly the peptide tag of the present disclosure can reduce, inhibit, or improve the aggregation tendency of a protein; increase, promote, or improve the non-aggregation property; or increase, promote, or improve the stability. For example, a gene encoding the peptide tag of the present disclosure is fused, for introduction into a cell, to the N terminal or the C terminal of a gene encoding an aggregating protein such as a scFv having an amino acid sequence of SEQ ID NO: 1, and thus, the aggregating protein fused with the peptide tag of the present disclosure can be expressed in the cell. An aggregation formed by the aggregating protein is observed

with an antibody against the aggregating protein, and thus, a rate (%) of cells having aggregations to cells expressing the aggregating protein can be calculated. This rate can be used for evaluating influence of the peptide tag of the present disclosure on the aggregation tendency, the non-aggregation tendency, and the stability of the protein.

[0342] The present disclosure provides a composition containing the peptide tag of the present disclosure. The peptide tag of the present disclosure can be linked to the protein of interest in a reaction solution by, for example, a click reaction. The click reaction can be a Huisgen reaction. One of the peptide tag and the protein of interest is modified with an alkyne and the other is modified with an azide compound, and thus, a 1,2,3-triazole ring is formed to obtain the link therebetween.

[0343] The present disclosure provides a composition containing the fusion protein of the peptide tag of the present invention and the protein of interest. The protein of interest can be an aggregating protein in one embodiment. The protein of interest can be an antigen-binding fragment of an antibody in one embodiment. The protein of interest can be a scFv in one embodiment.

[0344] The present disclosure provides an mRNA containing a nucleic acid encoding the fusion protein of the peptide tag of the present disclosure and the protein of interest, and a composition containing the mRNA. The present disclosure provides a vesicle containing an mRNA containing a nucleic acid encoding the fusion protein of the peptide tag of the present disclosure and the protein of interest, and a composition containing the vesicle.

[0345] The present disclosure provides a protein expression vector containing a nucleic acid encoding the fusion protein of the peptide tag of the present disclosure operably linked to a regulatory sequence and the protein of interest, and a composition containing the protein expression vector.

[0346] In one embodiment, the composition may further contain a pharmaceutically acceptable carrier, excipient, and/or additive. The composition can be a pharmaceutical composition in one embodiment.

[0347] In all embodiments of the present disclosure, the fusion protein having the peptide tag of the present disclosure linked thereto can be a non-natural protein.

[0348] In all embodiments of the present disclosure, the peptide and the protein can be respectively recombinant peptide and protein.

[0349] In all embodiments of the present disclosure, regarding a scFv having at least an amino acid sequence of SEQ ID NO: 1, the peptide tag of the present disclosure can reduce, inhibit, or improve aggregation tendency of the protein; increase, promote, or improve a non-aggregation property; or increase, promote, or improve stability.

[0350] A peptide, a protein, and a nucleic acid can be isolated, concentrated, or purified. Isolation means that one or more components of a system are separated from a given component. Purification means that a relative concentration of a given component is increased as compared with a concentration of one or more other components of a system. Concentration means that a concentration of a given component is increased.

[0351] One aspect of the present disclosure provides:

[0352] a method for acquiring (or selecting or identifying) an amino acid sequence (or a nucleic acid encoding the amino acid sequence), including acquiring an amino acid sequence in which:

[0353] (a) 5% or more and less than 45% of amino acids contained in the amino acid sequence are acidic amino acids; and

[0354] (b) 20% or more of the amino acids contained in the amino acid sequence are amino acids selected from the group consisting of F, P, Y, G, S, Q, N, and A.

[0355] One aspect of the present disclosure provides:

[0356] a method for acquiring (or selecting or identifying) an amino acid sequence (or a nucleic acid encoding the amino acid sequence) including:

[0357] acquiring an amino acid sequence satisfying any one of conditions (A) to (AE) and (AF) to (AU) described above, or any combination of these conditions.

[0358] In one aspect of the present disclosure, the method for acquiring an amino acid sequence may further include:

[0359] selecting or identifying an amino acid sequence of a peptide tag that, when a fusion protein of a peptide having the selected or identified amino acid sequence and a reference protein is expressed in a mammal cell (preferably in a human cell), provides reduction in a proportion of cells in which the fusion protein forms an aggregation (aggregation rate) (or the proportion which is not more than a predetermined value).

[0360] In one aspect of the present disclosure, the method for acquiring may further include obtaining a peptide having the amino acid sequence, or a nucleic acid encoding the peptide.

[0361] In the method for acquiring, the amino acid to be obtained may have a length of 10 to 200 amino acids (for example, a length of 10 to 90 amino acids).

[0362] One aspect of the present disclosure provides a method for acquiring (or selecting or identifying) an amino acid sequence (or a nucleic acid encoding the amino acid sequence) having a length of 10 to 200 amino acids (10 to 90 amino acids), including:

[0363] acquiring an amino acid sequence in which:

[0364] (a) 5% or more and less than 45% of amino acids contained in the amino acid sequence are acidic amino acids; and

[0365] (b) 20% or more of the amino acids contained in the amino acid sequence are amino acids selected from the group consisting of F, P, Y, G, S, Q, N, and A;

[0366] Selecting or identifying an amino acid sequence of a peptide tag that, when the fusion protein of a peptide having the selected or identified amino acid sequence and a reference protein (reference protein) is expressed in a mammal cell (preferably in a human cell), provides reduction of a proportion of cells in which a fusion protein forms an aggregation (aggregation rate) (or the proportion which is not more than a predetermined value); and

[0367] obtaining a peptide having the amino acid, or a nucleic acid encoding the peptide.

[0368] In this method, a peptide having a particularly excellent stabilizing action can be selected from peptides having a stabilizing action by increasing the extent of reduction or by reducing the predetermined value. In one aspect, the present method may further include expressing, in a mammal cell (preferably in a human cell), the fusion protein of the peptide having the selected or identified amino acid sequence and the reference protein. The present method may further include obtaining a nucleic acid encoding the selected or identified amino acid sequence. The predetermined value can be a numerical value based on a rate (%) of

cells having an aggregation to cells expressing the aggregating protein. For example, the predetermined value can be a value of 30% or less, a value of 20% or less, a value of 15% or less, a value of 10% or less, a value of 5% or less, a value of 3% or less, a value of 2% or less, or a value of 1% or less, or 0%. The predetermined value can be a value in a range of 0% to 10%, a value in a range of 10% to 20%, or a value in a range of 20 to 30%. When a peptide having a higher effect is desired to be acquired, the predetermined value is preferably smaller. In this embodiment, a peptide tag exhibiting a stronger effect can be selected from, for example, the above-described peptide tags of the present disclosure (for example, any one of the peptide tags of (A) to (Z), (AA) to (AE), and (AF) to (AU) described above).

[0369] The reference protein can be, for example, a protein (aggregating protein, such as a scFv) in which the proportion of cells having an aggregation formed therein by the fusion protein is more than 30%, 40% to 50% or less, 50% to 60%, 60% to 70%, 70% to 80%, 80% to 90%, 90% to 95%, 95% to 99%, 99% to 99.9%, or 99.9% or more. The reference protein needs not have special functionality or binding property to an antigen but is used simply for evaluating the reduction of the aggregation rate, and therefore, the CDR sequence thereof may be any sequence. The aggregation rate of the scFv may be varied depending on the amino acid sequence of the CDR. It is possible to search for an amino acid sequence of a peptide tag that provides the aggregation rate not more than the predetermined value in the scFv having the varied aggregation rate. In one preferable embodiment, the reference protein can be a protein having an amino acid sequence of SEQ ID NO: 1.

[0370] In one embodiment, the amino acid sequence to be obtained further satisfies the condition defined in (B) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (D) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (E) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (F) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (G) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (H) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (I) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (J) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (K) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (L) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (M) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (N) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (O) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (P) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (Q) described above. In one

embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (R) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (S) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (T) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (U) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (V) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (W) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (X) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (Y) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (Z) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (AA) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (AB) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (AC) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (AD) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (AE) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (AF) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (AG) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (AH) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (AI) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (AJ) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (AK) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (AL) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (AM) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (AN) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (AO) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (AP) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (AQ) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (AR) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (AS) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (AT) described above. In one embodiment, the amino acid sequence to be acquired further satisfies the condition defined in (AU) described above. In

gation rate of the fusion protein is not more than a predetermined value, which is lower than this aggregation rate by 40% or more.

[0473] In one preferable embodiment, the aggregation rate of the reference protein is more than 70%, and the aggregation rate of the fusion protein is not more than a rate by 50% or more.

[0474] In one preferable embodiment, the aggregation rate of the reference protein is more than 70%, and the aggregation rate of the fusion protein is not more than a predetermined value, which is lower than this aggregation rate by 60% or more.

[0475] In one preferable embodiment, the aggregation rate of the reference protein is more than 70%, and the aggregation rate of the fusion protein is not more than a predetermined value, which is lower than this aggregation rate by 70% or more.

[0476] In one preferable embodiment, the aggregation rate of the reference protein is more than 80%, and the aggregation rate of the fusion protein is not more than a predetermined value, which is lower than this aggregation rate by 10% or more.

[0477] In one preferable embodiment, the aggregation rate of the reference protein is more than 80%, and the aggregation rate of the fusion protein is not more than a predetermined value, which is lower than this aggregation rate by 20% or more.

[0478] In one preferable embodiment, the aggregation rate of the reference protein is more than 80%, and the aggregation rate of the fusion protein is not more than a rate by 30% or more.

[0479] In one preferable embodiment, the aggregation rate of the reference protein is more than 80%, and the aggregation rate of the fusion protein is not more than a predetermined value, which is lower than this aggregation rate by 40% or more.

[0480] In one preferable embodiment, the aggregation rate of the reference protein is more than 80%, and the aggregation rate of the fusion protein is not more than a predetermined value, which is lower than this aggregation rate by 50% or more.

[0481] In one preferable embodiment, the aggregation rate of the reference protein is more than 80%, and the aggregation rate of the fusion protein is not more than a predetermined value, which is lower than this aggregation rate by 60% or more.

[0482] In one preferable embodiment, the aggregation rate of the reference protein is more than 80%, and the aggregation rate of the fusion protein is not more than a predetermined value, which is lower than this aggregation rate by 70% or more.

[0483] In one preferable embodiment, the aggregation rate of the reference protein is more than 80%, and the aggregation rate of the fusion protein is not more than a rate by 80% or more.

[0484] In one preferable embodiment, the aggregation rate of the reference protein is more than 90%, and the aggregation rate of the fusion protein is not more than a predetermined value, which is lower than this aggregation rate by 10% or more.

[0485] In one preferable embodiment, the aggregation rate of the reference protein is more than 90%, and the aggre-

gation rate of the fusion protein is not more than a predetermined value, which is lower than this aggregation rate by 20% or more.

[0486] In one preferable embodiment, the aggregation rate of the reference protein is more than 90%, and the aggregation rate of the fusion protein is not more than a predetermined value, which is lower than this aggregation rate by 30% or more.

[0487] In one preferable embodiment, the aggregation rate of the reference protein is more than 90%, and the aggregation rate of the fusion protein is not more than a predetermined value, which is lower than this aggregation rate by 40% or more.

[0488] In one preferable embodiment, the aggregation rate of the reference protein is more than 90%, and the aggregation rate of the fusion protein is not more than a rate by 50% or more.

[0489] In one preferable embodiment, the aggregation rate of the reference protein is more than 90%, and the aggregation rate of the fusion protein is not more than a predetermined value, which is lower than this aggregation rate by 60% or more.

[0490] In one preferable embodiment, the aggregation rate of the reference protein is more than 90%, and the aggregation rate of the fusion protein is not more than a predetermined value, which is lower than this aggregation rate by 70% or more.

[0491] In one preferable embodiment, the aggregation rate of the reference protein is more than 90%, and the aggregation rate of the fusion protein is not more than a predetermined value, which is lower than this aggregation rate by 80% or more.

[0492] In one preferable embodiment, the aggregation rate of the reference protein is more than 90%, and the aggregation rate of the fusion protein is not more than a predetermined value, which is lower than this aggregation rate by 90% or more.

[0493] A nucleic acid encoding the obtained peptide tag is linked in-frame to a nucleic acid encoding the protein of interest, and thus, a nucleic acid encoding the fusion protein of the peptide tag and the protein of interest can be obtained. From the nucleic acid encoding the fusion protein, the fusion protein can be expressed. A protein expression vector containing a nucleic acid encoding a fusion protein operably linked to a regulatory sequence may be prepared.

[0494] In one aspect of the present disclosure, the method can be employed for determining whether or not a peptide tag having an amino acid sequence with a length of 10 to 200 amino acids (for example, 10 to 90 amino acids) has an effect of improving an aggregation property not less than a predetermined intensity against a tagged protein. In other words, in one aspect of the present disclosure, a method for determining whether or not a peptide tag having an amino acid sequence with a length of 10 to 90 amino acids has an effect of improving an aggregation property not less than a predetermined intensity against a tagged protein, and the method including:

[0495] acquiring an amino acid sequence in which:

[0496] (α) 5% or more and less than 45% of amino acids contained in the amino acid sequence are acidic amino acids; and

[0497] (β) 20% or more of the amino acids contained in the amino acid sequence are amino acids selected from the group consisting of F, P, Y, G, S, Q, N, and A;

[0498] determining that an amino acid sequence of a peptide tag that, when the fusion protein of a peptide having the selected or identified amino acid sequence and a reference protein is expressed in a mammal cell (preferably in a human cell), provides reduction of a proportion of cells in which a fusion protein forms an aggregation (aggregation rate) (or the proportion which is not more than a predetermined value), has the effect of improving an aggregation property (for example, an aggregation property not less than a predetermined intensity).

[0499] In one embodiment, the method may further include expressing, in a mammal cell (preferably in a human cell), the fusion protein of the peptide having the selected or identified amino acid sequence and the reference protein. In one embodiment, the amino acid sequence to be acquired can be selected from an amino acid sequence group. The details of the amino acid group are the same as those described above. In one aspect, the amino acid sequence to be acquired satisfies another one or more conditions. The conditions are the same as those described above. In one embodiment, the aggregation rate of the reference protein and the predetermined value are the same as those described above.

[0500] In one embodiment, in the method for selecting or identifying an amino acid sequence having a length of 10 to 200 amino acids (for example, 10 to 90 amino acids), the amino acid sequence group having a length of 10 to 200 amino acids (for example, 10 to 90 amino acids) can be a group of amino acid sequences encoded by the coding region of the human genome. In one embodiment, the amino acid sequence group having a length of 10 to 90 amino acids can be a group of amino acid sequences encoded by the coding region of the genome of a non-human living thing.

[0501] In one embodiment, the amino acid sequence having a length of 10 to 200 amino acids (for example, 10 to 90 amino acids) can be a neo-antigen. A neo-antigen was discovered as a mutant antigen newly caused by gene mutation peculiar to a cancer cell. The neo-antigen is not expressed in a non-cancer cell. It is expected that immunity can be induced specifically to cancer by inducing immunity to a neo-antigen by administering a peptide containing the neo-antigen. The neo-antigen can be different among cancer cells. The neo-antigen can be used, for example, for tagging a protein of interest to be expressed in a cell having the neo-antigen, and can be thus usable because it is a peptide originally expressed, and hence the cell is not or little adversely affected. The neo-antigen can be a naturally occurring mutant. The neo-antigen has one or more mutations selected from the group consisting of addition, insertion, deletion, and substitution in, for example, a wild type sequence thereof. For example, a neo-antigen of a human typically has one or more (for example, 1 to 10, for example, 1 to several, 1 to 5, 1 to 4, 1 to 3, 2 or 1) mutations selected from the group consisting of addition, insertion, deletion, and substitution in a wild type sequence of a human. The neo-antigen of a human can have, for example, 80% or more identity, 85% or more identity, 90% or more identity, or 95% or more identity to the wild type sequence of a human.

[0502] In one preferable embodiment, the reference protein is a scFv. In one embodiment, the scFv has the amino acid sequence of SEQ ID NO: 1.

[0503] Still another aspect of the present invention provides a method for modifying an amino acid sequence of a peptide tag, including:

[0504] preparing a peptide tag (that may be any one of the peptide tags disclosed herein) for producing a fusion protein; and

[0505] obtaining a modified amino acid sequence by substituting, with either of P and N, one or more (preferably, a plurality of) amino acids of the Element 2, 3, or 4 (preferably, any one of the Element 2, the Element 4, and A, G, Y, and F).

The method of this aspect may further include:

[0506] determining that an amino acid sequence of a peptide tag that, when the fusion protein of a peptide having a selected or identified amino acid sequence and a reference protein is expressed in a mammal cell (preferably in a human cell), provides reduction of a proportion of cells in which a fusion protein forms an aggregation (aggregation rate) (or the proportion which is not more than a predetermined value), has the effect of improving an aggregation property (for example, an aggregation property not less than a predetermined intensity).

In this manner, modification having particularly strong aggregation reducing action can be performed.

EXAMPLES

Example 1

[Method]

Construction of Gene Expression Vector

[0507] A gene fragment encoding a fusion protein containing a peptide tag and an aggregating protein was produced by Eurofins Genomics K.K. or VectorBuilder Japan, Inc. The thus synthesized gene fragment was cloned into a pEF-BOS vector (Mizushima and Nagata, Nucleic Acids Res. 1990 Sep. 11; 18 (17): 5322). As the aggregating protein, a protein (specifically, a scFv) having an amino acid sequence set forth in SEQ ID NO: 1 was used. This aggregating protein aggregates in the cytoplasm when expressed in a cell. The sequences of tags used here and SEQ ID NOs thereof were as shown in Table 1 below.

TABLE 1

List of Tags used in Experiment		
Tag Name	Amino Acid Sequence	SEQ ID NO:
Tag-1-1	AHSSSAESESTSDDSSSDS ESESSSSDSEGS	2
Tag-1-2	AHSLSAELESTIDSDCSDDW ESELSSSDSEGS	3
Tag-1-3	AQSSSAESESGSDSDSSDS ESESSSSDSEGS	4
Tag-4-1	NEYGYREAFDEDYEQQDEDFA EQDPDGNEAFEGEYDGPNQD EYPDEAQNFE	5

TABLE 1-continued

List of Tags used in Experiment		
Tag Name	Amino Acid Sequence	SEQ ID NO:
Tag-2-1	DEAGSSGAPADEAGSSGAPA DEAGSSGAPADEAGSSGAPA DEAGSSGAPAGS	6
Tag-2-2	DEVGISLAPTDENVISLAPT DEVGISLAPTDENVISLAPT DEVGISLAPTGS	7
Tag-2-3	DEVMISLWPTDEVMISLWPT DEVMISLWPTDEVMISLWPT DEVMISLWPTGSS	8
Tag-3-1	DEAGSSGAPADFGSSGAPA DFAGSSGAPADFGSSGAPA DFAGSSGAPAGS	9
Tag-3-2	DTAVSSIAPLDTAVSSIAPL DTAVSSIAPLDTAVSSIAPL DTAVSSIAPLGS	10
Tag-3-3	DTWVSLIAILDTWVSLIAIL DTWVSLIAILDTWVSLIAIL DTWVSLIAILGS	11
Myc	EQKLISEEDL	12

[0508] Specific content ratios of amino acids were as follows.

TABLE 2

Tag Name	Amino Acid Content Ratios in Each Peptide and Influence of Tag Addition on Intracellular Aggregation Formation of Protein							
	Element 1 (%)	Element 2 (%)	Element 3 (%)	Element 4 (%)	S (%)	G (%)	A (%)	Aggregation Rate (%)
Myc (control)	40.0	10.0	20.0	30.0	10.0	0	0	40.1
Tag-1-1	28.1	3.1	65.6	3.1	56.3	3.1	6.3	11.4
Tag-1-2	28.1	3.1	46.9	21.9	37.5	3.1	6.3	20.2
Tag-1-3	28.1	0.0	71.9	0.0	56.3	6.3	6.3	15.7
Tag-4-1	42.0	2.0	56.0	0.0	0.0	8.0	8.0	4.0
Tag-2-1	19.2	0.0	80.8	0.0	21.2	21.2	28.8	29.7
Tag-2-2	19.2	0.0	42.3	38.5	11.5	11.5	9.6	22.5
Tag-2-3	19.2	0.0	23.1	57.7	11.5	1.9	0.0	41.2
Tag-3-1	9.6	0.0	90.4	0.0	21.2	21.2	28.8	43.4
Tag-3-2	9.6	0.0	51.9	38.5	21.2	1.9	19.2	26.1
Tag-3-3	9.6	0.0	23.1	67.3	11.5	1.9	9.6	65.0

[0509] In Table 2, Element 1 refers to D and E, Element 2 refers to H, K, and R, Element 3 refers to F, P, Y, G, S, Q, N, and A, and Element 4 refers to the other amino acids.

[0510] The peptide tag was fused to the N terminal of an antibody fragment. The antibody fragment fused with the peptide tag was linked, for detection, to an HA tag of SEQ ID NO: 13 at the C terminal of the antibody fragment.

Measurement of Intracellular Aggregation Rate of Intracellular Antibody

(Cell Culture)

[0511] A HeLa cell derived from human cervical epithelial cancer was prepared. The HeLa cell was purchased from JCRB Cell Bank, National Institutes of Biomedical Innova-

tion, Health and Nutrition (JCRB9004), and cultured in DMEM (D-MEM, FUJIFILM Wako Pure Chemical Corporation, 4548995066251) containing 10% FBS.

(Transfection)

[0512] In a 35 mm glass bottom dish (IWAKI 3911-035, glass hole, inner diameter: 12 mm) coated with poly-L-lysine (Sigma-Aldrich P1399 Poly-L-lysine hydrobromide mol. Wt. 150000-300000), 4×10^5 HeLa cells were plated, and after 24 hours, the antibody fragment gene described above was introduced into the HeLa cells with Lipofectamine 3000 (Invitrogen, L3000-008) based on a use method provided by the manufacturer, and thus, a tagged antibody fragment was expressed in the cells. A culture fluid was removed 24 hours after the transfection, and the resultant cells were fixed with 4% PFA. 20 minutes after the fixation, the resultant cells were washed with PBS(-), and thereafter, the antibody expressed in the cells (hereinafter, simply referred to as the “intracellular antibody”) was observed by immunostaining to measure the aggregation rate thereof.

(Immunostaining)

[0513] The immunostaining was performed by a standard method. The cells were treated with 0.3% Triton X-100/PBS(-) for 2 minutes, and kept for 1 hour at room temperature in a blocking solution (1% BSA, 0.1% Triton X-100/PBS(-)). The cells were kept for 2 hours at room temperature in an anti-HA antibody (rabbit anti-Ha anti-

body: Sigma-Aldrich, H6908) diluted with the blocking solution, the resultant cells were washed with the blocking solution, and then the resultant cells were kept at room temperature in Alexa Fluor 488 Goat anti-rabbit IgG (H+L) (Invitrogen, A11034) for 2 hours. The resultant cells were washed with the blocking solution, kept for 15 minutes at room temperature in a nuclear staining probe (NucBlue Fixed cell stain ReadyProbes, Invitrogen, R37606), washed with PBS(-), and stored at 4°C until fluorescence imaging.

(Fluorescence Imaging)

[0514] The fluorescence imaging was performed with Keyence BZ-X700 or BZ-X800 using a 40x objective lens. An image including 200 or more cells per dish was acquired to count the number of cells having aggregation of the

intracellular antibody therein (intracellular antibody aggregating cells). The number of the intracellular antibody aggregating cells was normalized with the total number of cells expressing the intracellular antibody to quantify an intracellular aggregation rate of the intracellular antibody.

[Results]

[0515] The peptides having the amino acid sequences shown in Table 1 were fused with antigen-binding fragments (specifically, scFvs) of antibodies, and the thus obtained fusion proteins were expressed in the cytoplasm of the HeLa cell. The amino acid ratio in each peptide tag are shown in Table 2. The intracellular aggregation rates obtained based on fluorescence images were as shown in Table 2. In Table 2, a molecule provided with a Myc tag was used as a control, and the effect of each peptide tag was evaluated based on a difference in the aggregation rate from that of the Myc tag.

[0516] As shown in Table 2,

[0517] the aggregation rate tended to be reduced when the rate of the Element 3 was in a range of 40 to 75%;

[0518] the aggregation rate tended to be reduced as the rate of the Element 4 was lower;

[0519] the aggregation rate tended to be increased when the rate of alanine (A) was more than 10%;

[0520] the aggregation rate tended to be increased when the rate of glycine (G) was more than 10%; and

[0521] the intracellular aggregation rate tended to be reduced as the rate of the Element 1 was higher.

[0522] Serine (S) may not be present (see Tag-4-1), but did not adversely affect the reduction of the aggregation rate even present in a large amount (see Tag-1-1 to Tag-1-3).

[0523] In either case, a peptide tag having an acidic amino acid content less than 45% did not exhibit non-specific adsorption (particularly, non-specific adsorption due to negative charge of the peptide tag) to an intracellular protein or the like. It was suggested that the peptide tag of the present disclosure is useful from the viewpoint that free

intracellular localization of a protein of interest is not restricted peculiarly to the tag.

[0524] It is understood, through comparison between the Myc tag (control) and Tag-4-1, that reduction of the Element 4 with increase of the Element 3 makes a strong contribution to the aggregation rate reducing action of the peptide tag. Similarly, it is understood, also through comparison between Tag-1-2 and Tag-1-1, that the reduction of the Element 4 with increase of the Element 3 makes a contribution to the aggregation rate reducing action of the peptide tag.

[0525] In particular, when the rate of the Element 1 was 20% or more and less than 45%, that of the Element 2 was 10% or less, that of the Element 3 was 40 to 75%, that of the Element 4 was 10% or less, that of alanine was 10% or less, and that of glycine was 10% or less, the aggregation rate was favorably reduced.

[0526] The HeLa cell expressing the fusion protein of the aggregating protein and the tag (Tag4-1) was observed under a fluorescence microscope. As a negative control, a HeLa cell expressing an aggregating protein without Tag4-1 was observed under a fluorescence microscope. Results were as illustrated in FIG. 1. As illustrated in FIG. 1, when tagged with Tag4-1 (fused with Tag-4-1), the aggregating protein homogeneously distributed in the cytoplasm, but the aggregating protein without Tag4-1 formed an aggregation in the cytoplasms of most of cells.

Example 2: Experiment of Adding Amino Acid to Tag

[0527] Amino acid contents in a tag and the aggregation rate of a tagged protein were further analyzed. Serine (S) was randomly inserted into or added to Tag4-1 or Tag11-1, and the effect of the thus obtained modified tags on the protein aggregation rate was examined. The experiment was conducted in the same manner as described in Example above except that the tags were different. Results were as shown in Table 3. In Table 3, added or inserted amino acids are underlined.

TABLE 3

Sequences, S Contents, D/E Contents, and Aggregation Rates of Modified Tags				
Tag	sequence	Aggregation Rate	S (%)	D, E (%)
Tag4-1 (SEQ ID NO: 59)	NEYR A E F ED D YE Q Q D ED F AE Q DPDGNEA FE G EYDGP N Q D EY P DEAQNF E	4.57	0	42
Tag4-1-S10 (SEQ ID NO: 60)	NEGSY R E A F D SE D YE Q S Q D S ED F SAE Q DPD I GN S EA F E G SE Y <u>SD</u> GP N Q D EY P DEAQNF E	5.06	16.7	35
Tag4-1-S20 (SEQ ID NO: 61)	NEGSY R E A F D SE D YE Q S Q D S SE D FS A E Q D <u>SSPDGNS</u> <u>SEA</u> <u>FEGS</u> <u>SEY</u> <u>SD</u> GP N Q D EY P DE <u>ASSQNFS</u> <u>SS</u>	6.76	28.6	30
Tag11-1 (SEQ ID NO: 62)	YDN P Y F E P Q Y GF P PE E DE	14.48	0	15
Tag11-1-S10 (SEQ ID NO: 63)	<u>SYDSNPS</u> <u>YF</u> <u>SEPS</u> <u>QYGS</u> <u>FPP</u> <u>SE</u> <u>SED</u> <u>SE</u> <u>SE</u>	11.29	33.3	10
Tag11-1-S20 (SEQ ID NO: 64)	<u>SSYSDSNP</u> <u>SYF</u> <u>SEPS</u> <u>SSQ</u> <u>SSY</u> <u>GSS</u> <u>FPP</u> <u>SE</u> <u>SE</u> <u>DSSES</u> <u>DSS</u> <u>SE</u>	18.77	50	7.5

[0528] As shown in Table 3 as results, increase of S in a tag did not largely affect the aggregation rate of the tagged peptide. Although it is known that increase of acidic amino acids in a tag largely affects reduction of the aggregation rate of the tagged peptide (US2020/0157210A), on the contrary,

the reduction of the rate of acidic amino acids did not largely reduce the aggregation rate in this example.

[0529] Similarly, amino acid contents except for serine were changed to examine the influence of tagged peptides on the aggregation rate. In Tables 4 to 8, added or inserted amino acids are underlined.

TABLE 4

Sequences, Q Contents, D/E Contents, and Aggregation Rates of Modified Tags				
Tag	sequence	Aggregation Rate	Q (%)	D, E (%)
Tag 4-1	NEGYREAFDEDYEQQDEDFAEQDPDGNEAF EGEYDGPQNQDEYPDEAQNF	4.88	10	42
Tag4-1-Q10 (SEQ ID NO: 65)	<u>NEGOY</u> REAFD <u>QED</u> YE <u>QQQ</u> Q <u>QED</u> F <u>QAE</u> QDP D <u>GNO</u> E <u>A</u> F <u>E</u> G <u>Q</u> EYDGP <u>N</u> Q <u>QDE</u> Y <u>Q</u> P <u>D</u> E <u>A</u> QNF <u>QE</u>	8.31	25	35
Tag4-1-Q20 (SEQ ID NO: 66)	<u>NEGOY</u> R <u>EQAFD</u> <u>QED</u> YE <u>QQQ</u> D <u>QQED</u> F <u>QAE</u> Q D <u>QOP</u> D <u>G</u> N <u>Q</u> E <u>A</u> F <u>E</u> G <u>Q</u> EY <u>Q</u> D <u>G</u> P <u>N</u> Q <u>QDE</u> Y P <u>D</u> E <u>A</u> <u>QQQ</u> N <u>F</u> <u>QQQ</u> E	8.40	35.7	30
Tag 11-1	YDNPYFEPQYGFPEEDEDE	16.59	5	15
Tag11-1-Q10 (SEQ ID NO: 67)	<u>QYD</u> QNP <u>QYF</u> <u>QEP</u> <u>QQY</u> G <u>QFP</u> <u>P</u> <u>QEQ</u> <u>QED</u> QED <u>QE</u>	26.70	36.7	10
Tag11-1-Q20 (SEQ ID NO: 68)	<u>QYQD</u> QNP <u>QYF</u> <u>QEP</u> <u>QQQ</u> Y <u>GQF</u> <u>P</u> <u>QEP</u> <u>QED</u> <u>QO</u> E <u>QD</u> <u>QQQ</u> E	17.27	52.5	7.5

TABLE 5

Sequences, N Contents, D/E Contents, and Aggregation Rates of Modified Tags				
Tag	sequence	Aggregation Rate	N (%)	D, E (%)
Tag 4-1	NEGYREAFDEDYEQQDEDFAEQDPDGNEAF EGEYDGPQNQDEYPDEAQNF	4.88	8.00	42
Tag4-1-N10 (SEQ ID NO: 69)	<u>NEG</u> NY <u>REAFD</u> <u>NED</u> YE <u>QNQ</u> D <u>NED</u> F <u>NAE</u> QDP D <u>GN</u> N <u>E</u> A <u>F</u> E <u>G</u> N <u>EY</u> N <u>D</u> G <u>P</u> N <u>Q</u> D <u>E</u> Y <u>P</u> D <u>E</u> A <u>Q</u> N <u>F</u> <u>NE</u>	6.12	23.33	35
Tag4-1-N20 (SEQ ID NO: 70)	<u>NEG</u> NY <u>REN</u> A <u>FDN</u> <u>NED</u> YE <u>QNQ</u> D <u>NN</u> N <u>E</u> D <u>F</u> NAE <u>Q</u> D <u>NNP</u> D <u>GNN</u> NE <u>A</u> F <u>E</u> G <u>C</u> N <u>EY</u> N <u>D</u> G <u>P</u> N <u>Q</u> D <u>E</u> Y <u>P</u> D <u>EAN</u> Q <u>N</u> F <u>NN</u> NE	6.86	34.29	30
Tag 11-1	YDNPYFEPQYGFPEEDEDE	16.59	5.00	15
Tag11-1-N10 (SEQ ID NO: 71)	<u>NYD</u> NP <u>NPY</u> F <u>NEP</u> <u>NQY</u> G <u>NFP</u> <u>N</u> E <u>DN</u> E <u>D</u> <u>N</u> <u>E</u>	18.56	36.67	10
Tag11-1-N20 (SEQ ID NO: 72)	<u>NNY</u> ND <u>NNP</u> NY <u>F</u> <u>NEP</u> <u>NNQ</u> NN <u>Y</u> <u>GNN</u> F <u>P</u> <u>N</u> E <u>NED</u> <u>NN</u> <u>E</u> <u>Q</u> <u>NN</u> <u>N</u> E	17.68	52.50	7.5

TABLE 6

Sequences, P Contents, D/E Contents, and Aggregation Rates of Modified Tags				
Tag	sequence	Aggregation Rate	P (%)	D, E (%)
Tag 4-1	NEGYREAFDEDYEQQDEDFAEQDPDGNEAF EGEYDGPQNQDEYPDEAQNF	4.49	6.00	42

TABLE 6-continued

Sequences, P Contents, D/E Contents, and Aggregation Rates of Modified Tags					
Tag	sequence	Aggregation Rate	P (%)	D, E (%)	
Tag4-1-P10 (SEQ ID NO: 73)	NEGPYREAFDPEDYEQPOQDPEDFPAEQDPD <u>GNPEAFEGPEY</u> <u>PDGPNPQDEY</u> <u>PDEAQNFPE</u>	8.84	21.67	35	
Tag4-1-P20 (SEQ ID NO: 74)	NEGPYREAFDPEDYEQPOQDPEDFPAEQD <u>PPPDGNPPEAFEGPPEY</u> <u>PDGPNPQDEY</u> <u>PDEAQNFPPPE</u>	6.44	32.86	30	
Tag 11-1	YDNPYFEPQYGFPPPEEDEDE	17.75	20.00	15	
Tag11-1-P10 (SEQ ID NO: 75)	<u>PYDPNPPYFPEPPQYGLFPPEEPE</u> <u>PEDPDE</u>	15.57	46.67	10	
Tag11-1-P20 (SEQ ID NO: 76)	<u>PPYPDPNPPYFPEPPPQPPYGPFPPEPE</u> <u>DPPEPDPPPE</u>	18.12	60.00	7.5	

TABLE 7

Sequences, F Contents, D/E Contents, and Aggregation Rates of Modified Tags					
Tag	sequence	Aggregation Rate	F (%)	D, E (%)	
Tag 4-1	NEYREAFDEDYEQQDEDFAEQDPDGNEAF EGEYDGPQNQDEY	4.90	8.00	42	
Tag4-1-F10 (SEQ ID NO: 77)	NEYFREAFDFEDYEQFQDFEDFFAEQDPD <u>GNPEAFEGFEY</u> <u>PDGPNPQDEY</u> <u>PDEAQNF</u> <u>FFE</u>	35.13	34.29	35	
Tag4-1-F20 (SEQ ID NO: 78)	NEYEREFADDEDEYEQEQDDEEEDEFAEQD <u>EFPDGNNFEEAFEGF</u> <u>EY</u> <u>PDGPNPQDEY</u> <u>PDEAQNF</u> <u>FFF</u> <u>EE</u>	38.76	23.33	30	
Tag 11-1	YDNPYFEPQYGFPPPEEDEDE	16.59	10.00	15	
Tag11-1-F10 (SEQ ID NO: 79)	<u>FYDFNPFYFFEPFQYGFPPFEFEDFEDF</u> <u>FE</u>	43.58	55.00	10	
Tag11-1-F20 (SEQ ID NO: 80)	<u>FFYFDENPFYFFEPFQFFYGFPPFEFE</u> <u>DFFEFDF</u> <u>FFE</u>	27.43	40.00	7.5	

TABLE 8

Sequences, Y Contents, D/E Contents, and Aggregation Rates of Modified Tags					
Tag	sequence	Aggregation Rate	Y (%)	D, E (%)	
Tag 4-1	NEYREAFDEDYEQQDEDFAEQDPDGNEAF EGEYDGPQNQDEY	4.94	8.00	42	
Tag4-1-Y10 (SEQ ID NO: 81)	NEYYYREAFDYEDYEQYQDYEDFYAEQDPD <u>GNYEAFEGYEY</u> <u>YDGPNYQDEY</u> <u>PDEAQNFYE</u>	29.59	23.33	35	
Tag4-1-Y20 (SEQ ID NO: 82)	NEYYYREAFDYEDYEQYQDYEDFYAEQD <u>YYPDGNYYEAFEGYY</u> <u>YEDGPNYQDEY</u> <u>PDEAQNFYYE</u>	37.93	34.29	30	
Tag 11-1	YDNPYFEPQYGFPPPEEDEDE	17.75	15.00	15	
Tag11-1-Y10 (SEQ ID NO: 83)	<u>YYDYNPYYFYEPYQYGYFPPPEYEDYED</u> <u>YE</u>	17.28	43.33	10	

TABLE 8 -continued

Sequences, Y Contents, D/E Contents, and Aggregation Rates of Modified Tags					
Tag	sequence		Aggregation Rate	Y (%)	D, E (%)
Tag11-1-Y20 (SEQ ID NO: 84)	<u>YYYYDYNPYFYEPYYQYYYGYYFPPY</u> <u>EDYYEYDYYE</u>		30.53	57.50	7.5

[0530] As described above, when the influence on the protein aggregation rate of tags obtained by randomly adding or inserting 10 to 20 specific amino acids was examined, the addition or insertion of F and Y tended to worsen the aggregation rate of the tagged proteins. The adverse effect of the addition or insertion of the other amino acids on the aggregation rates of tagged proteins was restrictive even if D and E contents were reduced.

Example 3: Effect on Aggregation Rate of Protein by Amino Acid Content Change by Substitution of Amino Acids in Tag

[0531] N, P, S, Q, F, and Y were respectively substituted with other amino acids, and thus, attempts were made to specify, among these amino acids, an amino acid exhibiting the effect on the protein aggregation rate. In Tables 9 and 10, substituted amino acids are underlined.

TABLE 9

Sequences, N Contents, D/E Contents, and Aggregation Rates of Modified Tags							
N-terminus	scFv	C-terminus	Cell	sequence	Aggregation Rate	N (%)	D, E (%)
Tag 9-1	Y13-259	HA	HeLa	GNNQDSSSDNEADEASDDEONDGN (SEQ ID NO: 85)	8.14	20.00	44.00
Tag 4-8	Y13-259	HA	HeLa	NEGNREASDEDSEQQDEDNAEQDPDGNEANE GESDGPNQDENPDEAQMNSE (SEQ ID NO: 86)	3.67	16.00	42.00
Tag 4-9	Y13-259	HA	HeLa	NEGNREANDEDNEQQQDEDNAEQDPDGNEANE GENDGPNQDENPDEAQMNNE (SEQ ID NO: 87)	4.10	24.00	42.00
Tag 4-10	Y13-259	HA	HeLa	NEGNREANDEDNEQQQDEDNAEGDPDGNEANE GENDGPNQDENPDEAQMNNE (SEQ ID NO: 88)	5.51	32.00	42.00
Tag 4-11	Y13-259	HA	HeLa	NEGNREANDEDNEQQQDEDNAEQDQDGNEANE GENDGQNQDENPDEAQMNNE (SEQ ID NO: 89)	4.44	32.00	42.00
Tag 18-1	Y13-259	HA	HeLa	DNNESADDNNENPEDNNKNTDDNEENPNNNEN (SEQ ID NO: 90)	4.82	43.75	37.50
Tag 4-8	6E	Tag 4-8, CMA, HA	SHSY5Y	NEGNREASDEDSEQQDEDNAEQDPDGNEANE GESDGPNQDENPDEAQMNSE	0.96	16.00	42.00
Tag 18-1	6E	Tag 18-1, CMA, HA	SHSY5Y	DNNESADDNNENPEDNNKNTDDNEENPNNNEN	1.14	43.75	37.50

TABLE 10

Sequences, P Contents, N Contents, and Aggregation Rates of Modified Tags		Aggregation Rate	N (%)	P (%)
Tag 18-1	DNNE SADDNNNENP EDNNKNT DNEENPNNNN	4.76	43.75	6.25
Tag 18-1-NS7 (SEQ ID NO: 91)	DNSE SADDNS ENP EDNS KNT DSEENPSNS	6.37	21.88	6.25
Tag 18-1-NS14 (SEQ ID NO: 92)	DSSE SADSSS E PEDSK ST DSEESPSSSES	5.72	0.00	6.25
Tag 18-NQ7 (SEQ ID NO: 93)	DNQE SADDNQ ENP EDNQ KNT DGEENPONGEQ	6.60	21.88	6.25
Tag 18-NQ14 (SEQ ID NO: 94)	DQQE SADDQQ EQP EDQQ Q TDDQEEQPQQEQ	6.03	0.00	6.25
Tag 18-1-NF7 (SEQ ID NO: 95)	DNFES ADDNF ENP EDNF KNT DFEENP FN FEF	13.50	21.88	6.25
Tag 18-1-NF14 (SEQ ID NO: 96)	DFFES ADDFF EF PEDFF KFT DFEFP FFFF FEF	13.60	0.00	6.25
Tag 18-1-NP7 (SEQ ID NO: 97)	DNPES ADDNP ENP EDNP KNT DPEENP PN PEP	4.48	21.88	28.13
Tag 18-1-NP14 (SEQ ID NO: 98)	DPPES ADDP PEPP EDPP KPT DDPEE PPP PEP	3.18	0.00	50.00
Tag 18-1-NY7 (SEQ ID NO: 99)	DNYE SADDNY ENP EDNY KNT DYYEEN PY NY	11.60	21.88	6.25
Tag 18-1-NY14 (SEQ ID NO: 100)	DYYE SADDYY EYP EDYY KY TDDYE YPY YY	15.70	0.00	6.25
Tag 11-1	YDNP YFEP QY GFP PEED EDE	17.75	5.00	20.00
Tag 11-1-FYPN (SEQ ID NO: 101)	NDNNNNEN QNGNN NEEDEDE	12.46	50.00	0.00
Tag 11-1-FYN (SEQ ID NO: 102)	NDNP NN EPQ NGNP PEED EDE	5.79	30.00	20.00
45Tag1 (SEQ ID NO: 1041)	NDEYSDFEDSDFDG DYK DS DEDY KDDSEN FDDG FEE	13.91	3.71	0
45Tag1-m1 (SEQ ID NO: 1042)	PDEPPDPEDPDPDPKDPKDPEP DDPPE	5.70	0	42.86

[0532] It was found, based on Tables 9 and 10, that the aggregation rate reducing action of a tagged protein is the greatest when the content of P or N in the tag was higher, and subsequently, the aggregation rate reducing action was exhibited in the order that the content of S and Q was higher, and the content of F and Y was higher. 45Tag1-m1 peptide tag, which was produced by substituting, with P, all of the Element 3 in 45Tag1 having an acidic amino acid content of about 51%, largely improved the aggregation inhibiting action thereof through the amino acid substitution with P. In this manner, it was revealed that the increase of the aggregation rate reducing action by adding P and N does not depend on the acidic amino acid content. This result shows that most of constituent amino acids can be acidic amino acids, and N or P.

Example 4: Sequence Shuffle

[0533] Two tags, Tag4-8 (Scr1) and Tag4-8 (Scr2), were synthesized by randomly shuffling the amino acid sequence of Tag4-8, and were tested for the aggregation rate reducing action against the scFv in the same manner as described above. Results were as shown in Table 11.

TABLE 11

Sequence Shuffling		
	Sequence	Aggregation Rate
Tag4-8	NEG NRAS DED S EQ Q D E D N A E Q D P D GNE ANEG E S D G P N Q D E N P D E A Q N S E	3.93
Tag4-8 (Scr1) (SEQ ID NO: 103)	E NE S ED N D E E P N Q N A D D G D P N A N P A A E Q G G D S S D D E E C Q E N D Q S R E N E Q	3.25
Tag4-8 (Scr2) (SEQ ID NO: 104)	Q N E N N G D D Q D Q E E G S E E Q Q G E S D R S N E E S D N E P A A D D N A P E G A E P D D N N	4.52

[0534] As shown in Table 11, Tag4-8, Tag4-8 (Scr1) and Tag4-8 (Scr2) exhibited equivalent aggregation rate reducing actions. Even when only the order of amino acids was changed without changing the content ratios and the lengths of the amino acids, the aggregation rate reducing action was not affected.

Example 5: Use as Tag of Human-Derived Peptide

[0535] From human proteome database (Proteome ID: UP000005640), peptides having specific amino acid content ratios were all extracted, these peptides were randomly selected to be used as tags, and thus, the aggregation rate reducing action against a tagged protein was examined.

Extraction Condition 1:

- [0536] length: 20 to 70 amino acids
- [0537] group [D, E]: content of [30] or more
- [0538] group [D, E]: content of less than [45]
- [0539] group [H, K, R]: content of [5] or less
- [0540] group [C, T, V, L, I, W, M]: content of [5] or less

- [0541] group [G]: content of less than [10]
- [0542] group [A]: content of less than [10]
- [0543] group [F, Y]: content of [5] or less
- [0544] group [N]: content of [15] or more
- [0545] In the above-described extraction condition, the unit of each content is %. The amino acids are described by one letter codes.

TABLE 12

Aggregation Rates of scFvs having Tags Extracted under Extraction Condition 1 Added										
Tag Name	Sequence	Aggregation Rate %								
		DE Rate	N Rate	HKR Rate	CTVLIW M Rate	G Rate	A Rate	FY Rate		
Tag 18-1(465-1)	DNNESADDNNENPEDNNKNTDDNEENPNNNEN	4.68	37.50	43.75	3.13	3.13	0	3.13	0	
Tag 465-2	GNNQDSSSDSNEADEASDDEDNDGN	5.04	44.00	20	0	0	8	8	0	
Tag 465-3	SHGNQNQDSSSDSNEADEASDDEDNDGNEGDNE	3.98	43.75	18.75	3.13	0	9.38	6.25	0	
Tag 465-4	ESADDNNENPEDNNKNTDDNE	9.99	42.86	33.33	4.76	4.76	0	4.76	0	
Tag 465-5	NNENPEDNNKNTDDNEENPNN	13.94	33.33	47.62	4.76	4.76	0	0	0	
Tag 465-6	TDNNESADDNNENPEDNNKN	8.95	35.00	40	5	5	0	5	0	
Tag 465-7	DNNESADDNNENPEDNNKNT	7.77	35.00	40	5	5	0	5	0	
Tag 465-8	NNESADDNNENPEDNNKNTD	7.39	35.00	40	5	5	0	5	0	
Tag 465-9	FWGSHGNNQDSSSDSNEADEASDDEDNDGNE	7.36	38.71	16.13	3.23	3.23	9.68	6.45	3.23	
Tag 1121-1	KPNNSNAPNEDQEEEIQQSE	14.73	30	20	5	5	0	5	0	
Tag 1121-2	NNSNAPNEDQEEEIQQSEQH	13.76	30	20	5	5	0	5	0	
Tag 1121-3	SEGEQQQLKPNNSNAPNEDQEEE	14.15	31.82	18.18	4.55	4.55	4.55	4.55	0	
Tag 2408-1	EQLNFSDDEQGSNSPKENNSEDQ	7.56	33.33	16.67	4.17	4.17	4.17	0	4.17	
Tag 2408-2	EQLNFSDDEQGSNSPKENNSEDQG	7.53	32.00	16	4	4	8	0	4	
Tag 2408-3	EQLNFSDDEQGSNSPKENNSEDQGS	9.42	30.77	15.38	3.85	3.85	7.69	0	3.85	
Tag 6301-1	EENENDENSLSSSSDSSED	12.66	40.00	15	5	5	0	0	0	
Tag 6301-2	NENDENSLSSSSDSSEDKDE	8.85	40.00	15	5	5	0	0	0	
Tag 6626-1	KETNNNSNAQNPSEEEGEGQDE	12.50	33.33	19.05	4.76	4.76	9.52	4.76	0	
Tag 6626-2	ETNNNSNAQNPSEEEGEGQDED	8.04	38.10	19.05	0	4.76	9.52	4.76	0	
Tag 6626-3	KETNNNSNAQNPSEEEGEGGDED	13.56	36.36	18.18	4.55	4.55	9.09	4.55	0	
Tag 6915-1	ENANDSSDDSGEETDESFNP	9.52	40.00	15	0	5	5	5	5	
Tag 7128-1	DDNESNSESAENGWDGSNFSEE	12.10	34.78	17.39	0	4.35	8.7	4.35	4.35	
Tag 7128-2	SDDNESNSESAENGWDGSNFSEE	8.21	33.33	16.67	0	4.17	8.33	4.17	4.17	
Tag 7128-3	SSDDNESNSESAENGWDGSNFSEE	8.06	32.00	16	0	4	8	4	4	
Tag 7315-1	EENASSGDSEENTNSDHESE	10.74	40.00	15	5	5	5	5	0	
Tag 7315-2	SEENASSGDSEENTNSDHES	11.72	35.00	15	5	5	5	5	0	
Tag 7315-3	ENASSGDSEENTNSDHESEQ	10.07	35.00	15	5	5	5	5	0	
Tag 8482-1	DDENSENNWRNEYPEEESSDG	8.17	42.86	19.05	4.76	4.76	4.76	0	4.76	
Tag 8482-2	ENSENNWRNEYPEEESSDGDE	7.11	42.86	19.05	4.76	4.76	4.76	0	4.76	
Tag 8482-3	NSENNWRNEYPEEESSDGEDDS	5.08	40.91	18.18	4.55	4.55	4.55	0	4.55	

TABLE 12-continued

Aggregation Rates of scFvs having Tags Extracted under Extraction Condition 1 Added										
Tag Name	Sequence	Aggregation Rate %								
		DE Rate	N Rate	HKR Rate	CTVLIWM Rate	G Rate	A Rate	FY Rate		
Tag 8974-1	EQQNEASEENNDQQSQEVPE	7.84	35.00	15	0	5	0	5	0	
Tag 8974-2	QQNEASEENNDQQSQEVPEK	6.39	30.00	15	5	5	0	5	0	
Tag 9333-1	MQEDEFDQGNQEQQEDNSNAE	15.37	40.00	15	0	5	5	5	5	
Tag 9333-2	SKMQEDEFDQGNQEQQEDNSN	9.62	35.00	15	5	5	5	0	5	
Tag 9333-3	QEDEFDQGNQEQQEDNSNAEMEEENASN	7.13	40.74	18.52	0	3.7	3.7	7.41	3.7	
Tag 10381-1	PSENENSQSEDSEVGNDNSEN	5.59	33.33	19.05	0	4.76	9.52	0	0	
Tag 11717-1	EVEESNPSAKEDSNPNSSGE	10.48	30.00	15	5	5	5	5	0	
Tag 11717-2	VEESNPSAKEDSNPNSSGED	12.10	30.00	15	5	5	5	5	0	
Tag 12237-1	KEENSESPLNENSDESYSEE	9.98	40.00	15	5	5	0	0	5	
Tag 12809-1	QPGPNHEEDADSYENMDNPDPD	28.15	35.00	15	5	5	5	5	5	
Tag 12809-2	PNHEEDADSYENMDNPDPGPD	20.94	40.00	15	5	5	5	5	5	
Tag 12809-3	NHEEDADSYENMDNPDPGPDP	23.41	40.00	15	5	5	5	5	5	
Tag 12885-1	DDPNSSDESNGNDDANSESD	14.62	40.00	20	0	0	5	5	0	
Tag 12885-2	NSSDESNGNDDANSESDNNNS	17.00	30.00	30	0	0	5	5	0	
Tag 12885-3	DESGNGNDANSESDNNSSRGD	13.93	31.82	22.73	4.55	0	9.09	4.55	0	
Tag 12968-1	DNNENAGEDGDNDFSPSDEEL	13.42	42.86	19.05	0	4.76	9.52	4.76	4.76	
Tag 12968-2	AEELEEDNNENAGEDGDNDFSPS	10.18	43.48	17.39	0	4.35	8.7	8.7	4.35	
Tag 12968-3	DNNENAGEDGDNDFSPSDEELAN	14.63	39.13	21.74	0	4.35	8.7	8.7	4.35	
Tag 13648-1	NPADDPNNQGEDFEEAEGVREEN	15.03	41.67	16.67	4.17	4.17	4.17	8.33	4.17	
Tag 14056-1	NEENTEPGAESSENADDPNKD	13.22	38.10	19.05	4.76	4.76	4.76	9.52	0	
Tag 14056-2	SSENADDPNKDTSENADGQSDEN	11.45	34.78	17.39	4.35	4.35	4.35	8.7	0	
Tag 14056-3	ESSENADDPNKDTSENADGGSDEN	10.38	37.50	16.67	4.17	4.17	4.17	8.33	0	
Tag 14681-1	DRDPMEMEEQPSSENDSQN	5.63	40.00	15	5	5	0	0	0	
Tag 14681-2	PEMENEEQPSSENDSQNQSG	11.50	30.00	15	0	5	5	0	0	
Tag 14681-3	ENEEQPSSENDSQNQSGEQI	17.15	30.00	15	0	5	5	0	0	
Tag 14844-1	DSESANVSDKEAGSNENDDQN	12.13	33.33	19.05	4.76	4.76	4.76	9.52	0	
Tag 15481-1	NYNDGSQEDRDWQDDQSDNQ	9.92	35.00	15	5	5	5	0	5	
Tag 16043-1	RENTNEASSEGNSDDSEDE	13.45	40.00	15	5	5	5	5	0	
Tag 16043-2	ENTNEASSEGNSDDSEDER	11.56	40.00	15	5	5	5	5	0	
Tag 16400-1	QENDNGNETESEQPKESNENQ	15.9	33.33	23.81	4.76	4.76	4.76	0	0	
Tag 16400-2	ENDNGNETESEQPKESNENG	18.1	38.1	23.81	4.76	4.76	4.76	0	0	
Tag 16400-3	QENDNGNETESEQPKESNENQE	8.84	36.36	22.73	4.55	4.55	4.55	0	0	
Tag 16417-1	QNEENPGDDEEAKNQVNSESDSDSEE	12.18	40.00	16	4	4	4	4	0	
Tag 16417-2	NEENPGDDEEAKNQVNSESDSDSEES	11.00	40.00	16	4	4	4	4	0	

TABLE 12-continued

Aggregation Rates of scFvs having Tags Extracted under Extraction Condition 1 Added										
Tag Name	Sequence	Aggregation Rate %								
		DE Rate	N Rate	HKR Rate	CTVLIWM Rate	G Rate	A Rate	FY Rate		
Tag 16417-3	QNEENPGDEEAKNQVNSESDDSEES	12.37	38.46	15.38	3.85	3.85	3.85	3.85	3.85	0
Tag 18137-1	YNGGNANPRPANNEEEEDEEDE	7.47	40.91	22.73	4.55	0	9.09	9.09	4.55	
Tag 18137-2	NGGNANPRPANNEEEEDEEDEY	8.17	40.91	22.73	4.55	0	9.09	9.09	4.55	
Tag 18137-3	NGGNANPRPANNEEEEDEEDEYD	9.86	43.48	21.74	4.35	0	8.7	8.7	4.35	
Tag 18347-1	GASENEEDDDYNKPLDPNS	16.06	40.00	15	5	5	5	5	5	
Tag 18478-1	ILQNQKEAEEPQGPDSENSQEN	16.82	30.00	15	5		5	5	0	
Tag 18478-2	QNGKEAEEPQGPDSENSQENP	14.77	30.00	15	5	0	5	5	0	
Tag 18478-3	NGKEAEEPQGPDSENSQENPP	12.74	30.00	15	5	0	5	5	0	
Tag 20166-1	KESVSPENNEEGGNDNQDNEN	16.95	33.33	28.57	4.76	4.76	9.52	0	0	
Tag 20166-2	ESVSPENNEEGGNDNQDNENP	13.67	33.33	28.57	0	4.76	9.52	0	0	
Tag 20166-3	KESVSPENNEEGGNDNQDNENP	26.56	31.82	27.27	4.55	4.55	9.09	0	0	
Tag 41693-1	ASPQPRESDDENSNDNSNEC	7.83	30.00	15	5	5	0	5	0	
Tag 55443-1	NNSQDEDGFQELNENGNAKDE	19.76	33.33	23.81	4.76	4.76	9.52	4.76	4.76	
Tag 55443-2	NSQDEDGFQELNENGNAKDEN	22.25	33.33	23.81	4.76	4.76	9.52	4.76	4.76	
Tag 55443-3	NNSQDEDGFQELNENGNAKDEN	18.69	31.82	27.27	4.55	4.55	9.09	4.55	4.55	

[0546] In Table 12, the N content ratio in the amino acid sequences of the extracted tags is 15% or more, and the content ratios of the other amino acids are as described above in Extraction Condition 1. In order to confirm that the aggregation inhibiting action does not depend on a specific amino acid sequence, tags were randomly selected from the tags extracted under Extraction Condition 1. As for some tags, an amino acid sequence satisfying the extraction condition was additionally selected from another portion of the same protein. The aggregation rates of scFvs tagged with the selected amino acid sequences were tested, and results as shown in Table 12 were obtained. As shown in Table 12, the aggregation rates of the tagged scFvs were low as a whole. Accordingly, it is obvious that the aggregation inhibiting action of a peptide tag largely depends on the amino acid content ratios thereof, and depends merely weakly on the specific amino acid sequence itself.

[0547] It is noted that human-derived amino acid sequences that can be extracted under Extraction Condition 1 were as follows.

TABLE 12-2

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 1	
105	TDNNESADDNNENPEDNNKN (Tag465-6)
106	DNNESADDNNENPEDNNKNT (Tag465-7)
107	NNESADDNNENPEDNNKNTD (Tag465-8)
108	NESADDNNENPEDNNKNTDD
109	ESADDNNENPEDNNKNTDDN
110	SADDNNENPEDNNKNTDDNE
111	DNNENPEDNNKNTDDNEENP
112	NNENPEDNNKNTDDNEENPN
113	NENPEDNNKNTDDNEENPN
114	ENPEDNNKNTDDNEENPNNN
115	NPEDNNKNTDDNEENPNNN
116	PEDNNKNTDDNEENPNNNN
117	DNNESADDNNENPEDNNKNTD
118	NNESADDNNENPEDNNKNTDD
119	NESADDNNENPEDNNKNTDDN
120	ESADDNNENPEDNNKNTDDNE (Tag 465-4)
121	SADDNNENPEDNNKNTDDNEE

TABLE 12-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 1	
122	ADDNNENPEDNNKNTDDNEEN
123	DDNNENPEDNNKNTDDNEENP
124	DNNENPEDNNKNTDDNEENPN
125	NNENPEDNNKNTDDNEENPNN (Tag 465-5)
126	NENPEDNNKNTDDNEENPNNN
127	ENPEDNNKNTDDNEENPNNNE
128	NPEDNNKNTDDNEENPNNNN
129	DNNESADDNNENPEDNNKNTDD
130	NNESADDNNENPEDNNKNTDDN
131	NESADDNNENPEDNNKNTDDNE
132	SADDNNENPEDNNKNTDDNEEN
133	ADDNNENPEDNNKNTDDNEENP
134	DDNNENPEDNNKNTDDNEENPN
135	DNNENPEDNNKNTDDNEENPNN
136	NNENPEDNNKNTDDNEENPNNN
137	NENPEDNNKNTDDNEENPNNNE
138	ENPEDNNKNTDDNEENPNNNN
139	DNNESADDNNENPEDNNKNTDDN
140	NNESADDNNENPEDNNKNTDDNE
141	NESADDNNENPEDNNKNTDDNEE
142	ESADDNNENPEDNNKNTDDNEEN
143	SADDNNENPEDNNKNTDDNEENP
144	ADDNNENPEDNNKNTDDNEENPN
145	DDNNENPEDNNKNTDDNEENPNN
146	DNNENPEDNNKNTDDNEENPNNN
147	NNENPEDNNKNTDDNEENPNNNE
148	NENPEDNNKNTDDNEENPNNNN
149	HGNQNQDSSSDNEADEASDDEDN
150	DNNESADDNNENPEDNNKNTDDNE
151	NNESADDNNENPEDNNKNTDDNEE
152	NESADDNNENPEDNNKNTDDNEEN
153	ESADDNNENPEDNNKNTDDNEENP
154	SADDNNENPEDNNKNTDDNEENPN
155	ADDNNENPEDNNKNTDDNEENPN
156	DDNNENPEDNNKNTDDNEENPNN
157	DNNENPEDNNKNTDDNEENPNNNE

TABLE 12-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 1	
158	NNENPEDNNKNTDDNEENPNNNN
159	SHGNQNQDSSSDNEADEASDDEDN
160	DNNESADDNNENPEDNNKNTDDNEE
161	NNESADDNNENPEDNNKNTDDNEEN
162	NESADDNNENPEDNNKNTDDNEENP
163	ESADDNNENPEDNNKNTDDNEENPN
164	SADDNNENPEDNNKNTDDNEENPNN
165	ADDNNENPEDNNKNTDDNEENPNNN
166	DDNNENPEDNNKNTDDNEENPNNNE
167	DNNENPEDNNKNTDDNEENPNNNN
168	GSHGNQNQDSSSDNEADEASDDEDN
169	SHGNQNQDSSSDNEADEASDDEDND
170	HGNQNQDSSSDNEADEASDDEDNDG
171	GNNQDSSSDNEADEASDDEDNDGN (Tag 465-2)
172	DNNESADDNNENPEDNNKNTDDNEEN
173	NNESADDNNENPEDNNKNTDDNEENP
174	NESADDNNENPEDNNKNTDDNEENPN
175	ESADDNNENPEDNNKNTDDNEENPNN
176	SADDNNENPEDNNKNTDDNEENPNNN
177	ADDNNENPEDNNKNTDDNEENPNNNE
178	DDNNENPEDNNKNTDDNEENPNNNN
179	WGSHGNQNQDSSSDNEADEASDDEDN
180	GSHGNQNQDSSSDNEADEASDDEDND
181	SHGNQNQDSSSDNEADEASDDEDNDG
182	HGNQNQDSSSDNEADEASDDEDNDGN
183	DNNESADDNNENPEDNNKNTDDNEENP
184	NNESADDNNENPEDNNKNTDDNEENPN
185	NESADDNNENPEDNNKNTDDNEENPNN
186	ESADDNNENPEDNNKNTDDNEENPNNN
187	SADDNNENPEDNNKNTDDNEENPNNNE
188	ADDNNENPEDNNKNTDDNEENPNNNN
189	SHGNQNQDSSSDNEADEASDDEDNDGN
190	HGNQNQDSSSDNEADEASDDEDNDGNE
191	DNNESADDNNENPEDNNKNTDDNEENPN
192	NNESADDNNENPEDNNKNTDDNEENPNN
193	NESADDNNENPEDNNKNTDDNEENPNNN

TABLE 12-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 1	
194	ESADDNNENPEDNNKNTDDNEENPNNE
195	SADDNNENPEDNNKNTDDNEENPNNNEN
196	SHGNNQDSSSDNEADEASDDEDNDGNE
197	DNNESADDNNENPEDNNKNTDDNEENPNNE
198	NNESADDNNENPEDNNKNTDDNEENPNNN
199	NESADDNNENPEDNNKNTDDNEENPNNE
200	ESADDNNENPEDNNKNTDDNEENPNNE
201	DNNESADDNNENPEDNNKNTDDNEENPNNN
202	NNESADDNNENPEDNNKNTDDNEENPNNE
203	NESADDNNENPEDNNKNTDDNEENPNNE
204	DNNESADDNNENPEDNNKNTDDNEENPNNE
205	NNESADDNNENPEDNNKNTDDNEENPNNE
206	FWGSHGNNQDSSSDNEADEASDDEDNDGNE (Tag465-9)
207	SHGNNQDSSSDNEADEASDDEDNDGNE
208	GDN
209	DNNESADDNNENPEDNNKNTDDNEENPNNE (18-1)
210	SHGNNQDSSSDNEADEASDDEDNDGNE
211	GDNE (Tag 465-3)
212	LEDNNEEPRDPQSPPDPPE (Tag656-1)
213	EDNNEEPRDPQSPPDPPEF (Tag656-2)
214	DNNEEPRDPQSPPDPPEFG (Tag656-3)
215	EGEQQLKPNNNSNAPNEDQEE
216	GEQQLKPNNNSNAPNEDQEEE
217	KPNNNSNAPNEDQEEEIQQSEQ (Tag1121-1)
218	PNNNSNAPNEDQEEEIQQSEQ
219	SEGEQQLKPNNNSNAPNEDQEEE
220	(Tag1121-3)
221	ENASSGDSEENTNSDHES (Tag7135-2)
222	ENASSGDSEENTNSDHESEQ (Tag7135-3)
223	ENASSGDSEENTNSDHESEQ (Tag7135-4)
224	ENASSGDSEENTNSDHESEQ (Tag7135-5)
225	ENASSGDSEENTNSDHESEQ (Tag7135-6)
226	ENASSGDSEENTNSDHESEQ (Tag7135-7)
227	ENASSGDSEENTNSDHESEQ (Tag7135-8)
228	ENASSGDSEENTNSDHESEQ (Tag7135-9)
229	ENASSGDSEENTNSDHESEQ (Tag7135-10)
230	ENASSGDSEENTNSDHESEQ (Tag7135-11)
231	ENASSGDSEENTNSDHESEQ (Tag7135-12)
232	ENASSGDSEENTNSDHESEQ (Tag7135-13)
233	ENASSGDSEENTNSDHESEQ (Tag7135-14)
234	ENASSGDSEENTNSDHESEQ (Tag7135-15)
235	ENASSGDSEENTNSDHESEQ (Tag7135-16)
236	ENASSGDSEENTNSDHESEQ (Tag7135-17)
237	ENASSGDSEENTNSDHESEQ (Tag7135-18)
238	ENASSGDSEENTNSDHESEQ (Tag7135-19)
239	ENASSGDSEENTNSDHESEQ (Tag7135-20)
240	ENASSGDSEENTNSDHESEQ (Tag7135-21)
241	ENASSGDSEENTNSDHESEQ (Tag7135-22)
242	ENASSGDSEENTNSDHESEQ (Tag7135-23)
243	ENASSGDSEENTNSDHESEQ (Tag7135-24)
244	ENASSGDSEENTNSDHESEQ (Tag7135-25)
245	ENASSGDSEENTNSDHESEQ (Tag7135-26)
246	ENASSGDSEENTNSDHESEQ (Tag7135-27)
247	ENASSGDSEENTNSDHESEQ (Tag7135-28)
248	ENASSGDSEENTNSDHESEQ (Tag7135-29)
249	ENASSGDSEENTNSDHESEQ (Tag7135-30)
250	ENASSGDSEENTNSDHESEQ (Tag7135-31)
251	ENASSGDSEENTNSDHESEQ (Tag7135-32)
252	ENASSGDSEENTNSDHESEQ (Tag7135-33)
253	ENASSGDSEENTNSDHESEQ (Tag7135-34)
254	ENASSGDSEENTNSDHESEQ (Tag7135-35)
255	ENASSGDSEENTNSDHESEQ (Tag7135-36)

TABLE 12-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 1	
224	LNFSDDDEQGSNSPKENNSED
225	NFSDDDEQGSNSPKENNSEDQ
226	EQLNFSDDEQGSNSPKENNSE
227	QLNFSDDEQGSNSPKENNSED
228	LNFSDDDEQGSNSPKENNSEDQ
229	NFSDDDEQGSNSPKENNSEDQG
230	EQLNFSDDEQGSNSPKENNSED
231	QLNFSDDEQGSNSPKENNSEDQ
232	LNFSDDDEQGSNSPKENNSEDQG
233	NFSDDDEQGSNSPKENNSEDQGS
234	EQLNFSDDEQGSNSPKENNSEDQ (Tag2408-1)
235	EQLNFSDDEQGSNSPKENNSEDQG (Tag2408-2)
236	EQLNFSDDEQGSNSPKENNSEDQGS (Tag2408-3)
237	EENENDENSLSSSDSSED (Tag6301-1)
238	NENDENSLSSSDSSEDKDE (Tag6301-1)
239	KETNNNSNAQNPSEEEGEGQDE (Tag6626-1)
240	ETNNNSNAQNPSEEEGEGQDED (Tag6626-2)
241	KETNNNSNAQNPSEEEGEGQDED (Tag6626-3)
242	ENANDSSDDSGETDESFPN (Tag6915-1)
243	DDNESNSEAENGWDGSNFSE
244	DNESNSEAENGWDGSNFSEE
245	SDDNESNSEAENGWDGSNFSE
246	DDNESNSEAENGWDGSNFSEE (Tag7128-1)
247	SDDNESNSEAENGWDGSNFSEE (Tag7128-2)
248	SSDDNESNSEAENGWDGSNFSEE (Tag7128-3)
249	SEENASSGDSEENTNSDHES (Tag7135-2)
250	EENASSGDSEENTNSDHESE (Tag 7315-1)
251	ENASSGDSEENTNSDHESEQ (Tag7135-3)
252	DENSENNWRNEYPEEESSDG
253	ENSENNWRNEYPEEESSDG
254	NSENNWRNEYPEEESSDGDE
255	DDENSENNWRNEYPEEESSDG (Tag8482-1)

TABLE 12-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 1	
256	DENSENNWRNEYYPEEEESSDGD
257	ENSENNWRNEYYPEEEESSDGD (Tag8482-2)
258	NSENNWRNEYYPEEEESSDGDDED
259	NSENNWRNEYYPEEEESSDGDDEDS (Tag8482-3)
260	ENSENNWRNEYYPEEEESSDGDDEDS
261	EQQNEASEENNDQQSQEVPE (Tag8974-1)
262	QQNEASEENNDQQSQEVPEK (Tag8974-2)
263	SKMQEDEFDQGNQEQQEDNSN (Tag9333-2)
264	KMQEDEFDQGNQEQQEDNSNA
265	MQEDEFDQGNQEQQEDNSNAE (Tag 9333-1)
266	QEDEFDQGNQEQQEDNSNAEM
267	FDQGNQEQQEDNSNAEMEEEEN
268	EFDQGNQEQQEDNSNAEMEEEEN
269	FDQGNQEQQEDNSNAEMEEEENA
270	DQGNQEQQEDNSNAEMEEEENAS
271	QGNQEQQEDNSNAEMEEEENASN
272	EFDQGNQEQQEDNSNAEMEEEENA
273	FDQGNQEQQEDNSNAEMEEEENAS
274	DQGNQEQQEDNSNAEMEEEENASN
275	DEFDQGNQEQQEDNSNAEMEEEENA
276	EFDQGNQEQQEDNSNAEMEEEENAS
277	FDQGNQEQQEDNSNAEMEEEENASN
278	DEFDQGNQEQQEDNSNAEMEEEENAS
279	EFDQGNQEQQEDNSNAEMEEEENASN
280	QEDEFDQGNQEQQEDNSNAEMEEEENA
281	EDEFDQGNQEQQEDNSNAEMEEEENAS
282	DEFDQGNQEQQEDNSNAEMEEEENASN
283	QEDEFDQGNQEQQEDNSNAEMEEEENAS
284	EDEFDQGNQEQQEDNSNAEMEEEENASN
285	QEDEFDQGNQEQQEDNSNAEMEEEENASN (Tag9333-3)
286	PSENENSQSEDVGKDNDSEN (Tag10381-1)
287	EVEESNPSAKEDSNPNSSGE (Tag11717-1)
288	VEESNPSAKEDSNPNSSGED (Tag11717-2)

TABLE 12-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 1	
289	KEENSESPLNENSDESYSEE (Tag12237-1)
290	QPQPNHEEDADSYENMDNP (Tag12809-1)
291	PNHEEDADSYENMDNPDGPD (Tag12809-2)
292	NHEEDADSYENMDNPDGDP (Tag12809-3)
293	DDPNSSDESNGNDDANSESD (Tag 12885-1)
294	DPNSSDESNGNDDANSESDN
295	PNSSDESNGNDDANSESDNN
296	NSSDESNGNDDANSESDNN (Tag 12885-2)
297	SSDESNGNDDANSESDNNSS
298	SDESNGNDDANSESDNNSS
299	DESNGNDDANSESDNNSSR
300	MQGDDPNSSDESNGNDDANSE
301	QGDDPNSSDESNGNDDANSES
302	GDDPNSSDESNGNDDANSESD
303	DDPNSSDESNGNDDANSESDN
304	DPNSSDESNGNDDANSESDNN
305	DKSMQGDDPNSSDESNGNDDAN
306	SMQGDDPNSSDESNGNDDANSE
307	MQGDDPNSSDESNGNDDANSES
308	QGDDPNSSDESNGNDDANSESD
309	GDDPNSSDESNGNDDANSESDN
310	DDPNSSDESNGNDDANSESDNN
311	DPNSSDESNGNDDANSESDNN
312	DESNGNDDANSESDNNSSRGD (Tag12885-3)
313	DDKSMQGDDPNSSDESNGNDDAN
314	DKSMQGDDPNSSDESNGNDDANS
315	KSMQGDDPNSSDESNGNDDANSE
316	SMQGDDPNSSDESNGNDDANSES
317	MQGDDPNSSDESNGNDDANSESD
318	QGDDPNSSDESNGNDDANSESDN
319	GDDPNSSDESNGNDDANSESDNN
320	DDPNSSDESNGNDDANSESDNN
321	DPNSSDESNGNDDANSESDNNSS
322	SDESNGNDDANSESDNNSSRGD

TABLE 12-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 1	
323	DESNGNDDANSESDNNSSSRGDA
324	DDANSESDNNSSSRGDASYNDE
325	FDDKSMQGDDPNSSDESNGNDDAN
326	DDKSMQGDDPNSSDESNGNDDANS
327	DKSMQGDDPNSSDESNGNDDANSE
328	SMQGDDPNSSDESNGNDDANSESD
329	MQGDDPNSSDESNGNDDANSESDN
330	QGDDPNSSDESNGNDDANSESDNN
331	GDDPNSSDESNGNDDANSESDNNS
332	DDPNSSDESNGNDDANSESDNNSS
333	DFDDKSMQGDDPNSSDESNGNDDAN
334	FDDKSMQGDDPNSSDESNGNDDANS
335	DDKSMQGDDPNSSDESNGNDDANSE
336	DKSMQGDDPNSSDESNGNDDANSES
337	KSMQGDDPNSSDESNGNDDANSESD
338	SMQGDDPNSSDESNGNDDANSESDN
339	MQGDDPNSSDESNGNDDANSESDNN
340	QGDDPNSSDESNGNDDANSESDNNS
341	GDDPNSSDESNGNDDANSESDNNSS
342	DDPNSSDESNGNDDANSESDNNSSS
343	DFDDKSMQGDDPNSSDESNGNDDANS
344	FDDKSMQGDDPNSSDESNGNDDANSE
345	DDKSMQGDDPNSSDESNGNDDANSES
346	DKSMQGDDPNSSDESNGNDDANSESD
347	KSMQGDDPNSSDESNGNDDANSESDN
348	SMQGDDPNSSDESNGNDDANSESDNN
349	MQGDDPNSSDESNGNDDANSESDNNS
350	QGDDPNSSDESNGNDDANSESDNNSS
351	GDDPNSSDESNGNDDANSESDNNSSS
352	DDPNSSDESNGNDDANSESDNNSSSR
353	DKSMQGDDPNSSDESNGNDDANSESDN
354	DDKSMQGDDPNSSDESNGNDDANSESDN
355	DKSMQGDDPNSSDESNGNDDANSESDN
356	DDPNSSDESNGNDDANSESDNNSSRGD
357	FDDKSMQGDDPNSSDESNGNDDANSESDN

TABLE 12-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 1	
358	DDKSMQGDDPNSSDESNGNDDANSESDNN
359	DKSMQGDDPNSSDESNGNDDANSESDN
360	DDPNSSDESNGNDDANSESDNNSSRGDA
361	DESNGNDDANSESDNNSSSRGDASYNDE
362	DFDDKSMQGDDPNSSDESNGNDDANSESDDN
363	FDDKSMQGDDPNSSDESNGNDDANSESDNN
364	DDKSMQGDDPNSSDESNGNDDANSESDNNS
365	DKSMQGDDPNSSDESNGNDDANSESDNNS
366	DDPNSSDESNGNDDANSESDNNSSRGDASAS
367	SDESNGNDDANSESDNNSSSRGDASYNDE
368	DESNGNDDANSESDNNSSSRGDASYNDES
369	DFDDKSMQGDDPNSSDESNGNDDANSESDDN
370	FDDKSMQGDDPNSSDESNGNDDANSESDNN
371	DDKSMQGDDPNSSDESNGNDDANSESDNNSS
372	DFDDKSMQGDDPNSSDESNGNDDANSESDDNNS
373	FDDKSMQGDDPNSSDESNGNDDANSESDNNSS
374	DDKSMQGDDPNSSDESNGNDDANSESDNNSSS
375	DFDDKSMQGDDPNSSDESNGNDDANSESDDNNS
376	FDDKSMQGDDPNSSDESNGNDDANSESDNNSSS
377	DFDDKSMQGDDPNSSDESNGNDDANSESDDNNS
378	DDPNSSDESNGNDDANSESDNNSSRGDASYNDE
379	GDDPNSSDESNGNDDANSESDDNSSSRGDASYNDE
380	DDPNSSDESNGNDDANSESDNNSSRGDASYNDES
381	DSYDFDDKSMQGDDPNSSDESNGNDDANSESDDNNSSRGDASYNDES

TABLE 12-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 1	
382	SYDFDDKSMQGDPNSSDESGNDDANS ESDNNSSSRGDA
383	YDFDDKSMQGDPNSSDESGNDDANSE SDNNSSSRGDAS
384	DFDDKSMQGDPNSSDESGNDDANSES DNNSSSRGDASY
385	QGDPNSSDESNGNDDANSES DNNSSSR GDASYN SDESKD
386	GDPNSSDESNGNDDANSES DNNSSSRG DASYN SDESKDN
387	DDPNSSDESNGNDDANSES DNNSSSRGD ASYN SDESKDNG
388	DDKSMQGDPNSSDESNGNDDANSES NNSSSRGDASYN SDE
389	DFDDKSMQGDPNSSDESNGNDDANSES DNNSSSRGDASYN SDE
390	FDDKSMQGDPNSSDESNGNDDANSES NNSSSRGDASYN SDE
391	DDKSMQGDPNSSDESNGNDDANSES NNSSSRGDASYN SDE
392	DFDDKSMQGDPNSSDESNGNDDANSES DNNSSSRGDASYN SDE
393	DFDDKSMQGDPNSSDESNGNDDANSES DNNSSSRGDASYN SDE
394	DDPNSSDESNGNDDANSES DNNSSSRGD ASYN SDESKDNGNGSDSKGAEDDDSDST SDTN
395	DPNSSDESNGNDDANSES DNNSSSRGD ASYN SDESKDNGNGSDSKGAEDDDSDSTS DTNN
396	NSSDESNGNDDANSES DNNSSSRGDASY NSDESKDNGNGSDSKGAEDDDSDSTS NNSD
397	SSDESNGNDDANSES DNNSSSRGDASY SDESKDNGNGSDSKGAEDDDSDSTS NSDS
398	SDESNGNDDANSES DNNSSSRGDASY DESKDNGNGSDSKGAEDDDSDSTS SDSN
399	AEDDDSDSTS DTNNSDSNGNGNNGNDDN DKSDSGKGKSDSSDSDSSDSSNSDSSD SSDS
400	EDDDSDSTS DTNNSDSNGNGNNGNDDN DKSDSGKGKSDSSDSDSSDSSNSDSSD SSDS
401	DDDSSTS DTNNSDSNGNGNNGNDDN KS DSGKGKSDSSDSDSSDSSNSDSSD SDSDS
402	GDPNSSDESNGNDDANSES DNNSSSR GDASYN SDESKDNGNGSDSKGAEDDDSDS TS DTNN

TABLE 12-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 1	
403	DDPNSSDESNGNDDANSES DNNSSSRGD ASYN SDESKDNGNGSDSKGAEDDDSDST SDTN
404	QGDPNSSDESNGNDDANSES DNNSSSR GDASYN SDESKDNGNGSDSKGAEDDDSD STS DTNN
405	GDPNSSDESNGNDDANSES DNNSSSRGD ASYN SDESKDNGNGSDSKGAEDDDSDS TS DTNN
406	DDPNSSDESNGNDDANSES DNNSSSRGD ASYN SDESKDNGNGSDSKGAEDDDSDST SDTN
407	DPNSSDESNGNDDANSES DNNSSSRGD ASYN SDESKDNGNGSDSKGAEDDDSDSTS DTNN
408	MQGDPNSSDESNGNDDANSES DNNSS SRGDASYN SDESKDNGNGSDSKGAEDDD SDSTS DTNN
409	QGDPNSSDESNGNDDANSES DNNSSSRGD ASYN SDESKDNGNGSDSKGAEDDDSD SDSTS DTNN
410	GDPNSSDESNGNDDANSES DNNSSSRGD ASYN SDESKDNGNGSDSKGAEDDDSDS TS DTNN
411	DDPNSSDESNGNDDANSES DNNSSSRGD ASYN SDESKDNGNGSDSKGAEDDDSDST SDTN
412	DPNSSDESNGNDDANSES DNNSSSRGD ASYN SDESKDNGNGSDSKGAEDDDSDSTS DTNN
413	GDPNSSDESNGNDDANSES DNNSSSRGD ASYN SDESKDNGNGSDSKGAEDDDSDS TS DTNN
414	DDPNSSDESNGNDDANSES DNNSSSRGD ASYN SDESKDNGNGSDSKGAEDDDSDST SDTN
415	QGDPNSSDESNGNDDANSES DNNSSSR GDASYN SDESKDNGNGSDSKGAEDDDSD STS DTNN
416	GDPNSSDESNGNDDANSES DNNSSSRGD ASYN SDESKDNGNGSDSKGAEDDDSDS TS DTNN
417	DDPNSSDESNGNDDANSES DNNSSSRGD ASYN SDESKDNGNGSDSKGAEDDDSDST SDTN
418	EDDDSDSTS DTNNSDSNGNGNNGNDDN DKSDSGKGKSDSSDSDSSDSSNSDSSD SSDS
419	MQGDPNSSDESNGNDDANSES DNNSS SRGDASYN SDESKDNGNGSDSKGAEDDD SDSTS DTNN
420	QGDPNSSDESNGNDDANSES DNNSSSRGD ASYN SDESKDNGNGSDSKGAEDDDSDS TS DTNN

TABLE 12-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 1	
421	GDDPNSSDESNGNDDANSES DNNN SSSRGD DAS YNSDES KDG NGN GSD SKGA EDDDS DS TSD TN NSDS N
422	DDPN SSDES NGN DDAN SES DNNN SSSRGD A S YNSDES KDG NGN GSD SKGA EDDDS DS TSD TN NSDS N
423	AEDDDSDSTS DTNN SDN SNGN GN GNN DNN DKSD SGK GKS DSD SD SSS N S S D S S S S D S D S S D S N
424	EDDDSDSTS DTNN SDN SNGN GN GNN DNN DKSD SGK GKS DSD SD SSS N S S D S S S S D S D S S D S N S
425	LEEDD NN ENA GED GD ND FSP S P S
426	DNN ENA GED GD ND FSP S D EEL (Tag 12968-1)
427	NN ENA GED GD ND FSP S D EEL A
428	NEN A GED GD ND FSP S D EEL AN
429	DNN ENA GED GD ND FSP S D EEL A
430	NN ENA GED GD ND FSP S D EEL AN
431	AE LEED D NN ENA GED GD ND FSP S (Tag 12968-2)
432	DDNN ENA GED GD ND FSP S D EEL A
433	DNN ENA GED GD ND FSP S D EEL AN (Tag 12968-3)
434	DDNN ENA GED GD ND FSP S D EEL AN
435	EDDN ENA GED GD ND FSP S D EEL AN
436	NPADD PNN QGEDEFEEAEQVREEN (Tag 13648-1)
437	NEEN TEP GAESSEN ADD PN KD (Tag 14056-1)
438	EN ADD PN KDT SE NADG QSD EN
439	SE NAD DPN KDT SE NADG QSD EN
440	SSEN ADD DPN KDT SE NADG QSD EN (Tag 14056-2)
441	ESSEN ADD DPN KDT SE NADG QSD EN (Tag 14056-3)
442	DRDPEMENE EQP SSEND S QN (Tag 14681- 1)
443	RDPEMENE EQP SSEND S QN Q
444	DPEMENE EQP SSEND S QN Q S
445	PEMENE EQP SSEND S QN Q SG (Tag 14681- 2)
446	EMENE EQP SSEND S QN Q SG E
447	MENE EQP SSEND S QN Q SG EQ

TABLE 12-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 1	
448	E NEE QP SSEND S QN Q SG EQ I (Tag 14681- 3)
449	DSE SAN VSD KEAG S NEN DD QN (Tag 14844-1)
450	NYND GS QED RDW QDD Q S D N Q (Tag 15481-1)
451	RENT NEASSEG N S D D S E D E (Tag 16043- 1)
452	ENT NEASSEG N S D D S E D E R (Tag 16043- 2)
453	QEND NGNETE SEQPKES NEN
454	END NGNETE SEQPKES NEN Q
455	NDNG NETE SEQPKES NEN Q E
456	QEND NGNETE SEQPKES NEN Q (Tag 16400-1)
457	END NGNETE SEQPKES NEN Q E (Tag 16400-2)
458	QEND NGNETE SEQPKES NEN Q E (Tag 16400-3)
459	QNEEN PGDEEA KNQVN S ESD
460	NEEN PGDEEA KNQVN S ESD S
461	EEN PGDEEA KNQVN S ESD S D
462	EN PGDEEA KNQVN S ESD S D S
463	NPGDEEA KNQVN S ESD S D S E
464	QNEEN PGDEEA KNQVN S ESD S
465	NEEN PGDEEA KNQVN S ESD S D
466	QNEEN PGDEEA KNQVN S ESD S D
467	NEEN PGDEEA KNQVN S ESD S D S
468	QNEEN PGDEEA KNQVN S ESD S D S
469	NEEN PGDEEA KNQVN S ESD S D S E
470	QNEEN PGDEEA KNQVN S ESD S D S E
471	NEEN PGDEEA KNQVN S ESD S D S E E
472	QNEEN PGDEEA KNQVN S ESD S D S E E (Tag 16417-1)
473	NEEN PGDEEA KNQVN S ESD S D S E E S (Tag 16417-2)
474	QNEEN PGDEEA KNQVN S ESD S D S E E S (Tag 16417-3)
475	YNGGNANPRPAN NEE EDEED
476	NGGNANPRPAN NEE EDEDE
477	GGNANPRPAN NEE EDEE DEY

TABLE 12-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 1	
478	YNGGNANPRPANNEEEDEEDE (Tag 18137-1)
479	NGGNANPRPANNEEEDEEDEY (Tag 18137-2)
480	NGGNANPRPANNEEEDEEDEYD (Tag 18137-3)
481	GASENEEDDDYNKPPLDPNS (Tag 18347-1)
482	LQNQKEAEEPGPDSENSQEN (Tag 18478-1)
483	QNQKEAEEPGPDSENSQENP (Tag 18478-2)
484	NQKEAEEPGPDSENSQENPP (Tag 18478-3)
485	KESVSPENNEEGGNDNQDNEN (Tag 20166-1)
486	ESVSPENNEEGGNDNQDNENP (Tag 20166-2)
487	KESVSPENNEEGGNDNQDNENP (Tag 20166-3)

TABLE 12-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 1	
488	ASPQPREPSSDDENSDNSNEC (Tag 41693-1)
489	NNSQDEDGFQELNENGNAKDE (Tag 55443-1)
490	NSQDEDGFQELNENGNAKDEN (Tag 55443-2)
491	NNSQDEDGFQELNENGNAKDEN (Tag 55443-3)

Extraction Condition 2:

- [0548] length: 20 to 70 amino acids
- [0549] group [D, E]: content of [30] or more
- [0550] group [D, E]: content of less than [45]
- [0551] group [P]: content of [10] or more
- [0552] group [H, K, R]: content of [5] or less
- [0553] group [G]: content of less than [10]
- [0554] group [A]: content of less than [10]
- [0555] group [C, T, V, L, I, M, W]: content of [0]
- [0556] group [F, Y]: content of [0]
- [0557] In the above-described extraction condition, the unit of each content is. The amino acids are described by one letter codes.

TABLE 13

Aggregation Rates of scFvs having Tags Extracted under Extraction Condition 2 Added							
Tag Name	Sequence	Aggregation Rate %	DE	P	HKR	G	A
			Rate	Rate	Rate	Rate	Rate
tag47-1	SSDSPKDQSPPEDSGESEAD	9.48	35	15	5	5	5
tag1784-1	IGPEPEPEPEPEPEPEPEPEPE	15.91	42.86	47.62	0	4.76	4.76
tag1784-2	PEPEPEPEPEPEPEPEPEPEPEP	14.03	42.86	52.38	0	0	4.76
tag1784-3	EPEPEPEPEPEPEPEPEPEPEPK	23.67	42.86	47.62	4.76	0	4.76
tag2257-1	PEEDGSPDPPEPSPEPEPKPS	22.33	35	40	5	5	0
tag2257-2	EDPEEDGSPDPPEPSPEPEPKP	14.76	42.86	38.1	4.76	4.76	0
tag2257-3	IDPEEDGSPDPPEPSPEPEPKPS	17.21	38.1	38.1	4.76	4.76	0
tag4398-1	PDDDGSDSSPPSASPAESEP	35.71	33.33	23.81	0	4.76	9.52
tag4398-2	DDDGSDSSPPSASPACSCPQ	18.64	33.33	19.05	0	4.76	9.52
tag4398-3	PDDDGSDSSPPSASPAESEPQ	18.58	31.82	22.73	0	4.55	9.09
tag4898-1	PPPEBQGQGDAPPHQHEDEEPA	9.9	33.33	28.57	4.76	9.52	9.52
tag5533-1	IGAPSPAPSPDSDSDSDGEE	19.43	33.33	19.05	0	9.52	9.52
tag5533-2	APSPAPSPDSDSDSDGEEE	16.43	38.1	19.05	0	4.76	9.52
tag5533-3	IPSPAPSPDSDSDSDGEEEE	8.76	42.86	19.05	0	4.76	4.76
tag5601-1	PAPPPPPPPEEDPEQDSGPED	33.03	30	50	0	5	5
tag5601-2	APPPPPPPEEDPEQDSGPED	20.10	35	45	0	5	5
tag5601-3	PAPPPPPPPEEDPEQDSGPED	24.34	33.33	47.62	0	4.76	4.76

TABLE 13-continued

Aggregation Rates of scFvs having Tags Extracted under Extraction Condition 2 Added							
Tag Name	Sequence	Aggregation Rate %	DE Rate	P Rate	HKR Rate	G Rate	A Rate
tag6354-1	PPPPPQAPPEEENESEPEEPS	7.87	33.33	42.86	0	0	4.76
tag6354-2	PPPPQAPPEEENESEPEEPSG	10.68	33.33	38.1	0	4.76	4.76
tag6354-3	PPPPPQAPPEEENESEPEEPSG	7.98	31.82	40.91	0	4.55	4.55
tag6681-1	PQDSSSKSPEPSADESPDND	8.07	30	20	5	0	5
tag7124-1	IQSSDNSEDEEEPPDNADSKS	5.56	40	10	5	0	5
tag7702-1	EEEEQPGKAPDPQDPQDAESD	13.15	42.86	19.05	4.76	4.76	9.52
tag7702-2	EEEQPGKAPDPQDPQDAESDS	17.29	38.1	19.05	4.76	4.76	9.52
tag7702-3	EEEEQPGKAPDPQDPQDAESDS	17.57	40.91	18.18	4.55	4.55	9.09
tag8341-1	PPPSEEEGPEEPPKASPESE	23.65	35	35	5	5	5
tag8341-2	PPPSEEEGPEEPPKASPESE	23.58	33.33	38.1	1.76	1.76	4.76
tag8341-3	PPPSEEEGPEEPPKASPESEA	24.91	33.33	33.33	4.76	4.76	9.52
tag10102-1	DDAAEPESPSSPRSPSPEP	19.00	30	45	5	0	5
tag11508-1	SGEASSSEEPPSPDDKENQA	14.91	33.33	14.29	4.76	4.76	9.52
tag11508-2	GEASSSEEPPSPDDKENQAP	14.49	33.33	19.05	4.76	4.76	9.52
tag11508-3	SGEASSSEEPPSPDDKENQAP	13.03	31.82	18.18	4.55	4.55	9.09
tag13088-1	PPPPPEESSDSEPEAEPG	14.98	30	45	0	5	5
tag13088-2	PPPPPLESSDSEPLAEPGSP	16.42	30	40	0	5	5
tag13088-3	PPPEESSDSEPEAEPGSPQK	20.58	30	30	5	5	5
tag13619-1	SDPEPPDAGEDSKSENGENAP	11.52	33.33	19.05	4.76	9.52	9.52
tag14205-1	SDSESEDPPRNOASDSENNE	10.73	40	10	5	0	5
tag14666-1	GPGEDAEPDEDPQSEDSEAPS	10.21	42.86	19.05	0	9.52	9.52
tag14666-2	PGEDAEPDEDPQSEDSEAPSS	10.04	42.86	19.05	0	4.76	9.52
tag14666-3	GPGEDAEPDEDPOSEDSEAPSS	12.08	40.91	18.18	0	9.09	9.09
tag15430-1	ESESSSSDSEANEPSQSASPE	14.06	30.43	13.04	0	0	8.7
tag15430-2	SDSESESSSSDSEANEPSQSASPE	10.87	30.77	11.54	0	0	7.69
tag15430-3	DSESESSSSDSEANEPSQSASPE	7.54	30.77	15.38	0	0	7.69
tag16604-1	HQEDSEEEQEEAEAGASEPPP	12.72	43.48	17.39	4.35	4.35	8.7
tag16604-2	GHQEDSEEEQEEAEAGASEPPP	12.95	44	16	1	8	8
tag17053-1	PSQPPEEPDEAESPDPO	15.35	35	35	0	0	5
tag17053-2	PAPSQPPEEPDEAESPDPO	16.53	33.33	38.1	0	0	9.52
tag17053-3	DPAPSQPPEEPDEAESPDPO	14.03	36.36	36.36	0	0	9.09
tag17170	PGSQPQASSGPEAAAAEDDE	11.11	42.86	14.29	0	9.52	9.52
tag17514-1	SPDSQEEQKGESSASSPEEP	13.45	30	15	5	5	5

TABLE 13-continued

Aggregation Rates of scFvs having Tags Extracted under Extraction Condition 2 Added							
Tag Name	Sequence	Aggregation Rate %	DE Rate	P Rate	HKR Rate	G Rate	A Rate
tag17514-2PADSPDSQEEQKGESSIONSPEEP		15.02	30.43	17.39	4.35	4.35	8.7
tag17514-3ADSPDSQLEQKGESSIONSPEPLEP		16.14	34.78	13.04	4.35	4.35	8.7
tag17603-1PSPCDCSSSSSSSSCDCD		6.01	35	10	0	0	0
tag17603-2PRSPSPEDDESSSSSSSEDEE		5.96	30.43	13.04	4.35		0
tag17603-3PRSPSPEDDESSSSSSSEDEEE		7.22	33.33	12.5	4.17	0	0
tag18253-1SSSDSSSDSSEDDCAPSKP		3.59	35	10	5	0	5
tag18453-1PSPGSPRGQPQDODDDDEEE		6.12	42.86	19.05	4.76	9.52	0
tag18467-1PAGDGEAGPQQAEHDHPQNPPEDPNQDPPEDD		9.87	32.26	25.81	3.23	9.68	9.68
tag18467-2AGDGEAGPQQAEHDHPQNPPEDPNQDPPEDD		5.73	32.26	22.58	3.23	9.68	9.68
tag18467-3PAGDGEAGPQQAEHDHPQNPPEDPNQDPPEDD		8.73	31.25	25	3.13	9.38	9.38
tag18478-1QNQKEAEEPGPDSENSQENP		14.94	30	15	5	5	5
tag18478-2NQKEAEEPGPDSENSQENPP		11.47	30	20	5	5	5
tag19033-1DQNESQSPQEPEEGPSSEDDKA		12.69	38.1	14.29	4.76	4.76	4.76
tag19033-2QNESQSPQEPEEGPSSEDDKAE		12.4	38.1	14.29	4.76	4.76	4.76
tag19033-3NESQSPQEPEEGPSSEDDKAEG		12.28	38.1	14.29	4.76	9.52	4.76
tag29487-1PASSSSNPEEGPEEDREAESE		12.77	38.1	14.29	4.76	4.76	9.52
tag34831-1DKPEEEDEAQQPQGPQSGPEEEAE		7.58	43.48	17.39	4.35	4.35	8.7
tag34831-2KPEEEDEAQQPQGPQSGPEEEAE		8.02	43.48	17.39	4.35	4.35	8.7
tag34831-3PEEEDEAQQPQGPQSGPEEEAE		9.14	43.48	17.39	0	8.7	8.7
tag34858-1PEEEHAPGEDESSPQPSQPS		17.52	30	25	5	5	5

[0558] In Table 13, the P content ratio in the amino acid sequences of the extracted tags is 15% or more, and the content ratios of the other amino acids are as described above in Extraction Condition 2. In order to confirm that the aggregation inhibiting action does not depend on a specific amino acid sequence, tags were randomly selected from the tags extracted under Extraction Condition 2. As for some tags, an amino acid sequence satisfying the extraction condition was additionally selected from another portion of the same protein. The aggregation rates of scFvs tagged with the selected amino acid sequences were tested, and results as shown in Table 13 were obtained. As shown in Table 13, the aggregation rates of the tagged scFvs were low as a whole.

[0559] It is noted that human-derived amino acid sequences that can be extracted under Extraction Condition 2 were as follows.

TABLE 13-2

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 2		
SEQ ID NO:	Sequence	
492	SSDSPKDQSPPEDSGESEAD	(tag47-1)
493	PGPEPEPEPEPEPEPEPEPE	
494	GPEPEPEPEPEPEPEPEPEP	
495	PEPEPEPEPEPEPEPEPEPK	
496	EPEPEPEPEPEPEPEPEPKP	
497	PEPEPEPEPEPEPEPEPKPG	

TABLE 13-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 2	
SEQ ID NO:	Sequence
498	AGPGPEPEPEPEPEPEPAPPEP
499	GPGPEPEPEPEPEPEPAPPEPE
500	PGPEPEPEPEPEPEPAPPEPEP
501	GPEPEPEPEPEPEPEPAPPEPEPE (tag1784-1)
502	PEPEPEPEPEPEPAPPEPEPEP (tag1784-2)
503	EPEPEPEPEPEPAPPEPEPEPK (tag1784-3)
504	PEPEPEPEPEPAPPEPEPEPKP
505	EPEPEPEPEPAPPEPEPEPKPG
506	PEPEPEPEPAPPEPEPEPKPG
507	EPEPEPEPAPPEPEPEPKPGAG
508	AGPGPEPEPEPEPEPEPAPPEPE
509	GPGPEPEPEPEPEPEPAPPEPEP
510	PGPEPEPEPEPEPEPAPPEPEPE
511	GPEPEPEPEPEPEPAPPEPEPEP
512	PEPEPEPEPEPEPAPPEPEPEPK
513	EPEPEPEPEPEPAPPEPEPEPK
514	PEPEPEPEPEPAPPEPEPEPKPG
515	EPEPEPEPEPAPPEPEPEPKPG
516	PEPEPEPEPAPPEPEPEPKPGAG
517	AGPGPEPEPEPEPEPEPAPPEPEP
518	GPGPEPEPEPEPEPEPAPPEPEPE
519	PGPEPEPEPEPEPAPPEPEPEP
520	GPEPEPEPEPEPEPAPPEPEPEPK
521	PEPEPEPEPEPEPAPPEPEPEPKP
522	EPEPEPEPEPEPAPPEPEPEPKPG
523	PEPEPEPEPEPAPPEPEPEPKPG
524	EPEPEPEPEPEPAPPEPEPEPKPGAG
525	AGPGPEPEPEPEPEPEPAPPEPEPE
526	GPGPEPEPEPEPEPEPAPPEPEPEP
527	PGPEPEPEPEPEPAPPEPEPEPK
528	GPEPEPEPEPEPEPAPPEPEPEPKP
529	PEPEPEPEPEPEPAPPEPEPEPKPG
530	EPEPEPEPEPEPAPPEPEPEPKPG
531	PEPEPEPEPEPAPPEPEPEPKPGAG

TABLE 13-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 2	
SEQ ID NO:	Sequence
532	AGPGPEPEPEPEPEPEPAPPEPEPEP
533	GPGPEPEPEPEPEPEPAPPEPEPEPK
534	PGPEPEPEPEPEPEPAPPEPEPEPKP
535	GPEPEPEPEPEPEPAPPEPEPEPKPG
536	PEPEPEPEPEPEPAPPEPEPEPKPG
537	EPEPEPEPEPEPAPPEPEPEPKPGAG
538	AGPGPEPEPEPEPEPEPAPPEPEPEPK
539	GPGPEPEPEPEPEPEPAPPEPEPEPKP
540	PGPEPEPEPEPEPEPAPPEPEPEPKP
541	GPEPEPEPEPEPEPAPPEPEPEPKPG
542	PEPEPEPEPEPEPAPPEPEPEPKPGAG
543	AGPGPEPEPEPEPEPEPAPPEPEPEPKP
544	PGPEPEPEPEPEPEPAPPEPEPEPKPG
545	DPEEDGSPDPEPSPEPEPKP
546	PEEDGSPDPEPSPEPEPKPS (tag2257-1)
547	EDPEEDGSPDPEPSPEPEPKP (tag2257-2)
548	DPEEDGSPDPEPSPEPEPKPS (tag2257-3)
549	EDPEEDGSPDPEPSPEPEPKPS
550	DEDPEEDGSPDPEPSPEPEPKPS
551	PDDGSDSSPPSASPAESEP (tag4398-1)
552	DDDGSDDSSPPSASPAESEPQ (tag4398-2)
553	PDDGSDSSPPSASPAESEPQ (tag4398-3)
554	PPPEEQGQGDAPPQHEDEEPA (tag4898-1)
555	PSPAPSPDSDSDSDGEEE
556	GAPSPAPSPDSDSDSDSDGEEE (tag5533-1)
557	APSPAPSPDSDSDSDSDGEEE (tag5533-2)
558	PSPAPSPDSDSDSDSDGEEE (tag5533-3)
559	NGAPSPAPSPDSDSDSDSDGEE
560	GAPSPAPSPDSDSDSDSDGEEE
561	APSPAPSPDSDSDSDSDGEEEE
562	QNGAPSPAPSPDSDSDSDGEE
563	NGAPSPAPSPDSDSDSDSDGEEE
564	GAPSPAPSPDSDSDSDSDGEEEE
565	APSPAPSPDSDSDSDSDGEEEEEE

TABLE 13-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 2	
SEQ ID NO:	Sequence
566	QNGAPSPAPSPDSDSDSDGEEE
567	NGAPSPAPSPDSDSDSDGEEEE
568	GAPSPAPSPDSDSDSDGEEEEEE
569	QNGAPSPAPSPDSDSDSDGEEEEEE
570	NGAPSPAPSPDSDSDSDGEEEEEE
571	GAPSPAPSPDSDSDSDGEEEEEE
572	QNGAPSPAPSPDSDSDSDGEEEEEE
573	NGAPSPAPSPDSDSDSDGEEEEEE
574	QNGAPSPAPSPDSDSDSDGEEEEEE
575	NGAPSPAPSPDSDSDSDGEEEEEE
576	QNGAPSPAPSPDSDSDSDGEEEEEE
577	QNGAPSPAPSPDSDSDSDGEEEEEE GER
578	QNGAPSPAPSPDSDSDSDGEEEEEE GERD
579	EPPAPPPPPPPEEDPEQDSG
580	PAPPPPPPPEEDPEQDSGPE (tag5601-1)
581	APPAPPAPPPEEDPEQDSGPED (tag5601-2)
582	PAPPAPPAPPPEEDPEQDSGPED (tag5601-3)
583	EPPAPPPPPPPEEDPEQDSGPE
584	PPAPPPPPPPEEDPEQDSGPED
585	AEPPAPPPPPPPEEDPEQDSGPE
586	EPPAPPPPPPPEEDPEQDSGPED
587	AEPPAPPPPPPPEEDPEQDSGPED
588	PPPPPQAPPEEEENESEPEEP
589	PPPPQAPPEEEENESEPEEPS
590	PPPQAPPEEEENESEPEEPSG
591	PPPPPQAPPEEEENESEPEEPS (tag6354-1)
592	PPPPQAPPEEEENESEPEEPSG (tag6354-2)
593	PPPPPQAPPEEEENESEPEEPSG (tag6354-3)
594	PQDSSSKSPEPSADESPDND (tag6681-1)
595	QSSDNSEDEEPPDNADSKS (tag7124-1)
596	EEEEQPGKAPDPQDPQDAES
597	EEEEQPGKAPDPQDPQDAESD (tag7702-1)

TABLE 13-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 2	
SEQ ID NO:	Sequence
598	EEEQPGKAPDPQDPQDAESDS (tag7702-2)
599	EEEEQPGKAPDPQDPQDAESDS (tag7702-3)
600	EEEEEEQPGKAPDPQDPQDAESDS
601	PPPPPSEEEGPEEPPKASPE
602	PPPPSEEEGPEEPPKASPE
603	PPPSEEEGPEEPPKASPESE (tag8341-1)
604	PPPPSEEEGPEEPPKASPESE (tag8341-2)
605	PPPSEEEGPEEPPKASPESEA (tag8341-3)
606	PPPPPSEEEGPEEPPKASPESE
607	PPPPSEEEGPEEPPKASPESEA
608	PPPPPSEEEGPEEPPKASPESEA
609	DDAEEPESPSSPRSPSPEP (tag10102-1)
610	SGEASSSEEPPSPDDKENQ
611	SGEASSSEEPPSPDDKENQA (tag11508-1)
612	GEASSSEEPPSPDDKENQAP (tag11508-2)
613	SGEASSSEEPPSPDDKENQAP (tag11508-3)
614	PQPPPPPPPEESSDSEPEAE
615	QPPPPPPPEESSDSEPEAE
616	PPPPPPPEESSDSEPEAE
617	PPPPPEESSDSEPEAE
618	PPPPPEESSDSEPEAE
619	PPPPPEESSDSEPEAE
620	PPPEESSDSEPEAE
621	SDPEPPDAGEDSKSEGENAP (tag13619-1)
622	SDSESEDPPRNQASDSENEE (tag14205-1)
623	GPGEDAEPDEDPQSEDSEAPS (tag14666-1)
624	PGEDAEPDEDPQSEDSEAPS (tag14666-2)
625	GPGEDAEPDEDPQSEDSEAPS (tag14666-3)
626	ESESSSSDSEANEPEQSASPEPE (tag15430-1)
627	DSESESSSSDSEANEPEQSASPEPE

TABLE 13-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 2	
SEQ ID NO:	Sequence
628	SDSESESSSSDSEANEPEQSASAPEPE (tag15430-2)
629	DSESESSSSDSEANEPEQSASPEPEP (tag15430-3)
630	HQEDSEEEQEEEAEAGASEPPPP (tag16604-1)
631	GDHQEDSEEEQEEEAEAGASEPPPP (tag16604-2)
632	PSQPPEEPEPDEAESSPDPQ (tag17053-1)
633	DPAPSQPPEEPEPDEAESSPD
634	PAPSQPPEEPEPDEAESSPDP (tag17053-2)
635	APSQPPEEPEPDEAESSPDPQ
636	PSQPPEEPEPDEAESSPDPQA
637	DPAPSQPPEEPEPDEAESSPDP (tag17053-3)
638	PAPSQPPEEPEPDEAESSPDPQ
639	DPAPSQPPEEPEPDEAESSPDPQ
640	PGSQPQASSGPEAEAAAEDDE (tag17170)
641	DSPDSQEEQKGEASSASSPSEE
642	SPDSQEEQKGEASSASSPEEP (tag17514-1)
643	PDSQEEQKGEASSASSPEEPE
644	DSQEEQKGEASSASSPEEPE
645	DSPDSQEEQKGEASSASSPEEP
646	SPDSQEEQKGEASSASSPEEPE
647	PDSQEEQKGEASSASSPEEPE
648	PADSPDSQEEQKGEASSASSPSEE
649	ADSPDSQEEQKGEASSASSPEEP
650	DSPDSQEEQKGEASSASSPEEPE
651	SPDSQEEQKGEASSASSPEEPE
652	PADSPDSQEEQKGEASSASSPEEP (tag17514-2)
653	ADSPDSQEEQKGEASSASSPEEPE (tag17514-3)
654	DSPDSQEEQKGEASSASSPEEPE
655	PADSPDSQEEQKGEASSASSPEEPE
656	ADSPDSQEEQKGEASSASSPEEPEE
657	PADSPDSQEEQKGEASSASSPEEPEE

TABLE 13-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 2	
SEQ ID NO:	Sequence
658	SPSPDEDESSSSSSSSSEDE
659	PSPEDEDESSSSSSSSSEDEE (tag17603-1)
660	PRSPSPDEDESSSSSSSSSEDEE (tag17603-2)
661	PRSPSPDEDESSSSSSSSSEDEEE (tag17603-3)
662	SSSDSSSDSSEDDEAPSKP (tag18253-1)
663	PSPGSPRGQPQDQDDDEDDE (tag18453-1)
664	QQAEDHPQNPPEDPNQDPPE
665	QAEDHPQNPPEDPNQDPPED
666	AEDHPQNPPEDPNQDPPEDD
667	EDHPQNPPEDPNQDPPEDDS
668	QQAEDHPQNPPEDPNQDPPEDD
669	QAEDHPQNPPEDPNQDPPEDD
670	AEDHPQNPPEDPNQDPPEDDS
671	PQQAEDHPQNPPEDPNQDPPEDD
672	QQAEDHPQNPPEDPNQDPPEDD
673	QAEDHPQNPPEDPNQDPPEDDS
674	DGEAGPQQAEDHPQNPPEDPNQD
675	GPQQAEDHPQNPPEDPNQDPPEDD
676	PQQAEDHPQNPPEDPNQDPPEDD
677	QQAEDHPQNPPEDPNQDPPEDDS
678	GPQQAEDHPQNPPEDPNQDPPEDD
679	PQQAEDHPQNPPEDPNQDPPEDDS
680	EAGPQQAEDHPQNPPEDPNQDPPEDD
681	AGPQQAEDHPQNPPEDPNQDPPEDD
682	GPQQAEDHPQNPPEDPNQDPPEDDS
683	DGEAGPQQAEDHPQNPPEDPNQDPPE
684	GEAGPQQAEDHPQNPPEDPNQDPPEDD
685	EAGPQQAEDHPQNPPEDPNQDPPEDD
686	AGPQQAEDHPQNPPEDPNQDPPEDDS
687	DGEAGPQQAEDHPQNPPEDPNQDPPEDD
688	GEAGPQQAEDHPQNPPEDPNQDPPEDD
689	EAGPQQAEDHPQNPPEDPNQDPPEDDS
690	DGEAGPQQAEDHPQNPPEDPNQDPPEDD

TABLE 13-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 2	
SEQ ID NO:	Sequence
691	GEAGPQQAEQEDHPQNPPEDPNQDPPEDDS
692	DGEAGPQQAEQEDHPQNPPEDPNQDPPEDDS
693	PAGDGEAGPQQAEQEDHPQNPPEDPNQDPPEDD (tag18467-1)
694	AGDGEAGPQQAEQEDHPQNPPEDPNQDPPEDDS (tag18467-2)
695	PAGDGEAGPQQAEQEDHPQNPPEDPNQDPPEDDS (tag18467-3)
696	QNQKEAEEPGPDSENSQENPP (tag18478-1)
697	NQKEAEEPGPDSENSQENPP (tag18478-2)
698	DQNESQSPQEPEEGPSEDDK
699	QNESQSPQEPEEGPSEDDKA
700	NESQSPQEPEEGPSEDDKAE
701	DQNESQSPQEPEEGPSEDDKA (tag19033-1)
702	QNESQSPQEPEEGPSEDDKAE (tag19033-2)
703	NESQSPQEPEEGPSEDDKAE (tag19033-3)
704	ESQSPQEPEEGPSEDDKAE
705	SQSPQEPEEGPSEDDKAE
706	DQNESQSPQEPEEGPSEDDKAE
707	QNESQSPQEPEEGPSEDDKAE
708	NESQSPQEPEEGPSEDDKAE
709	DQNESQSPQEPEEGPSEDDKAE
710	QNESQSPQEPEEGPSEDDKAE
711	NESQSPQEPEEGPSEDDKAE
712	DQNESQSPQEPEEGPSEDDKAE
713	QNESQSPQEPEEGPSEDDKAE
714	DQNESQSPQEPEEGPSEDDKAE
715	QNESQSPQEPEEGPSEDDKAE
716	XEASSSEEPPSPDDKENQAP (tag25919-1)
717	PASSSSNPPEGPEEDREAESE (tag29487-1)
718	DKPEEEEDDEAQQPQPQSGPE
719	KPEEEEDDEAQQPQPQSGPE

TABLE 13-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 2	
SEQ ID NO:	Sequence
720	AGEGDKPEEEEDDEAQQPQPQSQS
721	EGDKPEEEEDDEAQQPQPQSGP
722	GDKPEEEEDDEAQQPQPQSGP
723	DKPEEEEDDEAQQPQPQSGP
724	KPEEEEDDEAQQPQPQSGP
725	PEEEEDDEAQQPQPQSGP
726	EEDDEAQQPQPQSGP
727	EDDEAQQPQPQSGP
728	DDEAQQPQPQSGP
729	DEAQQPQPQSGP
730	EAQQPQPQSGP
731	QQPQPQSGP
732	QPQPQSGP
733	EGDKPEEEEDDEAQQPQPQSGP
734	GDKPEEEEDDEAQQPQPQSGP
735	DKPEEEEDDEAQQPQPQSGP
736	KPEEEEDDEAQQPQPQSGP
737	QQPQPQSGP
738	EGDKPEEEEDDEAQQPQPQSGP
739	GDKPEEEEDDEAQQPQPQSGP
740	DKPEEEEDDEAQQPQPQSGP
741	KPEEEEDDEAQQPQPQSGP
742	PEEEEDDEAQQPQPQSGP
743	EGDKPEEEEDDEAQQPQPQSGP
744	GDKPEEEEDDEAQQPQPQSGP
745	KPEEEEDDEAQQPQPQSGP
746	EGDKPEEEEDDEAQQPQPQSGP
747	GDKPEEEEDDEAQQPQPQSGP
748	DKPEEEEDDEAQQPQPQSGP
749	KPEEEEDDEAQQPQPQSGP
750	PEEEHAPGEDESSPQPSQPS (tag34858-1)

TABLE 13-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 2	
SEQ ID NO:	Sequence
751	XPPPEESSDSEPEAEPGSPQ (tag)
752	PPPEESSDSEPEAEPGSPQK (tag)

Extraction Condition 3:

- [0560] length: 20 to 70 amino acids
- [0561] group [D, E]: content of [45] or more
- [0562] group [G]: content of less than [10]
- [0563] group [A]: content of less than [10]
- [0564] group [F, Y]: content of [0]
- [0565] group [C, M, L, I, W, T, V]: content of [0]
- [0566] group [P]: content of [15] or more

TABLE 14

Aggregation Rates of scFvs having Tags Extracted under Extraction Condition 3 Added						
Tag Name	Sequence	Aggregation Rate %	DE Rate	P Rate	G Rate	A Rate
Tag 167-1	KEPKEEKKDDDEEAKPKSSD	8.02	45	15	0	5
Tag 1034-1	SEEEKPPEEDKEEEEKKAP	8.91	55	15	0	5
Tag 1409-1	PAEEDEDDEPQEKEAGEPGRP	5.17	47.62	19.05	9.52	9.52
Tag 1784-1	EPEPEPEPEPEPEPEPEPEP	12.97	45	50	0	5
Tag 2257-1	EDPEEDGSPDPEPSPEPEPK	12.61	45	35	5	0
Tag 2257-2	DEDPEEDGSPDPEPSPEPEPK	11.50	47.62	33.33	4.76	0
Tag 2257-3	DEDPEEDGSPDPEPSPEPEPKP	8.62	45.45	36.36	4.55	0
Tag 2740-1	KPEDKDPDRDPEESKEPKEEK	13.22	45	20	0	0
Tag 2740-2	PEDKDPDRDPEESKEPKEEKQ	11.68	45	20	0	0
Tag 2740-3	EDKDPRDPEESKEPKEEKQR	8.79	45	15	0	0
Tag 3227-1	KRNDSEEEERERDEEQEPPP	20.63	50	15	0	0
Tag 4898-1	PEEEPDQDAPDEHEPSPSED	10.71	52.38	23.81	0	4.76
Tag 4898-2	EEEPDDQDAPDEHEPSPSED	7.01	52.38	19.05	0	9.52
Tag 4898-3	EEPDDQDAPDEHEPSPSEDAP	19.90	47.62	23.81	0	9.52
Tag 5533-1	PSPAPS PDS DSD SD GEEEEEEE	1.89	50	16.67	4.17	4.17
Tag 5533-2	APSPAPS PDS DSD SD GEEEEEEE	1.00	48	16	4	8
Tag 5533-3	PSPAPS PDS DSD SD GEEEEEEE	0.90	48	16	8	4
Tag 6236-1	EKNDEDEPQK PEDKG DPEGPE	10.19	47.62	19.05	9.52	0
Tag 6236-2	EKNDEDEPQK PEDKG DPEGPEA	8.40	45.45	18.18	9.09	4.55
Tag 6755-1	EDEEEEEEEEEEDEGPAPP	4.10	75	15	5	5
Tag 6755-2	DEEEEEEEEEEDEGPAPP	2.38	70	15	5	5
Tag 7167-1	GEREPDP D PDR DAS DGE DE KP	12.37	47.62	19.05	9.52	4.76
Tag 7167-2	EREPDP D PDR DAS DGE DE KP	8.77	47.62	23.81	4.76	4.76
Tag 7167-3	EGERE PD P D PDR DAS DGE DE KP	9.97	50	18.18	9.09	4.55
Tag 7702-1	EEEEEE QPG KAP DP QDP QDA ESD	13.30	45.45	18.18	4.55	9.09
Tag 8243-1	EPEEK Q EPEEK Q EPEEK Q KPE	11.90	47.62	19.05	0	0
Tag 8243-2	QEPEEK Q EPEEK Q EPEEK Q KPE	12.44	45.45	18.18	0	0

TABLE 14-continued

Aggregation Rates of scFvs having Tags Extracted under Extraction Condition 3 Added						
Tag Name	Sequence	Aggregation Rate %	DE Rate	P Rate	G Rate	A Rate
Tag 8243-3	EPEEKQEPEEKQEPPEEKQKPEA	12.60	45.45	18.18	0	4.55
Tag 8818-1	DDDDDDDDSPDPESPDDSESD	2.58	65	15	0	0
Tag 8818-2	DDDDDDSPDPESPDDSESDS	2.19	60	15	0	0
Tag 8818-3	DDSPDPESPDDSESDSESEK	8.33	50	15	0	0
Tag 9050-1	DPDQPREDPAEEEKEEKDAPE	15.60	52.38	19.05	0	9.52
Tag 9166-1	PENESEPCKHEEPKPEEKPEEE	17.13	50	22.73	0	0
Tag 9166-2	NESEPCKHEEPKPEEKPEEEEK	7.62	50	18.18	0	0
Tag 9166-3	PENESEPCKHEEPKPEEKPEEEE	13.00	52.17	21.74	0	0
Tag 9590-1	PPEEDPEEQAEENPEGEQPE	13.99	50	25	5	5
Tag 9590-2	KPPEEDPEEQAEENPEGEQPE	10.18	47.62	23.81	4.76	4.76
Tag 9590-3	KPPEEDPEEQAEENPEGEQPEE	9.67	50	22.73	4.55	4.55
Tag 9704-1	PDDDESEDHDDPDNAHESP	7.65	55	15	0	5
Tag 9749-1	PEPEPEPEPEPEPEPEPEPE	15.74	50	45	0	0
Tag 9749-2	GGEPEPEPEPEPEPEPEPEPE	12.24	47.83	39.13	8.7	0
Tag 9749-3	GGEPEPEPEPEPEPEPEPEPE	14.29	48	40	8	0
Tag 10346-1	PEEEPDQDAPDEHESPPPE	8.16	50	30	0	5
Tag 10346-2	EFEPDDQDAPDEHESPPPEDA	6.88	52.38	23.81	0	9.52
Tag 10346-3	PEEEPDQDAPDEHESPPPEDAP	8.97	47.83	30.43	0	8.7
Tag 11099-1	GPSSDDENEEESKPEKEDEP	7.21	50	15	5	0
Tag 11099-2	PSSDDENEEESKPEKEDEPQ	5.09	50	15	0	0
Tag 12127-1	SDDSDSEKRRPEEQEEEPQP	17.51	45	15	0	0
Tag 12127-2	DDSDSEKRRPEEQEEEPQPR	17.44	45	15	0	0
Tag 13036-1	PEEEDEEPGDPREGEEEEEDEPDP	3.17	68	20	8	0
Tag 13036-2	PEEEDEEPGDPREGEEEEEDEPPEAP	5.60	64.29	21.43	7.14	3.57
Tag 13036-3	AAPEEEDEEPGDPREGEEEEEDEPPEAPENG	6.19	55.88	17.65	8.82	8.82
Tag 14128-1	PPPSEGSDEEEEEDEEDEE	2.43	70	15	5	0
Tag 14128-2	KPPPSEGSDEEEEEDEEDEEERKP	10.86	60	16	4	0
Tag 14128-3	KPPPSEGSDEEEEEDEEDEEERKPQ	9.76	57.69	15.38	3.85	0
Tag 16549-1	EEEEEEEEEEEEEEAPP	4.95	75	20	0	5
Tag 16549-2	EEEEEEEEEEEEEEAPP	3.01	73.9	17.39	0	4.35
Tag 16549-3	PDDDEEDEEEEEEEEEEEAPP	0.89	81.82	15.15	0	3.03
Tag 16604-1	EDSEEEQSQQAAEGASEPPP	6.00	47.62	19.05	4.76	9.52
Tag 16604-2	QEDSEEEQSQQAAEGASEPPP	7.39	45.45	18.18	4.55	9.09
Tag 16604-3	DHQEDSEEEQSQQAAEGASEPPP	10.59	45.83	16.67	4.17	8.33
Tag 16741-1	DQSEEEEEEKHPKPAKPE	10.62	45	20	0	5

TABLE 14-continued

Aggregation Rates of scFvs having Tags Extracted under Extraction Condition 3 Added						
Tag Name	Sequence	Aggregation Rate %	DE Rate	P Rate	G Rate	A Rate
Tag 16991-1	PAPAHRPPEDEGEENEGEDE	14.97	47.62	19.05	9.52	9.52
Tag 16991-2	PAPAHRPPEDEGEENEGEDEE	7.12	50	18.18	9.09	9.09
Tag 17199-1	QENGQREEEEEKEPEAEP	9.09	50	15	5	5
Tag 17199-2	PAEGQENGQREEEEEKEPEAEP	10.29	45.83	16.67	8.33	8.33
Tag 17936-1	EQEPEPEPEPEPEPEPEPEP	6.37	50	45	0	0
Tag 17936-2	EPEPEPEPEPEPEPEPEPEQ	10.94	50	45	0	0
Tag 17936-3	EQEPEPEPEPEPEPEPEPEQ	16.61	50	40.91	0	0
Tag 18132-1	PEEEEEEEEEEPEASPPERK	3.36	60	20	0	5
Tag 18453-1	PSPGSPRGQPQDQDDDEDDEED	4.31	45.45	18.18	9.09	0
Tag 18453-2	RSPGSPRGQPQDQDDDEDDEDE	13.65	45.83	16.67	8.33	0
Tag 18453-3	PSPGSPRGQPQDQDDDEDDEDEA	6.54	45.83	16.67	8.33	4.17
Tag 18866-1	AEDDDEEDEEEEEEPDPDP	2.24	80	15	0	5
Tag 18866-2	EDDDEEDEEEEEEPDPDP	2.29	85	15	0	0
Tag 19350-1	KQEPPDPEEDKEENKDDSAS	14.73	45	15	0	5
Tag 19350-2	QEPPDPEEDKEENKDDSASK	14.12	45	15	0	5
Tag 19511-1	EDEDEDESSEEDSEDEEPPP	1.70	70	15	0	0
Tag 19511-2	PDDSRDEDEDEDESSEEDSEDEEPPP	4.37	65.38	15.38	0	0
Tag 19511-3	PKKEPDDSRDEDEDEDESSEEDSEDEEPPPKR	3.87	54.55	15.15	0	0
Tag 22900-1	PEEEAAEEEEEEERPKPSRP	10.93	52.38	19.05	0	9.52
Tag 22900-2	EQPEEEAAEEEEEEERPKPSRP	6.90	52.17	17.39	0	8.7
Tag 22900-3	EEEQPEEEAAEEEEEEERPKPSRP	9.67	56	16	0	8
Tag 34831-1	PEEEDEAQQPQPQSGPEEAEE	10.16	45.45	18.18	4.55	9.09
Tag 34831-2	EGDKPEEEDEAQQPQPQSGPEEEAEE	10.78	46.15	15.38	7.69	7.69
Tag 34831-3	KPEEEDEAQQPQPQSGPEEEAEGEEEAAERGP	8.33	45.45	15.15	9.09	9.09
Tag 39056-1	NNSEEEEDDDDEEEEPDKPP	0.44	85	15	0	0
Tag 39056-2	NSEEEEDDDDEEEEPDKPPA	0.99	65	15	0	5
Tag 39056-3	SEEEEDDDDEEEEPDKPPAN	1.23	65	15	0	5

[0567] In Table 14, the acidic amino acid ratio in the amino acid sequences of the extracted tags is 45% or more, the P content ratio is 15% or more, and the content ratios of the other amino acids are as described above in Extraction Condition 3. In order to confirm that the aggregation inhibiting action does not depend on a specific amino acid sequence, tags were randomly selected from the tags extracted under Extraction Condition 3. As for some tags, an amino acid sequence satisfying the extraction condition was additionally selected from another portion of the same

protein. The aggregation rates of scFvs tagged with the selected amino acid sequences were tested, and results as shown in Table 14 were obtained. As shown in Table 14, the aggregation rates of the tagged scFvs were low as a whole. It is understood, from Table 14-1, that a strategy of reducing the G, A, F, Y, C, M, L, I, W, T, and V contents and increasing the P content works on a peptide tag having a high acidic amino acid content.

[0568] It is noted that human-derived amino acid sequences that can be extracted under Extraction Condition 3 were as follows.

TABLE 14-2

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 3	
SEQ ID NO:	Sequence
753	KEPKEEKKKDDDEEAKPKSSD (Tag 167-1)
754	SEEEEKPPEDKEEEEEKKAP (Tag 1034-1)
755	PAEEEDDDPEQEKEAGEPGRP (Tag 1409-1)
756	PEPEPEPEPEPEPEPEPEPE
757	EPEPEPEPEPEPEAPEPEPEP (Tag 1784-1)
758	DEDPEEDGSPDPEPSPEPEP
759	EDPEEDGSPDPEPSPEPEPK (Tag 2257-1)
760	DEDPEEDGSPDPEPSPEPEPK (Tag 2257-2)
761	DEDPEEDGSPDPEPSPEPEPKP (Tag 2257-3)
762	KPEDKDPDRPEESKEPKEEK (Tag 2740-1)
763	PEDKDPDRPEESKEPKEEKQ (Tag 2740-2)
764	EDKDPDRPEESKEPKEEKQR (Tag 2740-3)
765	KRNDSEEEERERDEEQEPPP (Tag 3227-1)
766	PEEEPDQDAPDEHEPSPSE
767	EEEPDQDAPDEHEPSPSED
768	PEEEPDQDAPDEHEPSPSED (Tag 4898-1)
769	EEEPDQDAPDEHEPSPEDA (Tag 4898-2)
770	EEPDDQDAPDEHEPSPSEDAP (Tag 4898-3)
771	PEEEPDQDAPDEHEPSPEDA
772	EEEPDQDAPDEHEPSPSEDAP
773	PEEEPDQDAPDEHEPSPSEDAP
774	SPAPSPDSDDSDGDGEAAA
775	PAPSPDSDDSDGDGEAAA
776	PSPAPSPDSDDSDGDGEAAA
777	PSPAPSPDSDDSDGDGEAAA
778	APSPAPSPDSDDSDGDGEAAA
779	PSPAPSPDSDDSDGDGEAAA (Tag 5533-1)
780	APSPAPSPDSDDSDGDGEAAA (Tag 5533-2)

TABLE 14-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 3	
SEQ ID NO:	Sequence
781	PSPAPSPDSDDSDGDGEAAAAA (Tag 5533-3)
782	GAPSPAPSPDSDDSDGDGEAAA
783	APSPAPSPDSDDSDGDGEAAA
784	PSPAPSPDSDDSDGDGEAAA
785	EKNDEDEPQKPEDKGDPEGPE (Tag 6236-1)
786	EKNDEDEPQKPEDKGDPEGPEA (Tag 6236-2)
787	EDEEEEEEEEEEDEGPAPP (Tag 6755-1)
788	DEEEEEEEEEEDEGPAPPS (Tag 6755-2)
789	EREPDPPDDRDASDGEDEKPP
790	REPDPDPPDDRDASDGEDEKPP
791	GEREPDPPDDRDASDGEDEKPP (Tag 7167-1)
792	EREPDPPDDRDASDGEDEKPP (Tag 7167-2)
793	EGEREPDPPDDRDASDGEDEKPP (Tag 7167-3)
794	GEREPDPPDDRDASDGEDEKPP
795	EGEREPDPPDDRDASDGEDEKPP
796	EEEEQPGKAPDPQDPQDAESD (Tag 7702-1)
797	QEPEEKQOEPEEKQOEPEEKQK
798	EPEEKQOEPEEKQOEPEEKQKP
799	PEEKQOEPEEKQOEPEEKQKPE
800	EEKQOEPEEKQOEPEEKQKPEA
801	EPEEKQOEPEEKQOEPEEKQKPE (Tag 8243-1)
802	QEPEEKQOEPEEKQOEPEEKQKPE (Tag 8243-2)
803	EPEEKQOEPEEKQOEPEEKQKPEA (Tag 8243-3)
804	AGDDDDDDDDSPDPESPDDDS
805	GDDDDDDDDSPDPESPDDSE
806	DDDDDDDDSPDPESPDDSES
807	DDDDDDDDSPDPESPDDSES (Tag 8818-1)
808	DDDDDDSPDPESPDDSES (Tag 8818-2)
809	DDDDDSSPDPESPDDSES

TABLE 14-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 3	
SEQ ID NO:	Sequence
810	DDDDSPDPESPDDSESDSES
811	DDDSPDPESPDDSESDSESE
812	DDSPDPESPDDSESDSESEK (Tag 8818-3)
813	DSPDPESPDDSESDSESEKE
814	SPDPESPDDSESDSESEKEE
815	PDPESPDDSESDSESEKEES
816	DPDQPREDPAEEEKEEKDAPE (Tag 9050-1)
817	EGKPENESEPCKHEEEPCKPSEE
818	PENESEPCKHEEEPCKPSEEKPE
819	ENESEPCKHEEEPCKPSEEKPEE
820	NESEPCKHEEEPCKPSEEKPEEE
821	ESEPCKHEEEPCKPSEEKPEEEE
822	SEPCKHEEEPCKPSEEKPEEEK
823	PENESEPCKHEEEPCKPSEEKPEEE
824	ENESEPCKHEEEPCKPSEEKPEEE
825	NESEPCKHEEEPCKPSEEKPEEEE
826	ESEPCKHEEEPCKPSEEKPEEEK
827	KPENESEPCKHEEEPCKPSEEKPEE
828	PENESEPCKHEEEPCKPSEEKPEEE (Tag 9166-1)
829	ENESEPCKHEEEPCKPSEEKPEEEE
830	NESEPCKHEEEPCKPSEEKPEEEK (Tag 9166-2)
831	KPENESEPCKHEEEPCKPSEEKPEEE
832	PENESEPCKHEEEPCKPSEEKPEEEE (Tag 9166-3)
833	ENESEPCKHEEEPCKPSEEKPEEEK
834	EGKPENESEPCKHEEEPCKPSEEKPEE
835	GKPENESEPCKHEEEPCKPSEEKPEEE
836	KPENESEPCKHEEEPCKPSEEKPEEEE
837	PENESEPCKHEEEPCKPSEEKPEEEK
838	EGKPENESEPCKHEEEPCKPSEEKPEEE
839	GKPENESEPCKHEEEPCKPSEEKPEEEE
840	KPENESEPCKHEEEPCKPSEEKPEEEK
841	EGKPENESEPCKHEEEPCKPSEEKPEEEE

TABLE 14-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 3	
SEQ ID NO:	Sequence
842	GKPENESEPCKHEEEPCKPSEEKPEEEK
843	EGKPENESEPCKHEEEPCKPSEEKPEEEK
844	KPPEEDPEEQAEENPEGEQP
845	PPEEDPEEQAEENPEGEQP (Tag 9590-1)
846	PEEDPEEQAEENPEGEQPEE
847	KPPEEDPEEQAEENPEGEQP (Tag 9590-2)
848	PPEEDPEEQAEENPEGEQPEE
849	KPPEEDPEEQAEENPEGEQP (Tag 9590-3)
850	PDDDDSESHHDDPDNAHESP (Tag 9704-1)
851	GEPEPEPEPEPEPEPESEPE
852	EPEPEPEPEPEPEPESEPE
853	PEPEPEPEPEPEPESEPE
854	EPEPEPEPEPEPESEPE
855	PEPEPEPEPEPESEPE (Tag 9749-1)
856	GGEPEPEPEPEPEPESEPE
857	GEPEPEPEPEPEPESEPE
858	EPEPEPEPEPEPESEPE
859	PEPEPEPEPEPEPESEPE
860	EPEPEPEPEPEPESEPE
861	GGEPEPEPEPEPEPESEPE
862	GEPEPEPEPEPEPESEPE
863	EPEPEPEPEPEPESEPE
864	PEPEPEPEPEPESEPE
865	GGEPEPEPEPEPEPESEPE (Tag 9749-2)
866	GEPEPEPEPEPEPESEPE
867	EPEPEPEPEPEPESEPE
868	GGEPEPEPEPEPEPESEPE
869	GEPEPEPEPEPEPESEPE
870	GGEPEPEPEPEPEPESEPE (Tag 9749-3)
871	PEEEPPDDQQDAPDEHESPPPE (Tag 10346-1)
872	EEEPPDDQQDAPDEHESPPPE

TABLE 14-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 3	
SEQ ID NO:	Sequence
873	PEEEPDQDAPDEHESPPPED
874	EEEPDDQDAPDEHESPPPEDA (Tag 10346-2)
875	EEPDDQDAPDEHESPPPEDAP
876	PEEEPDQDAPDEHESPPPEDA
877	EEEPDDQDAPDEHESPPPEDAP
878	PEEEPDQDAPDEHESPPPEDAP (Tag 10346-3)
879	GPSSDDNEEESKPEKEDEP (Tag 11099-1)
880	PSSDDNEEESKPEKEDEPQ (Tag 11099-2)
881	SDDSDSEKRRPEEQEEEPQP (Tag 12127-1)
882	DDSDSEKRRPEEQEEEPQP (Tag 12127-2)
883	DPREGEEEEEDEPDPEAPE
884	PREGEEEEEDEPDPEAPEN
885	DEEPGDPREGEEEEEDEPDPD
886	EEPGDPREGEEEEEDEPDPE
887	EPGDPREGEEEEEDEPDPEA
888	PGDPREGEEEEEDEPDPEAP
889	GDPREGEEEEEDEPDPEAPE
890	DPRGEEEEEDEPDPEAPEN
891	PREGEEEEEDEPDPEAPENG
892	EDEEPGDPREGEEEEDEPDPEP
893	DEEPGDPREGEEEEEDEPDPE
894	EEPGDPREGEEEEEDEPDPEA
895	EPGDPREGEEEEEDEPDPEAP
896	PGDPREGEEEEEDEPDPEAPE
897	GDPREGEEEEEDEPDPEAPEN
898	DPRGEEEEEDEPDPEAPENG
899	PREGEEEEEDEPDPEAPENG
900	PEEEDEEPGDPREGEEEEDEPDPE
901	EEDEEPGDPREGEEEEDEPDPEP
902	EDEEPGDPREGEEEEDEPDPE
903	DEEPGDPREGEEEEDEPDPEA

TABLE 14-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 3	
SEQ ID NO:	Sequence
904	EEPGDPREGEEEEEDEPDPEAP
905	EPGDPREGEEEEEDEPDPEAPE
906	PGDPREGEEEEEDEPDPEAPEN
907	DPRGEEEEEDEPDPEAPENG
908	APEEEDEPGDPREGEEEEDEP
909	PEEEDEPGDPREGEEEEDEPD
910	EEEDEEPGDPREGEEEEDEPD
911	EDEEPGDPREGEEEEDEPDPE
912	EDEEPGDPREGEEEEDEPDPEA
913	DEEPGDPREGEEEEEDEPDPEAP
914	EEPGDPREGEEEEEDEPDPEAPE
915	EPGDPREGEEEEEDEPDPEAPEN
916	AAPEEEDEEPGDPREGEEEEDEP
917	APEEEDEEPGDPREGEEEEDEPD
918	PEEEDEEPGDPREGEEEEDEPD (Tag 13036-1)
919	EEEDEEPGDPREGEEEEDEPDPE
920	EDEEPGDPREGEEEEDEPDPEA
921	EDEEPGDPREGEEEEDEPDPEAP
922	DEEPGDPREGEEEEDEPDPEAPE
923	EEPGDPREGEEEEDEPDPEAPEN
924	AAPEEEDEEPGDPREGEEEEDEPD
925	APEEEDEEPGDPREGEEEEDEPD
926	PEEEDEEPGDPREGEEEEDEPDPE
927	EEEDEEPGDPREGEEEEDEPDPEA
928	EDEEPGDPREGEEEEDEPDPEAP
929	EDEEPGDPREGEEEEDEPDPEAPE
930	DEEPGDPREGEEEEDEPDPEAPEN
931	AAPEEEDEEPGDPREGEEEEDEPD
932	APEEEDEEPGDPREGEEEEDEPDPE
933	PEEEDEEPGDPREGEEEEDEPDPEA
934	EEDEEPGDPREGEEEEDEPDPEAP
935	EDEEPGDPREGEEEEDEPDPEAPE
936	EDEEPGDPREGEEEEDEPDPEAPEN
937	AAPEEEDEEPGDPREGEEEEDEPDPE

TABLE 14-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 3	
SEQ ID NO:	Sequence
938	APEEEDEEPGDPREGEAAAAAEDEPDPEA
939	PEEEDEEPGDPREGEAAAAAEDEPDPEAP (Tag 13036-2)
940	EEEDEEPGDPREGEAAAAAEDEPDPEAPE
941	EEDEEPGDPREGEAAAAAEDEPDPEAPEN
942	APEEEDEEPGDPREGEAAAAAEDEPDPEAP
943	PEEEDEEPGDPREGEAAAAAEDEPDPEAP
944	EEEDEEPGDPREGEAAAAAEDEPDPEAPE
945	APEEEDEEPGDPREGEAAAAAEDEPDPEAP
946	PEEEDEEPGDPREGEAAAAAEDEPDPEAPEN
947	PRGAAPEEEDEEPGDPREGEAAAAAEDEPDPE
948	RGAAAPEEEDEEPGDPREGEAAAAAEDEPDPE
949	AAPEEEDEEPGDPREGEAAAAAEDEPDPEAPE
950	APEEEDEEPGDPREGEAAAAAEDEPDPEAPEN
951	PEEEDEEPGDPREGEAAAAAEDEPDPEAPENG
952	EEEDEEPGDPREGEAAAAAEDEPDPEAPE
953	PRGAAPEEEDEEPGDPREGEAAAAAEDEPDPE
954	AAPEEEDEEPGDPREGEAAAAAEDEPDPEAPEN
955	APEEEDEEPGDPREGEAAAAAEDEPDPEAPENG
956	PEEEDEEPGDPREGEAAAAAEDEPDPEAPENG
957	AAPEEEDEEPGDPREGEAAAAAEDEPDPEAPENG
958	APEEEDEEPGDPREGEAAAAAEDEPDPEAPENG
959	AAPEEEDEEPGDPREGEAAAAAEDEPDPEAPENG (Tag 13036-3)
960	KPPSEGSDEEEEEEDEEDE
961	PPPSEGSDEEEEEEDEEDE (Tag 14128-1)
962	PPPSEGSDEEEEEEDEEDEERKP

TABLE 14-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 3	
SEQ ID NO:	Sequence
963	KPPSEGSDEEEEEEDEEDEERKP (Tag 14128-2)
964	PPPSEGSDEEEEEEDEEDEERKPQ
965	KPPSEGSDEEEEEEDEEDEERKPQ (Tag 14128-3)
966	EEEEEEEEEEEEEEAPP
967	EEEEEEEEEEEEEEAPP (Tag 16549-1)
968	EEEEEEEEEEEEEEAPP
969	EEEEEEEEEEEEEEAPP
970	EEEEEEEEEEEEEEAPP
971	EEEEEEEEEEEEEEAPP
972	EEEEEEEEEEEEEEAPP
973	EEEEEEEEEEEEEEAPP
974	EEEEEEEEEEEEEEAPP (Tag 16549-2)
975	EEEEEEEEEEEEEEAPP
976	EEEEEEEEEEEEEEAPP
977	EEEEEEEEEEEEEEAPP
978	EEEEEEEEEEEEEEAPP
979	EEEEEEEEEEEEEEAPP
980	EEEEEEEEEEEEEEAPP
981	PDDDEEDEEEEEEEEEEEAPP (Tag 16549-3)
982	EDSEESQEEEAEGASEAPP (Tag 16604-1)
983	QEDSEESQEEEAEGASEAPP (Tag 16604-2)
984	DHQEDSEESQEEEAEGASEAPP (Tag 16604-3)
985	DQSEEEEEEKHPKPAKPE (Tag 16741-1)
986	PAPAHRPPEDEGEENECEEDE (Tag 16991-1)
987	PAPAHRPPEDEGEENECEEDE (Tag 16991-2)
988	QENGQREEEEEKEPEAEPP (Tag 17199-1)
989	PAEGQENGQREEEEEKEPEAEPP (Tag 17199-2)

TABLE 14-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 3		
SEQ ID NO:	Sequence	
990	EQEPEPEPEPEPEPEPEPEP (Tag 17936-1)	
991	QEPEPEPEPEPEPEPEPEPE	
992	EPEPEPEPEPEPEPEPEPEQ (Tag 17936-2)	
993	EQEPEPEPEPEPEPEPEPEPE	
994	QEPEPEPEPEPEPEPEPEQ	
995	EQEPEPEPEPEPEPEPEPEQ (Tag 17936-3)	
996	PEEEEEEEEEEPPPERK (Tag 18132-1)	
997	PSPGSPRGQPQDQDDDEDDEED (Tag 18453-1)	
998	PSPGSPRGQPQDQDDDEDDEDE	
999	RSPGSPRGQPQDQDDDEDDEDE (Tag 18453-2)	
1000	PSPGSPRGQPQDQDDDEDDEDEA (Tag 18453-3)	
1001	AEDDDEEDEEEEEEPPDPDP (Tag 18866-1)	
1002	EDDDEEDEEEEEEPPDPDP (Tag 18866-2)	
1003	KQEPPDPEEDKEENKDDSAS (Tag 19350-1)	
1004	QEPPDPEEDKEENKDDSASK (Tag 19350-2)	
1005	EDEDEDESSEEDSEDEEPPP (Tag 19511-1)	
1006	DEDEDESSEEDSEDEEPPP	
1007	EDEDESSEEDSEDEEPPP	
1008	DEDESSEEDSEDEEPPP	
1009	PDDSRDEDEDEDESSEEDSEDEEPPP (Tag 19511-2)	
1010	PKKEPDDSRDEDEDEDESSEEDSEDEEPP	
1011	PKKEPDDSRDEDEDEDESSEEDSEDEEPP	
1012	PKKEPDDSRDEDEDEDESSEEDSEDEEPP	
1013	PKKEPDDSRDEDEDEDESSEEDSEDEEPP	
1014	PEEEAAEEEEEEERPKPSRP (Tag 22900-1)	

TABLE 14-2-continued

Examples of human-derived amino acid sequences that can be extracted under Extraction Condition 3		
SEQ ID NO:	Sequence	
1015	QPEEEAAEEEEEEERPKPSRP	
1016	EQPEEEAAEEEEEEERPKPSRP (Tag 22900-2)	
1017	EEQPEEEAAEEEEEEERPKPSRP	
1018	EEEQPEEEAAEEEEEEERPKPSRP (Tag 22900-3)	
1019	PEEEEDDEAQQPQPQSGPEEAEE (Tag 34831-1)	
1020	DKPEEEEDDEAQQPQPQSGPEEAEE	
1021	PEEEEDDEAQQPQPQSGPEEAEEGE	
1022	PEEEEDDEAQQPQPQSGPEEAEEGEE	
1023	EGDKPEEEDEAQQPQPQSGPEEAEE (Tag 34831-2)	
1024	DKPEEEEDDEAQQPQPQSGPEEAEEGE	
1025	KPEEEEDDEAQQPQPQSGPEEAEEGEE	
1026	PEEEEDDEAQQPQPQSGPEEAEEGEEE	
1027	PEEEDEAQQPQPQSGPEEAEEGEEE	
1028	KPEEEEDDEAQQPQPQSGPEEAEEGEEE	
1029	EEAERGP (Tag 34831-3)	
1030	NNSEEEDDDDEEEEPDKPP (Tag 39056-1)	
1031	NSEEEDDDDEEEEPDKPPA (Tag 39056-2)	
	SEEEDDDDEEEEPDKPPAN (Tag 39056-3)	

[0569] As shown by these examples, the protein aggregation rate reducing action of a peptide tag had low dependency on a specific amino acid sequence, also had low dependency on a protein from which it is derived, but had high dependency on amino acid contents.

Example 6: Addition to Various Proteins and Aggregation Inhibiting Action

[0570] In the above Example, Y13-259 was used as the scFv. In this example, a VHH antibody (a heavy chain variable domain of a heavy chain antibody) was used. As the VHH antibody, iDab #6 binding to Ras was used. As the tag, Tag4-8 was used. The other conditions were the same as those employed in Example 1. As a result, the aggregation rate of the VHH antibody not having the tag was 57.89%, the aggregation rate of the VHH antibody having Tag4-8 at the C terminal was 8.77%, and thus, a strong aggregation inhibiting action was exhibited by the tag addition.

[0571] When SHSY5Y (human dopamine-like cell) that is a neuroblastoma cell line was used as the cell, scFv-6E (6E)

was used as the scFv, and Tag4-8 or Tag18-1 was added as the tag to the C terminal and the N terminal, the aggregation rate of the Tag4-8 added scFv was 0.96%, the aggregation rate of the Tag18-1 added scFv was 1.14%, and thus, it was revealed that the tag addition makes a contribution to the low aggregation rate.

[0572] When D4 binding to botulinum toxin type A (SEQ ID NO: 1032: QVQLQQSGGGLVQPGGLSLRLSCAASGFTLDYYAIGWFRQAPGKEREGLVLCISSLGGSTNYADSVKGRFTISRDNAKNTVYLQMNSLKPED-TAVYYCAADDLRCGSNWSSYFRGS WGQGTQTVTSS) was used as the VHH antibody, and the Tag4-8 was added as the tag to the C terminal of the VHH antibody, the aggregation rate of the Tag4-8 added D4 was 7.8%, and the aggregation rate of D4 not having the tag was 81.5%. This result reveals that the tag addition makes a contribution to the reduction of the aggregation rate of the VHH antibody.

Example 7: Action Enhancement by Stabilization of Intracellular Antibody (Effect of Enhancement of Antibody Action on Amyloid Accumulation)

[0573] It is known that a central nervous system disease is caused by accumulation of amyloid in a nerve cell. When human α -synuclein fibril is extracellularly introduced into a nerve cell, synuclein fibril is formed with synuclein having a normal structure in the cell involved. When GFP-tagged synuclein has been expressed in the cell, the synuclein forms synuclein fibril together with the introduced human α -synuclein fibril, and the thus formed synuclein fibril can be observed with fluorescence of GFP. In this example, GFP-tagged synuclein was expressed in SHSY-5Y cell, and α -synuclein fibril (Cosmo Bio Co., Ltd., SYNO3) was extracellularly introduced into the SHSY-5Y cell. In this example, the ability of an antibody to reduce the synuclein fibril was tested. Specifically, as the antibody, scFv-6E binding to fibrillized synuclein (SEQ ID NO: 1033: AEVQLLESGGGLVQPGGLSLRLSCAASGFTFSSYAM-SWVRQAPGKGLEWVSYIASGGD TTNY-ADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYY-CAKGASAFDYWGQGTLVTTSSGGGGGGGGSGGGSTDIQMTQSPSSL-SASVGDRVTITCRASQSISSYLNWYQQ KPGKAPKLLIYAA-

SYLQSGVPSRFSGSGSGTDFLTISLQPEDFA-TYYCQQSSNDP YTFGQGKTVEIKR) was used. The scFv-6E is an antibody not having a significant binding property to a monomer or oligomer of synuclein, and such an antibody selective or specific to synuclein fibril is suitable for selectively removing synuclein fibril. To the scFv-6E, a tag having the aggregation rate reducing action (Tag4-8 or Tag18-1) and a degradation-inducing sequence (CMA (SEQ ID NO: 1034): MARVKKDQAEPLHRKFERQPPG) were added. An expression plasmid vector for the protein and the synuclein fibril were introduced into the cell respectively with X-trem GENE9 and Multifectam (Merck). The antibody was provided with an HA tag or a myc tag, and detected with an anti-HA tag antibody. The synuclein fibril was detected with an anti-phosphorylated α -synuclein antibody. These antibodies were specifically detected with Alexa 555 labeled antibody and Alexa 633 labeled antibody. Fluorescent stained cells were observed with Keyence BZ-X800.

[0574] The tag was added to the N terminal and the C terminal. In consideration that the aggregation rate of one

having Tag4-8 added to the N terminal and the C terminal was 0.96% in the SHSY-5Y cell, that the aggregation rate of one having TAG18-1 added to the N terminal and the C terminal was 1.14% in the SHSY-5 cell, and that the aggregation rate of the scFv-E6 having no tag added was 41.6% in the HeLa cell, it seems that favorable aggregation inhibition was exhibited.

[0575] When a scFv binds to synuclein fibril, the lysosome is caused to target the synuclein fibril for a degradation-inducing sequence to degrade the synuclein fibril, and as a result, the amount of synuclein fibril in the cell is expected to be reduced. It was evaluated whether or not the reduction amount of the synuclein fibril is increased when the aggregation rate of the scFv was reduced by using the tag to increase the amount of functional scFv in the cell.

[0576] The antibody was expressed in the cell where the synuclein fibril was formed, and the number of synuclein fibril-positive cells was counted. A positive rate of the synuclein fibril was compared between a cell in which the antibody was expressed and a cell in which it was not expressed. The result was obtained as a rate of synuclein fibril-positive cells in antibody-positive cells/a rate of synuclein fibril-positive cells in antibody-negative cells (P/N).

[0577] As illustrated in FIG. 2, phosphorylated synuclein was lost in the cell into which the tagged scFv-E6 was introduced. The P/N was as illustrated in FIG. 3. As illustrated in FIG. 3, both the scFv-E6 tagged with Tag4-8 and the scFv-E6 tagged with Tag18-1 largely reduced the rate of synuclein-positive cells.

Example 8: Enhancement of Action by Stabilization of Intracellular Antibody (Effect of Enhancement of Antibody Action for Recovering Function of Cftr)

[0578] The cystic fibrosis transmembrane conductance regulator (CFTR; UniprotKB/Swiss-Prot: P13569.3) is a negative ion channel expressed in epithelial membrane cells of the whole body, and abnormality thereof causes cystic fibrosis. F508 deletion mutation of CFTR (CFTRAF508) is known as the most common mutation of CFTR, in which the 508th phenylalanine is deleted due to deletion of three nucleotides. As a result, CFTRAF508 cannot be normally folded, strongly tends to form an aggregation, and is deemed to move onto the membrane in a smaller amount than the wild type. The scFv-C2 (SEQ ID NO: 1035: EVQLLESGG-GLVQPGGLSLRLSCAASGFTFSSYAMSWVRQAPGK-GLEWVSAISGSGGS TYYADSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYCAKMRLGLFDYWGQGTLVT VSSGGGGSGGGGGGGGEIVLTQSPGTLSSL-SPGERATLSCRASQSVSSSYLAWYQQK PGQAPRLLIYGASSRATGIPDRFSGSQSGTDFTLTISR-LEPEDFAVYYCQQRGDVPP TFGQGKTVEIKAAA) binds to the NBD1 domain of CFTR, and thus can inhibit formation of an aggregation by the NBD1 (Lovato et al., Protein Engineering, Design and Selection, 20(12): 607-614, 2007). The scFv-C2 exhibits an effect of increasing the amount of CFTRAF508 moving onto the membrane. In this example, it was evaluated whether or not aggregation of the mutant AF508 in the NBD1 domain in the HeLa cell can be inhibited by adding a tag (Tag18-1) to the scFv-C2. Plasmid vectors for expressing these proteins were introduced into HeLa cell with Lipo3000 (TM). An antibody was tagged with a myc tag to be detected with a rabbit anti-myc tag antibody, and the NBD1 was tagged with a His tag to be

detected with a mouse anti-His tag antibody. The mouse anti-His tag antibody was detected with Alexa 633-labeled anti-mouse IgG antibody, and the rabbit anti-myc tag antibody was detected with Alexa 488-labeled anti-rabbit IgG antibody. Fluorescent stained cells were observed with Keyence BZ-X800.

[0579] First, the aggregation rate of the scFv-C2 in HeLa cell was 82%. On the contrary, the aggregation rate of the scFv-C2 having Tag18-1 at the N terminal was 31%, and the aggregation rate of the scFv-C2 having Tag18-1 at the N terminal and the C terminal was 4.6% (see FIG. 3A and FIG. 3B). The aggregation rate of wild type NBD1 domain was 74% in a cell expressing scFv-C2, and the aggregation rate of the AF508 mutant of the NBD1 was 85% in a cell expressing scFv-C2. On the contrary, in a cell expressing scFv-C2 having Tag18-1 at the N terminal and the C terminal, the aggregation rate of the wild type NBD1 domain was 32%, and the aggregation rate of the AF508 mutant of NBD1 was 43%. In this manner, aggregation formation of the scFv-C2 itself could be inhibited by tagging the scFv-C2, and thus, the formation of an aggregation by the NBD1 could be inhibited. The inhibition of the aggregation of the NBD1 is expected to make a contribution to improvement of expression level of the NBD1 in cell membrane.

Example 9: Effect of Amino Acid Substitution in Existing Tag

[0580] Through Examples described above, it was revealed that amino acid substitution for satisfying the condition needed for the peptide tag of the present disclosure, particularly amino acid substitution with P or N increases the aggregation rate reducing action of the tag. In this example, with some amino acids of PEST sequence

substituted with N, the aggregation rates of the scFvs (Y14-259) having tags before and after the substitution were examined.

TABLE 15

Comparison of sequence between before and after substitution	
PEST(before substitution) SEQ ID NO: 1036	YPYDVPDYAGSPQPVEDGEDEF CTPMACEANIQSGDSAAPMSAV HRHRL
PEST(after substitution) SEQ ID NO: 1037	NNYDVPDNAGSPQPQEDGEDEF NNPQANEANQQSGDSNNPNSAV NRHNN

[0581] After these scFvs were expressed in a cell, the aggregation rates of the scFvs were evaluated in the same manner as in Example 1, and the aggregation rate of the scFv having the tag before the substitution was 50.3%, but the aggregation rate of the scFv having the tag after the substitution was 18.0%. In this manner, it was revealed that the amino acid substitution for satisfying the condition needed for the peptide tag of the present disclosure, particularly the amino acid substitution with P or N increases the effect of inhibiting protein aggregation of the tag.

[0582] As described so far, various proteins including antibodies form aggregation in a cell, and thus, the functions can be partially or entirely impaired. A protein tag for inhibiting the formation of an aggregation can inhibit the aggregation formation of these proteins, and thus, can cause the proteins to exhibit their actions to be originally exhibited. A tag in which an acidic amino acid ratio is relatively low can be helpful in a scene where a tag having a high acidic amino acid ratio is difficult to use.

SEQUENCE LISTING

```

Sequence total quantity: 1043
SEQ ID NO: 1      moltype = AA  length = 242
FEATURE          Location/Qualifiers
source           1..242
                mol_type = protein
                organism = synthetic construct
REGION          1..242
                note = Y13-259

SEQUENCE: 1
QVQLQQSGGG LVQPGRSLKL SCVVGFTFS NYGMNWRQQT PGKGLEWVAY ISSGSSYLYY 60
AETVKGRFTI SRDANKNTLY LQMTSLRSEED TALYYCARHE GTGTDFFDYW GQGTTTVTSS 120
GGGGSGGGGS GGGGSDIQLT QSPHSLASL GETVSIECLA SEGISNYLAW YQQKPGKSPQ 180
LLIYYASSLQ DGVPQRSTSGS GSCTQFSLKI SNMQPEDEVY YYCQQAYKYP STFGAGTKLE 240
IK                                         242

SEQ ID NO: 2      moltype = AA  length = 32
FEATURE          Location/Qualifiers
source           1..32
                mol_type = protein
                organism = synthetic construct
REGION          1..32
                note = Tag-1-1

SEQUENCE: 2
AHSSSAESES TSDSDSSSDS ESESSSSDSE GS

SEQ ID NO: 3      moltype = AA  length = 32
FEATURE          Location/Qualifiers
source           1..32
                mol_type = protein
                organism = synthetic construct
REGION          1..32
                note = Tag-1-2

```

-continued

SEQUENCE: 3	
AHSLSAELES TIDSDCSSDW ESELSSSDSE GS	32
SEQ ID NO: 4	moltype = AA length = 32
FEATURE	Location/Qualifiers
source	1..32
	mol_type = protein
	organism = synthetic construct
REGION	1..32
	note = Tag-1-3
SEQUENCE: 4	
AQSSSAESES GSDSDSSDS ESESSSSDSE GS	32
SEQ ID NO: 5	moltype = AA length = 50
FEATURE	Location/Qualifiers
source	1..50
	mol_type = protein
	organism = synthetic construct
REGION	1..50
	note = Tag-4-1
SEQUENCE: 5	
NEYRAREAFDE DYEQQDEDFA EQDPDGNEAF EGEYDGPNQD EYPDEAQNF	50
SEQ ID NO: 6	moltype = AA length = 52
FEATURE	Location/Qualifiers
source	1..52
	mol_type = protein
	organism = synthetic construct
REGION	1..52
	note = Tag-2-1
SEQUENCE: 6	
DEAGSSGAPA DEAGSSGAPA DEAGSSGAPA DEAGSSGAPA DEAGSSGAPA GS	52
SEQ ID NO: 7	moltype = AA length = 52
FEATURE	Location/Qualifiers
source	1..52
	mol_type = protein
	organism = synthetic construct
REGION	1..52
	note = Tag-2-2
SEQUENCE: 7	
DEVGISLAPT DEVGISLAPT DEVGISLAPT DEVGISLAPT GS	52
SEQ ID NO: 8	moltype = AA length = 52
FEATURE	Location/Qualifiers
source	1..52
	mol_type = protein
	organism = synthetic construct
REGION	1..52
	note = Tag-2-3
SEQUENCE: 8	
DEVMISLWPT DEVMISLWPT DEVMISLWPT DEVMISLWPT DEVMISLWPT GS	52
SEQ ID NO: 9	moltype = AA length = 52
FEATURE	Location/Qualifiers
source	1..52
	mol_type = protein
	organism = synthetic construct
REGION	1..52
	note = Tag-3-1
SEQUENCE: 9	
DFAGSSGAPA DFAGSSGAPA DFAGSSGAPA DFAGSSGAPA DFAGSSGAPA GS	52
SEQ ID NO: 10	moltype = AA length = 52
FEATURE	Location/Qualifiers
source	1..52
	mol_type = protein
	organism = synthetic construct
REGION	1..52
	note = Tag-3-2
SEQUENCE: 10	
DTAVSSIAPL DTAVSSIAPL DTAVSSIAPL DTAVSSIAPL DTAVSSIAPL GS	52
SEQ ID NO: 11	moltype = AA length = 52
FEATURE	Location/Qualifiers
source	1..52

-continued

REGION	mol_type = protein organism = synthetic construct 1..52 note = Tag-3-3	
SEQUENCE: 11	DTWVSLIAIL DTWVSLIAIL DTWVSLIAIL DTWVSLIAIL GS	52
SEQ ID NO: 12	moltype = AA length = 10	
FEATURE	Location/Qualifiers	
source	1..10	
REGION	mol_type = protein organism = synthetic construct 1..10 note = Myc tag	
SEQUENCE: 12	EQKLISEEDL	10
SEQ ID NO: 13	moltype = AA length = 9	
FEATURE	Location/Qualifiers	
source	1..9	
REGION	mol_type = protein organism = synthetic construct 1..9 note = HA tag	
SEQUENCE: 13	YPYDVPDYA	9
SEQ ID NO: 14	moltype = AA length = 15	
FEATURE	Location/Qualifiers	
source	1..15	
REGION	mol_type = protein organism = synthetic construct 1..15 note = S-tag	
SEQUENCE: 14	KETAAAKFER QHMDS	15
SEQ ID NO: 15	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
REGION	mol_type = protein organism = synthetic construct 1..20 note = Membrane-localization signal	
SEQUENCE: 15	KLNPPDESGP GCMSCCKCVLS	20
SEQ ID NO: 16	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
REGION	mol_type = protein organism = synthetic construct 1..20 note = extracellular scFv Tag	
SEQUENCE: 16	EFGGAPEFPK PSTPPGSSGL	20
SEQ ID NO: 17	moltype = AA length = 49	
FEATURE	Location/Qualifiers	
source	1..49	
REGION	mol_type = protein organism = synthetic construct 1..49 note = hMst1	
SEQUENCE: 17	DYEFLKSWTV EDLQKRLLL DPMMEQEIEE IRQKYQSKRQ PILDAIEAK	49
SEQ ID NO: 18	moltype = AA length = 49	
FEATURE	Location/Qualifiers	
source	1..49	
REGION	mol_type = protein organism = synthetic construct 1..49 note = hMST2	
SEQUENCE: 18	DFDFFLKNLSL EELQMRLKAL DPMMEREIEE LRQRYTAKRQ PILDAMDAK	49

-continued

```

SEQ ID NO: 19      moltype = AA  length = 49
FEATURE          Location/Qualifiers
source           1..49
mol_type = protein
organism = synthetic construct
REGION          1..49
note = hRaf1

SEQUENCE: 19
GEVNWDAFSM PELHNFLRIL QREEEEHLRQ ILQKYSYSRQ KIQEALHAS        49

SEQ ID NO: 20      moltype = AA  length = 49
FEATURE          Location/Qualifiers
source           1..49
mol_type = protein
organism = synthetic construct
REGION          1..49
note = hRaf5

SEQUENCE: 20
GEVEWDAFSI PELQNFLTIL EKEEQDKIQQ VQKKYDKFPRQ KLEEARRES        49

SEQ ID NO: 21      moltype = AA  length = 49
FEATURE          Location/Qualifiers
source           1..49
mol_type = protein
organism = synthetic construct
REGION          1..49
note = hSAV1

SEQUENCE: 21
HILKWELFQL ADLDTYQGML KLLFMKELEQ IVKMYEAYRQ ALLTELENR        49

SEQ ID NO: 22      moltype = AA  length = 61
FEATURE          Location/Qualifiers
source           1..61
mol_type = protein
organism = synthetic construct
REGION          1..61
note = T7C

SEQUENCE: 22
LEDPFQSGVM LGVASTVAAS PEEASVTSTE ETLTPAQEEA RTRAANKARK EAELAAATAE 60
Q               61

SEQ ID NO: 23      moltype = AA  length = 44
FEATURE          Location/Qualifiers
source           1..44
mol_type = protein
organism = synthetic construct
REGION          1..44
note = T7B

SEQUENCE: 23
LEDPEEASVT STEETLTPAQ EAARTRAANK ARKEAELAAA TAEQ                44

SEQ ID NO: 24      moltype = AA  length = 44
FEATURE          Location/Qualifiers
source           1..44
mol_type = protein
organism = synthetic construct
REGION          1..44
note = T7B1

SEQUENCE: 24
LEDPEEASVT STEETLTPAQ EAARTRPPNK ARKEAELAAA TAEQ                44

SEQ ID NO: 25      moltype = AA  length = 44
FEATURE          Location/Qualifiers
source           1..44
mol_type = protein
organism = synthetic construct
REGION          1..44
note = T7B2

SEQUENCE: 25
LEDPEEASVT STEETLTPAQ EAARTRGGNK ARKEAELAAA TAEQ                44

SEQ ID NO: 26      moltype = AA  length = 41
FEATURE          Location/Qualifiers
source           1..41
mol_type = protein

```

-continued

```

organism = synthetic construct
REGION          1..41
note = T7B3

SEQUENCE: 26
LEDPEEASVT STEETLTPAQ EAARTRAANK ARKEAELTAE Q           41

SEQ ID NO: 27      moltype = AA  length = 44
FEATURE          Location/Qualifiers
source           1..44
mol_type = protein
organism = synthetic construct
REGION          1..44
note = T7B4

SEQUENCE: 27
LEDPEEASVT STEETLTPAQ EAARTRAANK ARKEAELEAE TAEQ           44

SEQ ID NO: 28      moltype = AA  length = 44
FEATURE          Location/Qualifiers
source           1..44
mol_type = protein
organism = synthetic construct
REGION          1..44
note = T7B5

SEQUENCE: 28
LEDPEEASVT STEETLTPAQ EAARTRAAAK ARKEAELAAA TAEQ           44

SEQ ID NO: 29      moltype = AA  length = 41
FEATURE          Location/Qualifiers
source           1..41
mol_type = protein
organism = synthetic construct
REGION          1..41
note = T7B6

SEQUENCE: 29
LEDPEEASVT STEETLTPAQ EAARTRKARK EAELAAATAE Q           41

SEQ ID NO: 30      moltype = AA  length = 39
FEATURE          Location/Qualifiers
source           1..39
mol_type = protein
organism = synthetic construct
REGION          1..39
note = T7B7

SEQUENCE: 30
LEDPEEASVT STEETLTPAQ EAARTRAANK ARKEAELAA               39

SEQ ID NO: 31      moltype = AA  length = 40
FEATURE          Location/Qualifiers
source           1..40
mol_type = protein
organism = synthetic construct
REGION          1..40
note = T7B8

SEQUENCE: 31
LEDPEEASVT STEETLTPAQ EAARTRAANK ARKEAELAAA              40

SEQ ID NO: 32      moltype = AA  length = 44
FEATURE          Location/Qualifiers
source           1..44
mol_type = protein
organism = synthetic construct
REGION          1..44
note = T7B9

SEQUENCE: 32
LEDPEEASVT STEETLTPAQ EAAETEAANK ARKEAELEAE TAEQ           44

SEQ ID NO: 33      moltype = AA  length = 32
FEATURE          Location/Qualifiers
source           1..32
mol_type = protein
organism = synthetic construct
REGION          1..32
note = T7B10

SEQUENCE: 33
LEDPTPAQEAA ARTRAANKAR KEAEELAAATA EQ           32

```

-continued

```

SEQ ID NO: 34      moltype = AA  length = 22
FEATURE          Location/Qualifiers
source           1..22
mol_type = protein
organism = synthetic construct
REGION           1..22
note = T7A

SEQUENCE: 34
LEDPAANKAR KEAEELAAATA EQ                                22

SEQ ID NO: 35      moltype = AA  length = 22
FEATURE          Location/Qualifiers
source           1..22
mol_type = protein
organism = synthetic construct
REGION           1..22
note = T7A1

SEQUENCE: 35
LEDPERNKER KEAEELAAATA EQ                                22

SEQ ID NO: 36      moltype = AA  length = 22
FEATURE          Location/Qualifiers
source           1..22
mol_type = protein
organism = synthetic construct
REGION           1..22
note = T7A2

SEQUENCE: 36
LEDPERNKER KEAELEAAATA EQ                                22

SEQ ID NO: 37      moltype = AA  length = 22
FEATURE          Location/Qualifiers
source           1..22
mol_type = protein
organism = synthetic construct
REGION           1..22
note = T7A3

SEQUENCE: 37
LEDPERNKER KEAELEAAETA EQ                                22

SEQ ID NO: 38      moltype = AA  length = 43
FEATURE          Location/Qualifiers
source           1..43
mol_type = protein
organism = synthetic construct
REGION           1..43
note = T3

SEQUENCE: 38
LEDPAVWEAG KVVAKGVGTA DITATTSNGL IASCKVIVNA ATS        43

SEQ ID NO: 39      moltype = AA  length = 43
FEATURE          Location/Qualifiers
source           1..43
mol_type = protein
organism = synthetic construct
REGION           1..43
note = T3A

SEQUENCE: 39
LEDPAVWEAG KVVAKGVGTA DITATTSNGL IASSEEADNA ATS        43

SEQ ID NO: 40      moltype = AA  length = 89
FEATURE          Location/Qualifiers
source           1..89
mol_type = protein
organism = synthetic construct
REGION           1..89
note = Zif628

SEQUENCE: 40
ERPYACPVES CDRRFSRSDE LTRHIRIHTG QKPFQCRICM RNFSRSDHLT THIRHTGEK  60
PFACDICGRK FARSDERKRH TKIHLRQKD                           89

SEQ ID NO: 41      moltype = AA  length = 52
FEATURE          Location/Qualifiers
source           1..52
mol_type = protein
organism = synthetic construct

```

-continued

```

REGION          1..52
                note = HinR

SEQUENCE: 41
GRPRAITKHE QEQISRLLEK GHPRQQLAII FGIGVSTLYR YFPASSIKKR MN      52

SEQ ID NO: 42      moltype = AA length = 114
FEATURE          Location/Qualifiers
source           1..114
                mol_type = protein
                organism = synthetic construct
REGION          1..114
                note = TrpR
REGION          17..18
                note = misc_feature - Xaa can be any naturally occurring
                amino acid
REGION          35..36
                note = misc_feature - Xaa can be any naturally occurring
                amino acid
REGION          53..54
                note = misc_feature - Xaa can be any naturally occurring
                amino acid

SEQUENCE: 42
MAQQSPYSA MAEQRHXXQE WLRFVDLLKN AYQNXXDLHL PLLNLMLTPD ERXXEALGTR 60
VRIVEELLRG EMSQRELKNE LGAGIATITR GSNSLKAAPV ELRQWLEEV LKSD      114

SEQ ID NO: 43      moltype = AA length = 18
FEATURE          Location/Qualifiers
source           1..18
                mol_type = protein
                organism = synthetic construct
REGION          1..18
                note = NE-1

SEQUENCE: 43
TKENPRSNQE ESYDDNES                                         18

SEQ ID NO: 44      moltype = AA length = 18
FEATURE          Location/Qualifiers
source           1..18
                mol_type = protein
                organism = synthetic construct
REGION          1..18
                note = NE-8

SEQUENCE: 44
TKENPRTNQE ESYDDNES                                         18

SEQ ID NO: 45      moltype = AA length = 18
FEATURE          Location/Qualifiers
source           1..18
                mol_type = protein
                organism = synthetic construct
REGION          1..18
                note = NE-9

SEQUENCE: 45
TKENPRSNQD ESYDDNES                                         18

SEQ ID NO: 46      moltype = AA length = 18
FEATURE          Location/Qualifiers
source           1..18
                mol_type = protein
                organism = synthetic construct
REGION          1..18
                note = NE-10

SEQUENCE: 46
TKENPRSNQP PSYDDNES                                         18

SEQ ID NO: 47      moltype = AA length = 12
FEATURE          Location/Qualifiers
source           1..12
                mol_type = protein
                organism = synthetic construct
REGION          1..12
                note = PA12-tag

SEQUENCE: 47
GVAMPAGAEDD VV                                              12

SEQ ID NO: 48      moltype = AA length = 14

```

-continued

FEATURE	Location/Qualifiers
source	1..14
	mol_type = protein
	organism = synthetic construct
REGION	1..14
	note = PA14-tag
SEQUENCE: 48	
EGGVAMPAGAE DDVV	14
SEQ ID NO: 49	moltype = AA length = 123
FEATURE	Location/Qualifiers
REGION	1..123
	note = FATT-tag
source	1..123
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 49	
DYKDDDDVEA EESDNVDSAD AEEDDSDVWW GGADTDYADG SEDKVVEVAE EEEVAEVEEE	60
EADDDEDDED GDEVEEEAEE PYEATERTT SIATTTTTT ESVEEVYPGQ VGYPQVGYP	120
GQV	123
SEQ ID NO: 50	moltype = AA length = 36
FEATURE	Location/Qualifiers
source	1..36
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 50	
GGSAQLEKEL QALEKENAQL EWELQALEKE LAQGAT	36
SEQ ID NO: 51	moltype = AA length = 288
FEATURE	Location/Qualifiers
source	1..288
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 51	
GEGEGGGGGG EGEGGGGGGE GEGGGEGGGG EGEGGGGGEG EGGEGGGEGEG GEGEGGGEGG	60
GEGGEGEGGG EGGEGEGGGE GGEGEGGGG EGEGGGGGEG EGGEGGGEGGE GEGGGEGGEG	120
EGGGEGGGEGE CGGEGGGEGEG GGGEGGGEGG EGGEGGGEGGG EGEGEGGGGE GGEGEGGGEG	180
GEGEGGGGGG EGEGGGGGGE GEGGGEGGGG EGEGGGGGEGE GGGEGGGEGEG GGEGEGGEGG	240
GEGGEGEGGG EGGEGEGGGG GGEGEGGGG EGEGGGEGGG EGEGGGEGEG GGEGEGGEGG	288
SEQ ID NO: 52	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 52	
ASPAAPAPAS PAAAPAPSAPA A	21
SEQ ID NO: 53	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 53	
ASAAAPAAAS AAASAPSAAA A	21
SEQ ID NO: 54	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 54	
AASPAAPSAP PAAASPAAPS APPAA	25
SEQ ID NO: 55	moltype = AA length = 58
FEATURE	Location/Qualifiers
source	1..58
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 55	
VDNKFNKEQQ NAFYEILHLP NLNNEQRNAF IQSLKDDPSQ SANLAEAKK LNDAQAPK	58
SEQ ID NO: 56	moltype = AA length = 58
FEATURE	Location/Qualifiers
source	1..58

-continued

```

mol_type = protein
organism = synthetic construct
SEQUENCE: 56
VDNKFNFKEQQ NAEYEIEHLP NLNEEQENAF IQSLEDDPSQ SANLLAEAKK LNDAQAPK      58

SEQ ID NO: 57      moltype = AA length = 58
FEATURE
source           Location/Qualifiers
1..58
mol_type = protein
organism = synthetic construct
SEQUENCE: 57
VDNKFNFKEEE EAEEEIEHLP NLNEEQEEAF IESLEDDPSQ SANLLAEAKK LNDAQAPK      58

SEQ ID NO: 58      moltype = AA length = 20
FEATURE
source           Location/Qualifiers
1..20
mol_type = protein
organism = synthetic construct
SEQUENCE: 58
KLNPPDESGP GCMSCCKCVLS                                         20

SEQ ID NO: 59      moltype = AA length = 50
FEATURE
source           Location/Qualifiers
1..50
mol_type = protein
organism = synthetic construct
SEQUENCE: 59
NEYGYREAFD DYEQQDEDFA EQDPDGNEAF EGEYDGPQNQ EYPDEAQNF             50

SEQ ID NO: 60      moltype = AA length = 60
FEATURE
source           Location/Qualifiers
1..60
mol_type = protein
organism = synthetic construct
SEQUENCE: 60
NEGSYREAFD SEDYEQSQDS EDFSAEQDPD GNSEAFEGSE YSDGPNSQDE YPDEAQNF             60
70

SEQ ID NO: 61      moltype = AA length = 70
FEATURE
source           Location/Qualifiers
1..70
mol_type = protein
organism = synthetic construct
SEQUENCE: 61
NEGSYRESAF DSEDYEQSQD SSEDFSAEQD SSPDGNSSEA FEGSSEYSDG PNSQDEYPDE   60
70
ASSQNFSSSE

SEQ ID NO: 62      moltype = AA length = 20
FEATURE
source           Location/Qualifiers
1..20
mol_type = protein
organism = synthetic construct
SEQUENCE: 62
YDNPYFEPQY GFPPEEDEDE                                         20

SEQ ID NO: 63      moltype = AA length = 30
FEATURE
source           Location/Qualifiers
1..30
mol_type = protein
organism = synthetic construct
SEQUENCE: 63
SYDSNPSYFS EPSQYGSFPP SESEDSEDSE                                         30

SEQ ID NO: 64      moltype = AA length = 40
FEATURE
source           Location/Qualifiers
1..40
mol_type = protein
organism = synthetic construct
SEQUENCE: 64
SSYSDSNPSY FSEPSSQSSY GSSFPPSESE DSSESDSSSE                         40

SEQ ID NO: 65      moltype = AA length = 60
FEATURE
source           Location/Qualifiers
1..60
mol_type = protein
organism = synthetic construct
SEQUENCE: 65

```

-continued

```

NEGQYREAFD QEDYEQQQDQ EDFQAEQDPD GNQEAPEGQE YDGPNQQDEY QPDEAQNFQE 60

SEQ ID NO: 66      moltype = AA length = 70
FEATURE           Location/Qualifiers
source            1..70
mol_type = protein
organism = synthetic construct

SEQUENCE: 66
NEGQYREAFD DQEDYEQQQD QQEDFQAEQD QQPQDGNQQEA FEGQQEYQDG PNQQDEYPDE 60
AQQQNFQQQE          70

SEQ ID NO: 67      moltype = AA length = 30
FEATURE           Location/Qualifiers
source            1..30
mol_type = protein
organism = synthetic construct

SEQUENCE: 67
QYDQNPQYFQ EPQQYQGFPP QEQQEDQEQE                                30

SEQ ID NO: 68      moltype = AA length = 40
FEATURE           Location/Qualifiers
source            1..40
mol_type = protein
organism = synthetic construct

SEQUENCE: 68
QQYQDQNPQY PQEPQQQQYQ QQFPQPQEQE DQQEQDQQQE                                40

SEQ ID NO: 69      moltype = AA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = protein
organism = synthetic construct

SEQUENCE: 69
NEGNYREAFD NEDYEQNQDN EDFNAEQDPD GNNEAPEGNE YNDGPNNQDE YPDEAQNFNE 60

SEQ ID NO: 70      moltype = AA length = 70
FEATURE           Location/Qualifiers
source            1..70
mol_type = protein
organism = synthetic construct

SEQUENCE: 70
NEGNYRENAF DNEDYEQNQD NNEDFNAEQD NNPDGNNNEA FEGNNEYNDG PNNQDEYPDE 60
ANNQNFNNNE          70

SEQ ID NO: 71      moltype = AA length = 30
FEATURE           Location/Qualifiers
source            1..30
mol_type = protein
organism = synthetic construct

SEQUENCE: 71
NYDNNNPYFN EPNQYGNFPP NENEDNEDNE                                30

SEQ ID NO: 72      moltype = AA length = 40
FEATURE           Location/Qualifiers
source            1..40
mol_type = protein
organism = synthetic construct

SEQUENCE: 72
NNYNDNNNPY FNEPNNQNNY GNNFPNENE DNNENDNNNE                                40

SEQ ID NO: 73      moltype = AA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = protein
organism = synthetic construct

SEQUENCE: 73
NEGPYREAFD PEDYEQPQDP EDFPAEQDPD GNPEAPEGE YPDGPNPQDE YPDEAQNFPE 60

SEQ ID NO: 74      moltype = AA length = 70
FEATURE           Location/Qualifiers
source            1..70
mol_type = protein
organism = synthetic construct

SEQUENCE: 74
NEGPYREPAF DPEDYEQPQD PPEDFPAEQD PPPDGNPPEA FEGPPEYPDG PNPQDEYPDE 60
APPQNFPPPE          70

```

-continued

```

SEQ ID NO: 75      moltype = AA  length = 30
FEATURE
source
1..30
mol_type = protein
organism = synthetic construct
SEQUENCE: 75
PYDPNPYFP EPPQYGPFPPE PEPEDPEDPE                                30

SEQ ID NO: 76      moltype = AA  length = 40
FEATURE
source
1..40
mol_type = protein
organism = synthetic construct
SEQUENCE: 76
PPYPDPNPFY FPEPPQQPQY GPPFPPEPE DPPEPDPPPE                                40

SEQ ID NO: 77      moltype = AA  length = 60
FEATURE
source
1..60
mol_type = protein
organism = synthetic construct
SEQUENCE: 77
NEGFYREAFD PEDYEQFQDF EDFFAEQDPD GNFEAFEGFE YFDGPNFQDE YPDEAQNFFE 60

SEQ ID NO: 78      moltype = AA  length = 70
FEATURE
source
1..70
mol_type = protein
organism = synthetic construct
SEQUENCE: 78
NEGFYREPAF DFEDYEQFQD FFEDFFAEQD FFPDGNNFEA FEGFFEYFDG PNFQDEYPDE 60
AFFQNFFFFFFE                                70

SEQ ID NO: 79      moltype = AA  length = 30
FEATURE
source
1..30
mol_type = protein
organism = synthetic construct
SEQUENCE: 79
FYDFNPYFFF EPFQYGFPPP FEFEDFEDFE                                30

SEQ ID NO: 80      moltype = AA  length = 40
FEATURE
source
1..40
mol_type = protein
organism = synthetic construct
SEQUENCE: 80
FFYFDFNPFY FFEPPQQFFY GFFFFPFEFE DFFEFDFFFF                                40

SEQ ID NO: 81      moltype = AA  length = 60
FEATURE
source
1..60
mol_type = protein
organism = synthetic construct
SEQUENCE: 81
NEGYYREAFD YEDYEQYQDY EDFYAEQDPD GNYEAFEGYE YYDGPNYQDE YPDEAQNFYE 60

SEQ ID NO: 82      moltype = AA  length = 70
FEATURE
source
1..70
mol_type = protein
organism = synthetic construct
SEQUENCE: 82
NEGYYREYAF DYEDYEQYQD YYEDFYAEQD YYPDGNYYEA FEGYYEYYDG PNYQDEYPDE 60
AYYQNFYYYE                                70

SEQ ID NO: 83      moltype = AA  length = 30
FEATURE
source
1..30
mol_type = protein
organism = synthetic construct
SEQUENCE: 83
YYDYNPYYFY EPYQYGYFPP YEYEDYEDYE                                30

SEQ ID NO: 84      moltype = AA  length = 40

```

-continued

FEATURE	Location/Qualifiers
source	1..40
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 84	
YYYYDYNYYY FYEPYYQYYY GYYFPFYEYE DYYEYDYYYE	40
SEQ ID NO: 85	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 85	
GNNQDSSSD NEADEASDDE DNDGN	25
SEQ ID NO: 86	moltype = AA length = 50
FEATURE	Location/Qualifiers
source	1..50
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 86	
NEGNREASDE DSEQQDEDNA EQDPDGNEAN EGESDGPQD ENPDEAQNSE	50
SEQ ID NO: 87	moltype = AA length = 50
FEATURE	Location/Qualifiers
source	1..50
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 87	
NEGNREANDE DNEQQDEDNA EQDPDGNEAN EGENDGPQD ENPDEAQNNE	50
SEQ ID NO: 88	moltype = AA length = 50
FEATURE	Location/Qualifiers
source	1..50
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 88	
NEGNREANDE DNEQQDEDNA EQDPDGNEAN EGENDGPQD ENPDEAQNNE	50
SEQ ID NO: 89	moltype = AA length = 50
FEATURE	Location/Qualifiers
source	1..50
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 89	
NEGNREANDE DNEQQDEDNA EQDQDGNEAN EGENDGQND ENPDEAQNNE	50
SEQ ID NO: 90	moltype = AA length = 32
FEATURE	Location/Qualifiers
source	1..32
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 90	
DNNESADDNN ENPEDNNKNT DDNEENPNNN EN	32
SEQ ID NO: 91	moltype = AA length = 32
FEATURE	Location/Qualifiers
source	1..32
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 91	
DNSESADDNS ENPEDNSKNT DDSEENPSNS ES	32
SEQ ID NO: 92	moltype = AA length = 32
FEATURE	Location/Qualifiers
source	1..32
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 92	
DSSESADDSS ESPEDSSKST DDSEESPSSS ES	32
SEQ ID NO: 93	moltype = AA length = 32
FEATURE	Location/Qualifiers
source	1..32
	mol_type = protein
	organism = synthetic construct

-continued

SEQUENCE: 93 DNQESADDNQ ENPEDNQKNT DDQEENPQNQ EQ	32
SEQ ID NO: 94 FEATURE source moltype = AA length = 32 Location/Qualifiers 1..32 mol_type = protein organism = synthetic construct	
SEQUENCE: 94 DQQESADDQQ EQPEDQQQKQT DDQEEQPQQQ EQ	32
SEQ ID NO: 95 FEATURE source moltype = AA length = 32 Location/Qualifiers 1..32 mol_type = protein organism = synthetic construct	
SEQUENCE: 95 DNFESADDNF ENPEDNFKNT DDFEENPFNF EF	32
SEQ ID NO: 96 FEATURE source moltype = AA length = 32 Location/Qualifiers 1..32 mol_type = protein organism = synthetic construct	
SEQUENCE: 96 DPFESADDFF EFPEDFFKFT DDFEEFPFFF EF	32
SEQ ID NO: 97 FEATURE source moltype = AA length = 32 Location/Qualifiers 1..32 mol_type = protein organism = synthetic construct	
SEQUENCE: 97 DNPESADDNP ENPEDNPKNP DDPEENPPNP EP	32
SEQ ID NO: 98 FEATURE source moltype = AA length = 32 Location/Qualifiers 1..32 mol_type = protein organism = synthetic construct	
SEQUENCE: 98 DPPEESADDPP EPPPEDPPKPT DDPEEPPPPP EP	32
SEQ ID NO: 99 FEATURE source moltype = AA length = 32 Location/Qualifiers 1..32 mol_type = protein organism = synthetic construct	
SEQUENCE: 99 DNYESADDNY ENPEDNYKNT DDYEENPYNY EY	32
SEQ ID NO: 100 FEATURE source moltype = AA length = 32 Location/Qualifiers 1..32 mol_type = protein organism = synthetic construct	
SEQUENCE: 100 DYYESADDYY EYPEDYYKYT DDYEYPYYY EY	32
SEQ ID NO: 101 FEATURE source moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 101 NDNNNNENQN CNNNEEDEDE	20
SEQ ID NO: 102 FEATURE source moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 102 NDNPNNEPQN GNPPEEDEDE	20
SEQ ID NO: 103 moltype = AA length = 50	

-continued

FEATURE	Location/Qualifiers
source	1..50
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 103	
ENESEDNDEE EPNQNADDGD	PNANPAAEQG GDSSDDEEGQ ENDQSRNEQ
	50
SEQ ID NO: 104	moltype = AA length = 50
FEATURE	Location/Qualifiers
source	1..50
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 104	
QNEENNGDDQD QEEGSEEQQG	ESDRSNEESD NEPAADDNAP EGAEEPDDNN
	50
SEQ ID NO: 105	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 105	
TDNNESADDN NENPEDNNKN	
	20
SEQ ID NO: 106	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 106	
NNESADDNN ENPEDNNKNT	
	20
SEQ ID NO: 107	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 107	
NNESADDNNNE NPEDNNKNTD	
	20
SEQ ID NO: 108	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 108	
NESADDNNEN PEDNNKNTDD	
	20
SEQ ID NO: 109	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 109	
ESADDNNENP EDNNKNTDDN	
	20
SEQ ID NO: 110	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 110	
SADDNNENPE DNNNKNTDDNE	
	20
SEQ ID NO: 111	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 111	
DNNENPEDNN KNTDDNEENP	
	20
SEQ ID NO: 112	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct

-continued

SEQUENCE: 112 NNENPEDNNK NTDDNEENPN	20
SEQ ID NO: 113 FEATURE source moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 113 NENPEDNNKN TDDNEENPNN	20
SEQ ID NO: 114 FEATURE source moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 114 ENPEDNNKNT DDNEENPNNN	20
SEQ ID NO: 115 FEATURE source moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 115 NPEDNNKNTD DNEENPNNN	20
SEQ ID NO: 116 FEATURE source moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 116 PEDNNKNTDD NEENPNNNEN	20
SEQ ID NO: 117 FEATURE source moltype = AA length = 21 Location/Qualifiers 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 117 DNNESADDNN ENPEDNNKNT D	21
SEQ ID NO: 118 FEATURE source moltype = AA length = 21 Location/Qualifiers 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 118 NNESADDNNE NPEDNNKNTD D	21
SEQ ID NO: 119 FEATURE source moltype = AA length = 21 Location/Qualifiers 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 119 NESADDNNEN PEDNNKNTDD N	21
SEQ ID NO: 120 FEATURE source moltype = AA length = 21 Location/Qualifiers 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 120 ESADDNNENP EDNNKNTDDN E	21
SEQ ID NO: 121 FEATURE source moltype = AA length = 21 Location/Qualifiers 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 121 SADDNNENPE DNNKNTDDNE E	21
SEQ ID NO: 122 moltype = AA length = 21	

-continued

FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 122	
ADDNNENPED NNKNTDDNEE N	21
SEQ ID NO: 123	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 123	
DDNNENPEDN NKNTDDNEEN P	21
SEQ ID NO: 124	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 124	
DNNENPEDNN KNTDDNEENP N	21
SEQ ID NO: 125	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 125	
NNENPEDNNK NTDDNEENPN N	21
SEQ ID NO: 126	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 126	
NENPEDNNKN TDDNEENPN N	21
SEQ ID NO: 127	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 127	
ENPEDNNKNT DDNEENPNNN E	21
SEQ ID NO: 128	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 128	
NPEDNNKNTD DNEENPNNN N	21
SEQ ID NO: 129	moltype = AA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 129	
DNNESADDNN ENPEDNNKNT DD	22
SEQ ID NO: 130	moltype = AA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 130	
NNESADDNN NEPEDNNKNTD DN	22
SEQ ID NO: 131	moltype = AA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = protein
	organism = synthetic construct

-continued

```

SEQUENCE: 131
NESADDNNEN PEDNNKNTDD NE                                22

SEQ ID NO: 132      moltype = AA  length = 22
FEATURE          Location/Qualifiers
source           1..22
                  mol_type = protein
                  organism = synthetic construct

SEQUENCE: 132
SADDNNNPE DNNKNTDDNE EN                                22

SEQ ID NO: 133      moltype = AA  length = 22
FEATURE          Location/Qualifiers
source           1..22
                  mol_type = protein
                  organism = synthetic construct

SEQUENCE: 133
ADDNNENPED NNKNTDDNEE NP                                22

SEQ ID NO: 134      moltype = AA  length = 22
FEATURE          Location/Qualifiers
source           1..22
                  mol_type = protein
                  organism = synthetic construct

SEQUENCE: 134
DDNNENPEDN NKNTDDNEEN PN                                22

SEQ ID NO: 135      moltype = AA  length = 22
FEATURE          Location/Qualifiers
source           1..22
                  mol_type = protein
                  organism = synthetic construct

SEQUENCE: 135
DNENPEDNN KNTDDNEENP NN                                22

SEQ ID NO: 136      moltype = AA  length = 22
FEATURE          Location/Qualifiers
source           1..22
                  mol_type = protein
                  organism = synthetic construct

SEQUENCE: 136
NNENPEDNNK NTDDNEENPN NN                                22

SEQ ID NO: 137      moltype = AA  length = 22
FEATURE          Location/Qualifiers
source           1..22
                  mol_type = protein
                  organism = synthetic construct

SEQUENCE: 137
NENPEDNNKN TDDNEENPN NE                                22

SEQ ID NO: 138      moltype = AA  length = 22
FEATURE          Location/Qualifiers
source           1..22
                  mol_type = protein
                  organism = synthetic construct

SEQUENCE: 138
ENPEDNNKNT DDNEENPNNN EN                                22

SEQ ID NO: 139      moltype = AA  length = 23
FEATURE          Location/Qualifiers
source           1..23
                  mol_type = protein
                  organism = synthetic construct

SEQUENCE: 139
DNESADDNN ENPEDNNKNT DDN                                23

SEQ ID NO: 140      moltype = AA  length = 23
FEATURE          Location/Qualifiers
source           1..23
                  mol_type = protein
                  organism = synthetic construct

SEQUENCE: 140
NNESADDNNNE NPEDNNKNTD DNE                               23

SEQ ID NO: 141      moltype = AA  length = 23

```

-continued

FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 141	
NESADDNNEN PEDNNKNTDD NEE	23
SEQ ID NO: 142	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 142	
ESADDNNENP EDNNKNTDDN EEN	23
SEQ ID NO: 143	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 143	
SADDNNNPE DNNKNTDDNE ENP	23
SEQ ID NO: 144	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 144	
ADDNNENPED NNNKNTDDNEE NPN	23
SEQ ID NO: 145	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 145	
DDNNENPEDN NKNTDDNEEN PNN	23
SEQ ID NO: 146	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 146	
DNNENPEDNN KNTDDNEENP NNN	23
SEQ ID NO: 147	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 147	
NNENPEDNNK NTDDNEENPN NNE	23
SEQ ID NO: 148	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 148	
NENPEDNNKN TDDNEENPNN NEN	23
SEQ ID NO: 149	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 149	
HGNQNQDSSDS DNEADEASDD EDN	23
SEQ ID NO: 150	moltype = AA length = 24
FEATURE	Location/Qualifiers
source	1..24
	mol_type = protein
	organism = synthetic construct

-continued

SEQUENCE: 150 DNNESADDNN ENPEDNNKNT DDNE	24
SEQ ID NO: 151 FEATURE source moltype = AA length = 24 Location/Qualifiers 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 151 NNESADDNN NEPENKNTD DNEE	24
SEQ ID NO: 152 FEATURE source moltype = AA length = 24 Location/Qualifiers 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 152 NESADDNNEN PEDNNKNTDD NEEN	24
SEQ ID NO: 153 FEATURE source moltype = AA length = 24 Location/Qualifiers 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 153 ESADDNNENP EDNNKNTDDN EENP	24
SEQ ID NO: 154 FEATURE source moltype = AA length = 24 Location/Qualifiers 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 154 SADDNNENPE DNNKNTDDNE ENPN	24
SEQ ID NO: 155 FEATURE source moltype = AA length = 24 Location/Qualifiers 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 155 ADDNNENPED NNKNTDDNEE NPNN	24
SEQ ID NO: 156 FEATURE source moltype = AA length = 24 Location/Qualifiers 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 156 DDNNENPEDN NKNTDDNEEN PNNN	24
SEQ ID NO: 157 FEATURE source moltype = AA length = 24 Location/Qualifiers 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 157 DNNENPEDNN KNTDDNEENP NNNE	24
SEQ ID NO: 158 FEATURE source moltype = AA length = 24 Location/Qualifiers 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 158 NNENPEDNNK NTDDNEENPN NNEN	24
SEQ ID NO: 159 FEATURE source moltype = AA length = 24 Location/Qualifiers 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 159 SHGNNQDSSD SDNEADEASD DEDN	24
SEQ ID NO: 160 moltype = AA length = 25	

-continued

FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 160	
NNESADDNN ENPEDNNKNT DDNEE	25
SEQ ID NO: 161	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 161	
NNESADDNN NE NPEDNNKNTD DNEEN	25
SEQ ID NO: 162	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 162	
NESADDNNN PEDNNKNTDD NEENP	25
SEQ ID NO: 163	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 163	
ESADDNNNENP EDNNKNTDDN EENPN	25
SEQ ID NO: 164	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 164	
SADDNNNENPE DNNKNTDDNE ENPPN	25
SEQ ID NO: 165	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 165	
ADDNNENPED NNNKNTDDNEE NPNNN	25
SEQ ID NO: 166	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 166	
DDNNENPEDN NKNTDDNEEN PNNNE	25
SEQ ID NO: 167	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 167	
DNNENPEDNN KNTDDNEENP NNNEN	25
SEQ ID NO: 168	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 168	
GSHGNNQDSS DSDNEADEAS DDEDN	25
SEQ ID NO: 169	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct

-continued

SEQUENCE: 169 SHGNQNQDSSD SDNEADEASD DEDND	25
SEQ ID NO: 170 FEATURE source moltype = AA length = 25 Location/Qualifiers 1..25 mol_type = protein organism = synthetic construct	
SEQUENCE: 170 HGNQNQDSSDS DNEADEASDD EDNDG	25
SEQ ID NO: 171 FEATURE source moltype = AA length = 25 Location/Qualifiers 1..25 mol_type = protein organism = synthetic construct	
SEQUENCE: 171 GNNQNQDSSSD NEADEASDDE DNDGN	25
SEQ ID NO: 172 FEATURE source moltype = AA length = 26 Location/Qualifiers 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 172 DNNESADDNN ENPEDNNNKNT DDNEEN	26
SEQ ID NO: 173 FEATURE source moltype = AA length = 26 Location/Qualifiers 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 173 NNESADDNNE NPEDNNNKNTD DNEENP	26
SEQ ID NO: 174 FEATURE source moltype = AA length = 26 Location/Qualifiers 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 174 NESADDNNENP PEDNNNKNTDD NEENPN	26
SEQ ID NO: 175 FEATURE source moltype = AA length = 26 Location/Qualifiers 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 175 ESADDNNNENP EDNNNKNTDDN EENPNN	26
SEQ ID NO: 176 FEATURE source moltype = AA length = 26 Location/Qualifiers 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 176 SADDNNNENPE DNNNKNTDDNE ENPNNN	26
SEQ ID NO: 177 FEATURE source moltype = AA length = 26 Location/Qualifiers 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 177 ADDNNNENPED NNKNTDDNEE NPNNNN	26
SEQ ID NO: 178 FEATURE source moltype = AA length = 26 Location/Qualifiers 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 178 DDNNNENPEDN NKNTDDNEEN PNPNNN	26
SEQ ID NO: 179 moltype = AA length = 26	

-continued

FEATURE	Location/Qualifiers
source	1..26
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 179	
WGSHGNNQDS SDSDNEADEA SDDEDN	26
SEQ ID NO: 180	moltype = AA length = 26
FEATURE	Location/Qualifiers
source	1..26
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 180	
GSHGNNQDSS DSDNEADEAS DDEDN	26
SEQ ID NO: 181	moltype = AA length = 26
FEATURE	Location/Qualifiers
source	1..26
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 181	
SHGNNQDSSD SDNEADEASD DEDND	26
SEQ ID NO: 182	moltype = AA length = 26
FEATURE	Location/Qualifiers
source	1..26
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 182	
HGNQNQDSSS DNEADEASDD EDNDGN	26
SEQ ID NO: 183	moltype = AA length = 27
FEATURE	Location/Qualifiers
source	1..27
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 183	
NNESADDNN ENPEDNNKNT DDNEENP	27
SEQ ID NO: 184	moltype = AA length = 27
FEATURE	Location/Qualifiers
source	1..27
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 184	
NNESADDNNE NPEDNNKNTD DNEENP	27
SEQ ID NO: 185	moltype = AA length = 27
FEATURE	Location/Qualifiers
source	1..27
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 185	
NESADDNNEN PEDNNKNTDD NEENPNN	27
SEQ ID NO: 186	moltype = AA length = 27
FEATURE	Location/Qualifiers
source	1..27
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 186	
ESADDNNNENP EDNNKNTDDN EENPN	27
SEQ ID NO: 187	moltype = AA length = 27
FEATURE	Location/Qualifiers
source	1..27
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 187	
SADDNNENPE DNNKNTDDNE ENPN	27
SEQ ID NO: 188	moltype = AA length = 27
FEATURE	Location/Qualifiers
source	1..27
	mol_type = protein
	organism = synthetic construct

-continued

SEQUENCE: 188 ADDNNENPED NNKNTDDNEE NPNNNN	27
SEQ ID NO: 189 FEATURE source moltype = AA length = 27 Location/Qualifiers 1..27 mol_type = protein organism = synthetic construct	
SEQUENCE: 189 SHGNNQDSSD SDNEADEASD DEDNDGN	27
SEQ ID NO: 190 FEATURE source moltype = AA length = 27 Location/Qualifiers 1..27 mol_type = protein organism = synthetic construct	
SEQUENCE: 190 HGNNQDSSD DNEADEASDD EDNDGNE	27
SEQ ID NO: 191 FEATURE source moltype = AA length = 28 Location/Qualifiers 1..28 mol_type = protein organism = synthetic construct	
SEQUENCE: 191 DNNESADDNN ENPEDNNKNT DDNEENPN	28
SEQ ID NO: 192 FEATURE source moltype = AA length = 28 Location/Qualifiers 1..28 mol_type = protein organism = synthetic construct	
SEQUENCE: 192 NNESADDNNE NPEDNNKNTD DNEENPNN	28
SEQ ID NO: 193 FEATURE source moltype = AA length = 28 Location/Qualifiers 1..28 mol_type = protein organism = synthetic construct	
SEQUENCE: 193 NESADDNNEN PEDNNNKNTDD NEENPNNN	28
SEQ ID NO: 194 FEATURE source moltype = AA length = 28 Location/Qualifiers 1..28 mol_type = protein organism = synthetic construct	
SEQUENCE: 194 ESADDNNENP EDNNNKNTDDN EENPNNN	28
SEQ ID NO: 195 FEATURE source moltype = AA length = 28 Location/Qualifiers 1..28 mol_type = protein organism = synthetic construct	
SEQUENCE: 195 SADDNNENPE DNNKNTDDNE ENPNNNEN	28
SEQ ID NO: 196 FEATURE source moltype = AA length = 28 Location/Qualifiers 1..28 mol_type = protein organism = synthetic construct	
SEQUENCE: 196 SHGNNQDSSD SDNEADEASD DEDNDGNE	28
SEQ ID NO: 197 FEATURE source moltype = AA length = 29 Location/Qualifiers 1..29 mol_type = protein organism = synthetic construct	
SEQUENCE: 197 DNNESADDNN ENPEDNNKNT DDNEENPNN	29
SEQ ID NO: 198 moltype = AA length = 29	

-continued

FEATURE	Location/Qualifiers
source	1..29
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 198	
NNESADDNNE NPEDNNKNTD DNEENPNNN	29
SEQ ID NO: 199	moltype = AA length = 29
FEATURE	Location/Qualifiers
source	1..29
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 199	
NESADDNNEN PEDNNKNTDD NEENPNNNE	29
SEQ ID NO: 200	moltype = AA length = 29
FEATURE	Location/Qualifiers
source	1..29
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 200	
ESADDNNNENP EDNNKNTDDN EENPNNNN	29
SEQ ID NO: 201	moltype = AA length = 30
FEATURE	Location/Qualifiers
source	1..30
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 201	
NNESADDNNN ENPEDNNKNT DDNEENPNNN	30
SEQ ID NO: 202	moltype = AA length = 30
FEATURE	Location/Qualifiers
source	1..30
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 202	
NNESADDNNE NPEDNNKNTD DNEENPNNNE	30
SEQ ID NO: 203	moltype = AA length = 30
FEATURE	Location/Qualifiers
source	1..30
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 203	
NESADDNNEN PEDNNKNTDD NEENPNNNN	30
SEQ ID NO: 204	moltype = AA length = 31
FEATURE	Location/Qualifiers
source	1..31
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 204	
NNESADDNNN ENPEDNNKNT DDNEENPNNN E	31
SEQ ID NO: 205	moltype = AA length = 31
FEATURE	Location/Qualifiers
source	1..31
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 205	
NNESADDNNE NPEDNNKNTD DNEENPNNNE N	31
SEQ ID NO: 206	moltype = AA length = 31
FEATURE	Location/Qualifiers
source	1..31
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 206	
FWGSHGNQD SSDSDNEADE ASDDEDNDGN E	31
SEQ ID NO: 207	moltype = AA length = 31
FEATURE	Location/Qualifiers
source	1..31
	mol_type = protein
	organism = synthetic construct

-continued

SEQUENCE: 207 SHGNNQDSSD SDNEADEASD DEDNDGNEG D N	31
SEQ ID NO: 208 FEATURE source mol_type = protein organism = synthetic construct	
SEQUENCE: 208 DNNESADDNN ENPEDNNKNT DDNEENPNNN EN	32
SEQ ID NO: 209 FEATURE source mol_type = protein organism = synthetic construct	
SEQUENCE: 209 SHGNNQDSSD SDNEADEASD DEDNDGNEG D NE	32
SEQ ID NO: 210 FEATURE source mol_type = protein organism = synthetic construct	
SEQUENCE: 210 LEDNNEEPRD PQSPPDPPE	20
SEQ ID NO: 211 FEATURE source mol_type = protein organism = synthetic construct	
SEQUENCE: 211 EDNNEEPRDP QSPPDPPEF	20
SEQ ID NO: 212 FEATURE source mol_type = protein organism = synthetic construct	
SEQUENCE: 212 DNNEEPRDPQ SPPDPPEFG	20
SEQ ID NO: 213 FEATURE source mol_type = protein organism = synthetic construct	
SEQUENCE: 213 EGEQQLKPNN SNAPNEDQEE	20
SEQ ID NO: 214 FEATURE source mol_type = protein organism = synthetic construct	
SEQUENCE: 214 GEQQLKPNN S NAPNEDQEE	20
SEQ ID NO: 215 FEATURE source mol_type = protein organism = synthetic construct	
SEQUENCE: 215 KPNNSNAPNE DQE E E I Q Q S E	20
SEQ ID NO: 216 FEATURE source mol_type = protein organism = synthetic construct	
SEQUENCE: 216 PNNSNAPNED Q E E E I Q Q S E Q	20
SEQ ID NO: 217 moltype = AA length = 20	

-continued

FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 217	
NNNSNAPNEDQ EEEEIQQSEQH	20
SEQ ID NO: 218	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 218	
EGEQQLKPNN SNAPNEDQEE E	21
SEQ ID NO: 219	moltype = AA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 219	
SEGEQQQLKPN NSNAPNEDQEE EE	22
SEQ ID NO: 220	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 220	
EQLNFSDDDE QGSNSPKENN	20
SEQ ID NO: 221	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 221	
LNFSDDDEQG SNSPKENNSE	20
SEQ ID NO: 222	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 222	
NFSDDDEQGS NSPKENNSED	20
SEQ ID NO: 223	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 223	
FSDDDEQGSN SPKENNSEDQ	20
SEQ ID NO: 224	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 224	
LNFSDDDEQG SNSPKENNSE D	21
SEQ ID NO: 225	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 225	
NFSDDDEQGS NSPKENNSED Q	21
SEQ ID NO: 226	moltype = AA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = protein
	organism = synthetic construct

-continued

```

SEQUENCE: 226
EQLNFSDDDE QGSNSPKENN SE                                22

SEQ ID NO: 227      moltype = AA  length = 22
FEATURE          Location/Qualifiers
source           1..22
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 227
QLNFSDDDEQ GSNSPKENNNS ED                                22

SEQ ID NO: 228      moltype = AA  length = 22
FEATURE          Location/Qualifiers
source           1..22
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 228
LNFSDDDEQG SNSPKENNSE DQ                                22

SEQ ID NO: 229      moltype = AA  length = 22
FEATURE          Location/Qualifiers
source           1..22
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 229
NFSDDDEQGS NSPKENNSED QG                                22

SEQ ID NO: 230      moltype = AA  length = 23
FEATURE          Location/Qualifiers
source           1..23
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 230
EQLNFSDDDE QGSNSPKENN SED                                23

SEQ ID NO: 231      moltype = AA  length = 23
FEATURE          Location/Qualifiers
source           1..23
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 231
QLNFSDDDEQ GSNSPKENNNS EDQ                               23

SEQ ID NO: 232      moltype = AA  length = 23
FEATURE          Location/Qualifiers
source           1..23
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 232
LNFSDDDEQG SNSPKENNSE DQG                               23

SEQ ID NO: 233      moltype = AA  length = 23
FEATURE          Location/Qualifiers
source           1..23
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 233
NFSDDDEQGS NSPKENNSED QGS                               23

SEQ ID NO: 234      moltype = AA  length = 24
FEATURE          Location/Qualifiers
source           1..24
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 234
EQLNFSDDDE QGSNSPKENN SEDQ                               24

SEQ ID NO: 235      moltype = AA  length = 25
FEATURE          Location/Qualifiers
source           1..25
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 235
EQLNFSDDDE QGSNSPKENN SEDQG                             25

SEQ ID NO: 236      moltype = AA  length = 26

```

-continued

FEATURE	Location/Qualifiers
source	1..26
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 236	
EQLNFSDDDE QGSNSPKENN SEDQGS	26
SEQ ID NO: 237	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 237	
EKNENDENS LSSSSDSSED	20
SEQ ID NO: 238	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 238	
NENDENSLSS SSDSSEDKDE	20
SEQ ID NO: 239	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 239	
KETNNNSNAQN PSEEEGEGQD E	21
SEQ ID NO: 240	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 240	
ETNNNSNAQN PSEEEGEGQD D	21
SEQ ID NO: 241	moltype = AA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 241	
KETNNNSNAQN PSEEEGEGQD ED	22
SEQ ID NO: 242	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 242	
ENANDSSDDS GEETDESFNP	20
SEQ ID NO: 243	moltype = AA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 243	
DNESNSES AENGWDGSGSNF SE	22
SEQ ID NO: 244	moltype = AA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 244	
DNESNSES AENGWDGSGSNFS EE	22
SEQ ID NO: 245	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct

-continued

SEQUENCE: 245 SDDNESNSES AENGWDSGSN FSE	23
SEQ ID NO: 246 FEATURE source moltype = AA length = 23 Location/Qualifiers 1..23 mol_type = protein organism = synthetic construct	
SEQUENCE: 246 DDNESNSES AENGWDSGSNF SEE	23
SEQ ID NO: 247 FEATURE source moltype = AA length = 24 Location/Qualifiers 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 247 SDDNESNSES AENGWDSGSN FSEE	24
SEQ ID NO: 248 FEATURE source moltype = AA length = 25 Location/Qualifiers 1..25 mol_type = protein organism = synthetic construct	
SEQUENCE: 248 SSDDNESNSE SAENGWDSGS NFSEE	25
SEQ ID NO: 249 FEATURE source moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 249 SEENASSGDS EENTNSDHEES	20
SEQ ID NO: 250 FEATURE source moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 250 EENASSGDSE ENTNSDHESE	20
SEQ ID NO: 251 FEATURE source moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 251 ENASSGDSEE NTNSDHESEQ	20
SEQ ID NO: 252 FEATURE source moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 252 DENSENNWRN EYPEEESSDG	20
SEQ ID NO: 253 FEATURE source moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 253 ENSENNWRNE YPEEEESSDG	20
SEQ ID NO: 254 FEATURE source moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 254 NSENNWRNEY PEEESSDGDE	20
SEQ ID NO: 255 moltype = AA length = 21	

-continued

FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 255	
DDENSENNWR NEYPEEESSD G	21
SEQ ID NO: 256	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 256	
DENSENNWRN EYPEEESSDG D	21
SEQ ID NO: 257	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 257	
ENSENNWRNE YPEEESSDG D E	21
SEQ ID NO: 258	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 258	
NSENNWRNEY PEEESSDGDE D	21
SEQ ID NO: 259	moltype = AA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 259	
ENSENNWRNEY PEEESSDGDE DS	22
SEQ ID NO: 260	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 260	
ENSENNWRNE YPEEESSDGDE EDS	23
SEQ ID NO: 261	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 261	
EQQNEASEEN NDQQSQEVPE	20
SEQ ID NO: 262	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 262	
QQNEASEENN DQQSQEVPEK	20
SEQ ID NO: 263	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 263	
SKMQEDEFDQ GNQEQedNSN	20
SEQ ID NO: 264	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct

-continued

SEQUENCE: 264 KMQEDEFDQG NQEQednsna	20
SEQ ID NO: 265 FEATURE source moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 265 MQEDEFDQGN QEQednsnae	20
SEQ ID NO: 266 FEATURE source moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 266 QEDEFDQGNQ EQEDnsnaem	20
SEQ ID NO: 267 FEATURE source moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 267 FDQGNQEQed NSNAEMEEEN	20
SEQ ID NO: 268 FEATURE source moltype = AA length = 21 Location/Qualifiers 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 268 EFDQGNQEQE DNSNAEMEEE N	21
SEQ ID NO: 269 FEATURE source moltype = AA length = 21 Location/Qualifiers 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 269 FDQGNQEQed NSNAEMEEEN A	21
SEQ ID NO: 270 FEATURE source moltype = AA length = 21 Location/Qualifiers 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 270 DQGNQEQedN SNAEMEEENA S	21
SEQ ID NO: 271 FEATURE source moltype = AA length = 21 Location/Qualifiers 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 271 QGNQEQedNS NAEMEEENAS N	21
SEQ ID NO: 272 FEATURE source moltype = AA length = 22 Location/Qualifiers 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 272 EFDQGNQEQE DNSNAEMEEE NA	22
SEQ ID NO: 273 FEATURE source moltype = AA length = 22 Location/Qualifiers 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 273 FDQGNQEQed NSNAEMEEEN AS	22
SEQ ID NO: 274 moltype = AA length = 22	

-continued

FEATURE	Location/Qualifiers
source	1..22
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 274	
DQGNQEQEDN SNAEMEEENA SN	22
SEQ ID NO: 275	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 275	
DEFDQGNQEQQ EDNSNAEMEE ENA	23
SEQ ID NO: 276	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 276	
EFDQGNQEQQ DNSNAEMEEE NAS	23
SEQ ID NO: 277	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 277	
FDQGNQEQQED NSNAEMEEE ASN	23
SEQ ID NO: 278	moltype = AA length = 24
FEATURE	Location/Qualifiers
source	1..24
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 278	
DEFDQGNQEQQ EDNSNAEMEE ENAS	24
SEQ ID NO: 279	moltype = AA length = 24
FEATURE	Location/Qualifiers
source	1..24
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 279	
EFDQGNQEQQE DNSNAEMEEE NASN	24
SEQ ID NO: 280	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 280	
QEDEFDQGNQ EQEDNSNAEM EEENA	25
SEQ ID NO: 281	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 281	
EDEFDQGNQE QEEDNSNAEME EENAS	25
SEQ ID NO: 282	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 282	
DEFDQGNQEQQ EDNSNAEMEE ENASN	25
SEQ ID NO: 283	moltype = AA length = 26
FEATURE	Location/Qualifiers
source	1..26
	mol_type = protein
	organism = synthetic construct

-continued

SEQUENCE: 283 QEDEFDQGNQ EQEDNSNAEM EEEENAS	26
SEQ ID NO: 284 FEATURE source moltype = AA length = 26 Location/Qualifiers 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 284 EDEFDQGNQE QEDNSNAEME EENASN	26
SEQ ID NO: 285 FEATURE source moltype = AA length = 27 Location/Qualifiers 1..27 mol_type = protein organism = synthetic construct	
SEQUENCE: 285 QEDEFDQGNQ EQEDNSNAEM EEEENASN	27
SEQ ID NO: 286 FEATURE source moltype = AA length = 21 Location/Qualifiers 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 286 PSENENSQSE DSVGGDNDSE N	21
SEQ ID NO: 287 FEATURE source moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 287 EVEESNPSAK EDSNPNSSGE	20
SEQ ID NO: 288 FEATURE source moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 288 VEESNPSAKE DSNPNSSGED	20
SEQ ID NO: 289 FEATURE source moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 289 KEENSESPLN ENSDESYSEE	20
SEQ ID NO: 290 FEATURE source moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 290 QPGPNHEEDA DSYENMDNPD	20
SEQ ID NO: 291 FEATURE source moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 291 PNHEEDADSY ENMDNPDPGP	20
SEQ ID NO: 292 FEATURE source moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 292 NHEEDADSYE NMNDNPDPGPDP	20
SEQ ID NO: 293 moltype = AA length = 20	

-continued

FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 293	
DDPNSSDESN GNDDANSESD	20
SEQ ID NO: 294	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 294	
DPNSSDESNG NDDANSESDN	20
SEQ ID NO: 295	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 295	
PNSSDESGN DANSESDN	20
SEQ ID NO: 296	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 296	
NSSDESGNGND DANSESDNN	20
SEQ ID NO: 297	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 297	
SSDESGNGDD ANSESDNN	20
SEQ ID NO: 298	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 298	
SDESGNGNDA NSESDNN	20
SEQ ID NO: 299	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 299	
DESNGNDDAN SESDNN	20
SEQ ID NO: 300	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 300	
MQGDDPNSS E SNGNDDANS	21
SEQ ID NO: 301	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 301	
QGDDPNSSDE SNGNDDANSE	21
SEQ ID NO: 302	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct

-continued

```

SEQUENCE: 302
GDDPNSSDES NGNDDANSES D                                21

SEQ ID NO: 303      moltype = AA  length = 21
FEATURE           Location/Qualifiers
source            1..21
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 303
DDPNSSDESN GNDDANSESD N                                21

SEQ ID NO: 304      moltype = AA  length = 21
FEATURE           Location/Qualifiers
source            1..21
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 304
DPNSSDESNG NDDANSESDN N                                21

SEQ ID NO: 305      moltype = AA  length = 22
FEATURE           Location/Qualifiers
source            1..22
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 305
DKSMQGDDPN SSDESGNGNDD AN                                22

SEQ ID NO: 306      moltype = AA  length = 22
FEATURE           Location/Qualifiers
source            1..22
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 306
SMQGDDPNSS DESNGNDDAN SE                                22

SEQ ID NO: 307      moltype = AA  length = 22
FEATURE           Location/Qualifiers
source            1..22
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 307
MQGDDPNSSD ESGNGNDDANS ES                                22

SEQ ID NO: 308      moltype = AA  length = 22
FEATURE           Location/Qualifiers
source            1..22
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 308
QGDDPNSSDE SNGNDDANSE SD                                22

SEQ ID NO: 309      moltype = AA  length = 22
FEATURE           Location/Qualifiers
source            1..22
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 309
GDDPNSSDES NGNDDANSES DN                                22

SEQ ID NO: 310      moltype = AA  length = 22
FEATURE           Location/Qualifiers
source            1..22
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 310
DDPNSSDESN GNDDANSESD NN                                22

SEQ ID NO: 311      moltype = AA  length = 22
FEATURE           Location/Qualifiers
source            1..22
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 311
DPNSSDESNG NDDANSESDN NS                                22

SEQ ID NO: 312      moltype = AA  length = 22

```

-continued

FEATURE	Location/Qualifiers
source	1..22
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 312	
DESGNGNDAN SESDNNSSSR GD	22
SEQ ID NO: 313	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 313	
DDKSMQGDDP NSSDESNGND DAN	23
SEQ ID NO: 314	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 314	
DKSMQGDDPN SSDESNGNDD ANS	23
SEQ ID NO: 315	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 315	
KSMQGDDPNS SDESNGNDDA NSE	23
SEQ ID NO: 316	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 316	
SMQGDDPNSS DESNGNDAN SES	23
SEQ ID NO: 317	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 317	
MQGDDPNSSD ESNGNDDANS ESD	23
SEQ ID NO: 318	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 318	
QGDDPNSSDE SNGNDDANSE SDN	23
SEQ ID NO: 319	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 319	
GDDPNSSDES NGNDDANSES DNN	23
SEQ ID NO: 320	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 320	
DDPNSSDESN GNDDANSESD NNS	23
SEQ ID NO: 321	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct

-continued

```

SEQUENCE: 321
DPNSSDESNG NDDANSESDN NSS                                23

SEQ ID NO: 322      moltype = AA  length = 23
FEATURE           Location/Qualifiers
source            1..23
                  mol_type = protein
                  organism = synthetic construct

SEQUENCE: 322
SDESNGNDAA NSESDNNSSS RGD                                23

SEQ ID NO: 323      moltype = AA  length = 23
FEATURE           Location/Qualifiers
source            1..23
                  mol_type = protein
                  organism = synthetic construct

SEQUENCE: 323
SESDNGNDAN SESDNNSSSR GDA                                23

SEQ ID NO: 324      moltype = AA  length = 23
FEATURE           Location/Qualifiers
source            1..23
                  mol_type = protein
                  organism = synthetic construct

SEQUENCE: 324
DDANSESDNN SSSRGDASYN SDE                                23

SEQ ID NO: 325      moltype = AA  length = 24
FEATURE           Location/Qualifiers
source            1..24
                  mol_type = protein
                  organism = synthetic construct

SEQUENCE: 325
FDDKSMQGDD PNSSDESNGN DDAN                                24

SEQ ID NO: 326      moltype = AA  length = 24
FEATURE           Location/Qualifiers
source            1..24
                  mol_type = protein
                  organism = synthetic construct

SEQUENCE: 326
DKSMQGDDP NSSDESNGND DANS                                24

SEQ ID NO: 327      moltype = AA  length = 24
FEATURE           Location/Qualifiers
source            1..24
                  mol_type = protein
                  organism = synthetic construct

SEQUENCE: 327
DKSMQGDDPN SSDESNGNDD ANSE                                24

SEQ ID NO: 328      moltype = AA  length = 24
FEATURE           Location/Qualifiers
source            1..24
                  mol_type = protein
                  organism = synthetic construct

SEQUENCE: 328
SMQGDDPNSS DESNGNDAN SESD                                24

SEQ ID NO: 329      moltype = AA  length = 24
FEATURE           Location/Qualifiers
source            1..24
                  mol_type = protein
                  organism = synthetic construct

SEQUENCE: 329
MQGDDPNSSD ESNGNDDANS ESDN                                24

SEQ ID NO: 330      moltype = AA  length = 24
FEATURE           Location/Qualifiers
source            1..24
                  mol_type = protein
                  organism = synthetic construct

SEQUENCE: 330
QGDDPNSSDE SNGNDANSE SDNN                                24

SEQ ID NO: 331      moltype = AA  length = 24

```

-continued

FEATURE	Location/Qualifiers
source	1..24
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 331	
GDDPNSSDES NGNDDANSES DNNS	24
SEQ ID NO: 332	moltype = AA length = 24
FEATURE	Location/Qualifiers
source	1..24
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 332	
DDPNSSDESN GNDDANSESD NNSS	24
SEQ ID NO: 333	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 333	
FDDDKSMQGD DPNSSDESGN NDDAN	25
SEQ ID NO: 334	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 334	
FDDKSMQMGGD PNSSDESGN DDANS	25
SEQ ID NO: 335	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 335	
DKSMQMGGDP NSSDESGNND DANSE	25
SEQ ID NO: 336	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 336	
DKSMQMGGDPN SSDESGNDD ANSES	25
SEQ ID NO: 337	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 337	
KSMQMGGDPNS SDESNGNDDA NSESD	25
SEQ ID NO: 338	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 338	
SMQGDDPNSS DESNGNDDAN SESDN	25
SEQ ID NO: 339	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 339	
MQGDDPNSSD ESGNDDANS ESDNN	25
SEQ ID NO: 340	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct

-continued

SEQUENCE: 340		
QGDPPNSSDE SNGNDDANSE SDNNNS		25
SEQ ID NO: 341	moltype = AA length = 25	
FEATURE	Location/Qualifiers	
source	1..25	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 341		
GDDPNSSDES NGNDDANSES DNNSS		25
SEQ ID NO: 342	moltype = AA length = 25	
FEATURE	Location/Qualifiers	
source	1..25	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 342		
DDPNSSDESN GNDDANSESD NNSSS		25
SEQ ID NO: 343	moltype = AA length = 26	
FEATURE	Location/Qualifiers	
source	1..26	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 343		
DFDDKSMQGD DPNSSDESGN NDDANS		26
SEQ ID NO: 344	moltype = AA length = 26	
FEATURE	Location/Qualifiers	
source	1..26	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 344		
FDDKSMQGDD PNSSDESGN DDANSE		26
SEQ ID NO: 345	moltype = AA length = 26	
FEATURE	Location/Qualifiers	
source	1..26	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 345		
DKSMQGDDP NSSDESGNND DANSE		26
SEQ ID NO: 346	moltype = AA length = 26	
FEATURE	Location/Qualifiers	
source	1..26	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 346		
DKSMQGDDPN SSDESGNDD ANSESD		26
SEQ ID NO: 347	moltype = AA length = 26	
FEATURE	Location/Qualifiers	
source	1..26	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 347		
KSMQGDDPNS SDESNGNDDA NSESDN		26
SEQ ID NO: 348	moltype = AA length = 26	
FEATURE	Location/Qualifiers	
source	1..26	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 348		
SMQGDDPNSS DESNGNDDAN SESDNN		26
SEQ ID NO: 349	moltype = AA length = 26	
FEATURE	Location/Qualifiers	
source	1..26	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 349		
MQGDDPNSSD ESNGNDDANS ESDNNNS		26
SEQ ID NO: 350	moltype = AA length = 26	

-continued

FEATURE	Location/Qualifiers
source	1..26
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 350	
QGDPNNSDE SNGNDDANSE SDNNSS	26
SEQ ID NO: 351	moltype = AA length = 26
FEATURE	Location/Qualifiers
source	1..26
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 351	
GDDPNSSDES NGNDDANSES DNNSSS	26
SEQ ID NO: 352	moltype = AA length = 26
FEATURE	Location/Qualifiers
source	1..26
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 352	
DDPNSSDESN GNDDANSESD NNSSSR	26
SEQ ID NO: 353	moltype = AA length = 27
FEATURE	Location/Qualifiers
source	1..27
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 353	
DKSMQGDDPN SSDESNGND ANSESDN	27
SEQ ID NO: 354	moltype = AA length = 28
FEATURE	Location/Qualifiers
source	1..28
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 354	
DDKSMQGDDP NSSDESNGND DANSESDN	28
SEQ ID NO: 355	moltype = AA length = 28
FEATURE	Location/Qualifiers
source	1..28
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 355	
DKSMQGDDPN SSDESNGND ANSESDNN	28
SEQ ID NO: 356	moltype = AA length = 28
FEATURE	Location/Qualifiers
source	1..28
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 356	
DDPNSSDESN GNDDANSESD NNSSSRGD	28
SEQ ID NO: 357	moltype = AA length = 29
FEATURE	Location/Qualifiers
source	1..29
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 357	
FDDKSMQGDD PNSSDESNGN DDANSESDN	29
SEQ ID NO: 358	moltype = AA length = 29
FEATURE	Location/Qualifiers
source	1..29
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 358	
DDKSMQGDDP NSSDESNGND DANSESDNN	29
SEQ ID NO: 359	moltype = AA length = 29
FEATURE	Location/Qualifiers
source	1..29
	mol_type = protein
	organism = synthetic construct

-continued

SEQUENCE: 359	
DKSMQGDDPN SSDESGNDD ANSESDNN	29
SEQ ID NO: 360	moltype = AA length = 29
FEATURE	Location/Qualifiers
source	1..29
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 360	
DDPNSSDESN GNDDANSESD NNSSSRGDA	29
SEQ ID NO: 361	moltype = AA length = 29
FEATURE	Location/Qualifiers
source	1..29
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 361	
DESNGNDAN SESDNNSSSR GDASYN SDE	29
SEQ ID NO: 362	moltype = AA length = 30
FEATURE	Location/Qualifiers
source	1..30
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 362	
DFDDKSMQGD DPNSSDESNG NDDANSESDN	30
SEQ ID NO: 363	moltype = AA length = 30
FEATURE	Location/Qualifiers
source	1..30
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 363	
FDDKSMQGDD PNSSDESNGN DDANSESDNN	30
SEQ ID NO: 364	moltype = AA length = 30
FEATURE	Location/Qualifiers
source	1..30
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 364	
DKSMQGDDP NSSDESNGND DANSESDNN	30
SEQ ID NO: 365	moltype = AA length = 30
FEATURE	Location/Qualifiers
source	1..30
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 365	
DKSMQGDDPN SSDESGNDD ANSESDNNSS	30
SEQ ID NO: 366	moltype = AA length = 30
FEATURE	Location/Qualifiers
source	1..30
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 366	
DDPNSSDESN GNDDANSESD NNSSSRGDAS	30
SEQ ID NO: 367	moltype = AA length = 30
FEATURE	Location/Qualifiers
source	1..30
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 367	
SDESNGNDAA NSESDNNSSS RG DASYN SDE	30
SEQ ID NO: 368	moltype = AA length = 30
FEATURE	Location/Qualifiers
source	1..30
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 368	
DESNGNDAN SESDNNSSSR GDASYN SDES	30
SEQ ID NO: 369	moltype = AA length = 31

-continued

FEATURE	Location/Qualifiers
source	1..31
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 369	
DFDDKSMQGD DPNSSDESNG NDDANSESDN N	31
SEQ ID NO: 370	moltype = AA length = 31
FEATURE	Location/Qualifiers
source	1..31
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 370	
FDDKSMQGDD PNSSDESNGN DDANSESDNN S	31
SEQ ID NO: 371	moltype = AA length = 31
FEATURE	Location/Qualifiers
source	1..31
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 371	
DDKSMQGDDP NSSDESNGND DANSESDNN S	31
SEQ ID NO: 372	moltype = AA length = 32
FEATURE	Location/Qualifiers
source	1..32
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 372	
DFDDKSMQGD DPNSSDESNG NDDANSESDN NS	32
SEQ ID NO: 373	moltype = AA length = 32
FEATURE	Location/Qualifiers
source	1..32
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 373	
FDDKSMQGDD PNSSDESNGN DDANSESDNN SS	32
SEQ ID NO: 374	moltype = AA length = 32
FEATURE	Location/Qualifiers
source	1..32
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 374	
DDKSMQGDDP NSSDESNGND DANSESDNN SS	32
SEQ ID NO: 375	moltype = AA length = 33
FEATURE	Location/Qualifiers
source	1..33
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 375	
DFDDKSMQGD DPNSSDESNG NDDANSESDN NSS	33
SEQ ID NO: 376	moltype = AA length = 33
FEATURE	Location/Qualifiers
source	1..33
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 376	
FDDKSMQGDD PNSSDESNGN DDANSESDNN SSS	33
SEQ ID NO: 377	moltype = AA length = 34
FEATURE	Location/Qualifiers
source	1..34
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 377	
DFDDKSMQGD DPNSSDESNG NDDANSESDN NSSS	34
SEQ ID NO: 378	moltype = AA length = 35
FEATURE	Location/Qualifiers
source	1..35
	mol_type = protein
	organism = synthetic construct

-continued

SEQUENCE: 378 DDPNSSDESN GNDDANSESD NNSSSRGDAS YNSDE	35
SEQ ID NO: 379 moltype = AA length = 36 FEATURE Location/Qualifiers source 1..36 mol_type = protein organism = synthetic construct	
SEQUENCE: 379 GDDPNSSDES NGNDDANSES DNNSSSRGDA SYNSDE	36
SEQ ID NO: 380 moltype = AA length = 36 FEATURE Location/Qualifiers source 1..36 mol_type = protein organism = synthetic construct	
SEQUENCE: 380 DDPNSSDESN GNDDANSESD NNSSSRGDAS YNSDES	36
SEQ ID NO: 381 moltype = AA length = 40 FEATURE Location/Qualifiers source 1..40 mol_type = protein organism = synthetic construct	
SEQUENCE: 381 SYDFDDKSM QGDDPNSSDE SNGNDDANSE SDNNSSSRGD	40
SEQ ID NO: 382 moltype = AA length = 40 FEATURE Location/Qualifiers source 1..40 mol_type = protein organism = synthetic construct	
SEQUENCE: 382 SYDFDDKSMQG GDDPNSSDES NGNDDANSES DNNSSSRGDA	40
SEQ ID NO: 383 moltype = AA length = 40 FEATURE Location/Qualifiers source 1..40 mol_type = protein organism = synthetic construct	
SEQUENCE: 383 YDFDDKSMQG DDPNSSDESN GNDDANSESD NNSSSRGDAS	40
SEQ ID NO: 384 moltype = AA length = 40 FEATURE Location/Qualifiers source 1..40 mol_type = protein organism = synthetic construct	
SEQUENCE: 384 DFDDKSMQGD DPNSSDESNG NDDANSESDN NSSSRGDASY	40
SEQ ID NO: 385 moltype = AA length = 40 FEATURE Location/Qualifiers source 1..40 mol_type = protein organism = synthetic construct	
SEQUENCE: 385 QGDDPNSSDE SNGNDDANSE SDNNSSSRGD ASYNSDESKD	40
SEQ ID NO: 386 moltype = AA length = 40 FEATURE Location/Qualifiers source 1..40 mol_type = protein organism = synthetic construct	
SEQUENCE: 386 GDDPNSSDES NGNDDANSES DNNSSSRGDA SYNSDESKDN	40
SEQ ID NO: 387 moltype = AA length = 40 FEATURE Location/Qualifiers source 1..40 mol_type = protein organism = synthetic construct	
SEQUENCE: 387 DDPNSSDESN GNDDANSESD NNSSSRGDAS YNSDESKDNG	40
SEQ ID NO: 388 moltype = AA length = 42	

-continued

```

FEATURE          Location/Qualifiers
source           1..42
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 388
DDKSMQGDDP NSSDESNGND DANSESDNNSS SSRGDASYN S DE               42

SEQ ID NO: 389      moltype = AA length = 43
FEATURE          Location/Qualifiers
source           1..43
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 389
DFDDKSMQGD DPNSSDESNG NDDANSESDN NSSSRGDASY NSD               43

SEQ ID NO: 390      moltype = AA length = 43
FEATURE          Location/Qualifiers
source           1..43
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 390
FDDKSMQGDD PNNSDESNGN DDANSESDNN SSSRGDASYN SDE               43

SEQ ID NO: 391      moltype = AA length = 43
FEATURE          Location/Qualifiers
source           1..43
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 391
DDKSMQGDDP NSSDESNGND DANSESDNNSS SSRGDASYN DES               43

SEQ ID NO: 392      moltype = AA length = 44
FEATURE          Location/Qualifiers
source           1..44
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 392
DFDDKSMQGD DPNSSDESNG NDDANSESDN NSSSRGDASY NSDE               44

SEQ ID NO: 393      moltype = AA length = 45
FEATURE          Location/Qualifiers
source           1..45
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 393
DFDDKSMQGD DPNSSDESNG NDDANSESDN NSSSRGDASY NSDES               45

SEQ ID NO: 394      moltype = AA length = 60
FEATURE          Location/Qualifiers
source           1..60
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 394
DPNSSDESN GNDDANSESD NNSSSRGDAS YNSDESKDNG NGSDSKGAED DDS DSTSDTN 60

SEQ ID NO: 395      moltype = AA length = 60
FEATURE          Location/Qualifiers
source           1..60
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 395
DPNSSDESNG NDDANSESDN NSSSRGDASY NSDESKDNGN GSDSKGAEED DSD STSDTN 60

SEQ ID NO: 396      moltype = AA length = 60
FEATURE          Location/Qualifiers
source           1..60
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 396
NSSDESNGND DANSESDNNSS SSRGDASYN S DESKDNGNGS DSKGAEDDS DSTSDTNNSD 60

SEQ ID NO: 397      moltype = AA length = 60
FEATURE          Location/Qualifiers
source           1..60
                  mol_type = protein
                  organism = synthetic construct

```

-continued

```

SEQUENCE: 397
SSDESNGNDD ANSESDNNSS SRGDASYNSD ESKDNGNGSD SKGAEDDDSD STSDTNNNSDS 60

SEQ ID NO: 398      moltype = AA length = 60
FEATURE          Location/Qualifiers
source           1..60
mol_type = protein
organism = synthetic construct

SEQUENCE: 398
SDESNGNDDA NSESDNNSSS RGDASYNSD E SKDNGNGSDS KGAEDDDSDS TSDTNNNSDSN 60

SEQ ID NO: 399      moltype = AA length = 60
FEATURE          Location/Qualifiers
source           1..60
mol_type = protein
organism = synthetic construct

SEQUENCE: 399
AEDDDSDSTS DTNNNSDSNGN GNNGNDDNDK SDSGKGKSDS SDSDSSDSSN SSDSSDSSDS 60

SEQ ID NO: 400      moltype = AA length = 60
FEATURE          Location/Qualifiers
source           1..60
mol_type = protein
organism = synthetic construct

SEQUENCE: 400
EDDDSDSTSD TNNSDSNGNG NNGNDDNDKS DSGKGKSDSS DSDSSDSSNS SDSSDSSDSD 60

SEQ ID NO: 401      moltype = AA length = 60
FEATURE          Location/Qualifiers
source           1..60
mol_type = protein
organism = synthetic construct

SEQUENCE: 401
DDDSDSTS DNT NSDSNGNGN NGNDDNDKSD SGKGKSDSSD SDSSDSSNNS DSSDSSDSDS 60

SEQ ID NO: 402      moltype = AA length = 61
FEATURE          Location/Qualifiers
source           1..61
mol_type = protein
organism = synthetic construct

SEQUENCE: 402
GDDPNSSDES NGNDDANSES DNNSSRGDA YNSDESKDN GNGSDSKGAE DDDSDSTS DT 60
N                                         61

SEQ ID NO: 403      moltype = AA length = 61
FEATURE          Location/Qualifiers
source           1..61
mol_type = protein
organism = synthetic construct

SEQUENCE: 403
DPPNSSDES NGNDDANSES DNNSSRGDA YNSDESKDN GNGSDSKGAE DDDSDSTS DTN 60
N                                         61

SEQ ID NO: 404      moltype = AA length = 62
FEATURE          Location/Qualifiers
source           1..62
mol_type = protein
organism = synthetic construct

SEQUENCE: 404
QGDDPNSSDE SNGNDDANSE SDNNSSSRGD ASYNSDESKD NGNGSDSKGA EDDDSDSTS D 60
TN                                         62

SEQ ID NO: 405      moltype = AA length = 62
FEATURE          Location/Qualifiers
source           1..62
mol_type = protein
organism = synthetic construct

SEQUENCE: 405
GDDPNSSDES NGNDDANSES DNNSSRGDA YNSDESKDN GNGSDSKGAE DDDSDSTS DT 60
NN                                         62

SEQ ID NO: 406      moltype = AA length = 62
FEATURE          Location/Qualifiers
source           1..62
mol_type = protein
organism = synthetic construct

```

-continued

```

SEQUENCE: 406
DPPNSSDESN GNDDANSESD NNSSSRGDAS YNSDESKDNG NGSDSKGAED DDSDSTS DTN 60
NS 62

SEQ ID NO: 407      moltype = AA length = 62
FEATURE
source
1..62
mol_type = protein
organism = synthetic construct
SEQUENCE: 407
DPNSSDESNG NDDANSESDN NSSSRGDASY NSDESKDNGN GSDSKGAE DDSDSTS DTNN 60
SD 62

SEQ ID NO: 408      moltype = AA length = 63
FEATURE
source
1..63
mol_type = protein
organism = synthetic construct
SEQUENCE: 408
MQGDDPNSSD ESGNGDDANS ESDNNSSSRG DASYN SDESK DNGNGSDSKG AE DDSDSTS DTN 60
DTN 63

SEQ ID NO: 409      moltype = AA length = 63
FEATURE
source
1..63
mol_type = protein
organism = synthetic construct
SEQUENCE: 409
QGDPNNSDE SNGNDDANSE SDNNSSSRGD ASYNSDESKD NGNGSDSKGA EDDDDSDSTS 60
TNN 63

SEQ ID NO: 410      moltype = AA length = 63
FEATURE
source
1..63
mol_type = protein
organism = synthetic construct
SEQUENCE: 410
GDDPNSSDES NGNDDANSE DNNSSSRGDA SYNSDESKDN GNGSDSKGAE DDDSDSTS DT 60
NNS 63

SEQ ID NO: 411      moltype = AA length = 63
FEATURE
source
1..63
mol_type = protein
organism = synthetic construct
SEQUENCE: 411
DPPNSSDESN GNDDANSESD NNSSSRGDAS YNSDESKDNG NGSDSKGAED DDSDSTS DTN 60
NSD 63

SEQ ID NO: 412      moltype = AA length = 63
FEATURE
source
1..63
mol_type = protein
organism = synthetic construct
SEQUENCE: 412
DPNSSDESNG NDDANSESDN NSSSRGDASY NSDESKDNGN GSDSKGAE DDSDSTS DTNN 60
SDS 63

SEQ ID NO: 413      moltype = AA length = 64
FEATURE
source
1..64
mol_type = protein
organism = synthetic construct
SEQUENCE: 413
GDDPNSSDES NGNDDANSE DNNSSSRGDA SYNSDESKDN GNGSDSKGAE DDDSDSTS DT 60
NNSD 64

SEQ ID NO: 414      moltype = AA length = 64
FEATURE
source
1..64
mol_type = protein
organism = synthetic construct
SEQUENCE: 414
DPPNSSDESN GNDDANSESD NNSSSRGDAS YNSDESKDNG NGSDSKGAED DDSDSTS DTN 60
NSDS 64

```

-continued

```

SEQ ID NO: 415      moltype = AA length = 65
FEATURE           Location/Qualifiers
source            1..65
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 415
QGDDPNSSDE SNGNDDANSE SDNNSSSRGD ASYNSDESKD NGNGSDSKGA EDDDSSTSD 60
TNNSDS          65

SEQ ID NO: 416      moltype = AA length = 65
FEATURE           Location/Qualifiers
source            1..65
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 416
GDDPNSSDES NGNDDANSES DNNSSSRGDA SYNSDESKDN GNGSDSKGAE DDDSDSTS 60
NNSDS          65

SEQ ID NO: 417      moltype = AA length = 65
FEATURE           Location/Qualifiers
source            1..65
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 417
DDPNSSDESN GNDDANSESD NNSSSRGDAS YNSDESKDNG NGSDSKGAED DDSDSTS 60
NSDSN          65

SEQ ID NO: 418      moltype = AA length = 65
FEATURE           Location/Qualifiers
source            1..65
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 418
EDDDSDSTS TNNSDSNGNG NNGNDDNDKS DSGKGKSDSS DSDSSDSSNS SDSSDSSD 60
SSDSN          65

SEQ ID NO: 419      moltype = AA length = 66
FEATURE           Location/Qualifiers
source            1..66
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 419
MQGDDPNSSD ESNGNDDANS ESDNNSSSRG DASYNSDESK DNGNGSDSKG AEDDDSDSTS 60
DTNNSD          66

SEQ ID NO: 420      moltype = AA length = 66
FEATURE           Location/Qualifiers
source            1..66
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 420
QGDDPNSSDE SNGNDDANSE SDNNSSSRGD ASYNSDESKD NGNGSDSKGA EDDDSSTSD 60
TNNSDS          66

SEQ ID NO: 421      moltype = AA length = 66
FEATURE           Location/Qualifiers
source            1..66
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 421
GDDPNSSDES NGNDDANSES DNNSSSRGDA SYNSDESKDN GNGSDSKGAE DDDSDSTS 60
NNSDSN          66

SEQ ID NO: 422      moltype = AA length = 66
FEATURE           Location/Qualifiers
source            1..66
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 422
DDPNSSDESN GNDDANSESD NNSSSRGDAS YNSDESKDNG NGSDSKGAED DDSDSTS 60
NSDSNG          66

SEQ ID NO: 423      moltype = AA length = 66
FEATURE           Location/Qualifiers
source            1..66
                  mol_type = protein

```

-continued

```

SEQUENCE: 423          organism = synthetic construct
AEDDDSDSTS DTNNSDSNGN GNNGNDDNDK SDSGKGKSDS SDSDSSDSSN SSDSSDSSDS 60
DSSDSN                                         66

SEQ ID NO: 424          moltype = AA length = 66
FEATURE
source
1..66
mol_type = protein
organism = synthetic construct

SEQUENCE: 424          moltype = AA length = 66
EDDDSDSTSD TNNSDSNGNG NNGNDDNDKS DSGKGKSDSS DSDSSDSSNS SDSSDSSDSD 60
SSDSNS                                         66

SEQ ID NO: 425          moltype = AA length = 21
FEATURE
source
1..21
mol_type = protein
organism = synthetic construct

SEQUENCE: 425          moltype = AA length = 21
LEEDNNENA GEDGDNDFSP S
                           21

SEQ ID NO: 426          moltype = AA length = 21
FEATURE
source
1..21
mol_type = protein
organism = synthetic construct

SEQUENCE: 426          moltype = AA length = 21
DNNENAGEDG DNDfspSDEE L
                           21

SEQ ID NO: 427          moltype = AA length = 21
FEATURE
source
1..21
mol_type = protein
organism = synthetic construct

SEQUENCE: 427          moltype = AA length = 21
NNENAGEDGD NDFSPSDEEL A
                           21

SEQ ID NO: 428          moltype = AA length = 21
FEATURE
source
1..21
mol_type = protein
organism = synthetic construct

SEQUENCE: 428          moltype = AA length = 21
NENAGEDGDN DFSPSDEELA N
                           21

SEQ ID NO: 429          moltype = AA length = 22
FEATURE
source
1..22
mol_type = protein
organism = synthetic construct

SEQUENCE: 429          moltype = AA length = 22
DNNENAGEDG DNDfspSDEE LA
                           22

SEQ ID NO: 430          moltype = AA length = 22
FEATURE
source
1..22
mol_type = protein
organism = synthetic construct

SEQUENCE: 430          moltype = AA length = 22
NNENAGEDGD NDFSPSDEEL AN
                           22

SEQ ID NO: 431          moltype = AA length = 23
FEATURE
source
1..23
mol_type = protein
organism = synthetic construct

SEQUENCE: 431          moltype = AA length = 23
AELEEDDNNE NAGEDGDNDF SPS
                           23

SEQ ID NO: 432          moltype = AA length = 23
FEATURE
source
1..23
mol_type = protein
organism = synthetic construct

SEQUENCE: 432

```

-continued

DDNNENAGED GDNDFSPSDE ELA	23
SEQ ID NO: 433 moltype = AA length = 23 FEATURE Location/Qualifiers source 1..23 mol_type = protein organism = synthetic construct	
SEQUENCE: 433 DNNENAGEDG DNDFSPSDEE LAN	23
SEQ ID NO: 434 moltype = AA length = 24 FEATURE Location/Qualifiers source 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 434 SEENNENAGED GDNDFSPSDE ELAN	24
SEQ ID NO: 435 moltype = AA length = 25 FEATURE Location/Qualifiers source 1..25 mol_type = protein organism = synthetic construct	
SEQUENCE: 435 EDDNNENAGE DGDNDFSPSD EELAN	25
SEQ ID NO: 436 moltype = AA length = 24 FEATURE Location/Qualifiers source 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 436 NPADDPNNQG EDEFEEAEQV REEN	24
SEQ ID NO: 437 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 437 NEENTEPGAE SSENADDPNK D	21
SEQ ID NO: 438 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 438 ENADDPNKDT SENADGQSDE N	21
SEQ ID NO: 439 moltype = AA length = 22 FEATURE Location/Qualifiers source 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 439 SENADDPNKG TSENADGQSD EN	22
SEQ ID NO: 440 moltype = AA length = 23 FEATURE Location/Qualifiers source 1..23 mol_type = protein organism = synthetic construct	
SEQUENCE: 440 SSENADDPNK DTSENADGQS DEN	23
SEQ ID NO: 441 moltype = AA length = 24 FEATURE Location/Qualifiers source 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 441 ESSENADDPN KDTSENADGQ SDEN	24
SEQ ID NO: 442 moltype = AA length = 20 FEATURE Location/Qualifiers	

-continued

source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 442		
DRDPEMENEE QPSSENDSQN		20
SEQ ID NO: 443	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 443		
RDPPEMENEEQ PSSENDSQNQ		20
SEQ ID NO: 444	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 444		
DPEMENEEQP SSENDSDQNS		20
SEQ ID NO: 445	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 445		
PEMENEEQPS SENDSQNQSG		20
SEQ ID NO: 446	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 446		
EMENEEQPSS ENDSQNQSGE		20
SEQ ID NO: 447	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 447		
MENEEQPSS E ND SQNQSGEQ		20
SEQ ID NO: 448	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 448		
ENE EQPSS EN DSQNQSGEQI		20
SEQ ID NO: 449	moltype = AA length = 21	
FEATURE	Location/Qualifiers	
source	1..21	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 449		
DSE SANVSDK EA GS NENDDQ N		21
SEQ ID NO: 450	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 450		
NYNDGSQEDR DWQDDQSDNQ		20
SEQ ID NO: 451	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 451		

-continued

RENTNEASSE GNSSDDSEDE	20
SEQ ID NO: 452	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 452	
ENTNEASSEG NSSDDSEDER	20
SEQ ID NO: 453	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 453	
QENDNGNETE SEQPKESNEN	20
SEQ ID NO: 454	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 454	
ENDNGNETES EQPKESNENQ	20
SEQ ID NO: 455	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 455	
NDNGNETESE QPKESNENQ	20
SEQ ID NO: 456	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 456	
QENDNGNETE SEQPKESNEN Q	21
SEQ ID NO: 457	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 457	
ENDNGNETES EQPKESNENQ E	21
SEQ ID NO: 458	moltype = AA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 458	
QENDNGNETE SEQPKESNEN QE	22
SEQ ID NO: 459	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 459	
QNEENPGDEE AKNQVNSESD	20
SEQ ID NO: 460	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 460	
NEENPGDEEA KNQVNSESDS	20
SEQ ID NO: 461	moltype = AA length = 20
FEATURE	Location/Qualifiers

-continued

source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 461		
EENPGDEEAK NQVNSESDSD		20
SEQ ID NO: 462	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 462		
ENPGDEEAKN QVNSESDSDS		20
SEQ ID NO: 463	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 463		
NPGDEEAKNQ VNSESDSDSE		20
SEQ ID NO: 464	moltype = AA length = 21	
FEATURE	Location/Qualifiers	
source	1..21	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 464		
QNEENPGDEE AKNQVNSESD S		21
SEQ ID NO: 465	moltype = AA length = 21	
FEATURE	Location/Qualifiers	
source	1..21	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 465		
NEENPGDEEA KNQVNSESD D		21
SEQ ID NO: 466	moltype = AA length = 22	
FEATURE	Location/Qualifiers	
source	1..22	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 466		
QNEENPGDEE AKNQVNSESD SD		22
SEQ ID NO: 467	moltype = AA length = 22	
FEATURE	Location/Qualifiers	
source	1..22	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 467		
NEENPGDEEA KNQVNSESDS DS		22
SEQ ID NO: 468	moltype = AA length = 23	
FEATURE	Location/Qualifiers	
source	1..23	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 468		
QNEENPGDEE AKNQVNSESD SDS		23
SEQ ID NO: 469	moltype = AA length = 23	
FEATURE	Location/Qualifiers	
source	1..23	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 469		
NEENPGDEEA KNQVNSESDS DSE		23
SEQ ID NO: 470	moltype = AA length = 24	
FEATURE	Location/Qualifiers	
source	1..24	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 470		

-continued

QNEENPGDEE AKNQVNSESD SDSE	24
SEQ ID NO: 471 moltype = AA length = 24 FEATURE Location/Qualifiers source 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 471 NEENPGDEEA KNQVNSESDS DSEE	24
SEQ ID NO: 472 moltype = AA length = 25 FEATURE Location/Qualifiers source 1..25 mol_type = protein organism = synthetic construct	
SEQUENCE: 472 QNEENPGDEE AKNQVNSESD SDSEE	25
SEQ ID NO: 473 moltype = AA length = 25 FEATURE Location/Qualifiers source 1..25 mol_type = protein organism = synthetic construct	
SEQUENCE: 473 NEENPGDEEA KNQVNSESDS DSEES	25
SEQ ID NO: 474 moltype = AA length = 26 FEATURE Location/Qualifiers source 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 474 QNEENPGDEE AKNQVNSESD SDSEES	26
SEQ ID NO: 475 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 475 YNGGNANPRP ANNEEEDEE D	21
SEQ ID NO: 476 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 476 NGGNANPRPA NNEEEEDEED E	21
SEQ ID NO: 477 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 477 GGNANPRPAN NEEEDEDE Y	21
SEQ ID NO: 478 moltype = AA length = 22 FEATURE Location/Qualifiers source 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 478 YNGGNANPRP ANNEEEDEE DE	22
SEQ ID NO: 479 moltype = AA length = 22 FEATURE Location/Qualifiers source 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 479 GGGNANPRPA NNEEEEDEED EY	22
SEQ ID NO: 480 moltype = AA length = 23 FEATURE Location/Qualifiers	

-continued

source	1..23	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 480		
GGGNANPRPA NNNEEEDEED EYD		23
SEQ ID NO: 481	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 481		
GASENEEEDD DYNKPPLDPNS		20
SEQ ID NO: 482	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 482		
LQNQKEAEEP GPDSENSQEN		20
SEQ ID NO: 483	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 483		
QNQKEAEEPQ PDSENSQENP		20
SEQ ID NO: 484	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 484		
NQKEAEEPQG DSENSQENPP		20
SEQ ID NO: 485	moltype = AA length = 21	
FEATURE	Location/Qualifiers	
source	1..21	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 485		
KESVSPENNE EGGNDNQDNE N		21
SEQ ID NO: 486	moltype = AA length = 21	
FEATURE	Location/Qualifiers	
source	1..21	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 486		
EVSPENNEE GGNDNQDNEN P		21
SEQ ID NO: 487	moltype = AA length = 22	
FEATURE	Location/Qualifiers	
source	1..22	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 487		
KESVSPENNE EGGNDNQDNE NP		22
SEQ ID NO: 488	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 488		
ASPQPREPSD DENSDNSNEC		20
SEQ ID NO: 489	moltype = AA length = 21	
FEATURE	Location/Qualifiers	
source	1..21	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 489		

-continued

NNSQDEDGFQ ELNENGNAKD E	21
SEQ ID NO: 490 moltype = AA length = 21	
FEATURE Location/Qualifiers	
source 1..21	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 490	
NSQDEDGFQE LNENGNNAKDE N	21
SEQ ID NO: 491 moltype = AA length = 22	
FEATURE Location/Qualifiers	
source 1..22	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 491	
NNSQDEDGFQ ELNENGNAKD EN	22
SEQ ID NO: 492 moltype = AA length = 20	
FEATURE Location/Qualifiers	
source 1..20	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 492	
SSDSPKDQSP PEDSGESEAD	20
SEQ ID NO: 493 moltype = AA length = 20	
FEATURE Location/Qualifiers	
source 1..20	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 493	
PGPEPEPEPE PEPEPAPAPEPE	20
SEQ ID NO: 494 moltype = AA length = 20	
FEATURE Location/Qualifiers	
source 1..20	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 494	
GPEPEPEPEP EPEPAPAPEPEP	20
SEQ ID NO: 495 moltype = AA length = 20	
FEATURE Location/Qualifiers	
source 1..20	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 495	
PEPEPEPEPE PAPEPEPEPKP	20
SEQ ID NO: 496 moltype = AA length = 20	
FEATURE Location/Qualifiers	
source 1..20	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 496	
EPEPEPEPEP APEPEPEPKP	20
SEQ ID NO: 497 moltype = AA length = 20	
FEATURE Location/Qualifiers	
source 1..20	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 497	
PEPEPEPEPA PEPEPEPKPG	20
SEQ ID NO: 498 moltype = AA length = 21	
FEATURE Location/Qualifiers	
source 1..21	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 498	
AGPGPEPEPE PEPEPEPAPAPE P	21
SEQ ID NO: 499 moltype = AA length = 21	
FEATURE Location/Qualifiers	

-continued

```

source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 499
GPGPEPEPEP EPEPEPAPPEP E                                21

SEQ ID NO: 500      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 500
PGPEPEPEPE PEPEPAPPEPE P                                21

SEQ ID NO: 501      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 501
GPEPEPEPEP EPEPAPPEPEP E                                21

SEQ ID NO: 502      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 502
PEPEPEPEPE PEPAPEPEPE P                                21

SEQ ID NO: 503      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 503
EPEPEPEPEP EPAPEPEPEP K                                21

SEQ ID NO: 504      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 504
PEPEPEPEPE PAPEPEPEPK P                                21

SEQ ID NO: 505      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 505
EPEPEPEPEP APEPEPEPEPK G                                21

SEQ ID NO: 506      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 506
PEPEPEPEPA PEPEPEPKPG A                                21

SEQ ID NO: 507      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 507
EPEPEPEPAP EPEPEPKPG A                                21

SEQ ID NO: 508      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 508

```

-continued

AGPGPEPEPE PEPEPEPAPPE PE	22
SEQ ID NO: 509 moltype = AA length = 22 FEATURE Location/Qualifiers source 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 509 GPGPEPEPEP EPEPEPAPPEP EP	22
SEQ ID NO: 510 moltype = AA length = 22 FEATURE Location/Qualifiers source 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 510 PGPEPEPEPE PEPEPAPPEPE PE	22
SEQ ID NO: 511 moltype = AA length = 22 FEATURE Location/Qualifiers source 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 511 GPEPEPEPEP EPEPAPPEPEP EP	22
SEQ ID NO: 512 moltype = AA length = 22 FEATURE Location/Qualifiers source 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 512 PEPEPEPEPE PEPAPPEPEPE PK	22
SEQ ID NO: 513 moltype = AA length = 22 FEATURE Location/Qualifiers source 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 513 EPEPEPEPEP EPAPEPEPEP KP	22
SEQ ID NO: 514 moltype = AA length = 22 FEATURE Location/Qualifiers source 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 514 PEPEPEPEPE PAPEPEPEPK PG	22
SEQ ID NO: 515 moltype = AA length = 22 FEATURE Location/Qualifiers source 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 515 EPEPEPEPEP APEPEPEPKP GA	22
SEQ ID NO: 516 moltype = AA length = 22 FEATURE Location/Qualifiers source 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 516 PEPEPEPEPA PEPEPEPKPG AG	22
SEQ ID NO: 517 moltype = AA length = 23 FEATURE Location/Qualifiers source 1..23 mol_type = protein organism = synthetic construct	
SEQUENCE: 517 AGPGPEPEPE PEPEPEPAPPE PEP	23
SEQ ID NO: 518 moltype = AA length = 23 FEATURE Location/Qualifiers	

-continued

```

source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 518
GPGPEPEPEP EPEPEPEAPEP EPE                         23

SEQ ID NO: 519      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 519
PGPEPEPEPE PEPEPAPPEPE PEP                         23

SEQ ID NO: 520      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 520
GPEPEPEPEP EPEPAPPEPEP EPK                         23

SEQ ID NO: 521      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 521
PEPEPEPEPE PEPAPEPEPEPE PKP                         23

SEQ ID NO: 522      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 522
EPEPEPEPEP EPAPEPEPEP KPG                         23

SEQ ID NO: 523      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 523
PEPEPEPEPE PAPEPEPEPEPK PGA                         23

SEQ ID NO: 524      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 524
EPEPEPEPEP APEPEPEPEPK GAG                         23

SEQ ID NO: 525      moltype = AA  length = 24
FEATURE
source          1..24
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 525
AGPGPEPEPE PEPEPEPAPPE PEPE                         24

SEQ ID NO: 526      moltype = AA  length = 24
FEATURE
source          1..24
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 526
GPGPEPEPEP EPEPEPAPPEP EPEP                         24

SEQ ID NO: 527      moltype = AA  length = 24
FEATURE
source          1..24
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 527

```

-continued

PGPEPEPEPE PEPEPAPPEPE PEPK	24
SEQ ID NO: 528 moltype = AA length = 24 FEATURE Location/Qualifiers source 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 528 GPEPEPEPEP EPEPAPPEPEP EPKP	24
SEQ ID NO: 529 moltype = AA length = 24 FEATURE Location/Qualifiers source 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 529 SEPEPEPEPE PEPAPEPEPE PKPG	24
SEQ ID NO: 530 moltype = AA length = 24 FEATURE Location/Qualifiers source 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 530 EPEPEPEPEP EPAPEPEPEP KPGA	24
SEQ ID NO: 531 moltype = AA length = 24 FEATURE Location/Qualifiers source 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 531 PEPEPEPEPE PAPEPEPEPK PGAG	24
SEQ ID NO: 532 moltype = AA length = 25 FEATURE Location/Qualifiers source 1..25 mol_type = protein organism = synthetic construct	
SEQUENCE: 532 AGPGPEPEPE PEPEPEPAPPE PEPEP	25
SEQ ID NO: 533 moltype = AA length = 25 FEATURE Location/Qualifiers source 1..25 mol_type = protein organism = synthetic construct	
SEQUENCE: 533 GPGPEPEPEP EPEPEPAPPEP EPEPK	25
SEQ ID NO: 534 moltype = AA length = 25 FEATURE Location/Qualifiers source 1..25 mol_type = protein organism = synthetic construct	
SEQUENCE: 534 PGPEPEPEPE PEPEPAPPEPE PEPKP	25
SEQ ID NO: 535 moltype = AA length = 25 FEATURE Location/Qualifiers source 1..25 mol_type = protein organism = synthetic construct	
SEQUENCE: 535 GPEPEPEPEP EPEPAPPEPEP EPKPG	25
SEQ ID NO: 536 moltype = AA length = 25 FEATURE Location/Qualifiers source 1..25 mol_type = protein organism = synthetic construct	
SEQUENCE: 536 PEPEPEPEPE PEPAPEPEPE PKPGA	25
SEQ ID NO: 537 moltype = AA length = 25 FEATURE Location/Qualifiers	

-continued

```

source          1..25
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 537
EPEPEPEPEP  EPAPEPEPEP  KPGAG                         25

SEQ ID NO: 538      moltype = AA  length = 26
FEATURE
source          1..26
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 538
AGPGPEPEPE PEPEPEPAPPE PEPEPK                          26

SEQ ID NO: 539      moltype = AA  length = 26
FEATURE
source          1..26
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 539
GPGPEPEPEP  EPEPEPAPPEP  EPEPKP                         26

SEQ ID NO: 540      moltype = AA  length = 26
FEATURE
source          1..26
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 540
PGPEPEPEPE PEPEPAPPEPE PEPKPG                         26

SEQ ID NO: 541      moltype = AA  length = 26
FEATURE
source          1..26
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 541
GPEPEPEPEP  EPEPAPPEPEP  EPKPGA                         26

SEQ ID NO: 542      moltype = AA  length = 26
FEATURE
source          1..26
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 542
PEPEPEPEPE PEPAPEPEPEPE PKPGAG                         26

SEQ ID NO: 543      moltype = AA  length = 27
FEATURE
source          1..27
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 543
AGPGPEPEPE PEPEPEPAPPE PEPEPKP                         27

SEQ ID NO: 544      moltype = AA  length = 27
FEATURE
source          1..27
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 544
PGPEPEPEPE PEPEPAPPEPE PEPKPGA                         27

SEQ ID NO: 545      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 545
DPEEDGSPDP  EPSPEPEPKP                                20

SEQ ID NO: 546      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 546

```

-continued

PEEDGSPDPE PSPEPEPKPS	20
SEQ ID NO: 547 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 547 EDPEEDGSPD PEPSPEPEPK P	21
SEQ ID NO: 548 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 548 DPEEDGSPD EPSPEPEPKP S	21
SEQ ID NO: 549 moltype = AA length = 22 FEATURE Location/Qualifiers source 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 549 EDPEEDGSPD PEPSPEPEPK PS	22
SEQ ID NO: 550 moltype = AA length = 23 FEATURE Location/Qualifiers source 1..23 mol_type = protein organism = synthetic construct	
SEQUENCE: 550 DDEPEEDGSP DPEPSPEPEPK KPS	23
SEQ ID NO: 551 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 551 PDDDGSDSS PPSASPAESE P	21
SEQ ID NO: 552 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 552 DDDGSDDSSP PSASPAESEP Q	21
SEQ ID NO: 553 moltype = AA length = 22 FEATURE Location/Qualifiers source 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 553 PDDDGSDSS PPSASPAESE PQ	22
SEQ ID NO: 554 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 554 PPPEEQQQGD APPQHEDEEP A	21
SEQ ID NO: 555 moltype = AA length = 20 FEATURE Location/Qualifiers source 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 555 PSPAPSPDSD SDSDSDGEEE	20
SEQ ID NO: 556 moltype = AA length = 21 FEATURE Location/Qualifiers	

-continued

```

source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 556
GAPSPAPSPD SDSDSDSDGE E                                21

SEQ ID NO: 557      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 557
APSPAPSPDS DSDSDSDGEE E                                21

SEQ ID NO: 558      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 558
PSPAPSPDSD SDSDSDGEEE E                                21

SEQ ID NO: 559      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 559
NGAPSPAPSP DSDSDSDG EE                                22

SEQ ID NO: 560      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 560
GAPSPAPSPD SDSDSDSDGE EE                                22

SEQ ID NO: 561      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 561
APSPAPSPDS DSDSDSDGEE EE                                22

SEQ ID NO: 562      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 562
QNGAPSPAPS PDSDDSDSD GEE                                23

SEQ ID NO: 563      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 563
NGAPSPAPSP DSDSDSDSDG EEE                                23

SEQ ID NO: 564      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 564
GAPSPAPSPD SDSDSDSDGE EEE                                23

SEQ ID NO: 565      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 565

```

-continued

APSPAPSPDS DSDSDSDGEE EEE	23
SEQ ID NO: 566 moltype = AA length = 24 FEATURE Location/Qualifiers source 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 566 QNGAPSPAPS PDSDSDSDSD GEEE	24
SEQ ID NO: 567 moltype = AA length = 24 FEATURE Location/Qualifiers source 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 567 NGAPSPAPSP DSDSDSDSDG EEEE	24
SEQ ID NO: 568 moltype = AA length = 24 FEATURE Location/Qualifiers source 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 568 GAPSPAPSPD SDSDSDSDGE EEEE	24
SEQ ID NO: 569 moltype = AA length = 25 FEATURE Location/Qualifiers source 1..25 mol_type = protein organism = synthetic construct	
SEQUENCE: 569 QNGAPSPAPS PDSDSDSDSD GEEEE	25
SEQ ID NO: 570 moltype = AA length = 25 FEATURE Location/Qualifiers source 1..25 mol_type = protein organism = synthetic construct	
SEQUENCE: 570 NGAPSPAPSP DSDSDSDSDG EEEEE	25
SEQ ID NO: 571 moltype = AA length = 25 FEATURE Location/Qualifiers source 1..25 mol_type = protein organism = synthetic construct	
SEQUENCE: 571 GAPSPAPSPD SDSDSDSDGE EEEEE	25
SEQ ID NO: 572 moltype = AA length = 26 FEATURE Location/Qualifiers source 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 572 QNGAPSPAPS PDSDSDSDSD GEEEEEE	26
SEQ ID NO: 573 moltype = AA length = 26 FEATURE Location/Qualifiers source 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 573 NGAPSPAPSP DSDSDSDSDG EEEEEEE	26
SEQ ID NO: 574 moltype = AA length = 27 FEATURE Location/Qualifiers source 1..27 mol_type = protein organism = synthetic construct	
SEQUENCE: 574 QNGAPSPAPS PDSDSDSDSD GEEEEEEE	27
SEQ ID NO: 575 moltype = AA length = 27 FEATURE Location/Qualifiers	

-continued

```

source          1..27
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 575
NGAPSPAPSP DSDSDSDSDG EEEEEEE           27

SEQ ID NO: 576      moltype = AA  length = 28
FEATURE
source          1..28
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 576
QNGAPSPAPS PDSDDSDSDSD GEEEEEE           28

SEQ ID NO: 577      moltype = AA  length = 31
FEATURE
source          1..31
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 577
QNGAPSPAPS PDSDDSDSDSD GEEEEEEGE R       31

SEQ ID NO: 578      moltype = AA  length = 32
FEATURE
source          1..32
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 578
QNGAPSPAPS PDSDDSDSDSD GEEEEEEGE RD      32

SEQ ID NO: 579      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 579
EPPAPPPPPP PEEDPEQDSG           20

SEQ ID NO: 580      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 580
PAPPPPPPPE EDPEQDSGPE           20

SEQ ID NO: 581      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 581
APPPPPPPEE DPEQDSGPED           20

SEQ ID NO: 582      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 582
PAPPPPPPPE EDPEQDSGPE D           21

SEQ ID NO: 583      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 583
EPPAPPPPPP PEEDPEQDSG PE          22

SEQ ID NO: 584      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 584

```

-continued

PPAPPPPPP EEDPEQDSGP ED	22
SEQ ID NO: 585 FEATURE source SEQUENCE: 585 AEPAPPAPP PPEEDPEQDS GPE	23
SEQ ID NO: 586 FEATURE source SEQUENCE: 586 EPPAPPPPPP PEEDPEQDSG PED	23
SEQ ID NO: 587 FEATURE source SEQUENCE: 587 AEPAPPAPP PPEEDPEQDS GPED	24
SEQ ID NO: 588 FEATURE source SEQUENCE: 588 PPPPQQAPPE EENESEPEEP	20
SEQ ID NO: 589 FEATURE source SEQUENCE: 589 PPPPQAPPEE ENESEPEEPS	20
SEQ ID NO: 590 FEATURE source SEQUENCE: 590 PPPQAPPEEE NESEPEEPSG	20
SEQ ID NO: 591 FEATURE source SEQUENCE: 591 PPPPQQAPPE EENESEPEEP S	21
SEQ ID NO: 592 FEATURE source SEQUENCE: 592 PPPPQAPPEE ENESEPEEPS G	21
SEQ ID NO: 593 FEATURE source SEQUENCE: 593 PPPPQQAPPE EENESEPEEP SG	22
SEQ ID NO: 594 FEATURE	Length = 20 Location/Qualifiers

-continued

source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 594		
PQDSSSKSPE PSADESPDND		20
SEQ ID NO: 595	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 595		
QSSDNSEDEE EPPDNADSKS		20
SEQ ID NO: 596	moltype = AA length = 21	
FEATURE	Location/Qualifiers	
source	1..21	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 596		
EEEEEQPGKA PDPQDPQDAE S		21
SEQ ID NO: 597	moltype = AA length = 21	
FEATURE	Location/Qualifiers	
source	1..21	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 597		
EEEEQPGKAP DPQDPQDAES D		21
SEQ ID NO: 598	moltype = AA length = 21	
FEATURE	Location/Qualifiers	
source	1..21	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 598		
EEEQPGKAPD PQDPQDAESD S		21
SEQ ID NO: 599	moltype = AA length = 22	
FEATURE	Location/Qualifiers	
source	1..22	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 599		
EEEEQPGKAP DPQDPQDAES DS		22
SEQ ID NO: 600	moltype = AA length = 23	
FEATURE	Location/Qualifiers	
source	1..23	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 600		
EEEEEQPGKA PDPQDPQDAE SDS		23
SEQ ID NO: 601	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 601		
PPPPSEEEGP FEEPPKASPE		20
SEQ ID NO: 602	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 602		
PPPPSEEEGP FEEPPKASPES		20
SEQ ID NO: 603	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 603		

-continued

PPPSEEEGP EPPKASPESE	20
SEQ ID NO: 604	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 604	
PPPPSEEEGP EEPKASPE E	21
SEQ ID NO: 605	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 605	
PPPSEEEGP EPPKASPE A	21
SEQ ID NO: 606	moltype = AA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 606	
PPPPSEEEG PEEPKASPE SE	22
SEQ ID NO: 607	moltype = AA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 607	
PPPPSEEEGP EEPKASPE EA	22
SEQ ID NO: 608	moltype = AA length = 23
FEATURE	Location/Qualifiers
source	1..23
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 608	
PPPPSEEEG PEEPKASPE SEA	23
SEQ ID NO: 609	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 609	
DDAEEPESPP PPPRSPSPEP	20
SEQ ID NO: 610	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 610	
SGEASSSEEE PPSPDDKENQ	20
SEQ ID NO: 611	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 611	
SGEASSSEEE PPSPDDKENQ A	21
SEQ ID NO: 612	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 612	
GEASSSEEEP PSPDDKENQA P	21
SEQ ID NO: 613	moltype = AA length = 22
FEATURE	Location/Qualifiers

-continued

```

source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 613
SGEASSSEEE PPPSPDDKENQ AP                                22

SEQ ID NO: 614      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 614
PQPQRSTUVWXYZ ESSDSEPEAE                                         20

SEQ ID NO: 615      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 615
QPPPPPPEE SSDSEPEAEP                                         20

SEQ ID NO: 616      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 616
PPPPPPPEES SDSEPEAEPG                                         20

SEQ ID NO: 617      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 617
PPPPPEESS DSEPEAEPGS                                         20

SEQ ID NO: 618      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 618
PPPPEEESSD SEPEAEPGSP                                         20

SEQ ID NO: 619      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 619
PPPEEESSDS EPEAEPGSPQ                                         20

SEQ ID NO: 620      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 620
PPPEEESSDSE PEAEPGSPQK                                         20

SEQ ID NO: 621      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 621
SDPEPPDAGE DSKSENGENA P                                         21

SEQ ID NO: 622      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 622

```

-continued

SDSESEDPPR NQASDSENEE	20
SEQ ID NO: 623 moltype = AA length = 21	
FEATURE Location/Qualifiers	
source 1..21	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 623	
GPGEDAEPDE DPQSEDSEAP S	21
SEQ ID NO: 624 moltype = AA length = 21	
FEATURE Location/Qualifiers	
source 1..21	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 624	
PGEDAEPDED PQSEDSEAPS S	21
SEQ ID NO: 625 moltype = AA length = 22	
FEATURE Location/Qualifiers	
source 1..22	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 625	
GPGEDAEPDE DPQSEDSEAP SS	22
SEQ ID NO: 626 moltype = AA length = 23	
FEATURE Location/Qualifiers	
source 1..23	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 626	
ESESSSSDSE ANEPSQSASP EPE	23
SEQ ID NO: 627 moltype = AA length = 25	
FEATURE Location/Qualifiers	
source 1..25	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 627	
DSESESSSSD SEANEPSQSA SPEPE	25
SEQ ID NO: 628 moltype = AA length = 26	
FEATURE Location/Qualifiers	
source 1..26	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 628	
SDSESESSSS DSEANEPSQS ASPEPE	26
SEQ ID NO: 629 moltype = AA length = 26	
FEATURE Location/Qualifiers	
source 1..26	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 629	
DSESESSSSD SEANEPSQSA SPEPEP	26
SEQ ID NO: 630 moltype = AA length = 23	
FEATURE Location/Qualifiers	
source 1..23	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 630	
HQEDSEEEESQ EEEAEGASEP PPP	23
SEQ ID NO: 631 moltype = AA length = 25	
FEATURE Location/Qualifiers	
source 1..25	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 631	
GDHQEDSEEE SQEEEAEAGAS EPPPP	25
SEQ ID NO: 632 moltype = AA length = 20	
FEATURE Location/Qualifiers	

-continued

```

source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 632
PSQPPEEPEP DEAESSPDPQ                                20

SEQ ID NO: 633      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 633
DPAPSQPEEE PEPDEAESSP D                                21

SEQ ID NO: 634      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 634
PAPSQPEEP EPDEAESSPD P                                21

SEQ ID NO: 635      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 635
APSQPPEPE PDEAESSPDP Q                                21

SEQ ID NO: 636      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 636
PSQPPEEPEP DEAESSPDPQ A                                21

SEQ ID NO: 637      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 637
DPAPSQPEEE PEPDEAESSP DP                                22

SEQ ID NO: 638      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 638
PAPSQPEEP EPDEAESSPD PQ                                22

SEQ ID NO: 639      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 639
DPAPSQPEEE PEPDEAESSP DPQ                               23

SEQ ID NO: 640      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 640
PGSQPQASSG PEAEEEEEEDD E                                21

SEQ ID NO: 641      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 641

```

-continued

DSPDSQEEQK GESSASSPPEE	20
SEQ ID NO: 642	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 642	
SPDSQEEQKG ESSASSPPEP	20
SEQ ID NO: 643	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 643	
PDSQEEQKGE SSASSPPEPE	20
SEQ ID NO: 644	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 644	
DSQEEQKGES SASSPPEPEE	20
SEQ ID NO: 645	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 645	
SPDPSQEEQK GESSASSPPEE P	21
SEQ ID NO: 646	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 646	
SPDSQEEQKG ESSASSPPEP E	21
SEQ ID NO: 647	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 647	
PDSQEEQKGE SSASSPPEPE E	21
SEQ ID NO: 648	moltype = AA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 648	
PADSPDSQEE QKGESSIONSSP EE	22
SEQ ID NO: 649	moltype = AA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 649	
ADSPDSQEEQ KGESSASSSPE EP	22
SEQ ID NO: 650	moltype = AA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 650	
DSPDSQEEQK GESSASSPPE PE	22
SEQ ID NO: 651	moltype = AA length = 22
FEATURE	Location/Qualifiers

-continued

source	1..22	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 651		
SPDSQEEQKG ESSASSPEEP EE		22
SEQ ID NO: 652	moltype = AA length = 23	
FEATURE	Location/Qualifiers	
source	1..23	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 652		
PADSPDSQEE QKGESSIONSP EEP		23
SEQ ID NO: 653	moltype = AA length = 23	
FEATURE	Location/Qualifiers	
source	1..23	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 653		
ADSPDSQEEQ KGESSASSPE EPE		23
SEQ ID NO: 654	moltype = AA length = 23	
FEATURE	Location/Qualifiers	
source	1..23	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 654		
DSPDSQEEQK GESSIONSPE PEE		23
SEQ ID NO: 655	moltype = AA length = 24	
FEATURE	Location/Qualifiers	
source	1..24	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 655		
PADSPDSQEE QKGESSIONSP EEPE		24
SEQ ID NO: 656	moltype = AA length = 24	
FEATURE	Location/Qualifiers	
source	1..24	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 656		
ADSPDSQEEQ KGESSASSPE EPEE		24
SEQ ID NO: 657	moltype = AA length = 25	
FEATURE	Location/Qualifiers	
source	1..25	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 657		
PADSPDSQEE QKGESSIONSP EEEPE		25
SEQ ID NO: 658	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 658		
SPSPDEDESS SSSSSSEDE		20
SEQ ID NO: 659	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 659		
PSPDEDESSS SSSSSSEDEE		20
SEQ ID NO: 660	moltype = AA length = 23	
FEATURE	Location/Qualifiers	
source	1..23	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 660		

-continued

PRSPSPEDES SSSSSSSSE DEE	23
SEQ ID NO: 661 moltype = AA length = 24 FEATURE Location/Qualifiers source 1..24 mol_type = protein organism = synthetic construct	
SEQUENCE: 661 PRSPSPEDES SSSSSSSSE DEE	24
SEQ ID NO: 662 moltype = AA length = 20 FEATURE Location/Qualifiers source 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 662 SSSDSSSDSDS SEDDEAPSXP	20
SEQ ID NO: 663 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 663 PSPGSPRGQP QDQDDDEDDE E	21
SEQ ID NO: 664 moltype = AA length = 20 FEATURE Location/Qualifiers source 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 664 QQAEDHPQNP PEDPNQDPPE	20
SEQ ID NO: 665 moltype = AA length = 20 FEATURE Location/Qualifiers source 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 665 QAEDHPQNP PEDPNQDPPE	20
SEQ ID NO: 666 moltype = AA length = 20 FEATURE Location/Qualifiers source 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 666 AEDHPQNP PEDPNQDPPE	20
SEQ ID NO: 667 moltype = AA length = 20 FEATURE Location/Qualifiers source 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 667 EDHPQNP PEDPNQDPPE	20
SEQ ID NO: 668 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 668 QQAEDHPQNP PEDPNQDPPE D	21
SEQ ID NO: 669 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 669 QAEDHPQNP PEDPNQDPPE D	21
SEQ ID NO: 670 moltype = AA length = 21 FEATURE Location/Qualifiers	

-continued

```

source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 670
AEDHPQNPE DPNQDPPEDD S                                21

SEQ ID NO: 671      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 671
PQQAEADHPQN PPEDPNQDPP ED                                22

SEQ ID NO: 672      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 672
QQAEDHQPQNP PEDPNQDPPE DD                                22

SEQ ID NO: 673      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 673
QAEDHPQNPP EDPNQDPPE DS                                22

SEQ ID NO: 674      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 674
DGEAGPQQAE DHPQNPPEDP NQD                                23

SEQ ID NO: 675      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 675
GPQQAEDHQPQ NPPEDPNQDP PED                                23

SEQ ID NO: 676      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 676
PQQAEADHPQN PPEDPNQDPP EDD                                23

SEQ ID NO: 677      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 677
QQAEADHPQN PPEDPNQDPPE DDS                                23

SEQ ID NO: 678      moltype = AA  length = 24
FEATURE
source          1..24
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 678
GPQQAEDHQPQ NPPEDPNQDP PEDD                                24

SEQ ID NO: 679      moltype = AA  length = 24
FEATURE
source          1..24
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 679

```

-continued

PQQQAEDHPQN PPEDPNQDPP EDDDS	24
SEQ ID NO: 680 moltype = AA length = 25 FEATURE Location/Qualifiers source 1..25 mol_type = protein organism = synthetic construct	
SEQUENCE: 680 EAGPQQAEDH PQNPPEDPNQ DPPED	25
SEQ ID NO: 681 moltype = AA length = 25 FEATURE Location/Qualifiers source 1..25 mol_type = protein organism = synthetic construct	
SEQUENCE: 681 AGPQQAEDHP QNPPEPDNPQD PPEDD	25
SEQ ID NO: 682 moltype = AA length = 25 FEATURE Location/Qualifiers source 1..25 mol_type = protein organism = synthetic construct	
SEQUENCE: 682 GPQQAEDHPQ NPPEDPNQDP PEDDS	25
SEQ ID NO: 683 moltype = AA length = 26 FEATURE Location/Qualifiers source 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 683 DGEAGPQQAE DHPQNPPEDP NQDPPE	26
SEQ ID NO: 684 moltype = AA length = 26 FEATURE Location/Qualifiers source 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 684 GEAGPQQAEH HPQNPPEDPN QDPPED	26
SEQ ID NO: 685 moltype = AA length = 26 FEATURE Location/Qualifiers source 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 685 EAGPQQAEDH PQNPPEDPNQ DPPEDD	26
SEQ ID NO: 686 moltype = AA length = 26 FEATURE Location/Qualifiers source 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 686 AGPQQAEDHP QNPPEPDNPQD PPEDDS	26
SEQ ID NO: 687 moltype = AA length = 27 FEATURE Location/Qualifiers source 1..27 mol_type = protein organism = synthetic construct	
SEQUENCE: 687 DGEAGPQQAE DHPQNPPEDP NQDPPE	27
SEQ ID NO: 688 moltype = AA length = 27 FEATURE Location/Qualifiers source 1..27 mol_type = protein organism = synthetic construct	
SEQUENCE: 688 GEAGPQQAEH HPQNPPEDPN QDPPEDD	27
SEQ ID NO: 689 moltype = AA length = 27 FEATURE Location/Qualifiers	

-continued

source	1..27	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 689		
EAGPQQAEH PQNPPEDPNQ DPPEDDS		27
SEQ ID NO: 690	moltype = AA length = 28	
FEATURE	Location/Qualifiers	
source	1..28	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 690		
DGEAGPQQAE DHPQNPPEDP NQDPPEDD		28
SEQ ID NO: 691	moltype = AA length = 28	
FEATURE	Location/Qualifiers	
source	1..28	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 691		
GEAGPQQAEH HPQNPPEDPN QDPPEDDS		28
SEQ ID NO: 692	moltype = AA length = 29	
FEATURE	Location/Qualifiers	
source	1..29	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 692		
DGEAGPQQAE DHPQNPPEDP NQDPPEDDS		29
SEQ ID NO: 693	moltype = AA length = 31	
FEATURE	Location/Qualifiers	
source	1..31	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 693		
PAGDGEAPQ QAEDHPQNPP EDPNQDPPE D		31
SEQ ID NO: 694	moltype = AA length = 31	
FEATURE	Location/Qualifiers	
source	1..31	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 694		
AGDGEAGPQ AEDHPQNPP EDPNQDPPE DS		31
SEQ ID NO: 695	moltype = AA length = 32	
FEATURE	Location/Qualifiers	
source	1..32	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 695		
PAGDGEAPQ QAEDHPQNPP EDPNQDPPE DS		32
SEQ ID NO: 696	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 696		
QNKEAEEPG PDSENSQENP		20
SEQ ID NO: 697	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 697		
NQKEAEEPGP DSENSQENPP		20
SEQ ID NO: 698	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 698		

-continued

DQNESQSPQE PEEGPSEDDK	20
SEQ ID NO: 699 FEATURE source SEQUENCE: 699 QNESQSPQEP EEGPSEDDKA	20
SEQ ID NO: 700 FEATURE source SEQUENCE: 700 NESQSPQEPE EGPSEDDKAE	20
SEQ ID NO: 701 FEATURE source SEQUENCE: 701 DQNESQSPQE PEEGPSEDDKA A	21
SEQ ID NO: 702 FEATURE source SEQUENCE: 702 QNESQSPQEP EEGPSEDDKA E	21
SEQ ID NO: 703 FEATURE source SEQUENCE: 703 NESQSPQEPE EGPSEDDKAE G	21
SEQ ID NO: 704 FEATURE source SEQUENCE: 704 ESQSPQEPEE GPSEDDKAEG E	21
SEQ ID NO: 705 FEATURE source SEQUENCE: 705 SQSPQEPEEG PSEDDKAEGE E	21
SEQ ID NO: 706 FEATURE source SEQUENCE: 706 DQNESQSPQE PEEGPSEDDK AE	22
SEQ ID NO: 707 FEATURE source SEQUENCE: 707 QNESQSPQEP EEGPSEDDKA EG	22
SEQ ID NO: 708 FEATURE SEQUENCE: 708 QNESQSPQEP EEGPSEDDKA EG	22

-continued

```

source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 708
NESQSPQEP EGPSEDDKAE GE                                22

SEQ ID NO: 709      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 709
DQNESQSPQE PEEGPSEDDK AEG                               23

SEQ ID NO: 710      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 710
QNESQSPQEP EEGPSEDDKA EGE                               23

SEQ ID NO: 711      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 711
NESQSPQEP EGPSEDDKAE GEE                               23

SEQ ID NO: 712      moltype = AA  length = 24
FEATURE
source          1..24
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 712
DQNESQSPQE PEEGPSEDDK AEGE                             24

SEQ ID NO: 713      moltype = AA  length = 24
FEATURE
source          1..24
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 713
QNESQSPQEP EEGPSEDDKA EGEE                            24

SEQ ID NO: 714      moltype = AA  length = 25
FEATURE
source          1..25
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 714
DQNESQSPQE PEEGPSEDDK AECEE                           25

SEQ ID NO: 715      moltype = AA  length = 25
FEATURE
source          1..25
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 715
QNESQSPQEP EEGPSEDDKA EGEPEE                         25

SEQ ID NO: 716      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 716
XEASSSEEEP PSPDDKENQA P                                21

SEQ ID NO: 717      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 717

```

-continued

PASSSNPEE SPEEDRREAES E	21
SEQ ID NO: 718 moltype = AA length = 20 FEATURE Location/Qualifiers source 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 718 DKPEEDDEA QQPQPQSGPE	20
SEQ ID NO: 719 moltype = AA length = 20 FEATURE Location/Qualifiers source 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 719 KPEEEDEAQ QQPQPQSGPEE	20
SEQ ID NO: 720 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 720 AGE GDKPEEE DDEAQQPQPQ S	21
SEQ ID NO: 721 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 721 EGDKPEEEED EAQQPQPQSG P	21
SEQ ID NO: 722 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 722 GDKPEEEEDDE AQQQPQPQSGP E	21
SEQ ID NO: 723 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 723 DPKEEEDEAQ QQPQPQSGPE	21
SEQ ID NO: 724 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 724 KPKEEEDEAQ QQPQPQSGPEE A	21
SEQ ID NO: 725 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 725 PEEDDEAQO PQPQSGPEEA E	21
SEQ ID NO: 726 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 726 EEDDEAQQPQ PQSGPEEAEE G	21
SEQ ID NO: 727 moltype = AA length = 21 FEATURE Location/Qualifiers	

-continued

```

source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 727
EDDEAQQPQPO QSGPEEAEEG E                                21

SEQ ID NO: 728      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 728
DDEAQQPQPO SGPEEAEEG E                                21

SEQ ID NO: 729      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 729
DEAQQPQPOSG GPEEAEEGEE E                                21

SEQ ID NO: 730      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 730
EAQQPQPOQSG PEEAEEGEEE E                                21

SEQ ID NO: 731      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 731
QQQPQPOQSGPE EAEEGEEEEAA E                               21

SEQ ID NO: 732      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 732
QPQPQSGPEE AEEGEEEEAE R                                21

SEQ ID NO: 733      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 733
EGDKPEEEEDD EAQQPQPOQSG PE                                22

SEQ ID NO: 734      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 734
GDKPEEEEDDE AQQPQPOQSGP EE                                22

SEQ ID NO: 735      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 735
DKPEEEEDDEA QQPQPOQSGPE EA                                22

SEQ ID NO: 736      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 736

```

-continued

KPEEEEDDEAQ QPQPQSGPEE AE	22
SEQ ID NO: 737 moltype = AA length = 22	
FEATURE Location/Qualifiers	
source 1..22	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 737	
QQQPQPSGPE EAEEGEEEEA ER	22
SEQ ID NO: 738 moltype = AA length = 23	
FEATURE Location/Qualifiers	
source 1..23	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 738	
EGDKPEEEEDD EAQQPQPQSG PEE	23
SEQ ID NO: 739 moltype = AA length = 23	
FEATURE Location/Qualifiers	
source 1..23	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 739	
GDKPEEEEDDE AQQPQPQSGP EEA	23
SEQ ID NO: 740 moltype = AA length = 23	
FEATURE Location/Qualifiers	
source 1..23	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 740	
DKPEEEEDDEA QQQPQPSGPE EAE	23
SEQ ID NO: 741 moltype = AA length = 23	
FEATURE Location/Qualifiers	
source 1..23	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 741	
KPEEEEDDEAQ QPQPQSGPEE AEE	23
SEQ ID NO: 742 moltype = AA length = 23	
FEATURE Location/Qualifiers	
source 1..23	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 742	
PEEEEDDEAQ PQPQPSGPEA EEG	23
SEQ ID NO: 743 moltype = AA length = 24	
FEATURE Location/Qualifiers	
source 1..24	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 743	
EGDKPEEEEDD EAQQPQPQSG PEEA	24
SEQ ID NO: 744 moltype = AA length = 24	
FEATURE Location/Qualifiers	
source 1..24	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 744	
GDKPEEEEDDE AQQPQPQSGP EEA	24
SEQ ID NO: 745 moltype = AA length = 24	
FEATURE Location/Qualifiers	
source 1..24	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 745	
KPEEEEDDEAQ QPQPQSGPEE AEEG	24
SEQ ID NO: 746 moltype = AA length = 25	
FEATURE Location/Qualifiers	

-continued

source	1..25	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 746		
EGDKPEEEDD EAQQQPQPSG PEEAE		25
SEQ ID NO: 747	moltype = AA length = 25	
FEATURE	Location/Qualifiers	
source	1..25	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 747		
GDKPEEEDDE AQQPQPQSGP EEAEE		25
SEQ ID NO: 748	moltype = AA length = 25	
FEATURE	Location/Qualifiers	
source	1..25	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 748		
DKPEEEDDEA QQQPQPQSGPE EAEEG		25
SEQ ID NO: 749	moltype = AA length = 25	
FEATURE	Location/Qualifiers	
source	1..25	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 749		
KPEEEDDEAQ QQPQPQSGPSEE		25
SEQ ID NO: 750	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 750		
PEEEHAPGED ESSPQPSQPS		20
SEQ ID NO: 751	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 751		
XPPPEESSDS EPEAEPGSPQK		20
SEQ ID NO: 752	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 752		
PPPEESSDSE PEAEPGSPQK		20
SEQ ID NO: 753	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 753		
KEPKEEKKDD DEEAKPKSSD		20
SEQ ID NO: 754	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 754		
SEEKKPPEED KEEEEEKKAP		20
SEQ ID NO: 755	moltype = AA length = 21	
FEATURE	Location/Qualifiers	
source	1..21	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 755		

-continued

PAEEDDEDPE QEKEAGEPGR P	21
SEQ ID NO: 756	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 756	
PEPEPEPEPE PEPAPEPEPE	20
SEQ ID NO: 757	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 757	
EPEPEPEPEP EPAPEPEPEP	20
SEQ ID NO: 758	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 758	
DEDPEEDGSP DPEPSPEPEP	20
SEQ ID NO: 759	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 759	
DEDPEDGSPD PEPSPEPEPK	20
SEQ ID NO: 760	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 760	
DEDPEEDGSP DPEPSPEPEPK	21
SEQ ID NO: 761	moltype = AA length = 22
FEATURE	Location/Qualifiers
source	1..22
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 761	
DEDPEEDGSP DPEPSPEPEPK	22
SEQ ID NO: 762	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 762	
KPEDKDPRD EESKEPKEEK	20
SEQ ID NO: 763	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 763	
PEDKDPRDPE ESKEPKEEKQ	20
SEQ ID NO: 764	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 764	
EDKDPRDPEE SKEPKEEKQR	20
SEQ ID NO: 765	moltype = AA length = 20
FEATURE	Location/Qualifiers

-continued

```

source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 765
KRNDSEEEER ERDEEQEPPP                                20

SEQ ID NO: 766      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 766
PEEEPDDQDA PDEHEPSPSE                                20

SEQ ID NO: 767      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 767
EEEPDDQDAP DEHEPSPSED                                20

SEQ ID NO: 768      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 768
PEEEPDDQDA PDEHEPSPSE D                                21

SEQ ID NO: 769      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 769
EEEPDDQDAP DEHEPSPSED A                                21

SEQ ID NO: 770      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 770
EEPDDQDAP EHEPSPSEDA P                                21

SEQ ID NO: 771      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 771
PEEEPDDQDA PDEHEPSPSE DA                                22

SEQ ID NO: 772      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 772
EEEPDDQDAP DEHEPSPSED AP                                22

SEQ ID NO: 773      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 773
PEEEPDDQDA PDEHEPSPSE DAP                               23

SEQ ID NO: 774      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 774

```

-continued

SPAPSPDSDS DSDSDGEEEE	20
SEQ ID NO: 775 FEATURE source SEQUENCE: 775 PAPSPDSD SDSDGEEEEEE	20
SEQ ID NO: 776 FEATURE source SEQUENCE: 776 PSPAPSPDSD SDSDSDGEEE EE	22
SEQ ID NO: 777 FEATURE source SEQUENCE: 777 PSPAPSPDSD SDSDSDGEEE EEE	23
SEQ ID NO: 778 FEATURE source SEQUENCE: 778 APSPAPSPDS DSDSDSDGEE EEEE	24
SEQ ID NO: 779 FEATURE source SEQUENCE: 779 PSPAPSPDSD SDSDSDGEEE EEEE	24
SEQ ID NO: 780 FEATURE source SEQUENCE: 780 APSPAPSPDS DSDSDSDGEE EEEEE	25
SEQ ID NO: 781 FEATURE source SEQUENCE: 781 PSPAPSPDSD SDSDSDGEEE EEEEG	25
SEQ ID NO: 782 FEATURE source SEQUENCE: 782 GAPSPAPSPD SDSDSDSDGE EEEEEEE	26
SEQ ID NO: 783 FEATURE source SEQUENCE: 783 APSPAPSPDS DSDSDSDGEE EEEEEE	26
SEQ ID NO: 784 FEATURE SEQUENCE: 784 APSPAPSPDS DSDSDSDGEE EEEEEE	26

-continued

source	1..26	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 784		
PSPAPSPDSD SDSDSDGEEE EEEEGE		26
SEQ ID NO: 785	moltype = AA length = 21	
FEATURE	Location/Qualifiers	
source	1..21	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 785		
EKNDEDEPQK PEDKGDPGP E		21
SEQ ID NO: 786	moltype = AA length = 22	
FEATURE	Location/Qualifiers	
source	1..22	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 786		
EKNDEDEPQK PEDKGDPGP EA		22
SEQ ID NO: 787	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 787		
DEEEEEEEEEE EEEDEGPAPP		20
SEQ ID NO: 788	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 788		
DEEEEEEEEEE EEDEGPAPPSS		20
SEQ ID NO: 789	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 789		
EREPDPPDDR DASDGEDEKP		20
SEQ ID NO: 790	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 790		
REPDPPDDR ASDGEDEKPP		20
SEQ ID NO: 791	moltype = AA length = 21	
FEATURE	Location/Qualifiers	
source	1..21	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 791		
GEREPDPPDD RDASDGEDEK P		21
SEQ ID NO: 792	moltype = AA length = 21	
FEATURE	Location/Qualifiers	
source	1..21	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 792		
EREPDPPDDR DASDGEDEKP P		21
SEQ ID NO: 793	moltype = AA length = 22	
FEATURE	Location/Qualifiers	
source	1..22	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 793		

-continued

EGEREPDPPD DRDASDGEDE KP	22
SEQ ID NO: 794 FEATURE source SEQUENCE: 794	moltype = AA length = 22 Location/Qualifiers 1..22 mol_type = protein organism = synthetic construct
GEREPDPPDD RDASDGEDEK PP	22
SEQ ID NO: 795 FEATURE source SEQUENCE: 795	moltype = AA length = 23 Location/Qualifiers 1..23 mol_type = protein organism = synthetic construct
EGEREPDPPD DRDASDGEDE KPP	23
SEQ ID NO: 796 FEATURE source SEQUENCE: 796	moltype = AA length = 22 Location/Qualifiers 1..22 mol_type = protein organism = synthetic construct
EEEEEQPGKA PDPQDPQDAE SD	22
SEQ ID NO: 797 FEATURE source SEQUENCE: 797	moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct
QEPEEKQEPE EKQEPEEKQK	20
SEQ ID NO: 798 FEATURE source SEQUENCE: 798	moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct
EPEEKQEPEEE KQEPEEKQKP	20
SEQ ID NO: 799 FEATURE source SEQUENCE: 799	moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct
PEEKQEPEEK QEPEEKQKPE	20
SEQ ID NO: 800 FEATURE source SEQUENCE: 800	moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct
EKQEPEEKQ EPEEKQKPEA	20
SEQ ID NO: 801 FEATURE source SEQUENCE: 801	moltype = AA length = 21 Location/Qualifiers 1..21 mol_type = protein organism = synthetic construct
EPEEKQEPEEE KQEPEEKQKP E	21
SEQ ID NO: 802 FEATURE source SEQUENCE: 802	moltype = AA length = 22 Location/Qualifiers 1..22 mol_type = protein organism = synthetic construct
QEPEEKQEPE EKQEPEEKQK PE	22
SEQ ID NO: 803 FEATURE	moltype = AA length = 22 Location/Qualifiers

-continued

```

source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 803
EPEEKQEPEE KQEPEEKQKP EA                                22

SEQ ID NO: 804      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 804
AGDDDDDDDD SPDPESPDDDS                                         20

SEQ ID NO: 805      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 805
GDDDDDDDDDS PDPESPDDSE                                         20

SEQ ID NO: 806      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 806
DDDDDDDDDSP DPESPDDSES                                         20

SEQ ID NO: 807      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 807
DDDDDDDDSPD PESPDDSESD                                         20

SEQ ID NO: 808      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 808
DDDDDDSPDP ESPDDSESDS                                         20

SEQ ID NO: 809      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 809
DDDDDDSPDPE SPDDSESDSE                                         20

SEQ ID NO: 810      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 810
DDDDSPDPES PDDSESDSES                                         20

SEQ ID NO: 811      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 811
DDDSPDPESP DDSESDSESE                                         20

SEQ ID NO: 812      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 812

```

-continued

DDSPDPESP D	DSESDSE S E K E K	20
SEQ ID NO: 813	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 813		
DSPDPESP D D S E SDSE S E K E K E		20
SEQ ID NO: 814	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 814		
SPDPESP D D S E S DSE S E K E K E		20
SEQ ID NO: 815	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 815		
PDPE P PD S D S E K E K E K E S E		20
SEQ ID NO: 816	moltype = AA length = 21	
FEATURE	Location/Qualifiers	
source	1..21	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 816		
DPDQP R EDPA EEEKEEK D A P E		21
SEQ ID NO: 817	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 817		
EGK P EN E SEP KHEE P K P EE K PEE K PEE K		20
SEQ ID NO: 818	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 818		
PENE E SEP K H K E EEPK P EEK K PEE K PEE K		20
SEQ ID NO: 819	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 819		
ENE E SEP K H K E EEPK P EEK K PEE K PEE K		20
SEQ ID NO: 820	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 820		
NE E SEP K H K EEE PK P EEK K PEE K PEE K		20
SEQ ID NO: 821	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 821		
ESEPKH E EEP K PEE K PEEE K		20
SEQ ID NO: 822	moltype = AA length = 20	
FEATURE	Location/Qualifiers	

-continued

```

source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 822
SEPKHEEEPK PEEKPEEEEK                                         20

SEQ ID NO: 823      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 823
PENESEPKHEE EEPKPEEKPE E                                         21

SEQ ID NO: 824      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 824
ENESEPKHEE EPKPEEKPEE E                                         21

SEQ ID NO: 825      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 825
NESEPKHEEE PKPEEKPEEE E                                         21

SEQ ID NO: 826      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 826
ESEPKHEEEEP KPEEKPEEEE K                                         21

SEQ ID NO: 827      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 827
KPNENESEPKH EEEPKPEEKPE EE                                         22

SEQ ID NO: 828      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 828
PENESEPKHEE EEPKPEEKPE EE                                         22

SEQ ID NO: 829      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 829
ENESEPKHEE EPKPEEKPEE EE                                         22

SEQ ID NO: 830      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 830
NESEPKHEEE PKPEEKPEEE EK                                         22

SEQ ID NO: 831      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 831

```

-continued

KPENESEPKH EEEPKEEKPE EEE	23
SEQ ID NO: 832 moltype = AA length = 23	
FEATURE Location/Qualifiers	
source 1..23	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 832	
PENESEPKHE EEPKPEEKPE EEE	23
SEQ ID NO: 833 moltype = AA length = 23	
FEATURE Location/Qualifiers	
source 1..23	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 833	
ENESEPKHEE EPKPEEKPEE EEK	23
SEQ ID NO: 834 moltype = AA length = 24	
FEATURE Location/Qualifiers	
source 1..24	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 834	
EGKPENESEP KHEEKPKEE KPEE	24
SEQ ID NO: 835 moltype = AA length = 24	
FEATURE Location/Qualifiers	
source 1..24	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 835	
GKPENESEPK HEEEPKPEEK PEEE	24
SEQ ID NO: 836 moltype = AA length = 24	
FEATURE Location/Qualifiers	
source 1..24	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 836	
KPENESEPKH EEEPKEEKPE EEEE	24
SEQ ID NO: 837 moltype = AA length = 24	
FEATURE Location/Qualifiers	
source 1..24	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 837	
PENESEPKHE EEPKPEEKPE EEEEK	24
SEQ ID NO: 838 moltype = AA length = 25	
FEATURE Location/Qualifiers	
source 1..25	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 838	
EGKPENESEP KHEEKPKEE KPEEE	25
SEQ ID NO: 839 moltype = AA length = 25	
FEATURE Location/Qualifiers	
source 1..25	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 839	
GKPENESEPK HEEEPKPEEK PEEEE	25
SEQ ID NO: 840 moltype = AA length = 25	
FEATURE Location/Qualifiers	
source 1..25	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 840	
KPENESEPKH EEEPKEEKPE EEEEK	25
SEQ ID NO: 841 moltype = AA length = 26	
FEATURE Location/Qualifiers	

-continued

```

source          1..26
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 841
EGKPENESEP KHEEKPKEE KPEEEE                         26

SEQ ID NO: 842      moltype = AA  length = 26
FEATURE
source          1..26
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 842
GKPENESEPK HEEEPKPEEK PEEEEE                         26

SEQ ID NO: 843      moltype = AA  length = 27
FEATURE
source          1..27
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 843
EGKPENESEP KHEEKPKEE KPEEEEK                         27

SEQ ID NO: 844      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 844
KPPEEDPEEQ AEENPEGEQP                                20

SEQ ID NO: 845      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 845
PPEEDPEEQA EENPEGEQPE                               20

SEQ ID NO: 846      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 846
PEEDPEEQAE ENPEGEQPEE                               20

SEQ ID NO: 847      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 847
KPPEEDPEEQ AEENPEGEQP E                            21

SEQ ID NO: 848      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 848
PPEEDPEEQA EENPEGEQPE E                            21

SEQ ID NO: 849      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 849
KPPEEDPEEQ AEENPEGEQP EE                           22

SEQ ID NO: 850      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 850

```

-continued

PDDDESEDH DDPDNAHESP	20
SEQ ID NO: 851	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 851	
GEPEPEPEPE PEPEPESEPE	20
SEQ ID NO: 852	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 852	
EPEPEPEPEP EPEPESEPEP	20
SEQ ID NO: 853	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 853	
PEPEPEPEPE PEPESEPEPE	20
SEQ ID NO: 854	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 854	
EPEPEPEPEP EPESEPEPEP	20
SEQ ID NO: 855	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 855	
PEPEPEPEPE PESEPEPEPE	20
SEQ ID NO: 856	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 856	
GGEPEPEPEP EPEPEPESEP E	21
SEQ ID NO: 857	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 857	
GEPEPEPEPE PEPEPESEPE P	21
SEQ ID NO: 858	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 858	
EPEPEPEPEP EPEPESEPEP E	21
SEQ ID NO: 859	moltype = AA length = 21
FEATURE	Location/Qualifiers
source	1..21
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 859	
PEPEPEPEPE PEPESEPEPE P	21
SEQ ID NO: 860	moltype = AA length = 21
FEATURE	Location/Qualifiers

-continued

```

source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 860
EPEPEPEPEP EPESEPEPEP E                               21

SEQ ID NO: 861      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 861
GGEPEPEPEP EPEPEPESEP EP                             22

SEQ ID NO: 862      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 862
GEPEPEPEPE PEPEPESEPE PE                             22

SEQ ID NO: 863      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 863
EPEPEPEPEP EPEPESEPEP EP                            22

SEQ ID NO: 864      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 864
PEPEPEPEPE PEPEPESEPE PE                            22

SEQ ID NO: 865      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 865
GGEPEPEPEP EPEPEPESEP EPE                           23

SEQ ID NO: 866      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 866
GEPEPEPEPE PEPEPESEPE PEP                          23

SEQ ID NO: 867      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 867
EPEPEPEPEP EPEPESEPEP EPE                           23

SEQ ID NO: 868      moltype = AA  length = 24
FEATURE
source          1..24
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 868
GGEPEPEPEP EPEPEPESEP EPEP                         24

SEQ ID NO: 869      moltype = AA  length = 24
FEATURE
source          1..24
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 869

```

-continued

GEPEPEPEPE PEPEPESEPE PEPE	24
SEQ ID NO: 870 moltype = AA length = 25 FEATURE Location/Qualifiers source 1..25 mol_type = protein organism = synthetic construct	
SEQUENCE: 870 GGEPEPEPEP EPEPEPESEP EPEPE	25
SEQ ID NO: 871 moltype = AA length = 20 FEATURE Location/Qualifiers source 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 871 PEEEPDDQDA PDEHESPPPE	20
SEQ ID NO: 872 moltype = AA length = 20 FEATURE Location/Qualifiers source 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 872 EEEPDDQDAP DEHESPPPED	20
SEQ ID NO: 873 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 873 PEEEPDDQDA PDEHESPPPE D	21
SEQ ID NO: 874 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 874 EEEPDDQDAP DEHESPPPED A	21
SEQ ID NO: 875 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 875 EEPDDQDAP EHESPPPEDA P	21
SEQ ID NO: 876 moltype = AA length = 22 FEATURE Location/Qualifiers source 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 876 PEEEPDDQDA PDEHESPPPE DA	22
SEQ ID NO: 877 moltype = AA length = 22 FEATURE Location/Qualifiers source 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 877 EEEPDDQDAP DEHESPPPED AP	22
SEQ ID NO: 878 moltype = AA length = 23 FEATURE Location/Qualifiers source 1..23 mol_type = protein organism = synthetic construct	
SEQUENCE: 878 PEEEPDDQDA PDEHESPPPE DAP	23
SEQ ID NO: 879 moltype = AA length = 20 FEATURE Location/Qualifiers	

-continued

```

source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 879
GPSSDDENEE ESKPEKEDEP                                20

SEQ ID NO: 880      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 880
PSSDDENEEE SKPEKEDEPQ                                20

SEQ ID NO: 881      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 881
SDDSDSEKRR PEEQEEEPQP                                20

SEQ ID NO: 882      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 882
DDSDSEKRRP EEQEEEPQPR                                20

SEQ ID NO: 883      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 883
DPREGEEEEE EDEPDPEAPE                                20

SEQ ID NO: 884      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 884
PREGEEEEEEE DEPDPEAPEN                                20

SEQ ID NO: 885      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 885
DEEPPGDPREG EEEEEEEDEPD P                                21

SEQ ID NO: 886      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 886
EPPGDPREGEE EEEEEEEDEPDP E                                21

SEQ ID NO: 887      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 887
EPGDPREGEE EEEEDEPDP A                                21

SEQ ID NO: 888      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 888

```

-continued

PGDPREGE EE EDEPDPEA P	21
SEQ ID NO: 889 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 889	
GDPREGE EE EDEPDPEAP E	21
SEQ ID NO: 890 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 890	
D P REGE EE EE EDEPDPEAPE N	21
SEQ ID NO: 891 moltype = AA length = 21 FEATURE Location/Qualifiers source 1..21 mol_type = protein organism = synthetic construct	
SEQUENCE: 891	
P R EGEEEEEE DEPDPEAPE N G	21
SEQ ID NO: 892 moltype = AA length = 22 FEATURE Location/Qualifiers source 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 892	
DEE E EPGDPRE GEEEEEEDEP DP	22
SEQ ID NO: 893 moltype = AA length = 22 FEATURE Location/Qualifiers source 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 893	
DEEP G DPREG EEEEEEEDEPD PE	22
SEQ ID NO: 894 moltype = AA length = 22 FEATURE Location/Qualifiers source 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 894	
EEP G DPREGE EEEEEEDEPDP EA	22
SEQ ID NO: 895 moltype = AA length = 22 FEATURE Location/Qualifiers source 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 895	
EPGDPREGE E EEEEDEPDPE AP	22
SEQ ID NO: 896 moltype = AA length = 22 FEATURE Location/Qualifiers source 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 896	
PGDPREGE EE EDEPDPEA PE	22
SEQ ID NO: 897 moltype = AA length = 22 FEATURE Location/Qualifiers source 1..22 mol_type = protein organism = synthetic construct	
SEQUENCE: 897	
GDPREGE EE EE EDEPDPEAP EN	22
SEQ ID NO: 898 moltype = AA length = 22 FEATURE Location/Qualifiers	

-continued

```

source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 898
DPREGEEEEEEE EDEPDPEAPE NG                                22

SEQ ID NO: 899      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 899
PREGEEEEEEE DEPDPEAPEN GS                                22

SEQ ID NO: 900      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 900
PEEEDEEPGD PREGEEEEEEE DEP                                23

SEQ ID NO: 901      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 901
EDEEPGD PRE GEGEEEEEEDE PDP                                23

SEQ ID NO: 902      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 902
EDEEPGD PREE GEEEEEDEP DPE                                23

SEQ ID NO: 903      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 903
DEEPGDPREG EEEEEEEDEPD PEA                                23

SEQ ID NO: 904      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 904
EEPGDPREG EEEEDEPDP EAP                                23

SEQ ID NO: 905      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 905
EPGDPREGEE EEEEDEPDPE APE                                23

SEQ ID NO: 906      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 906
PGDPREGEEE EEEDEPDPEA PEN                                23

SEQ ID NO: 907      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 907

```

-continued

DPREGE EEEEEE EDEPDPEAPE NGS	23
SEQ ID NO: 908	moltype = AA length = 24
FEATURE	Location/Qualifiers
source	1..24
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 908	
APEEEDEEPG DPREG EEEEEE EDEP	24
SEQ ID NO: 909	moltype = AA length = 24
FEATURE	Location/Qualifiers
source	1..24
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 909	
PEEDEE PGD PREG EEEEEE DEPD	24
SEQ ID NO: 910	moltype = AA length = 24
FEATURE	Location/Qualifiers
source	1..24
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 910	
EEDEE PGD P REGE EEEEEE EPDP	24
SEQ ID NO: 911	moltype = AA length = 24
FEATURE	Location/Qualifiers
source	1..24
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 911	
EDEE PGD P REGE EEEEEE DEPE	24
SEQ ID NO: 912	moltype = AA length = 24
FEATURE	Location/Qualifiers
source	1..24
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 912	
EDEE PGD P REGE GEE EEEEEE DPEA	24
SEQ ID NO: 913	moltype = AA length = 24
FEATURE	Location/Qualifiers
source	1..24
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 913	
DEE PGD P REG EEE EEEEEE DEPD PEAP	24
SEQ ID NO: 914	moltype = AA length = 24
FEATURE	Location/Qualifiers
source	1..24
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 914	
E PGD P REG EEE EEEEEE DEPD EAPE	24
SEQ ID NO: 915	moltype = AA length = 24
FEATURE	Location/Qualifiers
source	1..24
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 915	
E PGD P REG EEE EEEEEE DEPD APEN	24
SEQ ID NO: 916	moltype = AA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 916	
AAPEEEDEEP GDPREG EEEEEE EEDEP	25
SEQ ID NO: 917	moltype = AA length = 25
FEATURE	Location/Qualifiers

-continued

```

source          1..25
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 917
APEEEDEEPG DPREGEEEEEE EDEPD                                25

SEQ ID NO: 918      moltype = AA  length = 25
FEATURE
source          1..25
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 918
PEEDEEPEPGD PREGEEEEEE DEPDP                                25

SEQ ID NO: 919      moltype = AA  length = 25
FEATURE
source          1..25
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 919
EEDEEPEPGDP REGEeeeeeee EPDPE                                25

SEQ ID NO: 920      moltype = AA  length = 25
FEATURE
source          1..25
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 920
EEDEEPEPGDPRE EGEEEEEEED EPDPE                                25

SEQ ID NO: 921      moltype = AA  length = 25
FEATURE
source          1..25
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 921
EDEEPEPGDPRE GEEEEEEEDEP DPEAP                                25

SEQ ID NO: 922      moltype = AA  length = 25
FEATURE
source          1..25
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 922
DEEPEPGDPREG EEEEEEEDEPD PEAPE                                25

SEQ ID NO: 923      moltype = AA  length = 25
FEATURE
source          1..25
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 923
EEPGDPREGE EEEEEEEDEPD EAPEN                                25

SEQ ID NO: 924      moltype = AA  length = 26
FEATURE
source          1..26
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 924
AAPEEEDEEP GDPREGEEEEEE EEDEPD                                26

SEQ ID NO: 925      moltype = AA  length = 26
FEATURE
source          1..26
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 925
APEEEDEEPG DPREGEEEEEE EDEPD                                26

SEQ ID NO: 926      moltype = AA  length = 26
FEATURE
source          1..26
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 926

```

-continued

PEEEDEEPEGD PREGEEEEEE DEPDPE	26
SEQ ID NO: 927 moltype = AA length = 26 FEATURE Location/Qualifiers source 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 927 EEDEEPEGPDP REGEeeeeed EPDPEA	26
SEQ ID NO: 928 moltype = AA length = 26 FEATURE Location/Qualifiers source 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 928 EDEEPEGPDR EGeeeeeeed PDPEAP	26
SEQ ID NO: 929 moltype = AA length = 26 FEATURE Location/Qualifiers source 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 929 EDEEPEGPDPRE GEEEEEEDEP DPEAPE	26
SEQ ID NO: 930 moltype = AA length = 26 FEATURE Location/Qualifiers source 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 930 DEEPEGPDPREG EEEEEEDEPD PEAPEN	26
SEQ ID NO: 931 moltype = AA length = 27 FEATURE Location/Qualifiers source 1..27 mol_type = protein organism = synthetic construct	
SEQUENCE: 931 AAPEEEDEEP GDPREGEeee EEDEPDP	27
SEQ ID NO: 932 moltype = AA length = 27 FEATURE Location/Qualifiers source 1..27 mol_type = protein organism = synthetic construct	
SEQUENCE: 932 AEEEEEDEEP DPREGEEEEE EDEPDPE	27
SEQ ID NO: 933 moltype = AA length = 27 FEATURE Location/Qualifiers source 1..27 mol_type = protein organism = synthetic construct	
SEQUENCE: 933 PEEEDEEPEGD PREGEEEEEE DEPDPEA	27
SEQ ID NO: 934 moltype = AA length = 27 FEATURE Location/Qualifiers source 1..27 mol_type = protein organism = synthetic construct	
SEQUENCE: 934 EEDEEPEGPDP REGEeeeeed EPDPEAP	27
SEQ ID NO: 935 moltype = AA length = 27 FEATURE Location/Qualifiers source 1..27 mol_type = protein organism = synthetic construct	
SEQUENCE: 935 EEDEEPEGPDR EGeeeeeeed PDPEAPE	27
SEQ ID NO: 936 moltype = AA length = 27 FEATURE Location/Qualifiers	

-continued

```

source          1..27
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 936
EDEEPGDPRE GEEEEEEDEP DPEAPEN                                27

SEQ ID NO: 937      moltype = AA  length = 28
FEATURE
source          1..28
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 937
AAPEEEDEEP GDPREGEEEE EEDEPDPE                                28

SEQ ID NO: 938      moltype = AA  length = 28
FEATURE
source          1..28
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 938
APEEEDEEPG DPREGEEEEEE EDEPDPEA                                28

SEQ ID NO: 939      moltype = AA  length = 28
FEATURE
source          1..28
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 939
PEEEDEEPGD PREGEEEEEE DEPDPEAP                                28

SEQ ID NO: 940      moltype = AA  length = 28
FEATURE
source          1..28
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 940
EEDEEPGDP REGEEEEEEED EPDPEAPE                                28

SEQ ID NO: 941      moltype = AA  length = 28
FEATURE
source          1..28
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 941
EEDEEPGDPR EGEEEEEEDE PDPEAPEN                                28

SEQ ID NO: 942      moltype = AA  length = 29
FEATURE
source          1..29
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 942
APEEEDEEPG DPREGEEEEEE EDEPDPEAP                                29

SEQ ID NO: 943      moltype = AA  length = 29
FEATURE
source          1..29
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 943
PEEEDEEPGD PREGEEEEEE DEPDPEAPE                                29

SEQ ID NO: 944      moltype = AA  length = 29
FEATURE
source          1..29
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 944
EEDEEPGDP REGEEEEEEED EPDPEAPEN                                29

SEQ ID NO: 945      moltype = AA  length = 30
FEATURE
source          1..30
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 945

```

-continued

APEEEDEEPG DPREGEEEE EDEPDPEAPE	30
SEQ ID NO: 946 moltype = AA length = 30	
FEATURE Location/Qualifiers	
source 1..30	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 946	
PEEDEEPGD PREGEEEEEE DEPDPEAPEN	30
SEQ ID NO: 947 moltype = AA length = 31	
FEATURE Location/Qualifiers	
source 1..31	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 947	
PRGAAAPEEE DEEPGDPREG EEEEEDEPD P	31
SEQ ID NO: 948 moltype = AA length = 31	
FEATURE Location/Qualifiers	
source 1..31	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 948	
RGAAAPEEE EEPGDPREGE EEEEEDEPDP E	31
SEQ ID NO: 949 moltype = AA length = 31	
FEATURE Location/Qualifiers	
source 1..31	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 949	
AAPEEEDEEP GDPREGEEEE EEDEPDPEAP E	31
SEQ ID NO: 950 moltype = AA length = 31	
FEATURE Location/Qualifiers	
source 1..31	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 950	
AAPEEEDEEP DPREGEEEE EDEPDPEAPE N	31
SEQ ID NO: 951 moltype = AA length = 31	
FEATURE Location/Qualifiers	
source 1..31	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 951	
PEEDEEPGD PREGEEEEEE DEPDPEAPEN G	31
SEQ ID NO: 952 moltype = AA length = 31	
FEATURE Location/Qualifiers	
source 1..31	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 952	
EEDEEPGDP REGEEEEEEED EPDPEAPENG S	31
SEQ ID NO: 953 moltype = AA length = 32	
FEATURE Location/Qualifiers	
source 1..32	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 953	
PRGAAAPEEE DEEPGDPREG EEEEEDEPD PE	32
SEQ ID NO: 954 moltype = AA length = 32	
FEATURE Location/Qualifiers	
source 1..32	
mol_type = protein	
organism = synthetic construct	
SEQUENCE: 954	
AAPEEEDEEP GDPREGEEE EEDEPDPEAP EN	32
SEQ ID NO: 955 moltype = AA length = 32	
FEATURE Location/Qualifiers	

-continued

source	1..32	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 955		
APEEEDEEPG DPREGEEEEEE EDEPDPEAPE NG		32
SEQ ID NO: 956	moltype = AA length = 32	
FEATURE	Location/Qualifiers	
source	1..32	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 956		
PEEDEEPEPGD PREGEEEEEE DEPDPEAPEN GS		32
SEQ ID NO: 957	moltype = AA length = 33	
FEATURE	Location/Qualifiers	
source	1..33	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 957		
AAPEEEDEEP GDPREGEEEEEE EEDEPDPEAP ENG		33
SEQ ID NO: 958	moltype = AA length = 33	
FEATURE	Location/Qualifiers	
source	1..33	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 958		
AAPEEEDEEP DPREGEEEEEE EDEPDPEAPE NGS		33
SEQ ID NO: 959	moltype = AA length = 34	
FEATURE	Location/Qualifiers	
source	1..34	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 959		
AAPEEEDEEP GDPREGEEEEEE EEDEPDPEAP ENGS		34
SEQ ID NO: 960	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 960		
KPPPSEGSDEE EEEEDEEDE		20
SEQ ID NO: 961	moltype = AA length = 20	
FEATURE	Location/Qualifiers	
source	1..20	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 961		
PPPSEGSDEE EEEEDEEDE		20
SEQ ID NO: 962	moltype = AA length = 24	
FEATURE	Location/Qualifiers	
source	1..24	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 962		
PPPSEGSDEE EEEEDEEDE ERKP		24
SEQ ID NO: 963	moltype = AA length = 25	
FEATURE	Location/Qualifiers	
source	1..25	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 963		
KPPPSEGSDEE EEEEDEEDE EERKP		25
SEQ ID NO: 964	moltype = AA length = 25	
FEATURE	Location/Qualifiers	
source	1..25	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 964		

-continued

PPPSEGSDEE EEEEDEEDEE ERKPQ	25
SEQ ID NO: 965 FEATURE source SEQUENCE: 965	moltype = AA length = 26 Location/Qualifiers 1..26 mol_type = protein organism = synthetic construct
KPPPSEGSDDE EEEEEDEEDE EERKPQ	26
SEQ ID NO: 966 FEATURE source SEQUENCE: 966	moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct
EEEEEEEEEE EEEEEEEAPP	20
SEQ ID NO: 967 FEATURE source SEQUENCE: 967	moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct
EEEEEEEEEE EEEEEAPP	20
SEQ ID NO: 968 FEATURE source SEQUENCE: 968	moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct
EEEEEEEEEE EEEEAPP	20
SEQ ID NO: 969 FEATURE source SEQUENCE: 969	moltype = AA length = 21 Location/Qualifiers 1..21 mol_type = protein organism = synthetic construct
EEEEEEEEEE EEEEEAPP P	21
SEQ ID NO: 970 FEATURE source SEQUENCE: 970	moltype = AA length = 21 Location/Qualifiers 1..21 mol_type = protein organism = synthetic construct
EEEEEEEEEE EEEEEAPP R	21
SEQ ID NO: 971 FEATURE source SEQUENCE: 971	moltype = AA length = 22 Location/Qualifiers 1..22 mol_type = protein organism = synthetic construct
EEEEEEEEEE EEEEEAPP PP	22
SEQ ID NO: 972 FEATURE source SEQUENCE: 972	moltype = AA length = 22 Location/Qualifiers 1..22 mol_type = protein organism = synthetic construct
EEEEEEEEEE EEEEEAPP PR	22
SEQ ID NO: 973 FEATURE source SEQUENCE: 973	moltype = AA length = 23 Location/Qualifiers 1..23 mol_type = protein organism = synthetic construct
EEEEEEEEEE EEEEEEEAP PPP	23
SEQ ID NO: 974 FEATURE	moltype = AA length = 23 Location/Qualifiers

-continued

```

source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 974
EEEEEEEEEE EEEEEEEAPP PPR                                23

SEQ ID NO: 975      moltype = AA  length = 24
FEATURE
source          1..24
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 975
EEEEEEEEEE EEEEEEEEEE PPPP                                24

SEQ ID NO: 976      moltype = AA  length = 24
FEATURE
source          1..24
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 976
EEEEEEEEEE EEEEEEEAP PPPR                                24

SEQ ID NO: 977      moltype = AA  length = 25
FEATURE
source          1..25
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 977
EEEEEEEEEE EEEEEEEEEE APPPP                                25

SEQ ID NO: 978      moltype = AA  length = 25
FEATURE
source          1..25
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 978
EEEEEEEEEE EEEEEEEEEE PPPPR                               25

SEQ ID NO: 979      moltype = AA  length = 26
FEATURE
source          1..26
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 979
EEEEEEEEEE EEEEEEEEEE EAPPPP                                26

SEQ ID NO: 980      moltype = AA  length = 26
FEATURE
source          1..26
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 980
EEEEEEEEEE EEEEEEEEEE APPPPR                               26

SEQ ID NO: 981      moltype = AA  length = 33
FEATURE
source          1..33
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 981
PDDDEEDEEE EEEEEEEEEE EEEEEEEAP PPP                                33

SEQ ID NO: 982      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 982
EDSEEEESQEE EAEGASEEPPP P                                21

SEQ ID NO: 983      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 983

```

-continued

QEDSEEEESQE EEAEGASEPP PP	22
SEQ ID NO: 984 FEATURE source SEQUENCE:	moltype = AA length = 24 Location/Qualifiers 1..24 mol_type = protein organism = synthetic construct
DHQEDSEEEES QEEEAEAGASE PPPP	24
SEQ ID NO: 985 FEATURE source SEQUENCE:	moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct
DQSEEEEEEE KHPPKPAKPE	20
SEQ ID NO: 986 FEATURE source SEQUENCE:	moltype = AA length = 21 Location/Qualifiers 1..21 mol_type = protein organism = synthetic construct
PAPAHRPPED EGEENEGEED E	21
SEQ ID NO: 987 FEATURE source SEQUENCE:	moltype = AA length = 22 Location/Qualifiers 1..22 mol_type = protein organism = synthetic construct
PAPAHRPPED EGEENEGEED EE	22
SEQ ID NO: 988 FEATURE source SEQUENCE:	moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct
QENGQREEEE EKEPEPEAEPPE	20
SEQ ID NO: 989 FEATURE source SEQUENCE:	moltype = AA length = 24 Location/Qualifiers 1..24 mol_type = protein organism = synthetic construct
PAEGQENGQR EEEEEEEKEPE AEPP	24
SEQ ID NO: 990 FEATURE source SEQUENCE:	moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct
EQEPEPEPEP EPEPEPEPEP	20
SEQ ID NO: 991 FEATURE source SEQUENCE:	moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct
QEPEPEPEPE PEPEPEPEPE	20
SEQ ID NO: 992 FEATURE source SEQUENCE:	moltype = AA length = 20 Location/Qualifiers 1..20 mol_type = protein organism = synthetic construct
EPEPEPEPEP EPEPEPEPEQ	20
SEQ ID NO: 993 FEATURE	moltype = AA length = 21 Location/Qualifiers

-continued

```

source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 993
EQEPEPEPEP EPEPEPEPEP E                                21

SEQ ID NO: 994      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 994
QEPEPEPEPE PEPEPEPEPE Q                                21

SEQ ID NO: 995      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 995
EQEPEPEPEP EPEPEPEPEP EQ                                22

SEQ ID NO: 996      moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 996
PEEEEEEEEEE EEPASPPERK                                20

SEQ ID NO: 997      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 997
PSPGSPRGQP QDQDDDEDDE ED                                22

SEQ ID NO: 998      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 998
PSPGSPRGQP QDQDDDEDDE EDE                               23

SEQ ID NO: 999      moltype = AA  length = 24
FEATURE
source          1..24
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 999
RSPSPGSPRGQ PQDQDDDEDDE EEDE                               24

SEQ ID NO: 1000     moltype = AA  length = 24
FEATURE
source          1..24
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 1000
PSPGSPRGQP QDQDDDEDDE EDEA                               24

SEQ ID NO: 1001     moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 1001
AEDDDEEDEE EEEEEPDPDP                                20

SEQ ID NO: 1002     moltype = AA  length = 20
FEATURE
source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 1002

```

-continued

EDDDDEEDEEE EEEEPDPDPE	20
SEQ ID NO: 1003	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 1003	
KQEPPDPEED KEENKDDSAS	20
SEQ ID NO: 1004	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 1004	
QEPPDPEEDK EENKDDSASK	20
SEQ ID NO: 1005	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 1005	
EDEDEDESSE EDSDEEDEPPP	20
SEQ ID NO: 1006	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 1006	
DEDEDESSEE DSEDEEPPPK	20
SEQ ID NO: 1007	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 1007	
EDEDESEED SEDEEPPPKR	20
SEQ ID NO: 1008	moltype = AA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 1008	
DEDESSEEDS EDEEPPPKRR	20
SEQ ID NO: 1009	moltype = AA length = 26
FEATURE	Location/Qualifiers
source	1..26
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 1009	
PDDSRDEDDED EDESSEEDSE DEEPPP	26
SEQ ID NO: 1010	moltype = AA length = 30
FEATURE	Location/Qualifiers
source	1..30
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 1010	
PKKEPDDSRD EDEDEDESSE EDSEDEEPPP	30
SEQ ID NO: 1011	moltype = AA length = 31
FEATURE	Location/Qualifiers
source	1..31
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 1011	
PKKEPDDSRD EDEDEDESSE EDSEDEEPPP K	31
SEQ ID NO: 1012	moltype = AA length = 32
FEATURE	Location/Qualifiers

-continued

```

source          1..32
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 1012
PKKEPDDSRD EDEDEDESSE EDSEDEEPPP KR                         32

SEQ ID NO: 1013      moltype = AA  length = 33
FEATURE
source          1..33
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 1013
PKKEPDDSRD EDEDEDESSE EDSEDEEPPP KRR                        33

SEQ ID NO: 1014      moltype = AA  length = 21
FEATURE
source          1..21
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 1014
PEEEAAEEEE EEEERPKPSR P                                     21

SEQ ID NO: 1015      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 1015
QPEEEAAEEE EEEEERPKPS RP                                    22

SEQ ID NO: 1016      moltype = AA  length = 23
FEATURE
source          1..23
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 1016
EQPEEEAAEE EEEEEERPKP SRP                                 23

SEQ ID NO: 1017      moltype = AA  length = 24
FEATURE
source          1..24
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 1017
EEQPEEEAAE EEEEEEEERPK PSRP                                24

SEQ ID NO: 1018      moltype = AA  length = 25
FEATURE
source          1..25
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 1018
EEBQPEEEAA EEEEEEEERP KPSRP                               25

SEQ ID NO: 1019      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 1019
PEEEEDDEAQO PQPQPSGPEA EE                                    22

SEQ ID NO: 1020      moltype = AA  length = 24
FEATURE
source          1..24
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 1020
DKPEEEEDDEA QQPQPQPSGPE EAEE                                24

SEQ ID NO: 1021      moltype = AA  length = 24
FEATURE
source          1..24
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 1021

```

-continued

PEEEEDDEAQQ PQPQSGPEEA EEEGE	24
SEQ ID NO: 1022 moltype = AA length = 25 FEATURE Location/Qualifiers source 1..25 mol_type = protein organism = synthetic construct	
SEQUENCE: 1022 PEEEEDDEAQQ PQPQSGPEEA EEEGE	25
SEQ ID NO: 1023 moltype = AA length = 26 FEATURE Location/Qualifiers source 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 1023 SGDKPEEEEDD EAQQQPQPQSG PEEAEE	26
SEQ ID NO: 1024 moltype = AA length = 26 FEATURE Location/Qualifiers source 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 1024 DKPEEEEDDEA QQQPQPQSGPE EAEEGE	26
SEQ ID NO: 1025 moltype = AA length = 26 FEATURE Location/Qualifiers source 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 1025 KPEEEEDDEAQ QPQPQSGPEE AEEEGE	26
SEQ ID NO: 1026 moltype = AA length = 26 FEATURE Location/Qualifiers source 1..26 mol_type = protein organism = synthetic construct	
SEQUENCE: 1026 PEEEEDDEAQQ PQPQSGPEEA EEEEEE	26
SEQ ID NO: 1027 moltype = AA length = 32 FEATURE Location/Qualifiers source 1..32 mol_type = protein organism = synthetic construct	
SEQUENCE: 1027 PEEEEDDEAQQ PQPQSGPEEA EEEEEEAE GP	32
SEQ ID NO: 1028 moltype = AA length = 33 FEATURE Location/Qualifiers source 1..33 mol_type = protein organism = synthetic construct	
SEQUENCE: 1028 KPEEEEDDEAQ QPQPQSGPEE AEEEEEEAE RGP	33
SEQ ID NO: 1029 moltype = AA length = 20 FEATURE Location/Qualifiers source 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 1029 NNSEEEEDDD DEEEEPDKPP	20
SEQ ID NO: 1030 moltype = AA length = 20 FEATURE Location/Qualifiers source 1..20 mol_type = protein organism = synthetic construct	
SEQUENCE: 1030 NSEEEEDDD DEEEEPDKPPA	20
SEQ ID NO: 1031 moltype = AA length = 20 FEATURE Location/Qualifiers	

-continued

```

source          1..20
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 1031
SEEEDDDDE EEEPDKPPAN                                         20

SEQ ID NO: 1032      moltype = AA  length = 125
FEATURE
source          1..125
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 1032
QVQLQQSGGG LVQPGGSLRL SCAASGFTLD YYAIGWFRQA PGKEREGLC ISSGGSTNY 60
ADSVVKGRFTI SRDNAKNTVY LQMNSLKPED TAVYYCAADD LRCGSNWSSY FRGSGWQGTQ 120
VTVSS                                         125

SEQ ID NO: 1033      moltype = AA  length = 241
FEATURE
source          1..241
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 1033
AEVQLLESGG GLVQPGGSLR LSCAASGFTF SSYAMSWVRQ APGKGLEWVS YIASGGDTTN 60
YADSVVKGRFTI ISRDNNSKNTL YLQMNLSLRAE DTAVYYCAKG ASAFTDWGQQ TLTVSSGGG 120
GSGGGGGSGGG GSTDIQMTQS PSSLASVGD RVTITCRASQ SISSYLNWYQ QKPGKAPKLL 180
IYAAASYLQSG VPSRFSGSGS GTDFTLTISS LQPEDFATYY CQQSSNDPYT FGQGTTKEIK 240
R                                         241

SEQ ID NO: 1034      moltype = AA  length = 22
FEATURE
source          1..22
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 1034
MARVKKDQAE PLHRKFERQP PG                                         22

SEQ ID NO: 1035      moltype = AA  length = 242
FEATURE
source          1..242
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 1035
EVQLLESGGG LVQPGGSLRL SCAASGFTFS SYAMSWVRQA PGKGLEWVSA ISGSGGSTYY 60
ADSVVKGRFTI SRDNNSKNTLY LQMNLSLRAED TAVYYCAKMR LGLFDYWGQQ TLTVSSGGG 120
GSGGGGGSGGG GEIVLTQSPG TLSLSPGERA TLSCRASQSV SSSYLAWYQQ KPGQAPRLI 180
YGASSRATGI PDRFSGSGSG TDFTLTISRL EPEDFAVYYC QQRGDVPPTF GQGTTKEIKA 240
AA                                         242

SEQ ID NO: 1036      moltype = AA  length = 49
FEATURE
source          1..49
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 1036
YPYDVPDYAG SPQPVEDGED EFCTPMACEA NIQSGDSAAP MSAVHRHRL                                         49

SEQ ID NO: 1037      moltype = AA  length = 49
FEATURE
source          1..49
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 1037
NNYDVPDNAG SPQPQEDGED EFNNNPQANE A NQQSGDSNNP NSAVNRHNN                                         49

SEQ ID NO: 1038      moltype = AA  length = 50
FEATURE
source          1..50
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 1038
MTRGNQRELA RQKNMKKQSD SVKGKRRDDG LSAAARKQRD SEIMQQKQKK                                         50

SEQ ID NO: 1039      moltype = AA  length = 83
FEATURE
source          1..83
               mol_type = protein

```

-continued

```

SEQUENCE: 1039          organism = synthetic construct
MSGPNGLGM PVEAGAEGEE DGFGEAEYAA INSMLDQINS CLDHLEEKND HLHARLQELL 60
ESNRQTRLEF QQQLGEAPSD ASP                         83

SEQ ID NO: 1040          moltype = AA length = 99
FEATURE
source
1..99
mol_type = protein
organism = synthetic construct

SEQUENCE: 1040
MAQGQRKFQA HKPAKSKTAA AASEKNRGP KGGRVIAPKK ARVVQQQKLK KNLEVGIRKK 60
IEHDVVMKAS SSLPKKLALL KAPAKKKGAA AATSSKTPS                         99

SEQ ID NO: 1041          moltype = AA length = 35
FEATURE
source
1..35
mol_type = protein
organism = synthetic construct

SEQUENCE: 1041
NDEYSDFEDS DFDGDYKDSD EDYKDDSENF DDGFE                               35

SEQ ID NO: 1042          moltype = AA length = 35
FEATURE
source
1..35
mol_type = protein
organism = synthetic construct

SEQUENCE: 1042
PDEPPDPEDP DPDPDPKDSD EDPKDDPEPP DDPPE                                35

SEQ ID NO: 1043          moltype = AA length = 35
FEATURE
source
1..35
mol_type = protein
organism = synthetic construct

SEQUENCE: 1043
NDEPPDPEDS DPDPDPKDSD EDPKDDSEPP DDGPE                               35

```

1. A peptide, wherein
 - (a) 5% or more and less than 45% of amino acids contained in an amino acid sequence thereof are acidic amino acids, and
 - (b) 20% or more of the amino acids contained in the amino acid sequence are amino acids selected from the group consisting of F, P, Y, G, S, Q, N, and A, and the peptide is capable of reducing an aggregation property in a cell of a protein linked to the peptide.
2. The peptide according to claim 1, wherein
 - (c) 30% or less of the amino acids contained in the amino acid sequence are amino acids selected from the group consisting of M, T, W, C, I, V, and L.
3. The peptide according to claim 1, wherein
 - (d) each of A and G constitutes less than 10% of the amino acids contained in the amino acid sequence.
4. The peptide according to claim 1, wherein
 - (a) 20% or more and less than 45% of the amino acids contained in the amino acid sequence are acidic amino acids,
 - (b) 30% or more and less than 70% of the amino acids contained in the amino acid sequence are amino acids selected from the group consisting of F, P, Y, G, S, Q, N, and A,
 - (c) 20% or less of the amino acids contained in the amino acid sequence are amino acids selected from the group consisting of M, T, W, C, I, V, and L, and
 - (d) each A and G constitutes less than 10% of the amino acids contained in the amino acid sequence.
5. A peptide having an amino acid sequence set forth in any one of SEQ ID NOs: 2 to 11.
6. A nucleic acid encoding the peptide according to claim 1.
7. A protein expression vector comprising: the nucleic acid according to claim 6 operably linked to a regulatory sequence; and a nucleic acid encoding a protein of interest in-frame to the nucleic acid according to claim 6.
8. The protein expression vector according to claim 7, wherein the protein of interest is an antibody, or an antigen-binding fragment of an antibody.
9. The protein expression vector according to claim 8, wherein the antigen-binding fragment of the antibody is a single chain Fv (scFv).
10. A fusion protein of the peptide according to claim 1 and a protein of interest.
11. The fusion protein according to claim 10, wherein the protein of interest is an antibody, or an antigen-binding fragment of an antibody.
12. The fusion protein according to claim 11, wherein the antigen-binding fragment of the antibody is a single chain Fv (scFv).
13. A protein-producing cell comprising: the nucleic acid according to claim 6 operably linked to a regulatory sequence; and a nucleic acid encoding a protein of interest in-frame to the nucleic acid according to claim 6.
14. A method for selecting or identifying an amino acid sequence, comprising:

- acquiring an amino acid sequence in which:
- (a) 5% or more and less than 45% of amino acids contained in the amino acid sequence are acidic amino acids; and
 - (b) 20% or more of the amino acids contained in the amino acid sequence are amino acids selected from the group consisting of F, P, Y, G, S, Q, N, and A; selecting or identifying an amino acid sequence of a peptide tag that, when a fusion protein of the peptide tag having the selected or identified amino acid sequence and a reference protein is expressed in a mammal cell, provides reduction of a proportion of cells in which the fusion protein forms an aggregation, or the proportion which is not more than a predetermined value; and obtaining the peptide tag having the amino acid sequence or a nucleic acid encoding the peptide tag.
- 15.** The method according to claim **14**, wherein the amino acid sequence to be acquired a peptide wherein 30% or more of the amino acids contained in the amino acid sequence are amino acids selected from the group consisting of F, P, Y, G, S, Q, N, and A.
- 16.** The method according to claim **14**, wherein the amino acid sequence to be acquired is a group of amino acid sequences encoded by coding regions of human genome.
- 17.** The method according to claim **14**, wherein the amino acid sequence to be acquired contains a neo-antigen.
- 18.** The peptide of claim **1**, wherein 30% or more of the amino acids contained in the amino acid sequence are amino acids selected from the group consisting of F, P, Y, G, S, Q, N, and A.
- 19.** The peptide according to claim **1**, wherein 20% or less of the amino acids contained in the amino acid sequence are amino acids selected from the group consisting of M, T, W, C, I, V, and L.
- 20.** The peptide according to claim **1**, wherein 15% or less of the amino acids contained in the amino acid sequence are amino acids selected from the group consisting of M, T, W, C, I, V, and L.
- 21.** The peptide according to claim **1**, wherein 10% or less of the amino acids contained in the amino acid sequence are amino acids selected from the group consisting of M, T, W, C, I, V, and L.
- 22.** The method of claim **14**, wherein the fusion protein of the peptide tag having the selected or identified amino acid sequence and a reference protein is expressed in a human cell.

* * * * *