



(51) International Patent Classification:

C07K 19/00 (2006.01) *A61K 38/26* (2006.01)
C12N 15/62 (2006.01) *A61P 3/10* (2006.01)

(21) International Application Number:

PCT/CN2012/001723

(22) International Filing Date:

24 December 2012 (24.12.2012)

(25) Filing Language:

English

(26) Publication Language:

English

(71) Applicant: **BEIJING ANXINHUAIDE BIOTECH. CO., LTD** [CN/CN]; Floor 2, Building No. 5, JiaJie Tech. Park, No. 26 Western Ring Southern Road, Beijing 100176 (CN).

(72) Inventors: **QIAN, Xinguo**; Link Park 4-1916, 8 Cultural Park Weastern Road, Economic and Technical Development Zone, Beijing 100176 (CN). **HONG, Wei**; ZhiLan-MingShi 11-1-702, Yuhua District, Shijiazhuang, Hebei 053600 (CN). **MA, Xiaoyu**; Building 22-422, Shuangqiao Farm 6 Jin, Chaoyang District, Beijing 100024 (CN).

(74) Agent: **CHINA PATENT AGENT (H.K.) LTD**; 22/F, Great Eagle Centre, 23 Harbour Road, Wanchai, Hong Kong (CN).

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

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

Published:

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

(54) Title: IMPROVING THE HALF LIFE OF A THERAPEUTIC POLYPEPTIDE BY FUSING WITH A TRIMERIC SCAFFOLD PROTEIN VIA A SPACER

(57) Abstract: A fusion protein comprises of a therapeutic polypeptide and a scaffold protein. The therapeutic polypeptide and the scaffold protein could be connected by a spacer polypeptide. The scaffold protein forms a stable trimeric structure in solution. Also provided are nucleic acids, vectors, pharmaceutical compositions and methods of improving the pharmacokinetic property.



Improving the half life of a therapeutic polypeptide by fusing with a
trimeric scaffold protein via a spacer

Field of the Invention

5 The present invention relates generally to a fusion protein with therapeutic efficacy. In particular, the present invention relates to a method of improving the half life of a therapeutic polypeptide by fusing with a trimeric scaffold protein via a spacer.

10 Background of the Invention

Most human proteins, except antibodies and human serum albumin (HSA), are quickly cleared from human circulation. Short plasma half life times are commonly due to fast renal clearance as well as to enzymatic degradation occurring during systemic circulation. Many therapeutic
15 polypeptides require long half life time to achieve their optimal efficacy. Increasing the *in vivo* residence times of therapeutic polypeptides could decrease their dosing frequencies.

PEGylation has been widely utilized to extend the half life of a therapeutic polypeptide (see review paper ¹⁻⁴, patents 1-9). PEGylation
20 changes the physical and chemical properties of the biomedical molecule, such as its conformation, electrostatic binding, and hydrophobicity, and results in an improvement in the pharmacokinetic behavior of the drug. In general, PEGylation improves drug solubility and decreases immunogenicity. PEGylation also increases drug stability and the retention time of the
25 conjugates in blood. However, PEGylation has severe consequences for the biological activities of the protein. The activity of the PEGylated protein usually reduces by 20-50 fold ^{2,5}(patents 1-9). In addition, the site for PEGylation needs to be carefully decided to avoid interfering with the active site of the therapeutic polypeptide. For some short peptides such as GLP-1,
30 PTH and Calcitonin, it would be difficult to choose the proper site for PEGylation without disturbing the biological activity of the peptides.

It has been reported that fusion of a therapeutic polypeptide with human

IgG Fc fragment or human serum albumin (HSA) may significantly increase the half life of the therapeutic polypeptide ⁶⁻⁹ (patents 10, 11, 12). However, recombinant fusion protein with IgG Fc fragment or HSA needs to be produced from eukaryotic systems such as mammalian cell lines or yeast cells, which significantly increases the cost of the recombinant protein.

Summary of the invention

In one aspect, the present invention provides a fusion protein comprising a therapeutic polypeptide fused to a scaffold protein which forms a homo-trimer in solution, wherein the therapeutic polypeptide is connected with the scaffold protein via a spacer, and the spacer comprises a flexible un-structured linker whose length is adjustable.

In some embodiments, the spacer further comprises a proteinous connecting moiety. In a particular embodiment, the proteinous connecting moiety is a proteinous sequence having an elongated shape, such as human Fibronectin type III domain. In some embodiments, the proteinous connecting moiety is connected with the therapeutic polypeptide and/or the scaffold protein via a flexible un-structured linker.

In some embodiments, the fusion protein of the invention comprises, from N-terminus to C-terminus, the therapeutic polypeptide, a first flexible un-structured linker (linker 1), a proteinous connecting moiety (preferably a proteinous sequence with an elongated shape), a second flexible un-structured linker (linker 2), and the scaffold protein. In some embodiments, the fusion protein of the invention comprises, from N-terminus to C-terminus, the therapeutic polypeptide, a proteinous connecting moiety (preferably a proteinous sequence with an elongated shape), the flexible un-structured linker, and the scaffold protein. In some embodiments, the fusion protein of the invention comprises, from N-terminus to C-terminus, the therapeutic polypeptide, the flexible un-structured linker, and the scaffold protein. In some embodiments, the fusion protein of the invention comprises, from N-terminus to C-terminus, the scaffold protein, a first flexible un-structured linker, a proteinous connecting moiety (preferably a proteinous sequence with an

elongated shape), a second flexible un-structured linker, and the therapeutic polypeptide. In some embodiments, the fusion protein of the invention comprises, from N-terminus to C-terminus, the scaffold protein, a proteinous connecting moiety (preferably a proteinous sequence with an elongated shape),
5 the flexible un-structured linker, and the therapeutic polypeptide. In some embodiments, the fusion protein of the invention comprises, from N-terminus to C-terminus, the scaffold protein, the flexible un-structured linker, and the therapeutic polypeptide.

In some embodiments, the scaffold protein of the invention is selected
10 from the group consisting of human collagen noncollagenous (NC) domains which form stable homo-trimers in solution, proteins which form homo-trimers in solution with C1q-like molecular structures, proteins which form homo-trimers in solution with TNF-like molecular structures, and proteins with C-type lectin-like domains (CTLD) which form homo-trimers in
15 solution.

In some embodiments, the therapeutic polypeptide is selected from the group consisting of human glucagon-like peptide-1 (GLP-1), Calcitonin, human Parathyroid hormone (PTH), G-CSF, GM-CSF, Interferon, VEGF receptors, TNF alpha receptors, RANK, Growth hormone, Erythropoietin,
20 single-chain Fv, single domain antibodies and functional variants thereof.

In particular embodiments of the invention, the fusion protein of the invention comprises: a therapeutic polypeptide selected from the group consisting of GLP-1, GLP1(A8G/G22E) and GLP1(A8G/G22E/R36S); a first flexible un-structured linker; a proteinous connecting moiety selected from the
25 group consisting of Fn7, Fn8 and TNCfn3; a second flexible un-structured linker; and a scaffold protein selected from the group consisting of COL18NC1, COL15NC1, COL19NC2, ACRP30 C1q-like domain, and MBL neck and CRD domain.

In a preferred embodiment of the invention, the fusion protein of the
30 invention comprises: GLP1(A8G/G22E/R36S), a first flexible un-structured linker, Fn8, a second flexible un-structured linker, and COL18NC1.

The present invention also provides a polynucleotide sequence encoding

the fusion protein, a pharmaceutical composition comprising the fusion protein and a pharmaceutically acceptable carrier, and an expression vector comprising the polynucleotide sequence and expression control elements.

In another aspect, the present invention provides a method of improving the pharmacokinetic property of a therapeutic polypeptide, comprising the step of fusing the therapeutic polypeptide to a scaffold protein which forms a homo-trimer in solution, wherein the therapeutic polypeptide is connected with the scaffold protein via a spacer, and the spacer comprises a flexible un-structured linker whose length is adjustable to adjust the apparent molecular size and/or the *in vivo* half life of the fusion protein.

Brief Description of the Drawings

Fig. 1 shows the schematic drawings illustrating the mechanisms of the present invention. The therapeutic polypeptide is shown as a red star in Fig. 1a and as a red helix in Fig. 1b, the spacer is shown in blue and the scaffold protein is shown as a grey sphere. a) The therapeutic polypeptide can be connected to the scaffold protein by a flexible, unstructured linker as the spacer. b) The therapeutic polypeptide can be connected to the scaffold protein by a first flexible unstructured linker (linker 1), a proteinous connecting moiety (preferably with an elongated shape, such as Fibronectin type III domain) and a second flexible unstructured linker (linker 2) as the spacer.

Fig. 2 shows the result of SDS-PAGE analysis of purified proteins.

GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1,
GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-20,
GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-30,
GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-54 and
GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-60 are shown in lane 1, 2, 3, 4 and 5. The molecular markers are labeled at the left of the 12% gel.

Fig. 3 shows the gel filtration chromatography profiles for purified
GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1,
GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-20,
GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-30,

GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-54 and
 GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-60 using the analytical column
 Superdex200 (GE Healthcare). In this figure, the profiles of these proteins are
 labeled as NC1, NC1-20, NC1-30, NC1-54 and NC1-60. The elution time for
 5 the molecular marker proteins (158Kd and 44Kd) are shown by arrows. The
 X-axis refers to elution time and the Y-axis refers to UV280 absorbance
 intensity.

Fig. 4 shows the results of cAMP assays for GLP-1 (7-37) peptide,
 GLP1(A8G/G22E/R36S)-Fn8-NC1 and
 10 GLP1(A8G/G22E/R36S)-Fn8-NC1-54. This assay is based on competitive
 binding technique. A monoclonal antibody specific for cAMP becomes bound
 to the goat anti-mouse antibody coated onto the microplate. Following a wash
 to remove excess monoclonal antibody, cAMP present in a sample competes
 with a fixed amount of horseradish peroxidase (HRP)-labeled cAMP for sites
 15 on the monoclonal antibody. This is followed by another wash to remove
 excess conjugate and unbound sample. A substrate solution is added to the
 wells to determine the bound enzyme activity. The color development is
 stopped and the absorbance is read at 450nm. The intensity of the color is
 inversely proportional to the concentration of cAMP in the sample. The Y-axis
 20 refers to the OD₄₅₀ obtained by the plate reader and the X-axis refers to the
 concentration of GLP-1 (7-37) peptide, GLP1(A8G/G22E/R36S)-Fn8-NC1
 and GLP1(A8G/G22E/R36S)-Fn8-NC1-54.

Fig. 5 shows the pharmacokinetics profiles of the GLP-1 containing
 proteins (GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1,
 25 GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-20,
 GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-30 and
 GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-54) in rats measured by use of the
 sandwich ELISA method. In this figure, the profiles of these proteins are
 labeled as NC1, NC20, NC30 and NC54.

30

Detailed description of the invention

In one aspect of the present invention, there is provided a method to

increase the half life of a therapeutic polypeptide by fusing the therapeutic polypeptide to a scaffold protein via a spacer. The scaffold protein can form a stable homo-trimer in solution. The therapeutic polypeptide is connected to the scaffold protein through a spacer. In some embodiments, the spacer may
5 comprise a flexible, un-structured linker whose length is adjustable (Fig. 1a). In some embodiments, the spacer may further comprise a proteinous connecting moiety, preferably a proteinous sequence with an elongated shape. The proteinous connecting moiety can be connected to the therapeutic polypeptide and/or the scaffold protein via a flexible, un-structured linker
10 whose length is adjustable (Fig. 1b). This novel method provided by the invention, termed as "Trident technology", can efficiently increase the radius of gyration (R_g) of the polypeptide molecule to extend its half life *in vivo*. Moreover, changing the length of the spacer can adjust the *in vivo* half life of the fusion protein in a tunable manner.

15 In preferred embodiments of the present invention, the method of the invention may have several major advantages over the traditional PEGylation method or Fc/HSA fusion method. 1. In the method of the invention, PEGylation on the polypeptide molecule is not essential, therefore the biological activity of the therapeutic polypeptide is fully retained. 2. Because
20 the scaffold protein forms a homo-trimer, the fusion protein of the therapeutic polypeptide with the scaffold protein may greatly increase the apparent size of the fusion protein to slow down renal filtration. Moreover, the trimer formation also renders the fusion protein tri-valency. This may greatly increase the activity of the therapeutic protein. 3. The length of the spacer
25 between the therapeutic polypeptide and the scaffold protein plays an important role in determining the *in vivo* half life of the fusion protein. The method of the invention provides a platform to fine tune the *in vivo* half life of a therapeutic polypeptide by varying the length of the spacer within the fusion protein. 4. The scaffold protein is preferably selected from human proteins,
30 usually from human extracellular proteins, therefore, no foreign protein sequences are introduced into the fusion protein. The immunogenicity of the fusion proteins generated using the method is low. 5. In many cases, the

recombinant fusion protein of the therapeutic polypeptide and the scaffold protein can be generated using *E.coli* expression system, which eliminates the need of the expensive chemical synthesis process for some therapeutic polypeptides or the need of using the eukaryotic expression systems.

5

The scaffold protein

Collagens are a diverse family of proteins that constitute the major structural component of the extracellular matrix¹⁰⁻¹². Collagen is composed of a triple helix, which generally consists of an identical chain ($\alpha 1$) or an
10 additional chain that differs slightly in its chemical composition ($\alpha 2$). Classification according to supramolecular structure assigns collagens to fibril, fibril-associated containing interrupted triple helicies (FACIT), beaded filament, anchoring fibril, network-forming, transmembrane or multiple triple helicies with interruptions (Multiplexin) families¹³. To date, 43 unique
15 α -chains that belong to 28 types of collagens (types I–XXVIII) have been discovered in vertebrates. The alpha chains of collagens consist of at least one triple helical collagenous domain of varying length and two noncollagenous (NC) domains of variable sequence, size, and shape that are positioned at the N and C terminus. The collagenous domains contain the G-X-Y repeats and
20 form the typical triple helix within the collagen molecule while some of the NC domains form homo-trimers to stabilize the collagen triple helix. Studies on classic fibril-forming collagens found that the extreme carboxy-terminal NC (NC1) domains were essential for trimerization¹⁴. The multiplexin family members also utilize NC1 domains for trimerization. On the other hand, Fibril
25 associated-collagens (FACIT) have recently been shown to trimerize via their NC2 domains (the second NC domain from the carboxy-terminal end)^{15,16}. The crystal structures of NC1 domains from the network-forming collagens IV, VIII and X indicated these NC1 domains constituted a stable homo-trimer with the clq-like molecular structure¹⁷⁻¹⁹. About 130 a.a. residues are present in
30 each monomer. In contrast, the NC1 domains from the multiplxin family members such as Collagen XV and XVIII formed the much smaller homo-trimer with the length of about 55 residues in each monomer^{14,20}. Both

types of the collagen NC domains form very stable homo-trimers in solution ($T_m > 60^\circ\text{C}$) and are suitable as the scaffold protein as described in the invention.

Many other human proteins or domains thereof which can form
5 homo-trimers in solution may also serve as the scaffold proteins in the method of the invention. The C1q family is characterized by a C-terminal conserved globular C1q domain (pfam ID: PF00386), which can form a stable homo-trimer²¹⁻²³. The C1q-like protein family includes, but not limited to, human C1q A chain, C1q B chain, C1q C chain, cb1n family members, human
10 EMILIN-1, multimerin, ACRP30/adiponectin, adipolin, resistin and resistin-like molecule (RELM) hormone family members. Tumor necrosis factor (TNF) refers to a cytokine that can induce cell apoptosis and inflammation^{24,25}. TNF family (Pfam ID: PF00229) members include, but not limited, human TNFalpha, TNFbeta, TRAIL, RANK ligand, Fas ligand, CD
15 30 ligand, CD40 ligand, CD27 ligand, OX40L and CD137. TNF family members form homo-trimers in solution and demonstrated the similar molecular structure as the C1q family members. Therefore, these two family members are also named as C1q/TNF-related proteins (CTRP)²².

The superfamily of proteins containing C-type lectin-like domains
20 (CTLDS, pfam ID: PF00059) is a large group of extracellular proteins with diverse functions including cell-cell adhesion, immune response to pathogens and apoptosis²⁶. A number of CTLD proteins contain a neck and a C-terminal C-type carbohydrate-recognition domain (CRD) and they form homo-trimer in solution. This type of CTLD includes mannan-binding lectin (MBL),
25 surfactant protein A (SP-A), surfactant protein D (SP-D), collectin liver 1 (CL-L1), collectin placenta 1 (CL-P1), conglutinin, collectin of 43 kDa (CL-43) and collectin of 46 kDa (CL-46), Langerin and Tetranectin²⁷⁻³⁰.

In this invention, the CTRP family members, including the C1q-like domains and TNF family members, have been utilized to fuse with the
30 therapeutic polypeptide to extend the *in vivo* half life of the fusion protein. Alternatively, the CTLD family members can be employed as the scaffold protein to drive the trimerization of the therapeutic polypeptides. Moreover,

the NC1 domain within Multiplexin type of human Collagen (such as collagen XV and XVIII) and NC2 domain within FACIT type of collagen (such as collagen IX, XII, XIV, XVI, XIX, XX, XXI, and XXII) may also serve as the scaffold proteins in the method of this invention. The therapeutic polypeptide
5 can be fused to either the N-terminus or the C-terminus of the scaffold protein. The trimer formation of the fusion protein can efficiently increase the radius of gyration (R_g) of the protein molecule. The fusion protein may demonstrate a much larger apparent size than a compact molecule with the same molecular weight. Therefore the fusion protein will show a much reduced clearing rate
10 by renal filtration and will exhibit an extended half life *in vivo*.

It will be appreciated by a skilled artisan that the term “scaffold protein” as used herein does not necessarily mean an entire wild type protein; a domain or functional variant thereof which can form a stable homo-trimer in solution and can therefore serve the purpose of the invention may also be used in our
15 method. In some embodiments of the invention, for example, the C-terminal C1q-like domain of human ACRP30 and the neck and CRD domain of MBL were used as a scaffold protein.

The therapeutic polypeptide

20 In some embodiments of this invention, the therapeutic polypeptide may be selected from, but not limited to, human glucagon-like peptide-1 (GLP-1), Calcitonin, human Parathyroid hormone (PTH), G-CSF, GM-CSF, Interferon, VEGF receptors, RANK, TNF receptors, growth hormone, Erythropoietin, single-chain Fv and single domain antibodies. In a preferred embodiment of
25 this invention, we use GLP-1 as one of the examples to illustrate how the method of the invention can significantly improve the pharmacokinetics property of GLP-1 and its mutants while retaining its biological activity. The natural incretin hormone glucagon-like peptide-1 (GLP-1) supports glucose homeostasis by enhancing glucose-dependent insulin secretion from β -cells
30 and suppressing inappropriately elevated postprandial glucagon secretion from α -cells. In addition, GLP-1 has been demonstrated to reduce appetite and food intake and inhibit gastric emptying, which may facilitate weight management

31,32. Therefore, GLP-1 remains to be a very promising therapeutic polypeptide for type 2 diabetes and weight loss. However, GLP-1 is a 30 residue polypeptide with a very short half life *in vivo*, which severely limits its applications. In the present invention, we demonstrated data to show that the
5 half life of GLP-1 can be significantly extended by use of the method of the invention.

One advantage of the method is that it can render tri-valency for the therapeutic polypeptide. It has been well documented that multivalency of protein can greatly enhance its affinity and avidity to binding partner³³⁻³⁵.
10 Antibody IgG is a Y-shaped molecule with bi-valency and utilizes two identical variable domains to interact with its ligand. The fusion protein generated using the method of the invention has tri-valency and therefore might behave better than the traditional human monoclonal antibody IgG in interacting with the ligand. For example, TNF alpha forms a homo-trimer in
15 solution and interact with three TNF receptors simultaneously. To inhibit the TNF alpha function, TNF receptor 2 (TNFR2, p75) have been fused to IgG Fc fragment to constitute Etanercept (Enbrel) to treat severe rheumatoid arthritis. However, one Enbrel molecule can only block two out of three possible binding sites located on TNF alpha homo-trimer. In contrast, our fusion
20 protein of TNFR2 and collagen XVIII NC1 domain generated (described below in example 15) can form a homo-trimer and block all three binding sites of TNFalpha while retaining a long half life *in vivo*.

The scaffold protein utilized in this invention can form homo-trimers by simultaneous self assembly. No inter-chain disulfide bonds are needed to drive
25 the trimerization. Many expression systems such as *E. coli*, yeast, insect cell and mammalian cell systems can be utilized to express the fusion proteins generated by the invention. In the sharp contrast, therapeutic monoclonal antibodies rely on the mammalian systems for mass productions.

30 The spacer

In this invention, the therapeutic polypeptide and the scaffold protein are connected by a spacer. In some embodiments, the spacer comprises a flexible

un-structured linker whose length is adjustable.

The term “flexible un-structured linker” refers to an amino acid sequence which is flexible in movement and which does not form any regular stable secondary and tertiary protein structures. The primary sequence of the
5 “flexible un-structured linker” is usually rich in G, S, A, T or P.

If the therapeutic polypeptide is a relatively large protein (such as Interferon, Growth hormone, Erythropoietin, G-CSF, or TNFR2, usually a protein with more than 100 amino acid residues), it may be directly fused to the scaffold protein through a flexible, un-structured linker as the spacer (Fig.
10 1a). In some other cases, the therapeutic polypeptide might be a short peptide (such as GLP-1 or PTH, usually a peptide not more than 100 residues). To efficiently utilize our method, the spacer between the therapeutic polypeptide and the scaffold protein may contain a proteinous connecting moiety, preferably a proteinous sequence with an elongated shape, such as the human
15 Fibronectin type III domain. The proteinous connecting moiety can be connected to the therapeutic polypeptide and/or the scaffold protein via a flexible, un-structured linker. The proteinous connecting moiety can stabilize and elongate the spacer region within the fusion protein and further increase the radius of gyration (R_g) of the fusion protein. In a particular embodiment of
20 this invention, the proteinous connecting moiety contains a protein domain Fn8. In some embodiments, the proteinous connecting moiety may comprise a whole protein, a truncated version of a protein, a protein domain or domains in tandem, or protein fragments. A skilled artisan will appreciate that the proteinous connecting moiety may comprise some non-proteinous
25 modifications which are not formed by amino acids, such as PEG.

The length of the flexible, unstructured linker may play an important role in determining the radius of gyration (R_g) and the *in vivo* half life of the fusion protein. The flexible, unstructured linker may contain sequences such as $(G4S)_n$, $(G3S)_n$, $(G2S2)_n$, where n is an integer, or other sequences that are
30 rich in G, S, A, T or P. The length of the flexible linker may vary from 1 to 300 amino acid residues, and particularly within the range of 5 to 100 amino acid residues. It has been reported that the un-structured stretches of

polypeptides may act like PEG molecule and increase the Rg of the protein molecule³⁶. By use of our method, a shorter flexible linker is needed to reach the desired Rg due to the trimer formation compared with the monomer. This may have great advantages for therapeutic proteins by reducing the immunogenicity.

In some embodiments of the invention, the spacer may consist of one flexible, un-structured linker. In some embodiments, the spacer may comprise a first flexible, un-structured linker, a proteinous connecting moiety (preferably a proteinous sequence with an elongated shape), and a second flexible, un-structured linker. It will be appreciated by a skilled artisan that the first flexible, un-structured linker and the second flexible, un-structured linker may be the same or different. In some embodiments of the invention, the proteinous connecting moiety may comprise a whole protein, a single protein domain, or several protein domains in tandem which may be the same or different with each other.

Our data clearly showed that varying the length of the flexible, unstructured linker or linkers can efficiently change the Rg of the molecule and control the *in vivo* half life of the engineered protein molecule. Therefore, our method can generate a recombinant protein with tunable *in vivo* half life by varying the flexible, unstructured linker length between the therapeutic polypeptide and the scaffold protein. This is advantageous compared with the traditional therapeutic IgG with fixed *in vivo* half life. In addition, our method may offer the fusion protein tri-valency for the ligand, in contrast, IgG only has bi-valency.

Fibronectin(Fn) is a high-molecular weight (~440kDa) glycoprotein of the extracellular matrix that binds to a number of proteins including integrins, collagen, fibrin and heparan sulfate proteoglycans (e.g. syndecans)³⁷. Fibronectin exists as a protein dimer, consisting of two nearly identical polypeptide chains linked by a pair of C-terminal disulfide bonds. Each fibronectin monomer has a molecular weight of 230–250 kDa and contains three types domains: type I, II, and III. Type I and type II are stabilized by intra-chain disulfide bonds, while fibronectin type III domains do not contain

any disulfide bonds³⁸. The Fibronectin type III domain is an evolutionary conserved protein domain that is widely found in animal proteins. The human fibronectin protein in which this domain was first identified contains 16 copies of this domain (Fn1 to Fn16). The fibronectin type III domain family (pfam ID: PF00041) member contains about 95 amino acids long and possesses a beta sandwich structure. Fibronectin type III domain forms a very stable domain structure with the melting temperature of ~70 °C as measured by DSC^{39,40}. Fibronectin type III domains are found in a wide variety of extracellular proteins. In human genome, fibronectin type III domain exists in many proteins including Tenascin, Usherin, Titin, tripartite motif (TRIM) family members, tissue factor, TIE1, TIE2, SPEG, SORL1, SDK1, ROBO1, ROBO2, SDK2, Receptor-type tyrosine-protein phosphatase, prolactin receptor, LICAM, NCAM1, NCAM2, myomesin 1, myomesin 2, Myosin-binding protein C, LIFR, Leptin receptor, Integrin, Insulin receptor, Contactin, Collagen, Cytokine receptor-like factor, Inteferon receptor, Growth hormone receptor, fibronectin, leucine rich transmembrane protein (FLRT) members, IL, ephrin type-A receptor, ephrin type-B receptor, IL-6R, gp130, IL11RA, IL12RB, IL20RB, IL23R, IL27RA and IL31RA etc. Therefore, using Fibronectin type III domain as part of the spacer in our method may have low immunogenicity. The Fibronectin type III domains can also be used in tandem fashion in the spacer. In addition, Fibronectin type III domain can be expressed in recombinant form using a number of expression systems including *E. coli*, using Fibronectin type III domain as part of the spacer in the method of the invention may greatly increase the expression yield of the fusion protein.

Summary for Trident technology

Improving the pharmacokinetic property of a therapeutic polypeptide may have major impacts on its application. In one aspect, the present invention provides a novel technique termed as “Trident technology” which allows the therapeutic polypeptide to fuse with a scaffold protein which forms homo-trimer through a spacer with variable length. The trimer formation may

greatly increase the R_g of the fusion molecule and improve the *in vivo* half life of the fusion protein. The scaffold proteins can be fused to the N-terminus or C-terminus of the therapeutic polypeptide. The length of the spacer between the therapeutic polypeptide and the scaffold protein can be modified to fine
5 tune the pharmacokinetics feature of the fusion protein. Moreover, the method of the invention can provide the therapeutic polypeptide with tri-valency, which may greatly increase the affinity and avidity of the fusion protein toward the ligand.

To further extend the *in vivo* half life of the fusion protein generated
10 by the method of the invention, the fusion protein can be further modified by PEGylation. The PEG moiety may have a molecular weight of between 2 and 100 kDa. For specific PEGylation, the Cys residue may need to be generated in the fusion protein using site-directed mutagenesis.

The terms “protein”, “polypeptide” and “peptide” as used here are
15 interchangeable, unless instructed to the contrary.

The term “functional variant” of a protein refers to a modified version of the native protein which comprises substitutions, deletions and/or additions of one or several amino acids, e.g., less than 15 amino acids, or preferably less than 10 or 5 amino acids, and which substantially retains the biological
20 activity of the native protein. Typically, conservative substitutions of amino acids are preferred which are well known to a skilled artisan. Deletions are preferably deletions of amino acids from regions not involved in the biological function of the protein. For example, GLP-1(A8G/G22E) is a functional variant of wild type GLP-1, which contains two substitutions of amino
25 acids and which substantially retains its biological activity such as increasing cAMP level.

Methods of Preparing fusion proteins generated by the Present Invention_

The fusion proteins of the present invention can be produced through the
30 application of recombinant DNA technology. Recombinant polynucleotide constructs encoding a fusion polypeptide of the present invention typically include an expression control sequence operably-linked to the coding

sequences of the fusion polypeptide, including naturally-associated or heterologous promoter regions. As such, another aspect of the invention includes vectors containing one or more nucleic acid sequences encoding a fusion polypeptide of the present invention. For recombinant expression of one or more polypeptides of the invention, the nucleic acid containing all or a portion of the nucleotide sequence encoding the fusion polypeptide is inserted into an appropriate cloning vector, or an expression vector (*i.e.*, a vector that contains the necessary elements for the transcription and translation of the inserted polypeptide coding sequence) by recombinant DNA techniques well known in the art and as detailed below. Methods for producing diverse populations of vectors have been described by Lerner *et al.*, U.S. Pat. No. 6,291,160; 6,680,192.

In general, expression vectors useful in recombinant DNA techniques are often in the form of plasmids. In some embodiments of the present invention, “plasmid” and “vector” can be used interchangeably as plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors that are not technically plasmids, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions. Preferably, the expression control sequences are eukaryotic promoter systems in vectors capable of transforming or transfecting eukaryotic host cells. Once the vector has been incorporated into the appropriate host, the host is maintained under conditions suitable for high level expression of the nucleotide sequences encoding the fusion polypeptide. These expression vectors are typically replicable in the host organisms either as episomes or as an integral part of the host chromosomal DNA. Commonly, expression vectors contain selection markers, *e.g.*, ampicillin-resistance or kanamycin-resistance, to permit detection of those cells transformed with the desired DNA sequences. Vectors can also encode a signal peptide, *e.g.*, pectate lyase, useful to direct the secretion of extracellular antibody fragments. *See* U.S. Pat. No. 5,576,195.

The recombinant expression vectors of the invention may comprise a nucleic acid encoding a fusion polypeptide in a form suitable for expression of

the nucleic acid in a host cell, which means that the recombinant expression vectors may include one or more regulatory sequences selected on the basis of the host cells to be used for expression that are operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector,
5 “operably-linked” is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term “regulatory sequence” is intended to include promoters, enhancers and other
10 expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, *e.g.*, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and
15 those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of polypeptide desired, *etc.*

20 Another aspect of the invention pertains to the fusion polypeptide-expressing host cells, which contain a nucleic acid encoding one or more fusion polypeptides. The recombinant expression vectors of the invention can be designed for expression of a fusion polypeptide in prokaryotic or eukaryotic cells. For example, a fusion polypeptide can be
25 expressed in bacterial cells such as *Escherichia coli*, insect cells, fungal cells, *e.g.*, yeast, or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and
30 translated *in vitro*, *e.g.* using T7 promoter regulatory sequences and T7 polymerase.

Expression of polypeptides in prokaryotes is most often carried out in *E.*

coli with vectors containing constitutive or inducible promoters directing the expression of the recombinant polypeptides. The vectors may add a number of amino acids to a polypeptide encoded therein, usually to the amino terminus of the recombinant polypeptide. Such vectors with extra amino acid residues typically serve three purposes: (i) to increase expression of recombinant polypeptide; (ii) to direct the recombinant protein to periplasmic space; and (iii) to aid in the purification of the recombinant polypeptide by acting as a ligand in affinity purification. Typical expression vectors serving this purpose include pGEX (GE Healthcare), pMAL (New England Biolabs), pET20b(Novagen), pET43b (Novagen), pET32b(Novagen) and pRIT5 (GE Healthcare).

Examples of suitable inducible *E. coli* expression vectors include pTrc vectors (Invitrogen), pQE (Qiagen) and pET vectors (Novagen). One strategy to maximize recombinant polypeptide expression is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in the expression host, *e.g.*, *E. coli*⁴¹. Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the fusion polypeptide expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerevisiae* include pYES2 (Invitrogen), pMFa⁴² and pJRY88⁴³. The fusion protein may also be expressed in *Pichia* system using the vectors pPICZ pGAPZ and pPIC9(InVitrogen). Alternatively, a fusion polypeptide can be expressed in insect cells using baculovirus expression vectors or using the stable insect cell lines. Baculovirus systems available for expression of polypeptides in cultured insect cells (*e.g.*, SF9 cells) include the BaculoGold system (BD Biosciences), BaculoDirect system (Invitrogen) and BacVector system (Novagen). The stable insect expression systems include, but not limited to, DES system (Invitrogen) and InsectDirect (Novagen).

In another embodiment, a nucleic acid encoding a fusion polypeptide of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include, *e.g.*, but are not limited to, pcDNA3.1 (Invitrogen), pSecTag (invitrogen), and pTriEx series

vectors (Novagen). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems
5 for both prokaryotic and eukaryotic cells useful for expression of the fusion polypeptide of the present invention, please see, *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

10 In the eukaryotic expression systems, the recombinant fusion protein can be expressed in the cytoplasm. Or alternatively, the fusion protein can be secreted into the medium by adding an N-terminal secretion signal.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention is to be introduced. The terms
15 "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the
20 parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, a fusion polypeptide can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells. Mammalian cells are a preferred host for expressing nucleotide segments encoding immunoglobulins or fragments
25 thereof. See Winnacker, *From Genes To Clones*, (VCH Publishers, NY, 1987). A number of suitable host cell lines capable of secreting intact heterologous proteins have been developed in the art, and include Chinese hamster ovary (CHO) cell lines, 293 cells, various COS cell lines, HeLa cells, L cells and myeloma cell lines. Preferably, the cells are nonhuman. Expression vectors for
30 these cells can include expression control sequences, such as an origin of replication, a promoter, an enhancer, and necessary processing information sites, such as ribosome binding sites, RNA splice sites, polyadenylation sites,

and transcriptional terminator sequences. Preferred expression control sequences are promoters derived from endogenous genes, cytomegalovirus, SV40, adenovirus, bovine papillomavirus, and the like. Other suitable host cells are known to those skilled in the art.

5 For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate a foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along
10 with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding the fusion polypeptide or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can
15 be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

Once expressed, the fusion polypeptides are purified from culture media and/or host cells. Purification of recombinant polypeptides is well known in the art and includes ammonium sulfate precipitation, affinity chromatography
20 purification technique, column chromatography, ion exchange purification technique, gel filtration and the like (see generally Scopes, *Protein Purification* (Springer-Verlag, N.Y., 1982).

Formulation of Pharmaceutical Compositions

25 The present invention envisions treating a disease, for example, type II diabetes, in a mammal by the administration of the fusion protein compositions of the present invention. Administration of the fusion protein in accordance with the present invention may be continuous or intermittent, depending, for example, upon the recipient's physiological condition, whether
30 the purpose of the administration is therapeutic or prophylactic, and other factors known to skilled practitioners. The administration of the vaccines of the invention may be essentially continuous over a preselected period of time

or may be in a series of spaced doses. Both local and systemic administration is contemplated.

The pharmaceutical composition of the present invention may be delivered via various routes and to various sites in a mammal body to achieve a particular effect. One skilled in the art will recognize that although more than one route can be used for administration, a particular route can provide a more immediate and more effective reaction than another route. Local or systemic delivery can be accomplished by administration comprising application or instillation of the formulation into body cavities, inhalation or insufflation of an aerosol, or by parenteral introduction, comprising intramuscular, intravenous, peritoneal, subcutaneous, intradermal, as well as topical administration.

The amount of the administered fusion protein of the present invention will vary depending on various factors including, but not limited to, the particular disease, the weight, the physical condition, and the age of the mammal, and whether prevention or treatment is to be achieved. Such factors can be readily determined by the clinician employing animal models or other test systems which are well known to the art. Generally, the amount of the fusion protein of the present invention to be administered to a mammal subject may vary in the range of 1ng/kg to 100mg/kg of the subject body weight. In an embodiment of the invention, the amount of administration was from 0.5mg/kg to 1.0mg/kg.

When the fusion proteins of the invention are prepared for administration, they are preferably combined with a pharmaceutically acceptable carrier to form a pharmaceutical formulation, or unit dosage form. Commonly used pharmaceutically acceptable carriers are well known to a skilled artisan in the field of pharmacy. The total active ingredients in such formulations include from 0.1 to 99.9% by weight of the formulation. The active ingredient for administration may be present as a powder or as granules; as a solution, a suspension or an emulsion.

Thus, the therapeutic agent may be formulated for parenteral administration (*e.g.*, by injection, for example, bolus injection or continuous

infusion) and may be presented in unit dosage form in ampules, pre-filled syringes, small volume infusion containers or in multi-dose containers with an added preservative. The active ingredients may take such forms as suspensions, solutions, or emulsions in oily or aqueous vehicles, and may contain
5 formulatory agents such as suspending, stabilizing and/or dispersing agents. Alternatively, the active ingredients may be in powder form, obtained by aseptic isolation of sterile solid or by lyophilization from solution, for re-constitution with a suitable vehicle, *e.g.*, sterile, pyrogen-free water, before use.

10 In general, water, suitable oil, saline, aqueous dextrose (glucose), and related sugar solutions and glycols such as propylene glycol or polyethylene glycols are suitable carriers for parenteral solutions. Solutions for parenteral administration contain the active ingredient, suitable stabilizing agents and, if necessary, buffer substances. Antioxidizing agents such as sodium bisulfate,
15 sodium sulfite or ascorbic acid, either alone or combined, are suitable stabilizing agents. Also used are citric acid and its salts and sodium ethylenediaminetetraacetic acid (EDTA). In addition, parenteral solutions can contain preservatives such as benzalkonium chloride, methyl- or propyl-paraben and chlorobutanol. Suitable pharmaceutical carriers are
20 described in *Remington's Pharmaceutical Sciences*, a standard reference text in this field.

 Additionally, standard pharmaceutical methods can be employed to control the duration of action. These are well known in the art and include control release preparations and can include appropriate macromolecules, for
25 example polymers, polyesters, polyamino acids, polyvinyl, pyrrolidone, ethylenevinylacetate, methyl cellulose, carboxymethyl cellulose or protamine sulfate. The concentration of macromolecules as well as the methods of incorporation can be adjusted in order to control release. Additionally, the agent can be incorporated into particles of polymeric materials such as
30 polyesters, polyamino acids, hydrogels, poly (lactic acid) or ethylenevinylacetate copolymers. In addition to being incorporated, these agents can also be used to trap the compound in microcapsules.

The terms “pharmaceutically-acceptable,” “physiologically-tolerable,” and grammatical variations thereof, as they refer to compositions, carriers, and reagents, are used interchangeably and represent that the materials are capable of administration to or upon a subject without the production of undesirable physiological effects to a degree that would prohibit administration of the composition.

Preferred examples of such carriers include, but are not limited to, water, saline, Ringer's solutions, and dextrose solution. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and compounds for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or compound is incompatible with the fusion polypeptides, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. The fusion polypeptides compositions of the present invention can be administered by parenteral, topical, intravenous, oral, subcutaneous, intraarterial, intradermal, transdermal, rectal, intracranial, intraperitoneal, intranasal; intramuscular route or as inhalants. The fusion polypeptides can optionally be administered in combination with other agents that are at least partly effective in treating various diseases including various actin- or microfilament-related diseases.

Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial compounds such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating compounds such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and compounds for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be

enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, *e.g.*, water, ethanol, polyol (*e.g.*, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, *e.g.*, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal compounds, *e.g.*, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic compounds, *e.g.*, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition a compound which delays absorption, *e.g.*, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the fusion polypeptides in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the binding agent into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yield a powder of the

active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof. The agents of this invention can be administered in the form of a depot injection or implant preparation which can be formulated in such a manner as to permit a sustained or pulsatile release of
5 the active ingredient.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the binding agent can be incorporated with excipients and used in the form of tablets, troches, or
10 capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding compounds, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any
15 of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating compound such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening compound such as sucrose or
20 saccharin; or a flavoring compound such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the fusion polypeptides are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a
25 nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, *e.g.*, for transmucosal administration,
30 detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the fusion polypeptides are

formulated into ointments, salves, gels, or creams as generally known in the art.

The fusion polypeptides can also be prepared as pharmaceutical compositions in the form of suppositories (*e.g.*, with conventional suppository
5 bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

Examples

The present invention is further illustrated by the following examples,
10 which should not be construed as limiting in any way.

Example 1. Construction of an expression vector of GLP-1 fused with human Fibronectin type III domain 7 (Fn7) and human collagen XVIII NC1 domain (COL18NC1)

In this example, GLP-1 polypeptide was fused to the N-terminus of
15 the human collagen XVIII NC1 domain (COL18NC1) which forms a stable homo-trimer. To further extend the Rg of the molecule, the human Fibronectin type III domain 7 (Fn7) was connected to the GLP-1 polypeptide and COL18NC1 through two flexible, unstructured linkers 1 and 2 (linker 1: GGGSGGGG, linker 2: GGGSGG). The pET29b vector
20 (Novagen) was used to construct a recombinant plasmid containing the GLP-1-Fn7-COL18NC1 fusion gene. First, human COL18NC1 was cloned into pET29b by *Bam*HI and *Xho*I to result in the pET29b-COL18NC1 vector. PCR reaction was carried out using human collagen XVIII cDNA as the template by using the following primers:

25 Col18NC1-forward:

CGGGATCCGGTGGCGGCGCCTCCTCAGGGGTGAGG

Col18NC1-reverse:

CCGCTCGAGTTACCCTCGTGGGAGTGGTGTCCGGGCCTCC

The PCR product was digested by restrictive enzyme *Bam*HI and *Xho*I
30 (Fermentas) and ligated into the pET29b vector by use of T4 ligase (Fermentas). The sequence of resulted pET29b-COL18NC1 vector was confirmed by DNA sequencing.

PCR reaction was carried out using human Fibronectin cDNA as the template by using the following primers:

Glp1-Fn7-Forward:

GGAATTCCATATGCATGCCGAAGGGACTTTTACCAGTGATGTAAGTTCTTAT
 5 TTGGAAGGTCAAGCTGCAAAAGAATTCATTGCTTGGCTGGTGAAAGGCCGTG
 GTGGTGGCGGCTCTGGTGGCGGTGGCACACCATTGTCTCCACCAACAACTT
 GCATCTG

Glp1-Fn7-Reverse:

CGGGATCCACCACCAGCTGGGATGATGGTATCAGAGATAGGGACACTTTCC
 10 The PCR product was digested by restrictive enzyme *NdeI* and *BamHI*
 (Fermentas) and ligated into the digested pET29b-COL18NC1 vector. The
 optimized DNA sequence of human GLP-1 (7-37) was included in the primer
 named as Glp1-Fn7-Forward. The cloned GLP-1-Fn7-COL18NC1 fusion gene
 was confirmed by DNA sequencing. The protein sequence of
 15 GLP-1-Fn7-COL18NC1 was listed as SEQ ID NO: 1.

SEQ ID NO: 1, GLP-1-Fn7-COL18NC1 protein sequence. Fn7 is connected to GLP-1 and COL18NC1 by use of two flexible unstructured linkers 1 and 2. The flexible, unstructured linker 1 between GLP-1 and Fn7 (GGGSGGGG) is underlined. The flexible, unstructured linker 2 (GGGSGG) between Fn7 and COL18NC1 is also underlined. GLP-1 sequence is in italic.

HAEGTFTSDVSSYLEGQAAKEFLAWLVKGRGGGSGGGG

TPLSPPTNLHLEANPDTGVLTVSWERSTTPDITGYRITTTPTNGQQGNSLEEVVHADQSSCTFDNLSPGLEYNVSVYTVKD

DKESVPISDTIIPAGGGSGGGASSGVRLWATRQAMLGQVHEVPEGWLIFVAEQEELYVRVQNGFRKVQLEARTPLPRG

Example 2. Cloning of GLP-1 fused with Fibronectin type III domain 8 (Fn8) and human collagen XVIII NC1 domain (COL18NC1)

20 Other human Fibronectin type III domains may also act as part of the spacer between the therapeutic polypeptide and the scaffold protein in our method. In this example, we showed that Fibronectin type III domain 8 (Fn8) can be utilized as part of the spacer between GLP-1 and collagen XVIII NC1 domain. The gene encoding human Fibronectin type III
 25 domain 8 (Fn8) was amplified by PCR using the following primers and

human Fibronectin cDNA as the template:

Fn8-Forward:

GGAATTCCATATGCATGCCGAAGGGACTTTTACCAGTGATGTAAGTTCTTAT
 TTGGAAGGTCAAGCTGCAAAAGAATTCAATTGCTTGGCTGGTGAAGGCCGTG
 5 GTGGTGGCGGCTCTGGTGGCGGTGGCTCTGCTGTTCCCTCCTCCCACTGACCT
 GCGATTC

Fn8-Reverse:

CGGGATCCACCACCACCTGTTTTCTGTCTTCCTCTAAGAGGTGTGC

The PCR product was digested by *Nde*I and *Bam*HI. The digested
 10 insert was ligated into the digested vector pET29b-COL18NC1 (generated
 in example 1). The resulted vector was named as
 pET29b-GLP-1-Fn8-COL18NC1. The protein sequence of the fusion
 protein GLP-1-Fn8-COL18NC1 was listed as SEQ ID NO: 2.

SEQ ID NO: 2, GLP-1-Fn8-COL18NC1 protein sequence, GLP-1 sequence is in italic. Fn8 is connected to
 GLP-1 and COL18NC1 by use of two flexible unstructured linkers 1 and 2. The flexible, unstructured linker 1
 between GLP-1 and Fn8 (GGGSGGGGS) is underlined. The flexible, unstructured linker 2 (GGGSGG)
 between Fn8 and COL18NC1 is also underlined.

HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRGGGSGGGGSAVPPPTDLRFTNIGPDTMRVTWAPPPSIDLTNFLVRYSPVK
 NEEDVAELSISPSDNAVVLTNLLPGTEYVVSVSVEQHESTPLRGRQKTG
GGGSGGGASSGVRLWATRQAMLGQVHEVPEGWLIFVAEQEELYVRVQNGFRKVQLEARTPLPRG

15

Example 3. Expression and Purification of fusion protein GLP-1-Fn8-COL18NC1

The constructed expression vector pET29b-Glp1-Fn8-COL18NC1,
 after sequencing confirmation, was used to transform *Escherichia coli*
 20 BL21 (DE3) for protein expression (for detailed protocols of the
 transformation, see *Molecular Cloning: A Laboratory Manual*). A single
 colony was selected from the culture dish, and placed into a 10 ml LB
 liquid medium with kanamycin (final concentration, 50 µg/ml), then
 shaken at 37 °C at 220 rpm overnight. 1L LB culture was inoculated and
 25 allowed to grown until OD₆₀₀ reached 0.4-1.0. Isopropyl thiogalactoside

(IPTG) was added to a final concentration of 0.2mM. After a successive culture at 30 °C for overnight, cells were collected by centrifugation. The cells were diluted 1: 20 with 20mM Tris, NaCl 50mM, 2mM EDTA, pH 8.0, and, after a thorough mix, disrupted by sonication. Insoluble precipitates were removed by centrifugation at 13,000 RCF for 30 min. The proteins of interest were present in the supernatant, with the expressed product comprising 20% of soluble proteins. 50ml of the supernatant was loaded on a HiTrap Q column (5ml) (GE healthcare). The fusion protein of GLP1-Fn8-COL18NC1 was eluted with about 0.3 M NaCl in the buffer. The eluted protein was further purified by gel filtration column S-200 (GE Healthcare), and the buffer was replaced with PBS (pH 7.5). The final product was confirmed by SDS-PAGE electrophoresis.

Example 4: GLP-1 mutants fused with Fn8-COL18NC1

The mutations of A8G/G22E, A8V/G22E, A8S/G22E and A8G/G22E/R36S within the Glp-1 sequence may increase its resistance to protease digestion, reduce immunogenicity and boost its biological activity^{6,44}. In this example, we constructed a vector to fuse GLP-1 (A8G/G22E) with Fibronectin type III domain 8 and collagen XVIII NC1 domain. The PCR reaction was carried out using vector pET29b-GLP1-Fn8-COL18NC1 prepared in example 2 as the template with the following primers:

Glp-1(A8G/G22E)-Fn8-forward:

GGAATTCCATATGCATGGCGAAGGGACTTTTACCAGTGATGTAAGTTCTTAT
TTGGAAGAGCAAGCTGCAAAAGAATTCATTGC

Coll18NC1-reverse:

CCGCTCGAGTTACCCTCGTGGGAGTGGTGTCCGGGCCTCC

The PCR product was digested by *Nde*I and *Xho*I (Fermentas). The digested insert was ligated into the digested vector of pET29b. The resulted vector was named as pET29b-GLP1(A8G/G22E)-Fn8-COL18NC1. The expression and purification protocol of the fusion protein of GLP-1(A8G/G22E)-Fn8-COL15NC1 was the same as described in example 3. Other mutations, such as A8V/G22E, A8S/G22E and

A8G/G22E/R36S, within the GLP-1 can be generated using the Quikchange II site-directed mutagenesis kit (Agilent) using the GLP-1(A8G/G22E)-Fn8-COL15NC1 gene as the template. The protein sequences of these GLP1 mutants fused with Fn8-COL18NC1 were listed as SEQ ID NO: 3-6. The expression and purification protocol of these fusion proteins can be carried out using similar protocols described in example 3.

SEQ ID NO: 3, GLP-1(A8G/G22E)-Fn8-COL18NC1 protein sequence, the GLP-1 mutation sites (A8G/G22E) are underlined.

HQEGTFTSDVSSYLEEQAAKEFI~~AWL~~VKGRGGGGSGGGGSAVPPPTDLRFTNIGPDTMRVTWAPPPSIDLTNFLVRYSPVK
NEEDVAELSISPSDNAVVLTNLLPGTEYVVSVSSVYEQHESTPLRGRQKTG
GGGSGGGASSGVRLWATRQAMLGQVHEVPEGWLIFVAEQEELYVRVQNGFRKVQLEARTPLPRG

SEQ ID NO: 4, GLP-1(A8V/G22E)-Fn8-COL18NC1 protein sequence, the GLP-1 mutation sites (A8V/G22E) are underlined.

HVEGTFTSDVSSYLEEQAAKEFI~~AWL~~VKGRGGGGSGGGGSAVPPPTDLRFTNIGPDTMRVTWAPPPSIDLTNFLVRYSPVK
NEEDVAELSISPSDNAVVLTNLLPGTEYVVSVSSVYEQHESTPLRGRQKTG
GGGSGGGASSGVRLWATRQAMLGQVHEVPEGWLIFVAEQEELYVRVQNGFRKVQLEARTPLPRG

10

SEQ ID NO: 5, GLP-1(A8S/G22E)-Fn8-COL18NC1 protein sequence, the GLP-1 mutation sites (A8S/G22E) are underlined.

HSEGTFTSDVSSYLEEQAAKEFI~~AWL~~VKGRGGGGSGGGGSAVPPPTDLRFTNIGPDTMRVTWAPPPSIDLTNFLVRYSPVK
NEEDVAELSISPSDNAVVLTNLLPGTEYVVSVSSVYEQHESTPLRGRQKTG
GGGSGGGASSGVRLWATRQAMLGQVHEVPEGWLIFVAEQEELYVRVQNGFRKVQLEARTPLPRG

SEQ ID NO: 6, GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1 protein sequence, the GLP-1 mutation sites (A8G/G22E/R36S) are underlined.

HQEGTFTSDVSSYLEEQAAKEFI~~AWL~~VKGGGGGSGGGGSAVPPPTDLRFTNIGPDTMRVTWAPPPSIDLTNFLVRYSPVK
NEEDVAELSISPSDNAVVLTNLLPGTEYVVSVSSVYEQHESTPLRGRQKTG
GGGSGGGASSGVRLWATRQAMLGQVHEVPEGWLIFVAEQEELYVRVQNGFRKVQLEARTPLPRG

Example 5: Cloning and expression of GLP-1(A8G/G22E) fused with Tenascin C fibronectin type III domain 3 (TNCfn3) and human collagen XVIII NC1 (COL18NC1)

5 In this example, we demonstrated that the fibronectin type III domain that can be used as part of the spacer for our method is not limited within human fibronectin. Other suitable fibronectin type III domain may alternatively be utilized in the spacer in the method of the invention. Here we showed that a fibronectin type III domain from human Tenascin C can
10 be utilized as part of the spacer between the therapeutic polypeptide and the scaffold protein as well.

In this example, the human Tenascin C fibronectin type III domain 3 (TNCfn3) was connected to the GLP-1 polypeptide and COL18NC1 through two flexible, unstructured linkers 1 and 2 (linker 1: GGGSGGGGS,
15 linker 2: GGGSGG). The gene encoding TNCfn3 was amplified by PCR using human Tenascin C cDNA as template using the following primers :

TNCfn3-Forward:

GGAATTCCATATGCATGGCGAAGGGACTTTTACCAGTGATGTAAGTTCTTAT
TTGGAAGAGCAAGCTGCAAAAGAATTCATTGCTTGGCTGGTGAAGGCCGTG
20 GTGGTGGCGGCTCTGGTGGCGGTGGCTCTCGCTTGGATGCCCCAGCCAGAT
CGAGG

TNCfn3-Reverse:

CGGGATCCACCAACCGCCTGTTGTGAAGGTCTCTTTGGCTG

The PCR product was digested by *NdeI* and *BamHI* (Fermentas). The
25 digested insert was ligated into the digested vector pET29b-GLP1(A8G/G22E)-Fn8-COL18NC1 as prepared in example 4. The resulted vector was named as pET29b-GLP1(A8G/G22E)-TNCfn3-COL18NC1. The expression and purification protocol of the fusion protein
30 GLP1(A8G/G22E)-TNCfn3-COL18NC1 (SEQ ID NO:7) was the same as described in example 3.

SEQ ID NO: 7, GLP-1-TNCfn3-COL18NC1 protein sequence, the TNCfn3 sequence is in italic. TNCfn3 is connected to GLP-1 and COL18NC1 by use of two flexible unstructured linkers 1 and 2. The flexible, unstructured linker 1 between GLP-1 and TNCfn3 (GGGSGGGGS) is underlined. The flexible, unstructured linker 2 (GGGSGG) between TNCfn3 and COL18NC1 is also underlined.

HAEGTFTSDVSSYLEGQAAKEFIAWLVKGRGGGSGGGGS

RLDAPSQIEVKDVTDTTALITWFKPLAEIDGIELTYGIKDVPGDRTTIDLTEDENQYSIGNLKPDEYEVS
LI
SRRGDMSSNPAKETFTTG

GGGSGGGASSGVRLWATRQAMLGQVHEVPEGWLIFVAEQEELYVRVQNGFRKVQLEARTPLPRG

Example 6: Cloning and expression of GLP-1(A8G/G22E) fused with Fn8 and human collagen XV NC1 (COL15NC1)

In this example, we presented data to show that the NC1 domain from collagen XV can alternatively be utilized as the scaffold protein in our method. It has been reported that human collagen XV NC1 domain, like the human collagen XVIII NC1 domain, forms a stable homo-trimer^{14,20}. To generate the expression vector encoding the fusion protein of GLP1(A8G/G22E)-Fn8-COL15NC1, PCR reaction was carried out using human collagen XV cDNA as the template by using the following primers: Col15NC1-forward:

CGGGATCCGGTGGCAACCTGGTCACAGCATTCAGCAACATGG

Col15NC1-reverse:

CCGCTCGAGTTAGGCAGGAATGGGGATCAGTTCTCCCAG

The PCR product was digested by restrictive enzyme *Bam*HI and *Xho*I (Fermentas) and ligated into the digested pET29b-GLP1(A8G/G22E)-Fn8-COL18NC1 vector as prepared in example 4 by use of T4 ligase (Fermentas). The resulted vector was named as pET29b-GLP1(A8G/G22E)-Fn8-COL15NC1 and the sequence of the vector was confirmed by DNA sequencing. The protein sequence of GLP1(A8G/G22E)-Fn8-COL15NC1 was listed as SEQ ID NO:8. The expression and purification protocol of the fusion protein GLP1(A8G/G22E)-Fn8-COL15NC1 (SEQ ID NO:8) was the same as described in example 3.

SEQ ID NO: 8, GLP-1(A8G/G22E)-Fn8-COL15NC1 protein sequence, the sequence of COL15NC1 is underlined. The flexible unstructured linker 1 (GGGSGGGGS) between GLP-1 and Fn8 is in italic. The flexible unstructured linker 2 (GGGSGG) between Fn8 and COL15NC1 is also in italic.

HGEGTFTSDVSSYLEEQAAKEFIAWLVKGRGGGGSGGGGSVPPPTDLRFTNIGPDTMRVTWAPPSIDLTNFLVRYSPVK
NEEDVAELISPSDNAVVLNLLPGTEYVVSVSVEQHESTPLRGRQKTG
GGGSGGNLVTAFSNMDDMLQKAHLVIEGTFIYLRDSTEFFIRVRDGGWKKLQLGELIPIA

Example 7: Cloning and expression of GLP-1(A8G/G22E) fused with Fn8 and human collagen XIX NC2 domain (COL19NC2)

5 In this example, we demonstrated that the NC2 domain from collagen XIX can be utilized as the scaffold protein in our method. It has been shown that human collagen XIX NC2 domain (COL19NC2) forms a highly stable homo-trimer¹⁶. To generate the expression vector encoding the fusion protein of GLP1(A8G/G22E)-Fn8-COL19NC2, PCR reaction was
10 carried out using human collagen XIX cDNA as the template by using the following primers:

Col19NC2-forward:

CGGGATCCGGTGGCGGCATTCCGGCTGATGCAGTTTC

Col19NC2-reverse:

15 CCGCTCGAGTTAAGGTCTCCCATAAGCTTGGGCAGCCAAC

The PCR product was digested by restrictive enzyme *Bam*HI and *Xho*I (Fermentas) and ligated into the digested pET29b-GLP1(A8G/G22E)-Fn8-COL18NC1 vector as prepared in example 4 by use of T4 ligase (Fermentas). The resulted vector was named as
20 pET29b-GLP1(A8G/G22E)-Fn8-COL19NC2 and the sequence of the vector was confirmed by DNA sequencing. The protein sequence of GLP1(A8G/G22E)-Fn8-COL19NC2 was listed as SEQ ID NO:9. The expression and purification protocol of the fusion protein GLP1(A8G/G22E)-Fn8-COL19NC2 (SEQ ID NO:9) was similar as that
25 was described in example 3.

SEQ ID NO: 9, GLP-1(A8G/G22E)-Fn8-COL19NC2 protein sequence, the sequence of COL19NC2 is underlined. The flexible unstructured linker 1 (GGGSGGGGS) between GLP-1 and Fn8 is in *Italic*. The flexible unstructured linker 2 (GGGSGG) between Fn8 and COL19NC2 is also in *italic*.

HGEGTFTSDVSSYLEEQAAKEFIAWLVKGRGGGGSGGGGSVPPPTDLRFTNIGPDTMRVTWAPPSIDLTNFLVRYSPVK
NEEDVAELSISSPDNAVVLTNLLPGTEYVVS SVSSVYEQHESTPLRGRQKTG
GGGSGGGIPADAVSFEEIKKYINQEVLRIFEERMAVFLSQLKLPAAMLAAQAYGRP

Example 8: Cloning and expression of GLP-1(A8G/G22E) fused with Fn8 and human ACRP30 C-terminal C1q-like domain

ACRP30 (also referred to as Adiponectin, GBP-28, apM1 and AdipoQ) is a protein hormone that modulates a number of metabolic processes, including glucose regulation and fatty acid catabolism⁴⁵. ACRP30 contains a C-terminal globular domain that forms a homo-trimer with typical C1q-like structure⁴⁶. In this example, we demonstrated that the C1q-like domain, such as the ACRP30 C1q-like domain, can be utilized as the scaffold protein in our method. To generate the expression vector encoding the fusion protein of GLP1(A8G/G22E)-Fn8-ACRP30, PCR reaction was carried out using human ACRP30 cDNA as the template by using the following primers:

ACRP30-forward:

CGGGATCCGGTGGCGTATACCGCTCAGCATTTCAGTGTGG

ACRP30-reverse:

CCGCTCGAGTTATCAGTTGGTGTCATGGTAGAGAAG

The PCR product was digested by restrictive enzyme *Bam*HI and *Xho*I (Fermentas) and ligated into the digested pET29b-GLP1(A8G/G22E)-Fn8-COL18NC1 vector as prepared in example 4 by use of T4 ligase (Fermentas). The resulted vector was named as pET29b-GLP1(A8G/G22E)-Fn8-ACRP30 and the sequence of the vector was confirmed by DNA sequencing. The protein sequence of GLP1(A8G/G22E)-Fn8-ACRP30 was listed as SEQ ID NO:10. The expression and purification protocol of the fusion protein GLP1(A8G/G22E)-Fn8-ACRP30 (SEQ ID NO:10) is similar as described

in example 3.

SEQ ID NO: 10, GLP-1(A8G/G22E)-Fn8-ACRP30 protein sequence, the sequence of ACRP30 C1q-like domain is underlined. The flexible unstructured linker 1 (GGGSGGGGS) between GLP-1 and Fn8 is in italic. The flexible unstructured linker 2 (GGGSGG) between Fn8 and ACRP30 is also in italic.

HGEGTFTSDVSSYLEEQAAKEFIAWLVKGRGGGGSGGGSAVPPPTDLRFTNIGPDTMRVTWAPPPSIDLTNFLVRYSPVK
NEEDVAELSISPSDNAVVLTNLLPGTEYVVSVSVEQHESTPLRGRQKTG

*GGGSGG*GVYRSAFSVGLETRVTVPNVPIRFTKIFYNQNNHYDGSTGKFYCNIPGLYYFSYHITVYMKDVKV

SLFKKDKAVLFTYDQYQEKNVDAQSGSVLLHLEVGQVWLQVYGDGDHNGLYADNVNDSTFTGFLLY

HDTN

Example 9: Cloning and expression of GLP-1(A8G/G22E) fused with Fn8
5 and human mannose-binding lectin (MBL) neck and CRD domain.

Human mannose-binding lectin (MBL) is a calcium-dependent lectin that plays important roles in innate immune defense^{47,48}. MBL belongs to the C-type lectin-like domain (CTLD) family and the long helix neck region and the C-terminal carbohydrate recognition domain (CRD) of MBL form a
10 homo-trimer²⁹. In this example, we demonstrated that the human MBL neck and CRD domain can be utilized as the scaffold protein in our method. To generate the expression vector encoding the fusion protein of GLP1(A8G/G22E)-Fn8-MBL, PCR reaction was carried out using human MBL cDNA as the template by using the following primers:

15 MBL-forward:

CGGGATCCGGTGGCGGTGATAGTAGCCTGGCTGCC

MBL-reverse:

CCGCTCGAGTCAGATAGGGAACACAGACGGCCAG

The PCR product was digested by restrictive enzyme *Bam*HI and *Xho*I
20 (Fermentas) and ligated into the digested pET29b-GLP1(A8G/G22E)-Fn8-COL18NC1 vector as prepared in example 4 by use of T4 ligase (Fermentas). The resulted vector was named as pET29b-GLP1(A8G/G22E)-Fn8-MBL and the sequence of the vector was

confirmed by DNA sequencing. The protein sequence of GLP1(A8G/G22E)-Fn8-MBL was listed as SEQ ID NO:11. The expression and purification protocol of the fusion protein GLP1(A8G/G22E)-Fn8-MBL (SEQ ID NO:11) was similar as described in example 3.

SEQ ID NO: 11, GLP-1(A8G/G22E)-Fn8-MBL protein sequence, the sequence of the neck and CRD domain of MBL is underlined. The flexible unstructured linker 1 (GGGSGGGGS) between GLP-1 and Fn8 is in italic. The flexible unstructured linker 2 between Fn8 and MBL is italicized.

HGEGTFTSDVSSYLEEQAAKEFIAWLVKGRGGGSGGGGS AVPPPTDLRFTNIGPDTMRVTWAPPSIDL TNFLVRYSPVK
NEEDVAELSISPSDNAVVL TNLLPGTEYVVS VSSVYE QHESTPLRGRQKTG

GGGSGG

GDSSLAASERKALQTEMARIKKWLTFSLGKQVGNKFFL

TNGEIMTFEKVKALCVKEQASVATPRNAAENGAIQNLIKEEAFLGITDEKTEGQFVDLTGNRLTYTNWNE

GEPNNAGSDEDCVLLKNGQWNDVPCSTSHLAVCEFP

Example 10. Between the therapeutic polypeptides and the scaffold proteins, various spacer lengths can be used to generate the fusion proteins with the desired radius of gyration (Rg)

In this example, we presented data to demonstrate that the flexible unstructured linkers with various lengths can be utilized in our method to adjust the Rg of the fusion protein. It is well established that a protein with a larger Rg may exhibit a longer half life *in vivo*. Therefore, the method of the invention may adjust the *in vivo* half life of the therapeutic polypeptide in a tunable fashion.

In the fusion protein of GLP-1(A8G/G22E)-Fn8-COL18NC1 described in example 4, the flexible unstructured linker 2 between Fn8 and COL18NC1 contains six residues (GGGSGG). To generate the flexible linkers with different lengths, we have synthesized the genes
GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-20,
GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-30,
GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-54 and

GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-60. In these genes, the length of the flexible linker 2 between Fn8 and COL18NC1 contained 20, 30, 54 and 60 residues, respectively. The protein sequences of GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-20, GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-30, GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-54 and GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-60 are listed as SEQ ID NO: 12-14 and 21. The synthetic genes of GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-20, GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-30, GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-54 and GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-60 were grafted to the pET29b by *NdeI* and *XhoI* for protein expressions. The expression and purification of these fusion proteins were carried out using similar protocols described in example 3. Fig. 2 showed the purified proteins of GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1, GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-20, GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-30, GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-54 and GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-60 using SDS-PAGE analysis. To estimate the Rg of the fusion proteins, the purified proteins were loaded on an analytical gel filtration column Superdex200 (GE Healthcare). Fig. 3 showed the chromatography profiles for these fusion proteins.

The gel filtration data clearly showed that varying the length of the flexible, unstructured linker between Fn8 and COL18NC1 can significantly change the apparent molecular size of the fusion proteins in solution (Fig. 3). GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-60 trimer exhibited an apparent molecular weight of ~170Kd but its genuine molecular weight is ~80Kd. GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-54 trimer exhibited an apparent molecular weight of ~160Kd while its genuine molecular weight is ~78Kd. GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1 trimer exhibited an apparent molecular weight of ~100Kd while its genuine

molecular weight is ~68Kd. GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-20, GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-30 exhibited larger apparent molecular weight than their genuine molecular weight as well. Therefore, our method can provide the therapeutic polypeptide with a larger Rg which exhibited increased apparent molecular size on gel filtration profile for longer *in vivo* half life. Moreover, the flexible linker between the therapeutic polypeptide and the scaffold protein may adjust the Rg of the fusion molecule in a tunable manner.

SEQ ID NO: 12, protein sequence of GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-20, the flexible unstructured linker 2 (20 residues) between Fn8 and COL18NC1 is underlined. The flexible unstructured linker 1 between GLP-1 and Fn8 (GGGSGG) is also underlined.

HGEGTFTSDVSSYLEEQAAKEFIAWLVKSGGGGGSGGAVPPPTDLRFTNIGPDTMRVTWAPPPSIDLTNFLVRYSPVKN
EEDVAELSISPSDNAVVLTNLLPGTEYVVS SVSSVYEQHESTPLRGRQKTGGGGSGGGSGGGSGGGGS
GASSGVRLWATRQAMLGQVHEVPEGWLIFVAEQEELYVRVQNGFRKVQLEARTPLPRG

10

SEQ ID NO: 13, protein sequence of GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-30, the flexible unstructured linker 2 (30 residues) between Fn8 and COL18NC1 is underlined. The flexible unstructured linker 1 between GLP-1 and Fn8 (GGGSGG) is also underlined.

HGEGTFTSDVSSYLEEQAAKEFIAWLVKSGGGGGSGGAVPPPTDLRFTNIGPDTMRVTWAPPPSIDLTNFLVRYSPVKN
EEDVAELSISPSDNAVVLTNLLPGTEYVVS SVSSVYEQHESTPLRGRQKTG
GGGGSGGGGSASSASTGGPSGGGGSGGGGS
GASSGVRLWATRQAMLGQVHEVPEGWLIFVAEQEELYVRVQNGFRKVQLEARTPLPRG

SEQ ID NO: 14, protein sequence of GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-54, the flexible unstructured linker 2 (54 residues) between Fn8 and COL18NC1 is underlined. The flexible unstructured linker 1 between GLP-1 and Fn8 (GGGSGG) is also underlined.

HGEGTFTSDVSSYLEEQAAKEFIAWLVKSGGGGGSGGAVPPPTDLRFTNIGPDTMRVTWAPPPSIDLTNFLVRYSPVKN
EEDVAELSISPSDNAVVLTNLLPGTEYVVS SVSSVYEQHESTPLRGRQKTG
GGGGSGGGGSTASSASTGGPSGGGGSGGGGSAPSSGSTSGGTAAGGGGSGGGGS
GASSGVRLWATRQAMLGQVHEVPEGWLIFVAEQEELYVRVQNGFRKVQLEARTPLPRG

HGEFTFTSDVSSYLEEQA AKEFI AVL VKSGGGGGG AVPPPTDLRFTNIGPDTMRVTWAPPSIDL TNFLVRYSPVKN
EEDVAELSISPSDN AVVL TNLLPGTEYVVSVSSVYEQHESTPLRGRQKTG
GGGSGGGSGGGSTASSASTKGPSGGGSGGGSGGGGSA PSSKSTSGGTAAGGGSGGGSGGGG
GASSGVRLWATRQAMLGQVHEVPEGWLIFVAEQEELYVRVQNGFRKVQLEARTPLPRG

Through binding and activating a specific G protein-coupled receptor (GLP-1 receptor), GLP-1 stimulates the signaling pathway to increase cAMP level in cells. Therefore, measuring the cytoplasmic cAMP level can be an accurate method to evaluate the biological activity of GLP-1. Chinese Hamster Ovary (CHO) cells stably transfected with human GLP-1 receptor (GLP-1R) were generated and named as S-CHO cells. S-CHO cells were propagated in DMEM medium with 10% FCS containing 0.05mg/ml G418. Before analysis, SCHO cells were grown to 70–80% confluence in 6-well plates at 37°C. The cells were treated 0.2mM 3-isobutyl-1-methylxanthine (IBMX). Cells were incubated with GLP-1 fusion proteins at various concentrations of 1nM, 3nM, 10nM, 33nM, 100nM for 15 min at 37°C. The cells were then lysed by use of cold lysis buffer. The supernatants of the cell extracts were used for cAMP level determinations. The Parameter cAMP ELISA kit from R&D Systems was utilized to measure the cAMP concentrations in the cell lysates. The EC50 values of the GLP-1 fusion proteins were generated by using the software Origin. The GLP-1 (7-37) peptide (Anaspec) and BSA were used as positive and negative controls. Fig. 4 showed the results of cAMP assays for GLP1 (7-37) peptide, GLP1(A8G/G22E)-Fn8-COL18NC1 fusion protein and GLP1(A8G/G22E)-Fn8-COL18NC1-54 fusion protein. The data indicated that the fusion of Glp1(A8G/G22E) and the scaffold protein did not affect GLP-1 activity. As a matter of fact, if the lengths of spacers between GLP-1 mutations and scaffold proteins are increased, the activities of GLP-1 improved. We reasoned that a longer flexible, unstructured linker may provide the GLP-1

peptide with more freedom to interact with the GLP-1 receptor. The EC50 values of a number of GLP-1 containing fusion proteins are listed in Table 1.

Table 1. The activities of GLP-1 and GLP-1 containing fusion proteins

Proteins	EC50 (nM) measured by cAMP assay
GLP-1(7-37) peptide	3.8
GLP-1-Fn7-COL18NC1	2.2
GLP-1-Fn8-COL18NC1	2.1
GLP-1-TNCfn3- COL18NC1	2.3
GLP1(A8G/G22E)-Fn8- COL18NC1	1.3
GLP1(A8V/G22E)-Fn8- COL18NC1	1.3
GLP1(A8S/G22E)-Fn8-COL18NC1	1.3
GLP1(A8V/G22E/R36S)-Fn8-COL18NC1	1.3
GLP1(A8G/G22E)-Fn8-COL15NC1	1.2
GLP1(A8G/G22E)-Fn8-COL19NC2	1.3
GLP1(A8G/G22E)-Fn8-ACRP30	1.5
GLP1(A8G/G22E)-Fn8-MBL	1.5
GLP1(A8G/G22E/R36S)-Fn8-COL15NC1-20	1.3
GLP1(A8G/G22E/R36S)-Fn8-COL15NC1-30	1.4
GLP1(A8G/G22E/R36S)-Fn8-COL15NC1-54	0.5
GLP1(A8G/G22E/R36S)-Fn8-COL15NC1-60	0.5

5

Example 12 : The pharmacokinetics studies for
 GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1,
 GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-20,
 GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-30,
 10 GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-54.

To evaluate the pharmacokinetics profiles for the GLP-1 containing fusion proteins generated using the method of the invention, we purified GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1,

GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-20,
GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-30,
GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-54 in PBS buffer, pH 7.2. These
fusion proteins were administered on SD(Sprague-Dawley) rats by
5 intraperitoneal injections at the concentration of 0.66mg, 0.72mg,
0.75mg,0.78mg/kg animal respectively. Blood samples were taken at various
time points after injections such as 0-min, 30-min, 1-hour, 2-hour, 4-hour,
8-hour, 24-hour, 48-hour, 3-day, 4-day, 5-day. 7-day, 10-day.. The serum
samples were centrifuged and kept at -80°C freezer.

10 The GLP-1 concentrations within the samples were examined by use of
the sandwich ELISA method. The mouse monoclonal antibody against GLP-1
peptide at the concentration of 1ug/ml (sc57510, Santa Cruz Biotech) was
coated on ELISA plate for 1 hour at room temperature. Then the plate was
washed by PBST buffer three times and the wells were blocked by PBS with
15 10% FBS for 1 hour at room temperature. The plate was washed three times
before the serum samples containing GLP-1 fusion proteins were added. The
serum samples could be diluted to 20-50 folds before use. The ELISA plate
was incubated with the serum samples at room temperature for 1 hour and then
washed by PBST buffer five times. Then Fibronectin rabbit ployclonal
20 antibody (Abcam, ab299) at the concentration of 1ug/ml in PBST buffer was
added to the wells. The plate was washed extensively after incubation of 1
hour at room temperature. The secondary antibody, Goat anti-rabbit IgG HRP
conjugated antibody (Beijing ZSGB-Bio company, ZB 5301), was added into
the wells and the color was developed using TMB
25 (3,3',5,5'-tetramethylbenzidine, BD biosciences, Cat 555214). The plate
reader (Bio –Rad microplate reader Model 680) was utilized to obtain the
OD450 readings. This method has been calibrated using purified proteins first.

Fig. 5 showed the pharmacokinetics profiles of the GLP-1 containing
proteins by use of the sandwich ELISA method described above. The
30 pharmacokinetics parameters were obtained by using the WinNonlin software
(Table 2). The data clearly showed that the GLP-1 containing fusion proteins
generated by use of the method of the invention exhibited much extended *in*

in vivo half life possibly due to their enlarged Rg. The data also showed that the flexible, unstructured linker between the therapeutic polypeptide and the scaffold protein may adjust the *in vivo* half life of the fusion molecules in a tunable manner.

- 5 Table 2. Pharmacokinetics parameters for the GLP-1 containing fusion proteins in Sprague-Dawley rat. GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1, GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-20, GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-30, GLP-1(A8G/G22E/R36S)-Fn8-COL18NC1-54 were shown in abbreviation as
10 NC1, NC1-20, NC1-30, NC1-54.

		NC1	NC1-20	NC1-30	NC1-54
T1/2	hour	5.22	6.45	7.22	9.10
Cmax	ug/ml	1.86	2.26	2.15	2.64
AUC(0-t)	ug*h/ml	24.8	36.21	47.58	86.64
AUCinf	ug*h/ml	26.53	37.24	47.6	86.67
V	ml/kg	568.04	530.16	479.81	350.86
CL	ml/h/kg	75.96	57.03	46.04	27.06

Example 13. Cloning of human parathyroid hormone (PTH) fused with COL18NC1

- 15 Human parathyroid hormone (PTH) has been shown to stimulate the bone mass significantly. The peptide PTH(1-34) and full length PTH(1-84) are currently used in the treatment of osteoporosis. The half life of human PTH(1-34) and PTH(1-84) in circulation is quite short, which limits their applications⁴⁹. In this invention, we applied the method of the invention on
20 human PTH(1-34) and PTH(1-84) by fusing human PTH(1-34) and PTH(1-84) with human collagen XVIII NC1 (COL18NC1). The synthetic genes encoding human PTH(1-34) and PTH(1-84) were digested by *NdeI* and *BamHI* and ligated into the digested vector pET29b- COL18NC1 (generated in example 1) to construct the fusion genes PTH(1-34)-COL18NC1 and
25 PTH-COL18NC1. The protein sequences of PTH(1-34)-COL18NC1 and

PTH-COL18NC1 are listed as SEQ ID NO: 15 and 16. The recombinant PTH(1-34)-COL18NC1 and PTH-COL18NC1 have been prepared and the purities of the proteins were examined by SDS-PAGE analysis (purity>95%). The biological activities of these proteins were measured by use of the cAMP assay. PTH can interact with the PTH receptor to stimulate the cAMP production. The cAMP assay was carried out as described in example 11 except that the rat osteosarcoma cell line UMR-106 (ATCC) was utilized in the assay. The results indicated that the biological activity of PTH(1-34)-COL18NC1 and PTH-COL18NC1 were similar as that of the PTH(1-34) polypeptide. The data showed that the fusion proteins PTH(1-34)-COL18NC1 and PTH-COL18NC1 retained the biological activity to interact with PTH receptor to stimulate the signaling pathway.

SEQ ID NO: 15, PTH(1-34)-COL18NC1 protein sequence. The unstructured linker region GGGGSGG between PTH(1-34) and COL18NC1 is underlined.

SVSEIQLMHNLGKHLNSMERVEWLRKKLQDVHNF

GGGGSGGGASSGVRLWATRQAMLGQVHEVPEGWLIFVAEQEELYVRVQNGFRKVQLEARTPLPRG

15

SEQ ID NO: 16, PTH-COL18NC1 protein sequence. The linker region between PTH and COL18NC1 is underlined. The unstructured linker region GGGGSGG is underlined.

SVSEIQLMHNLGKHLNSMERVEWLRKKLQDVHNFVALGAPLAPRDAGSQRPKKEDNVLVESHEKSLGEADKADVNLTAKSQ

GGGGSGGGASSGVRLWATRQAMLGQVHEVPEGWLIFVAEQEELYVRVQNGFRKVQLEARTPLPRG

Example 14. Cloning of Calcitonin fused with COL18NC1

Calcitonin inhibits bone removal by osteoclasts (bone remodeling cells) and promotes bone formation by osteoblasts. This leads to a net increase in bone mass. It has been shown that Salmon calcitonin is more effective than human calcitonin. The Salmon calcitonin (sCT) is currently used in the treatment of osteoporosis. The half life of calcitonin in human circulation is short (~50 minutes), which limits its applications. In this invention, we fused

salmon calcitonin with the human collagen XVIII NC1 (COL18NC1). The synthetic gene encoding salmon calcitonin was digested by *NdeI* and *BamHI* and ligated into the digested vector pET29b- COL18NC1 (generated in example 1) to construct the fusion gene sCT-COL18NC1. The protein
 5 sequence of sCT-COL18NC1 was listed as SEQ ID NO: 17.

SEQ ID NO: 17, sCT-COL18NC1 protein sequence. The unstructured linker region GGGGSGG between sCT and COL18NC1 is underlined.

CSNLSTCVLGKLSQELHKLQTYPRNTGSGTP

GGGGSGGGGASSGVRLWATRQAMLGQVHEVPEGWLIFVAEQEELYVRVQNGFRKVQLEARTPLPRG

Example 15: Construction of the fusion protein of TNFR2 and COL18NC1

TNF Receptor (TNFR2, or p75) has been fused to IgG1 Fc fragment
 10 to constitute a fusion protein Etanercept (Enbrel). Etanercept has been successfully utilized to treat severe active rheumatoid arthritis by blocking the TNF alpha functions⁵⁰. In this example, we applied the method of the invention on TNFR2 to generate the TNFR2-COL18NC1 fusion protein. TNFR2-COL18NC1 fusion protein is tri-valent and can block all three
 15 binding sites of TNFalpha while retaining a long half life *in vivo*. On the other hand, one Etanercept molecule can only block two out of three possible binding sites located on TNF alpha homo-trimer. A flexible unstructured linker of 60 residues was utilized to connect TNFR2 and COL18NC1 in the fusion protein. The gene that encoded the protein sequence ID: 18 was
 20 synthesized and sub-cloned into pET29b by *NdeI* and *XhoI*.

SEQ ID NO: 18, TNFR2-COL18NC1 protein sequence. The flexible unstructured linker between TNFR2 and COL18NC1 is underlined.

LPAQVAFTPYAPEPGSTCRLREYYDQTAQMCCSKCSPGQHAKVFCTKTSDTVCDSCEDST
 YTQLWNWVPECLSCGSRSSDQVETQACTREQNRICTCRPGWYCALSKQEGCRLCAPLRK
 CRPGFGVARPGTETSDVVCKPCAPGTFSNTTSSSTDICRPHQICNVVAIPGNASMDAVCTS
 TSPT
GGGSGGGSGGGSTASSASTKGPSGGGSGGGSGGGAPSSKSTSGGTAAGGGSGGGSGGG
 GASSGVRLWATRQAMLGQVHEVPEGWLIFVAEQEELYVRVQNGFRKVQLEARTPLPRG

The protein expression of TNFR2-COL18NC1 was carried out as described in example 3. Most of the fusion protein TNFR2-COL18NC1 was expressed as the inclusion bodies. The inclusion bodies were washed extensively by use of 50mM Tris buffer (pH8.0), NaCl 150mM, EDTA 5mM, Tween20 0.1%. The inclusion bodies were dissolved in 50mM Tris buffer (pH8.0), Urea 8M, DTT 50mM, NaCl 150mM. The denatured protein was cleared by high speed centrifugation (15,000g) for 15 minutes and the protein concentration was adjusted to 0.5mg/ml. The protein refolding was carried out by dialysis against 50mM Tris buffer (pH 8.5), NaCl 150mM, 1mM EDTA, 2mM Cystine, PEG3350 0.1%, L-Arginine 50mM and Tween20 0.05%. The refolding was allowed at 4°C for three days. The refolded TNFR2-COL18NC1 fusion protein was further purified by use of Hitrap SP HP column and gel filtration column Superdex 200 (GE Healthcare). The purified protein was kept in 20mM Hepes buffer (pH 7.5), NaCl 150mM. The purity of the fusion protein was examined by SDS-PAGE electrophoresis (purity >95%).

The biological activity of TNFR2-COL18NC1 was measured by its
20 ability to block the TNFalpha signaling. Our data indicated that
TNFR2-COL18NC1 can inhibit the cell killing activity of TNFalpha for
L929 cells as efficiently as TNFR2-IgG1 Fc fusion protein.

Example 16: Construction of the fusion protein VEGFR1R2 and COL18NC1

Vascular endothelial growth factor (VEGF) plays a critical role during
5 normal embryonic angiogenesis and also in the pathological angiogenesis such
as cancer. Numerous studies suggested that inhibiting VEGF functions may be
an efficient treatment for cancer patients. VEGF-Trap was created by fusing
the second Ig domain of VEGF receptor 1 (VEGFR1) with the third Ig domain
of VEGF receptor 2 (VEGFR2) and the Fc fragment of IgG⁵¹. VEGF-trap has
10 high affinity to VEGF and has shown promising anti-cancer efficacy in
clinical trials. In this example, we constructed a novel fusion protein
VEGFR1R2 that consists of human VEGFR1 second Ig domain and human
VEGFR2 third Ig domain. Then we applied the method of the invention to
VEGFR1R2 to generate a tri-valent VEGFR1R2-COL18NC1 fusion
15 protein. A flexible unstructured linker of 60 residues was utilized to
connect the VEGFR1R2 and COL18NC1 in the construct.

The synthetic gene that encodes the human VEGFR1R2-COL18NC1
was grafted into the digested vector pET29b by use of *NdeI* and *XhoI*. The
protein sequence of VEGFR1R2-COL18NC1 fusion protein was listed as
20 SEQ ID NO: 19.

The expression and refolding of VEGFR1R2-COL18NC1 was carried
out using the similar protocol as described in example 15. The purity of
the fusion protein was examined by SDS-PAGE electrophoresis (purity
>95%). The biological activity of VEGFR1R2-COL18NC1 fusion protein
25 was shown by its ability to interact with VEGF. Our data from SPR by use
of Biacore indicated that VEGFR1R2-COL18NC1 fusion protein can bind
VEGF with the similar affinity as the VEGFR1R2-IgG Fc fusion protein.

SEQ ID NO: 19, VEGFR1R2-COL18NC1 fusion protein sequence. The flexible unstructured linker region between VEGFR1R2 and COL18NC1 is underlined.

SDTGRPFVEMYSEIPEIIHMTGRELVIPCRVTSPNITVTLKKFPLDTLIPDGKRIIWDNRKGFISNATYKEIGLLT
 CEATVNGHLYKTNLTHRQGYRIYDVVLSPSHGIELSVGEKLVNCTARTELVGIDFNWEYPSSKHQHKKLVRDLK
 TQSGSEMKKFLSTLTIDGITSQGLYTCAASSGLMTKKNSTFVRVHE
GGGSGGGSGGGSTASSASTKGPSGGGSGGGSGGGSSAPSSKSTSGGTAAGGGSGGGSGGGG
 GASSGVRLWATRQAMLGQVHEVPEGWLIFVAEQEELYVRVQNGFRKVQLEARTPLPRG

Example 17: Construction of the fusion protein of human RANK and COL18NC1

5 In many bone loss conditions, RANK Ligand (RANKL) overwhelms the body's natural defenses against bone destruction⁵². Therefore, blocking the RANKL functions has been utilized to treat osteoporosis. For example, the monoclonal antibody Denosumab is designed to target RANKL, and it has been approved by FDA to treat osteoporosis. It has been shown that RANK
 10 extra-cellular domain can interact with RANKL with high affinity ($K_d \sim 60 \text{ pM}$)⁵³. RANKL forms a homo-trimer in solution and RANK extra-cellular domain forms a monomer. In this example, RANK was clustered into a homo-trimer by use of the method of the invention to interact with RANKL. A flexible unstructured linker of 60 residues was utilized to connect RANK extra-cellular
 15 domain with COL18NC1 in this example.

The RANK-COL18NC1 fusion gene was synthesized and sub-cloned into pET29b for expression. The protein sequence of RANK-COL18NC1 fusion protein was listed as SEQ ID NO: 20.

20 The expression and refolding of RANK-COL18NC1 was carried out using the similar protocol as described in example 15. The purity of the fusion protein was examined by SDS-PAGE electrophoresis (purity >95%). The biological activity of RANK-COL18NC1 fusion protein was shown by its ability to interact with RANK ligand. Our data from SPR by use of Biacore indicated that RANK-COL18NC1 fusion protein can interact with

RANK ligand with the similar affinity as the RANK extracellular domain-IgG Fc fusion protein.

SEQ ID NO: 20, RANK-COL18NC1 fusion protein sequence. The flexible unstructured linker region between RANK and COL18NC1 is underlined.

QIAPPCTSEKHYEHLGRCCNKCEPGKYMSSKCTTTSDSVCLPCGPDEYLDWNEEDKCLLHKVCDTGKALVAVVAGNSTT
PRRCCTAGYHWSQDCECCRRNTECAPGLGAQHPLQLNKDTVCKPCLAGYFSDAFSSTDKCRPWTNCTFLGKRVEHHGTE
KSDAVCSSSLPA
GGSGGGSGGGSTASSASTKGPSGGSGGGSGGGSGGGAPSSKSTSGGTAAGGGSGGGSGGG
GASSGVRLWATRQAMLGQVHEVPEGWLIFVAEQEELYVRVQNGFRKVQLEARTPLPRG

Cited patents:

1. United States Patent 7557183 Polyethylene glycol linked GLP-1 compounds
2. United States Patent 8093356 Pegylated human interferon polypeptides
- 5 3. United States Patent 8058398 Modified G-CSF polypeptide
4. United States Patent 7052686 Pegylated interleukin-10
5. United States Patent 8030269 Calcitonin drug-oligomer conjugates, and uses thereof
6. United States Patent Application 20090325865 Liquid Formulations of
10 Pegylated Growth Hormone
7. United States Patent 8053561 Pegylated factor VIII
8. United States Patent 8053410 Pegylated factor VII glycoforms
9. United States Patent Application 20090312236 PEGYLATED
INSULIN LISPRO COMPOUNDS
- 15 10. United States Patent 7271149 GLP-1 fusion proteins
11. United States Patent 6946134 Albumin fusion proteins
12. United States Patent 7785599 Albumin fusion proteins

Cited literatures:

1. Molineux, G. Pegylation: engineering improved biopharmaceuticals for
5 oncology. *Pharmacotherapy* 23, 3S-8S (2003).
2. Crawford, J. Clinical uses of pegylated pharmaceuticals in oncology.
Cancer Treat Rev 28 Suppl A, 7-11 (2002).
3. Pasut, G. & Veronese, F.M. PEG conjugates in clinical development or
use as anticancer agents: an overview. *Adv Drug Deliv Rev* 61, 1177-88
10 (2009).
4. Veronese, F.M. & Mero, A. The impact of PEGylation on biological
therapies. *BioDrugs* 22, 315-29 (2008).
5. Eliason, J.F. Pegylated cytokines: potential application in
immunotherapy of cancer. *BioDrugs* 15, 705-11 (2001).
- 15 6. Glaesner, W. et al. Engineering and characterization of the long-acting
glucagon-like peptide-1 analogue LY2189265, an Fc fusion protein. *Diabetes
Metab Res Rev* 26, 287-96.
7. Picha, K.M. et al. Protein engineering strategies for sustained
glucagon-like peptide-1 receptor-dependent control of glucose homeostasis.
20 *Diabetes* 57, 1926-34 (2008).
8. Bush, M.A. et al. Safety, tolerability, pharmacodynamics and
pharmacokinetics of albiglutide, a long-acting glucagon-like peptide-1
mimetic, in healthy subjects. *Diabetes Obes Metab* 11, 498-505 (2009).
9. Matthews, J.E. et al. Pharmacodynamics, pharmacokinetics, safety, and
25 tolerability of albiglutide, a long-acting glucagon-like peptide-1 mimetic, in
patients with type 2 diabetes. *J Clin Endocrinol Metab* 93, 4810-7 (2008).
10. Ricard-Blum, S. & Ruggiero, F. The collagen superfamily: from the
extracellular matrix to the cell membrane. *Pathol Biol (Paris)* 53, 430-42
(2005).
- 30 11. Khoshnoodi, J., Cartiailler, J.P., Alvares, K., Veis, A. & Hudson,
B.G. Molecular recognition in the assembly of collagens: terminal
noncollagenous domains are key recognition modules in the formation of triple

helical protomers. *J Biol Chem* 281, 38117-21 (2006).

12. Gordon, M.K. & Hahn, R.A. Collagens. *Cell Tissue Res* 339, 247-57.

13. Shoulders, M.D. & Raines, R.T. Collagen structure and stability.
5 *Annu Rev Biochem* 78, 929-58 (2009).

14. Wirz, J.A., Boudko, S.P., Lerch, T.F., Chapman, M.S. & Bachinger, H.P. Crystal structure of the human collagen XV trimerization domain: a potent trimerizing unit common to multiplexin collagens. *Matrix Biol* 30, 9-15.

15. Boudko, S.P. et al. The NC2 domain of collagen IX provides chain selection and heterotrimerization. *J Biol Chem* 285, 23721-31.

16. Boudko, S.P., Engel, J. & Bachinger, H.P. Trimerization and triple helix stabilization of the collagen XIX NC2 domain. *J Biol Chem* 283, 34345-51 (2008).

17. Kvansakul, M., Bogin, O., Hohenester, E. & Yayon, A. Crystal
15 structure of the collagen alpha1(VIII) NC1 trimer. *Matrix Biol* 22, 145-52 (2003).

18. Bogin, O. et al. Insight into Schmid metaphyseal chondrodysplasia from the crystal structure of the collagen X NC1 domain trimer. *Structure* 10, 165-73 (2002).

19. Sundaramoorthy, M., Meiyappan, M., Todd, P. & Hudson, B.G. Crystal structure of NC1 domains. Structural basis for type IV collagen assembly in basement membranes. *J Biol Chem* 277, 31142-53 (2002).

20. Boudko, S.P. et al. Crystal structure of human collagen XVIII trimerization domain: A novel collagen trimerization Fold. *J Mol Biol* 392,
25 787-802 (2009).

21. Ghai, R. et al. C1q and its growing family. *Immunobiology* 212, 253-66 (2007).

22. Schaffler, A. & Buechler, C. CTRP family: linking immunity to metabolism. *Trends Endocrinol Metab* 23, 194-204.

23. Kishore, U. & Reid, K.B. C1q: structure, function, and receptors. *Immunopharmacology* 49, 159-70 (2000).

24. Farrah, T. & Smith, C.A. Emerging cytokine family. *Nature* 358,

26 (1992).

25. Cabal-Hierro, L. & Lazo, P.S. Signal transduction by tumor necrosis factor receptors. *Cell Signal* 24, 1297-305.

26. Zelensky, A.N. & Gready, J.E. The C-type lectin-like domain
5 superfamily. *FEBS J* 272, 6179-217 (2005).

27. Feinberg, H., Powlesland, A.S., Taylor, M.E. & Weis, W.I. Trimeric structure of langerin. *J Biol Chem* 285, 13285-93.

28. Hakansson, K., Lim, N.K., Hoppe, H.J. & Reid, K.B. Crystal
10 structure of the trimeric alpha-helical coiled-coil and the three lectin domains
of human lung surfactant protein D. *Structure* 7, 255-64 (1999).

29. Ng, K.K. et al. Orientation of bound ligands in mannose-binding proteins. Implications for multivalent ligand recognition. *J Biol Chem* 277, 16088-95 (2002).

30. Nielsen, B.B. et al. Crystal structure of tetranectin, a trimeric
15 plasminogen-binding protein with an alpha-helical coiled coil. *FEBS Lett* 412, 388-96 (1997).

31. Holst, J.J., Vilsboll, T. & Deacon, C.F. The incretin system and its role in type 2 diabetes mellitus. *Mol Cell Endocrinol* 297, 127-36 (2009).

32. Baggio, L.L. & Drucker, D.J. Biology of incretins: GLP-1 and GIP.
20 *Gastroenterology* 132, 2131-57 (2007).

33. Connaris, H., Crocker, P.R. & Taylor, G.L. Enhancing the receptor affinity of the sialic acid-binding domain of *Vibrio cholerae* sialidase through multivalency. *J Biol Chem* 284, 7339-51 (2009).

34. Kim, D., Yan, Y., Valencia, C.A. & Liu, R. Heptameric Targeting
25 Ligands against EGFR and HER2 with High Stability and Avidity. *PLoS One* 7, e43077.

35. O'Reilly, M.K. & Paulson, J.C. Multivalent ligands for siglecs. *Methods Enzymol* 478, 343-63.

36. Schellenberger, V. et al. A recombinant polypeptide extends the in
30 vivo half-life of peptides and proteins in a tunable manner. *Nat Biotechnol* 27, 1186-90 (2009).

37. Pankov, R. & Yamada, K.M. Fibronectin at a glance. *J Cell Sci* 115,

3861-3 (2002).

38. Mao, Y. & Schwarzbauer, J.E. Fibronectin fibrillogenesis, a cell-mediated matrix assembly process. *Matrix Biol* 24, 389-99 (2005).

39. Leahy, D.J., Aukhil, I. & Erickson, H.P. 2.0 A crystal structure of a
5 four-domain segment of human fibronectin encompassing the RGD loop and synergy region. *Cell* 84, 155-64 (1996).

40. de Pereda, J.M., Wiche, G. & Liddington, R.C. Crystal structure of a tandem pair of fibronectin type III domains from the cytoplasmic tail of integrin $\alpha 6 \beta 4$. *EMBO J* 18, 4087-95 (1999).

10 41. Wada, K., Wada, Y., Ishibashi, F., Gojobori, T. & Ikemura, T. Codon usage tabulated from the GenBank genetic sequence data. *Nucleic Acids Res* 20 Suppl, 2111-8 (1992).

42. Kurjan, J. & Herskowitz, I. Structure of a yeast pheromone gene (MF α): a putative α -factor precursor contains four tandem copies of
15 mature α -factor. *Cell* 30, 933-43 (1982).

43. Schultz, L.D. et al. Expression and secretion in yeast of a 400-kDa envelope glycoprotein derived from Epstein-Barr virus. *Gene* 54, 113-23 (1987).

44. Burcelin, R. et al. Encapsulated, genetically engineered cells,
20 secreting glucagon-like peptide-1 for the treatment of non-insulin-dependent diabetes mellitus. *Ann N Y Acad Sci* 875, 277-85 (1999).

45. Diez, J.J. & Iglesias, P. The role of the novel adipocyte-derived hormone adiponectin in human disease. *Eur J Endocrinol* 148, 293-300 (2003).

25 46. Shapiro, L. & Scherer, P.E. The crystal structure of a complement-1q family protein suggests an evolutionary link to tumor necrosis factor. *Curr Biol* 8, 335-8 (1998).

47. Ip, W.K., Takahashi, K., Ezekowitz, R.A. & Stuart, L.M. Mannose-binding lectin and innate immunity. *Immunol Rev* 230, 9-21 (2009).

30 48. Takahashi, K. & Ezekowitz, R.A. The role of the mannose-binding lectin in innate immunity. *Clin Infect Dis* 41 Suppl 7, S440-4 (2005).

49. Bieglmayer, C., Prager, G. & Niederle, B. Kinetic analyses of

parathyroid hormone clearance as measured by three rapid immunoassays during parathyroidectomy. *Clin Chem* 48, 1731-8 (2002).

5 50. Aaltonen, K.J. et al. Systematic review and meta-analysis of the efficacy and safety of existing TNF blocking agents in treatment of rheumatoid arthritis. *PLoS One* 7, e30275.

51. Stewart, M.W. Aflibercept (VEGF-TRAP): the next anti-VEGF drug. *Inflamm Allergy Drug Targets* 10, 497-508.

52. McClung, M.R. et al. Denosumab in postmenopausal women with low bone mineral density. *N Engl J Med* 354, 821-31 (2006).

10 53. Liu, C. et al. Structural and functional insights of RANKL-RANK interaction and signaling. *J Immunol* 184, 6910-9.

CLAIMS

1. A fusion protein comprising a therapeutic polypeptide fused to a scaffold protein which forms a homo-trimer in solution, wherein the
5 therapeutic polypeptide is connected with the scaffold protein via a spacer, and the spacer comprises a flexible un-structured linker whose length is adjustable.

2. The fusion protein of claim 1, wherein the spacer further comprises a proteinous connecting moiety.

3. The fusion protein of claim 2, wherein the proteinous connecting
10 moiety is a proteinous sequence having an elongated shape.

4. The fusion protein of any of claims 1-3, wherein the proteinous connecting moiety is connected with the therapeutic polypeptide and/or the scaffold protein via a flexible un-structured linker.

5. The fusion protein of any of claims 1-4, wherein the fusion protein
15 comprises, from N-terminus to C-terminus, the therapeutic polypeptide, a first flexible un-structured linker, a proteinous connecting moiety, a second flexible un-structured linker, and the scaffold protein.

6. The fusion protein of any of claims 1-4, wherein the fusion protein
20 comprises, from N-terminus to C-terminus, the therapeutic polypeptide, a proteinous connecting moiety, the flexible un-structured linker, and the scaffold protein.

7. The fusion protein of any of claims 1-4, wherein the fusion protein
25 comprises, from N-terminus to C-terminus, the scaffold protein, a first flexible un-structured linker, a proteinous connecting moiety, a second flexible un-structured linker, and the therapeutic polypeptide.

8. The fusion protein of any of claims 1-4, wherein the fusion protein
comprises, from N-terminus to C-terminus, the scaffold protein, a proteinous connecting moiety, the flexible un-structured linker, and the therapeutic polypeptide.

30 9. The fusion protein of any of claims 1-8, wherein the scaffold protein is selected from the group consisting of human collagen noncollagenous (NC) domains which form stable homo-trimers in solution, proteins which form

homo-trimers in solution with C1q-like molecular structures, proteins which form homo-trimers in solution with TNF-like molecular structures, and proteins with C-type lectin-like domains (CTLD) which form homo-trimers in solution.

5 10. The fusion protein of any of claims 1-8, wherein the scaffold protein is selected from the group consisting of the NC1 domain within Multiplexin type of human Collagen, NC2 domain within FACIT type of collagen, human C1q A chain, C1q B chain, C1q C chain, cb1n family members, human EMILIN-1, multimerin, ACRP30/adiponectin, adipolin, resistin, resistin-like
10 molecule (RELM) hormone family members, human TNFalpha, TNFbeta, TRAIL, RANK ligand, Fas ligand, CD 30 ligand, CD40 ligand, CD27 ligand, OX40L, CD137, mannan-binding lectin (MBL), surfactant protein A (SP-A), surfactant protein D (SP-D), collectin liver 1 (CL-L1), collectin placenta 1 (CL-P1), conglutinin, collectin of 43 kDa (CL-43) and collectin of 46 kDa
15 (CL-46), Langerin, Tetranectin and functional variants thereof.

11. The fusion protein of any of claims 1-10, wherein the therapeutic polypeptide is connected to the N-terminus or C-terminus of the scaffold protein.

12. The fusion protein of any of claims 1-11, wherein the therapeutic
20 polypeptide is selected from the group consisting of human glucagon-like peptide-1 (GLP-1), Calcitonin, human Parathyroid hormone (PTH), G-CSF, GM-CSF, Interferon, VEGF receptors, TNF alpha receptors, RANK, Growth hormone, Erythropoietin, single-chain Fv, single domain antibodies and functional variants thereof.

25 13. The fusion protein of any of claims 1-12, wherein the flexible un-structured linker is rich in G, S, A, T or P.

14. The fusion protein of claim 13, wherein the flexible un-structured linker is selected from the group consisting of (G4S)_n, (G3S)_n, (G2S2)_n, where n is an integer, GGGS, GGGSGGGG, GGGSGGGGS, GSGG,
30 GGGSGGG, GGGGSGGG, GGGSGG, GGGGSGG, GGGGSGGGGSGGGGSGGGGS, GGGGSGGGGSASSASTGGPSGGGGSGGGGS,

GGGGSGGGGSTASSASTGGPSGGGGSGGGGSAPSSGSTSGGTAAAGGG
 GSGGGGS and
 GGGSGGGSGGGSTASSASTKGPSGGGGSGGGSGGGGSAPSSKSTSGGTAA
 GGGSGGGSGGGGS.

5 15. The fusion protein of any of claims 1-14, wherein the length of the flexible un-structured linker is within the range of 5 to 100 amino acid residues.

16. The fusion protein of any of claims 1-15, wherein the proteinous connecting moiety is human Fibronectin type III domain.

10 17. The fusion protein of claim 16, wherein the human Fibronectin type III domain is selected from the group consisting of Tenascin, Usherin, Titin, tripartite motif (TRIM) family members, tissue factor, TIE1, TIE2, SPEG, SORL1, SDK1, ROBO1, ROBO2, SDK2, Receptor-type tyrosine-protein phosphatase, prolactin receptor, L1CAM, NCAM1, NCAM2, myomesin 1,
 15 myomesin 2, Myosin-binding protein C, LIFR, Leptin receptor, Integrin, Insulin receptor, Contactin, Collagen, Cytokine receptor-like factor, Inteferon receptor, Growth hormone receptor, fibronectin, leucine rich transmembrane protein (FLRT) members, IL, ephrin type-A receptor, ephrin type-B receptor, IL-6R, gp130, IL11RA, IL12RB, IL20RB, IL23R, IL27RA and IL31RA.

20 18. The fusion protein of any of claims 1-17, wherein the fusion protein comprises: a therapeutic polypeptide selected from the group consisting of GLP-1, GLP1(A8G/G22E) and GLP1(A8G/G22E/R36S); a first flexible un-structured linker; a proteinous connecting moiety selected from the group consisting of human fibronectin type III domain 7 (Fn7), human fibronectin
 25 type III domain 8 (Fn8), and human Tenascin C fibronectin type III domain 3 (TNCfn3); a second flexible un-structured linker; and a scaffold protein selected from the group consisting of COL18NC1, COL15NC1, COL19NC2, ACRP30 C1q-like domain, and MBL neck and CRD domain.

30 19. The fusion protein of any of claims 1-17, wherein the fusion protein comprises: GLP-1(A8G/G22E/R36S), a first flexible un-structured linker, Fn8, a second flexible un-structured linker, and COL18NC1.

20. The fusion protein of claim 1, wherein the fusion protein is selected

from the group consisting of SEQ ID NO: 1-21.

21. A polynucleotide sequence encoding the fusion protein of any of claims 1-20.

22. An expression vector comprising the polynucleotide sequence of
5 claim 21 and expression control elements.

23. A pharmaceutical composition comprising the fusion protein of any of claims 1-20 and a pharmaceutically acceptable carrier.

24. A method of improving the pharmacokinetic property of a therapeutic polypeptide, comprising the step of fusing the therapeutic polypeptide to a
10 scaffold protein which forms a homo-trimer in solution, wherein the therapeutic polypeptide is connected with the scaffold protein via a spacer, and the spacer comprises a flexible un-structured linker whose length is adjustable to adjust the apparent molecular size and/or the *in vivo* half life of the fusion protein.

15 25. The method of claim 24, wherein the spacer further comprises a proteinous connecting moiety..

26. The method of claim 25, wherein the proteinous connecting moiety is a proteinous sequence having an elongated shape.

27. The method of any of claims 24-26, wherein proteinous connecting
20 moiety is connected with the therapeutic polypeptide and/or the scaffold protein via a flexible un-structured linker.

28. The method of any of claims 24-27, wherein the fusion protein comprises, from N-terminus to C-terminus, the therapeutic polypeptide, a first flexible un-structured linker, a proteinous connecting moiety, a second
25 flexible un-structured linker, and the scaffold protein.

29. The method of any of claims 24-28, wherein the scaffold protein is selected from the group consisting of human collagen noncollagenous (NC) domains which form stable homo-trimers in solution, proteins which form
30 homo-trimers in solution with C1q-like molecular structures, proteins which form homo-trimers in solution with TNF-like molecular structures, and proteins with C-type lectin-like domains (CTLD) which form homo-trimers in solution.

35. The method of any of claims 24-34, wherein the length of the flexible

un-structured linker is within the range of 5 to 100 amino acid residues.

36. The method of any of claims 24-35, wherein the proteinous connecting moiety is human Fibronectin type III domain.

37. The method of claim 36, wherein the human Fibronectin type III domain (Fn) is selected from the group consisting of Tenascin, Usherin, Titin, tripartite motif (TRIM) family members, tissue factor, TIE1, TIE2, SPEG, SORL1, SDK1, ROBO1, ROBO2, SDK2, Receptor-type tyrosine-protein phosphatase, prolactin receptor, L1CAM, NCAM1, NCAM2, myomesin 1, myomesin 2, Myosin-binding protein C, LIFR, Leptin receptor, Integrin, Insulin receptor, Contactin, Collagen, Cytokine receptor-like factor, Inteferon receptor, Growth hormone receptor, fibronectin, leucine rich transmembrane protein (FLRT) members, IL, ephrin type-A receptor, ephrin type-B receptor, IL-6R, gp130, IL11RA, IL12RB, IL20RB, IL23R, IL27RA and IL31RA.

38. The method of any of claims 24-37, wherein the fusion protein obtained thereby comprises: a therapeutic polypeptide selected from the group consisting of GLP-1, GLP1(A8G/G22E) and GLP1(A8G/G22E/R36S); a first flexible un-structured linker; a proteinous connecting moiety selected from the group consisting of Fn7, Fn8 and TNCfn3; a second flexible un-structured linker; and a scaffold protein selected from the group consisting of COL18NC1, COL15NC1, COL19NC2, ACRP30 C1q-like domain, and MBL neck and CRD domain.

39. The method of any of claims 24-37, wherein the fusion protein obtained thereby comprises: GLP1(A8G/G22E/R36S), a first flexible un-structured linker, Fn8, a second flexible un-structured linker, and COL18NC1.

40. The method of claim 24, wherein the fusion protein obtained thereby is selected from the group consisting of SEQ ID NO: 1-21.

1/5

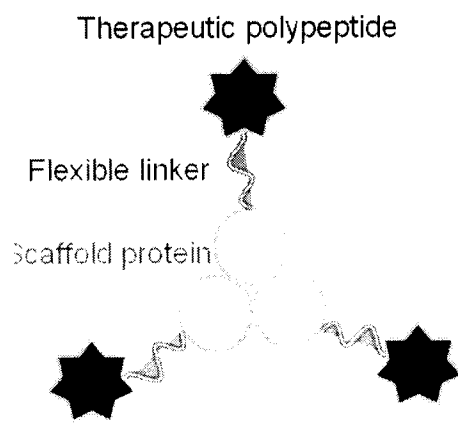


Fig. 1a

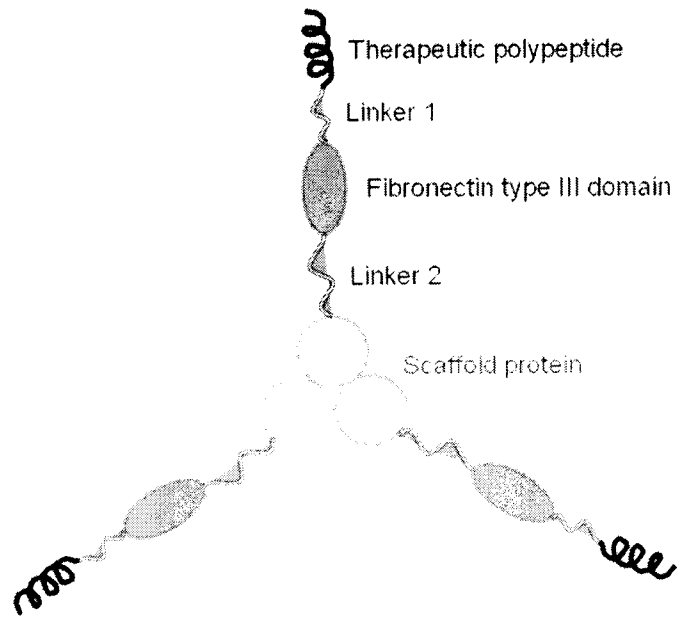


Fig. 1b

2/5

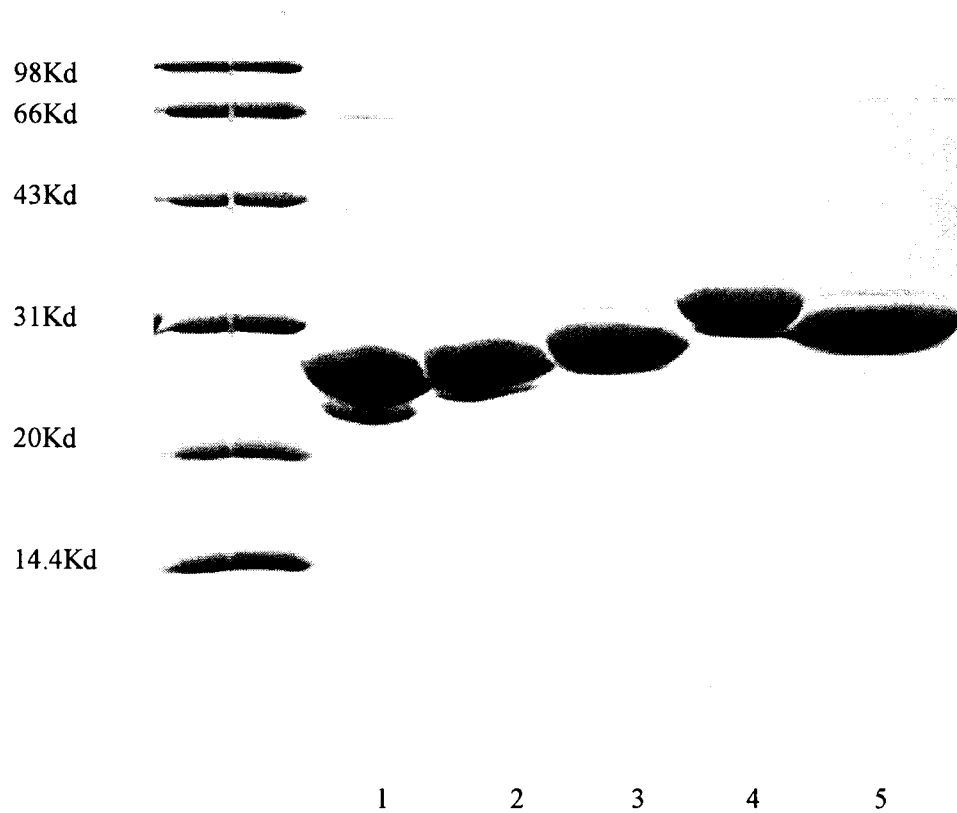


Fig. 2

3/5

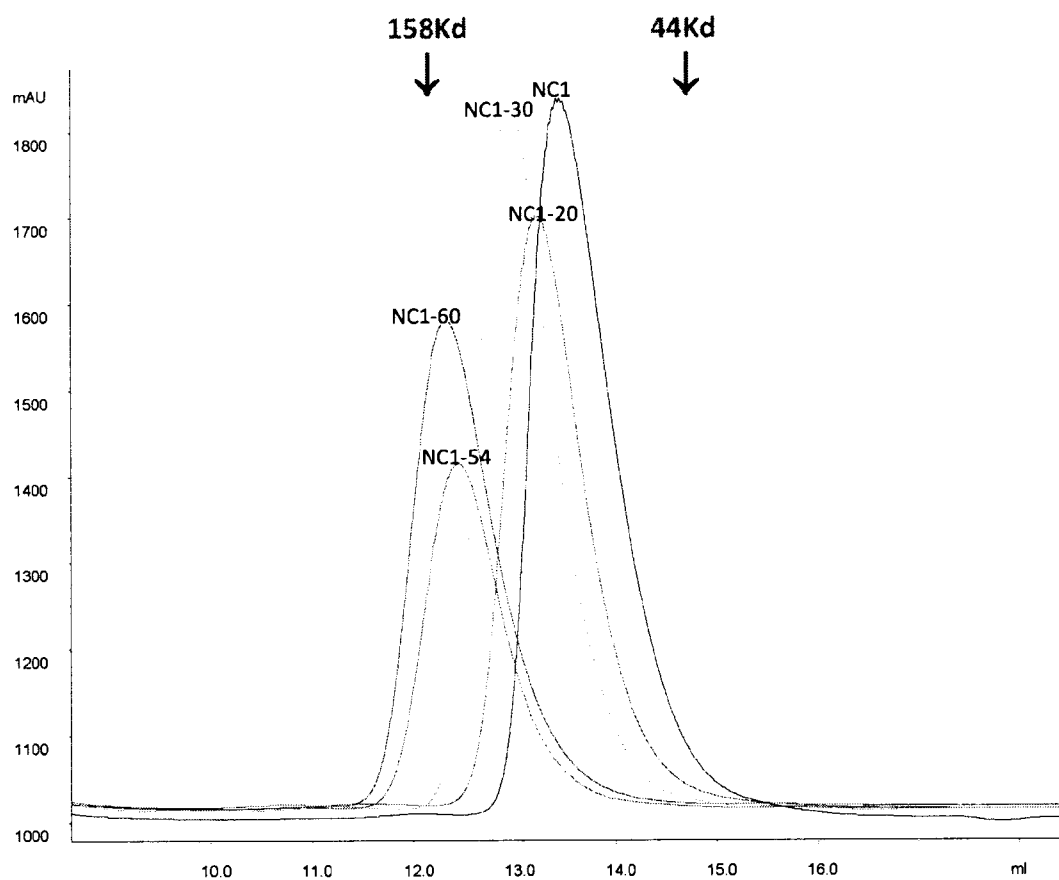


Fig. 3

4/5

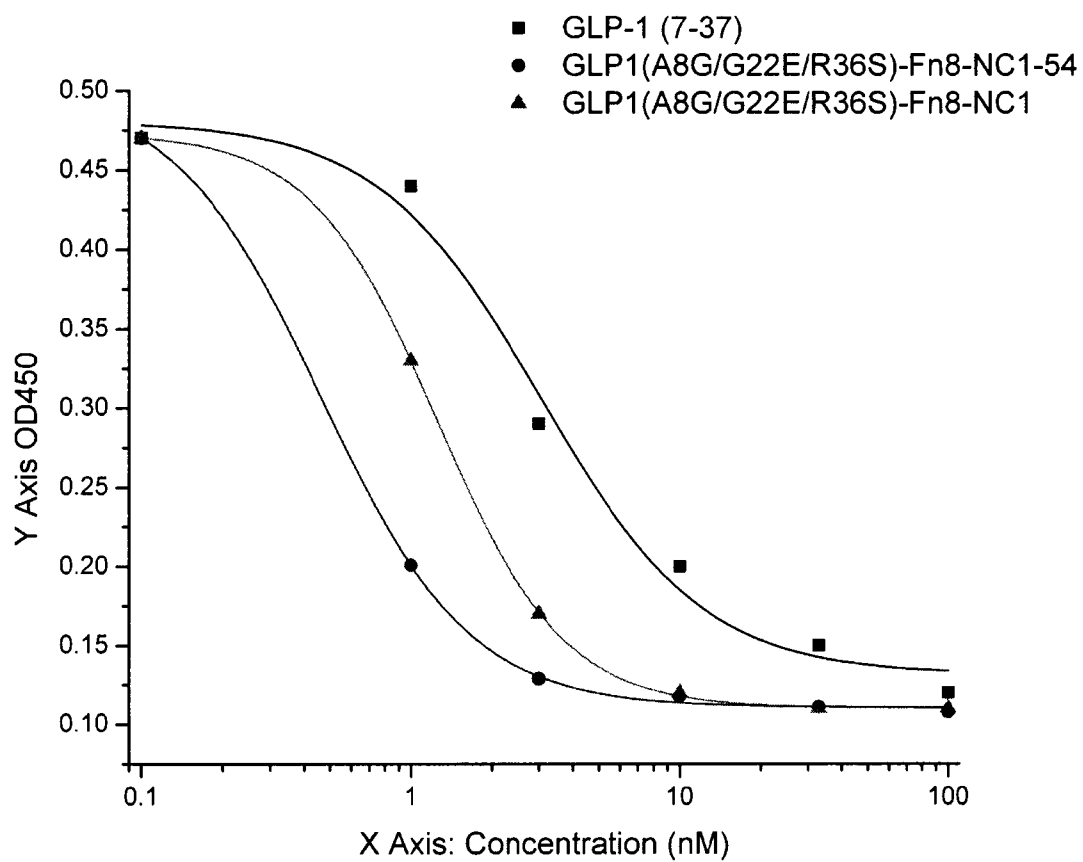


Fig. 4

5/5

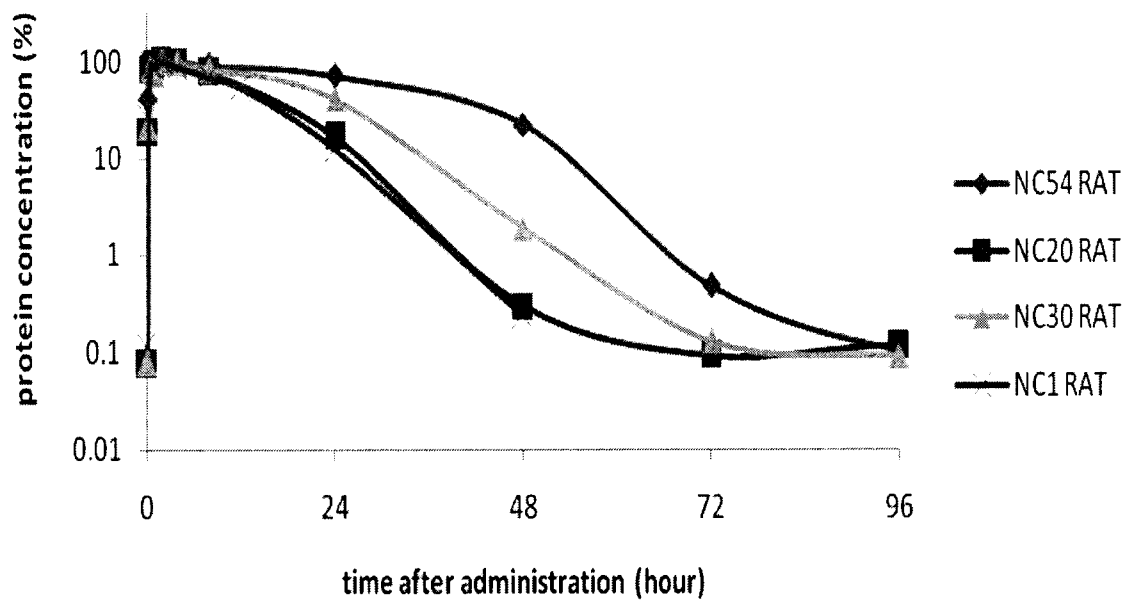


Fig. 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2012/001723

A. CLASSIFICATION OF SUBJECT MATTER

See the extra sheet

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: C07K, C12N, A61K, A61P

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

DWPI, SIPOABS, STN and keywords: GLP-1, glucagon-like peptide-1, calcitonin, human parathyroid hormone, PTH, G-CSF, GM-CSF, interferon, VEGF receptor, TNF receptor, RANK, growth hormone, erythropoietin, noncollageous, NC1, NC2, clq, cbln, emilin-1, multimerin acrp30, adiponectin, adipolin, resistin, TNF, RANK, TRAIL, fas, cd, ox401, cd137, mannan-binding lectin, mbl, surfactant, collectin placenta 1, cl-pl, conglutinin, collectin, langerin, tetranectin, fibronectin, tenascin, usherin, titin, tripartite motif, trim, tiel, tie2, speg, sorl1 etc

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 2065402 A1 (IND TECH RES INST) 03 June 2009 (03.06.2009) See description , paragraph [0050] to [0060], [0080] to [0082], [0119], [0127], [0136] to [0139]	1, 11-15, 22-24, 31-35
Y		2-10, 16-21, 25-30, 36-40

☒ Further documents are listed in the continuation of Box C.☒ See patent family annex.

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

Date of the actual completion of the international search 22 September 2013 (22.09.2013)	Date of mailing of the international search report 03 Oct. 2013 (03.10.2013)
Name and mailing address of the ISA/CN The State Intellectual Property Office, the P.R.China 6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China 100088 Facsimile No. 86-10-62019451	Authorized officer LIU, Rui Telephone No. (86-10)62411107

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2012/001723

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 102153652 A (ZHEJIANG UNIVERSITY) 17 August 2011 (17.08.2011) See description, paragraph [0007] to [0012]	10, 18-20, 30, 38-40
Y	WO 2011/140086 A2 (BRISTOL-MYERS SQUIBB COMPANY) 10 November 2011 (10.11.2011) See description, page 1, line 31 to page 2, line 7, page 5, line 8 to page 6, line 30	2-9, 16-21, 25-30, 36-40
A	BOUDKO, S. P. et al. Trimerization and triple helix stabilization of the collagen XIX NC2 domain. The journal of biological chemistry, December 2008, vol. 283, no.49, pages 34345-34351, ISSN 1083-351X See the whole document	1-40
A	TSILIBARY, E. C. et al. The role of the main noncollageous domain (NC1) in type IV collagen self-assembly. The journal of cell biology, December 1986, vol. 103, no. 6, pages 2467-2473, ISSN 1540-8140 See the whole document	1-40

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2012/001723

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
EP 2065402 A1	03.06.2009	CN 101445559 A	03.06.2009
		TW 200922944 A	01.06.2009
		CN 101445559B B	12.09.2012
		TW I374891B B	21.10.2012
		EP 1798240 A1	20.06.2007
		CA 2569062 A1	15.06.2007
		JP 2007181458 A	19.07.2007
		US 2007264687 A1	15.11.2007
		AU 2006252078 A1	05.07.2007
		US 2008176247 A1	24.07.2008
		AU 2006252078B B2	25.09.2008
		TW 200734354 A	16.09.2007
		JP 2009131237 A	18.06.2009
		CA 2618374 A1	30.05.2009
		AU 2008200282 A1	18.06.2009
		EP 1798240 B1	27.04.2011
		DE 602006021520E E	09.06.2011
		EP 2336170 A2	22.06.2011
		JP 2011217760 A	04.11.2011
		EP 2336170 A3	25.04.2012
		TW I373475B B	01.10.2012
		ES 2365744T T3	10.10.2011
CN 102153652 A	17.08.2011	None	
WO 2011140086 A2	10.11.2011	US 2011305663 A1	15.12.2011
		WO 2011140086 A9	29.12.2011
		WO 2011140086 A3	13.09.2012
		TW 201138808 A	16.11.2011
		AU 2011248273 A1	20.12.2012

International application No.
PCT/CN2012/001723

Form PCT/ISA /210 (patent family annex) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2012/001723

Continuation of : CLASSIFICATION OF SUBJECT MATTER

C07K 19/00 (2006.01) i

C12N 15/62 (2006.01) i

A61K 38/26 (2006.01) i

A61P 3/10 (2006.01) i