(19) DANMARK

(10) **DK/EP 3660045 T3**

Patent- og
Varemærkestyrelsen

(12) # Oversættelse af europæisk patentskrift

(51) Int.Cl.: ***C 07 K 14/78 (2006.01)***

(45) Oversættelsen bekendtgjort den: **2024-09-09**

(80) Dato for Den Europæiske Patentmyndigheds
bekendtgørelse om meddelelse af patentet: **2024-08-07**

(86) Europæisk ansøgning nr.: **19211389.2**

(86) Europæisk indleveringsdag: **2019-11-26**

(87) Den europæiske ansøgnings publiceringsdag: **2020-06-03**

(30) Prioritet: **2018-11-28 CN 201811438582**

(84) Designerede stater: **AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR**

(73) Patenthaver: **Shanxi Jinbo Bio-Pharmaceutical Co., Ltd., 18 Jinbo Street , Taiyuan Tanghuai Park , Shanxi Comprehensive Reform Demonstration Zone, Shanxi 030032, Kina**

(72) Opfinder: **YANG, Xia, c/o Shanxi Jinbo Bio-Pharmaceutical Co., Ltd., 18 Jinbo Street, Taiyuan Tanghuai Park, Shanxi Comprehensive Reform Demonstration Zone, Shanxi, 030032, Kina**

(74) Fuldmægtig i Danmark: **Plougmann Vingtoft A/S, Strandvejen 70, 2900 Hellerup, Danmark**

(54) Benævnelse: **POLYPEPTID, FREMGANGSMÅDE TIL FREMSTILLING DERAF OG ANVENDELSE DERAF**

(56) Fremdragne publikationer:

EP-A1- 2 976 352
EP-B1- 2 976 352
CN-A- 103 122 027
CN-B- 109 593 126
DATABASE Geneseq [online] 5 December 2013 (2013-12-05), "Human collagen type III peptide.", retrieved from EBI accession no. GSP:BAX15724 Database accession no. BAX15724
DATABASE Geneseq [online] 5 December 2013 (2013-12-05), "Recombinant collagen protein.", retrieved from EBI accession no. GSP:BAX15728 Database accession no. BAX15728
DATABASE Geneseq [online] 11 July 2019 (2019-07-11), "Human collagen like polypeptide (TE16C) with N-terminal sequence, SEQ 5.", retrieved from EBI accession no. GSP:BGE85516 Database accession no. BGE85516
C HUA ET AL: "Characterization by high-resolution crystal structure analysis of a triple-helix region of human collagen type III with potent cell adhesion activity.", BIOCHEM BIOPHYS RES COMMUN., vol. 508, no. 4, 11 December 2018 (2018-12-11), pages 1018 - 1023, XP055677661, DOI: 10.1016/j.bbrc.2018.12.018
J YAO ET AL: "Design, expression and characterization of collagen-like proteins based on the cell adhesive and crosslinking sequences derived from native collagens.", J BIOCHEM, vol. 136, no. 5, 1 November 2004 (2004-11-01), pages 643 - 649, XP055677659
YONG Y PENG ET AL: "Constructs for the expression of repeating triple-helical protein domains; Repeating triple-helical domains", BIOMEDICAL MATERIALS, INSTITUTE OF PHYSICS PUBLISHING, BRISTOL, GB, vol. 4, no. 1, 1 February 2009 (2009-02-01), pages 15006, XP020157716, ISSN: 1748-605X, DOI: 10.1088/1748-6041/4/1/015006

Fortsættes ...

HAN RUNLIN ET AL: "Assessment of prokaryotic collagen-like sequences derived from streptococcal Scl1 and Scl2 proteins as a source of recombinant GXY polymers", APPLIED MICROBIOLOGY AND BIOTECHNOLOGY, SPRINGER BERLIN HEIDELBERG, BERLIN/HEIDELBERG, vol. 72, no. 1, 22 March 2006 (2006-03-22), pages 109 - 115, XP037008420, ISSN: 0175-7598, [retrieved on 20060322], DOI: 10.1007/S00253-006-0387-5

Description

Technical Field

5 **[0001]** The invention belongs to the technical field of genetic engineering, and relates to a polypeptide, a process for the production thereof, and a use thereof.

Background Art

10 **[0002]** Collagen is a type of protein widely distributed in connective tissues of the human body, and is also the most abundant protein in the human body and can account for 25% to 35% of total proteins. Its main function is to maintain the extracellular environment, maintain the normal physiological functions of tissues and organs, and repair the body damage. Collagen is a natural biological resource, which has the biological histocompatibility, support elasticity to cells and degradability that other polymer materials cannot achieve. Therefore, collagen may be widely used in industries 15 such as medicine and cosmetics.

[0003] However, natural collagen is insoluble in water and is not uniform in nature, and therefore it is difficult to be utilized by the human body and often needs to be treated by chemical means. Moreover, all the collagen products currently available on the market are taken from tissues of animals such as pigs, cattle and fish, and it is difficult to avoid viral infection, which, together with their incompatibility with the human body, may lead to immune rejection and allergy 20 symptoms. If collagen is extracted from human placenta raw materials, it not only has a limited source, but also faces severe punishment from the law. Therefore, the current collagen can only be used in cosmetics and health care products, and the original biological function of collagen cannot be exerted at all.

[0004] Structurally, the structure of the natural collagen of the human body is very complicated, resulting in that it is very difficult to express and mass-produce human collagen by conventional means. The natural collagen molecule has 25 a special supercoiled structure consisting of three polypeptide chains which have no intrachain hydrogen bonds and are supported only by interchain hydrogen bonds. This helical structure is a left-handed helix with three amino acid residues as the basic repeat, and these three amino acid residues are typically Gly-X-Pro. Gly is essential for the formation of hydrogen bonds in collagen; it has no side chains itself, resulting in that collagen piles up tightly. At a higher structural level, collagen supercoils will further associate to form collagen fibrils. In organisms, the synthesis and modification of 30 collagen begins with procollagen and undergoes many chemical changes such as hydroxylation, glycosylation, and cross-linking, and are regulated by a variety of biological enzymes. In addition to the collagen chain, procollagen also contains spherical heads and tails. Without these heads and tails, the collagen chain would not fold into the correct triple helix, leading to loss of the biological activity of collagen. Therefore, collagen prepared according to the original gene sequence is unlikely to spontaneously organize *in vitro* to form the correct spatial structure. Such difficulties have seriously 35 hindered the development and production of human collagen.

[0005] The traditional process for the production of collagen is to treat animal-derived tissues by acid, alkali, and enzymolysis methods and extract collagen derivatives. The collagen extracted by these methods has lost the original biological activity and thus cannot be applied to the biomedical field to perform the true function. With the development of modern biotechnology, people continue to try to utilize transgenic techniques to prepare recombinant human collagen 40 in animal, plant and microbial expression systems, which solves many shortcomings of the traditional extraction processes. Foreign research institutes have obtained milk comprising human collagen by cultivating mice containing a human collagen gene, whereas the cost of such production is too high and the production cycle is too long, so that it cannot be put into mass production.

[0006] Chinese Patent Invention No. 201210482543.2 discloses a recombinant human collagen and a process for 45 producing the same, and it discloses a recombinant human collagen obtained by linking a type III collagen peptide segment via a linker and adding a stable sequence at C-terminus.

[0007] However, there is a need in the art for more recombinant human collagens with better characteristics.

Summary of the Invention

50

[0008] The present invention provides the following contents:

1. A polypeptide comprising 16 repeats of a sequence set forth in SEQ ID No. 1, the repeats of the sequence are directly linked, wherein the polypeptide does not comprise a sequence set forth in SEQ ID No. 2.
55
2. A polynucleotide encoding the polypeptide according to item 1.
3. An expression vector comprising the polynucleotide according to item 2, and optionally comprising a nucleic acid sequence encoding SEQ ID No. 4, wherein the nucleic acid sequence encoding SEQ ID No. 4 is directly linked to the 5' end of the coding nucleic acid sequence of the polypeptide, preferably the expression vector comprises the

nucleic acid sequence of SEQ ID No. 5.

4. A host cell comprising the expression vector according to item 3, wherein the host cell is preferably *Escherichia coli.*

5. A process for the production of the polypeptide according to item 1, comprising:

(1) cultivating the host cell according to item 4 in a production medium and producing the polypeptide;
(2) harvesting the polypeptide, and optionally digesting the polypeptide, preferably digesting the polypeptide with TEV protease; and
(3) purifying the polypeptide by a Ni column and/or an anion exchange chromatography;
wherein optionally the process for the production does not include an additional step of removing endotoxin; wherein preferably the purified polypeptide is substantially free of endotoxin or contains less than 5 EU/ml endotoxin.

6. A composition comprising the polypeptide according to item 1, wherein the composition is preferably a medical device, a tissue engineering product, a cosmetic or a health care product, preferably the polypeptide is in the form of an aqueous solution of the polypeptide, preferably the composition is free of a component that prevents degradation of the polypeptide, and preferably the composition is a composition for long-term use, the long-term use being preferably more than half a year of use.

7. Use of the polypeptide according to item 1 in the manufacture of a composition, preferably a medical device, a tissue engineering product, a cosmetic, or a health supplement, wherein preferably the polypeptide is in the form of an aqueous solution of the polypeptide, preferably the composition is free of a component that prevents degradation of a polypeptide, and the composition is a composition for long-term use, the long-term use being more than half a year of use.

8. Use of the polypeptide according to item 1 for promoting cell adhesion.

9. Use of an amino acid sequence of SEQ ID No. 4 or the expression vector according to the present invention for the production of the polypeptide according to the present invention.

[0009] The polypeptide of the present invention has the following advantages: its cell adhesion is stronger than the existing recombinant type III collagens; it has higher stability in an aqueous solution, as demonstrated by degradation in more than half a year, thus eliminating the need to add a reagent for prevent polypeptide degradation; most of the endotoxin may be removed by purification by a Ni column and an anion exchange chromatography from the product of the polypeptide, and the purified polypeptide product is substantially free of endotoxin or contains 5 EU/ml or less endotoxin; the polypeptide of the present invention is obtained from a host cell in higher yield and purity.

Brief Description of the Drawings

[0010]

Figure 1: Mass spectrometry peak results of the expressed product;
Figure 2: Comparison of cell adhesion activities between different recombinant type III collagens;
Figure 3: Cell expression levels of different recombinant type III collagens;
Figure 4: Purification of different recombinant type III collagens;
Figure 5: Stability of different recombinant type III collagens;
Figure 6: Residual endotoxin of different recombinant type III collagens;
Figure 7: Structure analysis of recombinant type III collagen TE16C.

Detailed Description

[0011] Further description is provided below to facilitate an understanding of the invention.

[0012] As used herein, "medical instrument" refers to apparatus, devices, appliances, *in vitro* diagnostic reagents and calibrators, materials, and other similar or related articles that are used directly or indirectly in the human body.

[0013] As used herein, "tissue engineering product" refers to a product used in tissue engineering. Tissue engineering is an emerging discipline that combines cell biology and materials science to construct tissues or organs *in vitro* or *in vivo.*

[0014] The present invention is based, in part, on the following findings by the inventors: when a polypeptide comprising a plurality of repeat sequences of SEQ ID No. 1 is expressed in *E. coli,* in order to increase the gelling property of the recombinantly expressed polypeptide, it is often necessary to add a sequence such as the hinge-region amino acids GPPGPCCGGG (SEQ ID No. 2) at the C-terminus of the polypeptide to aid gelling (Reference: Journal of Biochemistry. 2004; 136: 643-649) because the protein of interest is a truncated protein and lacks the hinge protein structure of the full-length protein. Therefore, a polypeptide sequence comprising SEQ ID NO. 3 and SEQ ID No. 2, for example, the

sequence T16a of SEQ ID No. 9, was designed; however, it was found that when the polypeptide of the sequence such as T16a was cultured in a shake flask or a fermentation tank, most of the polypeptide of interest formed a gelatinous precipitate which could not be dissolved and purified, so the yield was very low. Previously, in order to settle this problem, through addition of non-collagen amino acid linkers between the repeat sequences (SEQ ID No. 1), a polypeptide comprising a plurality of such repeat sequences was obtained; besides, the modification of such a polypeptide was also mainly concentrated on the linker amino acid residues and C-terminal hinge region amino acids, also known as a C-terminal stable sequence.

[0015] Surprisingly, after analyzing the crystal structure of SEQ ID No. 1, the inventors found that the region of SEQ ID No. 1 can form a very stable collagen trimer structure without the involvement of a hinge region. Therefore, the inventors continued to modify the polypeptide sequence and found that in the process of expressing a polypeptide of the sequence set forth in SEQ ID No. 5, which comprises the sequence set forth in SEQ ID No. 4 and the sequence set forth in SEQ ID No. 3, a polypeptide comprising the sequence set forth in SEQ ID No. 3 may be obtained in large quantities, and the involvement of the C-terminal stable sequence is no further required. Moreover, at this time, a polypeptide does not form a colloidal structure prematurely while being recombinantly expressed, thereby not affecting subsequent protein purification. Furthermore, the inventors have surprisingly found that compared with the recombinant human collagen (SEQ ID No. 6) in Chinese Patent Invention No. 201210482543.2, the polypeptide of the present invention is more stable in aqueous solution, may be obtained from host cells in higher yield and purity, and has significantly lower endotoxin after purification by a Ni column and an anion exchange column.

GERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGI
PGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPG
FRGPAGPNGIPGEKGPAGERGAPRSGERGAPGFRGPAGPNGIPGEKGPAGE
RGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNG
IPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPRSGERG
APGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPA
GERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGP
NGIPGEKGPAGERGAPRSGERGAPGFRGPAGPNGIPGEKGPAGERGAPGE
RGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKG
PAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPRSPEFGPPGPCCG
GG (SEQ ID NO. 6).

[0016] In the present invention, the used repeat sequence of SEQ ID No. 1 is GERGAPGFRGPAGPNGIPGEKG-PAGERGAP (SEQ ID No. 1). The polypeptide of the present invention may comprise a plurality of repeat sequences set forth in SEQ ID No. 1, provided that there is no linker between the repeat sequences, the number of repeat sequences is 16, i.e., a polypeptide comprising the sequence set forth in SEQ ID No. 3:

GERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGI
PGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPG
FRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGER
GAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGI
PGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPG
FRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGER
GAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGI
PGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPG
FRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGER
GAPGERGAPGFRGPAGPNGIPGEKGPAGERGAP (SEQ ID No. 3)

[0017] The polypeptide sequence of the invention may comprise the C-terminal sequence GPPGPCCGGG (SEQ ID No. 2), which embodiment is not covered by the claimed invention. When expressing a polypeptide comprising a plurality of repeat sequences, those skilled in the art typically consider adding said sequence to increase the stability of the expressed polypeptide. Preferably, the polypeptide sequence may not comprise the C-terminal sequence GPPGPC-CGGG (SEQ ID No. 2), maintaining sequence identity to human type III collagen without introducing additional amino acids.

[0018] When the polypeptide sequence of the present invention is expressed, the sequence of ENLYFQ (SEQ ID No. 4) should be added at its N-terminus, which can be cut by TEV protease to directly obtain the sequence of SEQ ID No. 3. Preferably, the ENLYFQ (SEQ ID No. 4) sequence is directly linked to the N-terminus, to get the sequence of SEQ ID No. 5:

ENLYFQGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGP
AGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPG
ERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEK
GPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGP
AGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPG

ERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEK
GPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGP
AGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPG
ERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEK
GPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAP (SEQ ID NO. 5).

[0019] In the present invention, polypeptides are recombinant type III collagens, which are used interchangeably herein with recombinant human collagens.

[0020] In the present invention, recombination of human collagens may be carried out by a conventional method in the art. For example, they may be produced as follows: (1) construction of *E. coli* genetically engineered bacteria; (2)

fermentation culture of *E. coli* genetically engineered bacteria; (3) induction and expression of recombinant human collagen; and (4) purification and optional digestion of recombinant human collagen.

[0021] In step (1), the construction of *E. coli* genetically engineered bacteria may be carried out as follows: (1) obtaining a gene fragment of interest; (2) inserting the obtained gene fragment of interest into PET-32a expression vector to obtain a recombinant expression plasmid; (3) transforming the recombinant expression plasmid into *E. coli* competent cell BL21 (DE3), and screening positive *E. coli* genetically engineered bacteria.

[0022] In steps (2) and (3), the fermentation culture of *E. coli* genetically engineered bacteria and the induction and expression of recombinant human collagen may be carried out as follows: (1) picking up optimum single colonies of *E. coli* genetically engineered bacteria from LAB plate, placing in 10 mL of LB medium, and culturing at 37°C and 220 rpm for 12-16 hours; (2) inoculating bacterial solution into 2xYT medium at 1:100 for scale-up culture, and culturing at 37°C for about 3 hours, and when OD_{600} is between 0.4 and 0.6, adding IPTG to a final concentration of 0.5 mM for induction, continuing to culture at 16°C for 20 hours, and collecting bacteria by centrifugation.

[0023] In step (4), the purification and digestion of recombinant human collagen polypeptide may be carried out as follows: (1) resuspending the bacteria with phosphate buffer (40 mM NaH_2PO_3, 500 mM NaCl, pH 7.8) prior to ultrasonication, and collecting the supernatant by centrifugation; (2) using NI-NTA affinity column to bind to recombinant human collagen, and after washing out impurity proteins with 10 mM imidazole, adding TEV protease, performing on-column enzymatic digestion at 4°C for 16 hours, and finally obtaining the collagen polypeptide of interest.

[0024] The host cell may be a eukaryotic cell, such as a fungus and yeast, a prokaryotic cell, such as an *Enterobacteriaceae* bacterium. It will be appreciated that those skilled in the art can replace the above-mentioned *E. coli* strain with other expression strains as a host cell.

[0025] The present invention further provides a nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide of the present invention. The nucleic acid may be DNA or cDNA. The nucleic acid molecule may consist essentially of a nucleic acid sequence encoding the peptide of the present invention, or may consist solely of a nucleic acid sequence encoding the peptide of the present invention. Such a nucleic acid molecule may be synthesized by methods known in the art. Due to the degeneracy of genetic code, those skilled in the art will appreciate that nucleic acid molecules of different nucleic acid sequences may encode the same amino acid sequence.

[0026] The present invention further provides a vector comprising the nucleic acid sequence of the present invention. Suitable vectors are known in the field of vector construction, including choices of promoters and other regulatory elements, such as enhancer elements. The vector of the present invention comprises a sequence suitable for introduction into a cell. For example, the vector may be an expression vector; in the vector, the coding sequence of the polypeptide is under the control of its own cis-acting regulatory element, and the design of the vector facilitates gene integration or gene replacement in a host cell.

[0027] Those skilled in the art will appreciate that in the present invention, the term "vector" includes a DNA molecule, for example, a plasmid, a phage, virus or other vectors; it comprises one or more heterologous or recombinant nucleic acid sequences. Suitable phages and viral vectors include, but are not limited to, lambda-phage, EMBL phage, simian virus, bovine papilloma virus, Epstein-Barr virus, adenovirus, herpes virus, murine sarcoma virus, murine mammary tumor virus, lentivirus, and the like.

[0028] The polypeptide of the present invention comprises a sequence set forth in SEQ ID No. 3, or a sequence in which one or more amino acids are substituted, deleted, inserted and/or added in the sequence set forth in SEQ ID No. 3 (which embodiment is not covered by the claimed invention), as long as the polypeptide of the present invention retains cell adhesion effect of the amino acid sequence of SEQ ID No. 3. "More" may be 2, 3, 4, 5, 6, 7, 8, 9, 10 or 11.

[0029] Amino acid addition refers to addition of an amino acid(s) at the C-terminus or the N-terminus of an amino acid sequence, *e.g.* SEQ ID No. 3, as long as the polypeptide of the present invention retains cell adhesion effect of the amino acid sequence of SEQ ID No. 3 (which embodiment is not covered by the claimed invention).

[0030] Amino acid substitution refers to replacement of an amino acid residue at a position in an amino acid sequence, *e.g.* the sequence of SEQ ID No. 3, with another amino acid residue, as long as the polypeptide of the present invention retains cell adhesion effect of the amino acid sequence of SEQ ID No. 3 (which embodiment is not covered by the claimed invention).

[0031] Amino acid insertion refers to insertion of an amino acid residue(s) at an appropriate position(s) in an amino acid sequence, *e.g.* the sequence of SEQ ID No. 3, and all or part of the inserted amino acid residues may also be adjacent to each other, or the inserted amino acid residues are not adjacent to each other, as long as the polypeptide of the present invention retains cell adhesion effect of the amino acid sequence of SEQ ID No. 3 (which embodiment is not covered by the claimed invention). The positions of the inserted amino acid herein are not between repeat sequences.

[0032] Amino acid deletion means that 1, 2, 3 or more amino acids may be deleted from an amino acid sequence, *e.g.* the sequence of SEQ ID No. 3, as long as the polypeptide of the present invention retains cell adhesion effect of the amino acid sequence of SEQ ID No. 3 (which embodiment is not covered by the claimed invention).

[0033] In the present invention, the substitution may be a conservative amino acid substitution, meaning that 3, more preferably 2 or 1 amino acid is replaced by an amino acid having similar or same property to form a peptide compared

to the amino acid sequence of SEQ ID No. 3. These conservative variant peptides may be produced by amino acid replacements according to Table 1.

Initial residue	Representative substitution	Preferred substitution
Ala (A)	Val:Leu:Ile	Val
Arg (R)	Lys:Gln:Asn	Lys
Asn (N)	Gln:His:Lys:Arg	Gln
Asp (D)	Glu	Glu
Cys (C)	Ser	Ser
Gln (Q)	Asn	Asn
Glu (E)	Asp	Asp
Gly (G)	Pro:Ala	Ala
His (H)	Asn:Gln:Lys:Arg	Arg
Ile (I)	Leu:Val:Met:Ala:Phe	Leu
Leu (L)	Ile:Val:Met:Ala:Phe	Ile
Lys (K)	Arg:Gln:Asn	Arg
Met (M)	Leu:Phe:Ile	Leu
Phe (E)	Leu:Val:Ile:Ala:Tyr	Leu
Pro (P)	Ala	Ala
Ser CS)	Thr	Thr
Thr (T)	Ser	Ser
Trp (W)	Tyr: Phe	Tyr
Tyr (Y)	Trp;Phe;Thr;Ser	Phe
Val (V)	Ile:Leu:Met:Phe:Ala	Leu

[0034] As used herein, the terms "medium stringency conditions", "medium-high stringency conditions", "high stringency conditions" or "very high stringency conditions" describe conditions for nucleic acid hybridization and washing. Guidance for performing hybridization reactions is described in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Both a method involving water and a method not involving water are described in this reference, and any of them may be used. For example, specific hybridization conditions are as follows: (1) low stringency hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by two washes in 0.2 X SSC, 0.1% SDS at at least 50°C (for low stringency hybridization conditions, wash temperature can be raised to 55°C); (2) medium stringency conditions are hybridization in 6X SSC at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 60°C; (3) high stringency conditions are hybridization in 6X SSC at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 65°C; and preferably (4) very high stringency conditions are 0.5 M sodium phosphate, 7% SDS, followed by one or more washes in 0.2X SSC, 1% SDS at 65°C, then at 65°C.

Examples

[0035] The following examples are provided to illustrate the invention. Those skilled in the art will appreciate that the examples are merely illustrative but not restrictive. The present invention is only limited by the scope of the appended claims.

Example 1: Construction and Expression of Recombinant Human Collagen Polypeptides

Construction and expression of TE16C gene expression vector

[0036]

1. The full-length protein sequence of human collagen TE16C used in Example 1 is the sequence set forth in SEQ ID No. 5 and has a full length of 486 aa, and its corresponding gene has a full length of 1458 bp. After codon optimization for the codon of *E. coli* (nucleotide sequence: gaaaacctgtatttccagggtgaacgtggtgcaccaggttttcgtggtccg-gcaggtccgaatggaattccgggtga gaaaggaccggctggtgagcgtggtgcgccgggtgaacgtggagcgcctggttttcgtggcccagcaggtccga acggtattcctggtgaaaaaggtccggcgggagagcgtggtgcaccgggtgaacgcggtgcaccgggatttcgtg gtccagcaggaccgaatggtatccctggtgaaaaaggaccggcaggtgagcgtggagcgccaggtgaacgtgg cgcaccgggttttcgtggaccggcaggcccgaatggtattccgggtgaaaaaggcccggcaggtgaacgtggtgc cccgggtgaacgtggtgcgcctggatttcgtggcccggcaggaccgaacggtatccctggagaaaaaggtcctgc aggtgagcgcggtgcgccgggcgagcgtggtgcccctggttttcgcggtccggcaggccctaatggtattcctgg agaaaaaggccctgcaggtgaacgcggagcaccgggtgagcgtggcgcacctggttttcgtggtcctgcaggcc cgaacggtattccgggcgaaaaaggtccagcaggtgaacgtggtgctccgggtgaacgtggtgcacctggatttc gcggtcctgctggtccgaatggtattccaggtgaaaaaggtccggcaggagagcgtggagcaccgggagaacgt ggtgcaccgggctttcgtggtccggccggtcctaacggtatcccaggtgaaaaaggtccggccggcgagcgtggc gcccctggtgagcgtggtgctcctggttttcgtggtccggctggtccgaacggaattcctggtgagaaaggtccggc tggcgaacgtggtgcaccgggtgaacgtggtgcaccgggtttccgtggtccggcgggtcctaatggtatcccgggt gaaaaaggtccggcaggtgaacgtggtgcaccgggtgaacgtggtgcaccgggttttcgcggaccggcaggacc taatggtattccgggagaaaaaggacctgcgggtgaacgtggtgcaccgggtgaacgtggtgcaccgggttttcgt ggtccggcaggtcctaatggaattcctggagagaaaggacctgcaggtgaacgtggtgcaccgggtgaacgtggt gcaccgggttttcgtggtccggcaggtccaaatggtattccgggtgaaaaaggtccggcaggtgaacgtggtgcac cgggtgaacgtggtgcaccgggttttcgtggtccggcaggtccgaatggcattcctggtgaaaaaggtccggcagg tgaacgtggtgcaccgggtgaacgtggtgcaccgggttttcgtggtccggcaggtccgaatggtattccgggtgaa aaaggtccggcaggtgaacgtggt-gcaccg, SEQ ID No. 7), Shanghai HuaGen Biotech Co., Ltd. was commissioned to synthesize a gene fragment, and the synthesized TE16C gene fragment was inserted into PET32a expression vector *via* restriction sites of *BamH* I (NEB Company, cat. No.: R0136L) and *Xho* I (NEB Company, Cat. No.: R0146L).

2. The successfully constructed expression plasmid was transformed into *E. coli* competent cell BL21 (DE3) (Merck Company). The specific procedure was as follows: 1: 1 μL of the plasmid was taken and placed into 100 μL of *E. coli* competent cells BL21 (DE3) and allowed to stand on ice for 30 min. 2: This mixture was heat-activated in a 42°C water bath for 90s, then quickly placed on ice and allowed to stand for 2 min. 3: 600 μL of non-resistant LB liquid medium was added to the mixture, and cultured at 37°C and 220 rpm for 1h. 4: 200 μL of the bacterial solution was uniformly spread onto an LB plate with ampicillin resistance (10 g/L peptone, 5 g/L yeast extract, 10 g/L sodium chloride, 15 g/L agar, and 100 μg/mL ampicillin). 5: The plate was inverted and cultured in a 37°C incubator, and cultured for about 20h to grow clear colonies.

3. Monoclonal colonies were picked up from the transformed LB plate and cultured in 10 mL of LB (containing 100 μg/mL ampicillin) medium for 12h-16h, then transferred to 2xYT medium (16g/L peptone, 10g/L yeast extract, and 5g/L sodium chloride) at a ratio of 1:100 for scale-up culture, and cultured at 37°C and 220rpm until the OD600 of the bacterial solution was between 0.4 and 0.6, and 0.5mM IPTG (Sigma Company, Cat. No.: I5502-1G) was added to a final concentration of 0.5 mM for induction expression, and induction conditions were culture at 18°C and 180 rpm for 20 h. Finally, bacteria were collected by centrifugation, and were stored at -20°C or immediately entered into the next step for purification.

4. (1L) The bacterial pellets were resuspended in about 50 mL of phosphate buffer (pH 7.8) (40 mM sodium dihydrogen phosphate, 500 mM sodium chloride), then broken by high-pressure bacteria breaking equipment (Scientz Biotechnology Co.,LTD.), and centrifuged at 13,000 rpm for 30 min to separate soluble proteins from inclusion bodies.

5. A Ni-NTA (Qiagen Company, Cat. No.: 30210) affinity column was equilibrated with 5 column volumes of binding buffer (40 mM NaH_2PO_3, 500 mM NaCl, pH 7.8). Then, the protein supernatant was added to the column and incubated at 4°C for 0.5-1 h to make the recombinant protein of interest fully bind to the column. Impurity proteins were washed out with 200 mL of washing buffer (10 mM imidazole, 40 mM NaH_2PO_3, 500 mM NaCl, pH 7.8) containing 10 mM imidazole (Sigma Company). Finally, an appropriate amount of His-tagged TEV protease (Sigma, SA T4455) was added, and after incubation at 4°C for 16 h, the flow-through fluid, i.e. the collagen of interest separated from vector protein was collected. The resulting product was dialyzed overnight and lyophilized to dry powder for use.

6. The resulting TE16C protein was measured for purity by SDS-PAGE. The specific procedure was as follows. 40 μL of the purified protein solution was taken, added with 10 μL of 5x protein loading buffer (250 mM Tris-HCl (pH: 6.8), 10% SDS, 0.5% bromophenol blue, 50% glycerol, and 5% β-mercaptoethanol), placed in boiling water at 100°C and boiled for 10 min. The resulting solution was added to SDS-PAGE protein gel at 10 μL per well, and run at 80 V for 2 h. Then the gel was stained with Coomassie Brilliant Blue staining solution (0.1% Coomassie Brilliant Blue R-250, 25% isopropyl alcohol, and 10% glacial acetic acid) for 20min, and then decolorized with protein decolorization solution (10% acetic acid, 5% ethanol). Finally, protein activity was measured as compared to a human native collagen.

Construction and expression of HC16 gene expression vector

[0037]

1. The full-length protein sequence of human collagen HC16 as a control used in Example 1 is the sequence set forth in SEQ ID No. 6 and has a full length of 501 aa, and its corresponding gene has a full length of 1503 bp. After codon optimization for the codon of *E. coli* (nucleotide sequence: ggcgagcgtggtgcacctggttttcgtggccctgcaggcccgaat-
ggcatcccgggtgaaaaaggcccggcagg cgaacgtggcgcccctggtgaacgcggcgcacctggtttccgtggcccggcaggtc-
ctaacggtatcccgggcg aaaagggtcctgcaggcgagcgtggcgccccgggtgaacgcggtgcccctggctttcgcggtcctgccggccct
aacggcattccgggtgagaaaggtcctgccggtgagcgcggtgcccctggtgagcgcggcgcaccgggctttcg
tggcccggccggtcctaatggtattcctggcgagaagggtccggcaggtgaacgcggtgcacctagatccggcga
gcgtggtgcacctggttttcgtggccctgcaggcccgaatggcatcccgggtgaaaaaggcccggcaggcgaac
gtggcgcccctggtgaacgcggcgcacctggtttccgtggcccggcaggtcctaacggtatcccgggcgaaaag
ggtcctgcaggcgagcgtggcgccccgggtgaacgcggtgcccctggctttcgcggtcctgccggccctaacgg
cattccgggtgagaaaggtcctgccggtgagcgcggtgcccctggtgagcgcggcgcaccgggctttcgtggcc
cggccggtcctaatggtattcctggcgagaagggtccggcaggtgaacgcggtgcacctagatccggcgagcgtg
gtgcacctggttttcgtggccctgcaggcccgaatggcatcccgggtgaaaaaggcccggcaggcgaacgtggc
gcccctggtgaacgcggcgcacctggtttccgtggcccggcaggtcctaacggtatcccgggcgaaaagggtcct
gcaggcgagcgtggcgccccgggtgaacgcggtgcccctggctttcgcggtcctgccggccctaacggcattcc
gggtgagaaaggtcctgccggtgagcgcggtgcccctggtgagcgcggcgcaccgggctttcgtggcccggcc
ggtcctaatggtattcctggcgagaagggtccggcaggtgaacgcggtgcacctagatccggcgagcgtggtgca
cctggttttcgtggccctgcaggcccgaatggcatcccgggtgaaaaaggcccggcaggcgaacgtggcgcccct
ggtgaacgcggcgcacctggtttccgtggcccggcaggtcctaacggtatcccgggcgaaaagggtcctgcagg
cgagcgtggcgccccgggtgaacgcggtgcccctggctttcgcggtcctgccggccctaacggcattccgggtga
gaaaggtcctgccggtgagcgcggtgcccctggtgagcgcggcgcaccgggctttcgtggcccggccggtccta
atggtattcctggcgagaagggtccggcaggtgaacgcggtgcacctagatctccggaattcggcccgcctggtcc ttgttgtggcggcggc, SEQ ID No. 8), Shanghai HuaGen Biotech Co., Ltd. was commissioned to synthesize a gene fragment, and the synthesized HC16 gene fragment was inserted into PET32a expression vector *via* restriction sites of *BamH* I (NEB Company, Cat. No.: R0136L) and *Xho* I (NEB Company, Cat. No.: R0146L).
2. The successfully constructed expression plasmid was transformed into *E. coli* competent cell BL21 (DE3) (Merck Company). The specific process is as described above.
3. Monoclonal colonies were picked up from the transformed LB plate and cultured in 10 mL of LB (containing 100 μg/mL ampicillin) medium for 12h-16h, then transferred to 2xYT medium (16g/L peptone, 10g/L yeast extract, and 5g/L sodium chloride) at a ratio of 1:100 for scale-up culture, and cultured at 37°C and 220rpm until the OD600 of the bacterial solution was between 0.4 and 0.6, and 0.5mM IPTG (Sigma Company, Cat. No.: I5502-1G) was added to a final concentration of 0.5 mM for induction expression, and induction conditions were culture at 18°C and 180 rpm for 20 h. Finally, bacteria were collected by centrifugation, and were stored at -20°C or immediately entered into the next step for purification.
4. (1L) The bacterial pellets were resuspended in about 50 mL of phosphate buffer (pH 7.8) (40 mM sodium dihydrogen phosphate, 500 mM sodium chloride), then broken by high-pressure bacteria breaking equipment (Scientz Biotechnology Co.,LTD.), and centrifuged at 13,000 rpm for 30 min to separate soluble proteins from inclusion bodies.
5. A Ni-NTA (Qiagen Company, Cat. No.: 30210) affinity column was equilibrated with 5 column volumes of binding buffer (40 mM NaH_2PO_3, 500 mM NaCl, pH 7.8). Then, the protein supernatant was added to the column and incubated at 4°C for 0.5-1 h to make the recombinant protein of interest fully bind to the column. Impurity proteins were washed out with 200 mL of washing buffer (10 mM imidazole, 40 mM NaH_2PO_3, 500 mM NaCl, pH 7.8) containing 10 mM imidazole (Sigma Company). Finally, an appropriate amount of His-tagged prescission protease (Ppase, for short) (Sigma, SAE0045) was added, and after incubation at 4°C for 16 h, the flow-through fluid, i.e. the collagen of interest separated from the vector protein, was collected. The resulting product was dialyzed overnight and lyophilized to dry powder for use.
6. The resulting HC16 protein was measured for purity by SDS-PAGE. The specific procedure was as follows. 40 μL of the purified protein solution was taken, added with 10 μL of 5x protein loading buffer (250 mM Tris-HCl (pH: 6.8), 10% SDS, 0.5% bromophenol blue, 50% glycerol, and 5% β-mercaptoethanol), placed in boiling water at 100°C and boiled for 10 min. The resulting solution was added to SDS-PAGE protein gel at 10 μL per well, and run at 80 V for 2 h. Then the gel was stained with Coomassie Brilliant Blue staining solution (0.1% Coomassie Brilliant Blue R-250, 25% isopropyl alcohol, and 10% glacial acetic acid) for 20min, and then decolorized with protein decolorization solution (10% acetic acid, 5% ethanol). Finally, protein activity was measured as compared to a human native collagen. The HC16 protein was verified by the same method as in Chinese Patent Invention 201210482543.2, showing the correct protein size.

Construction and expression of polypeptide T16a containing repeat sequences and a C-terminal stable sequence

[0038]

1. Human collagen T16a used in Example 1, which differs from the TE16C gene or the HC16 gene in that T16a comprises both SEQ ID No. 3 having no linker amino acid and the hinge region amino acid SEQ ID No. 2, has a full length of 490 aa, and its sequence is:

GERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGI PGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPG FRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGER GAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGI PGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPG FRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGER GAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGI PGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPG FRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGER GAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGPPGPCCGGG (SEQ ID No. 9)

Its corresponding gene has a full length of 1407 bp. After codon optimization for the codon of *E. coli* (nucleotide sequence: ggtgaacgtggtgcaccaggttttcgtggtccggcaggtccgaatggaattccgggtgagaaaggaccggctggt gagcgtggtgcgccgggtgaacgtggagcgcctggttttcgtggcccagcaggtccgaacggtattcctggtgaaa aaggtccggcgggagagcgtggtgcaccgggtgaacgcggtgcaccgggatttcgtggtccagcaggaccgaat ggtatccctggtgaaaaaggaccggcaggtgagcgtggagcgccaggtgaacgtggcgcaccgggttttcgtgg accggcaggcccgaatggtattccgggtgaaaaaggcccggcaggtgaacgtggtgccccgggtgaacgtggtg cgcctggatttcgtggcccggcaggaccgaacggtatccctggagaaaaaggtcctgcaggtgagcgcggtgcg ccgggcgagcgtggtgcccctggttttcgcggtccggcaggccctaatggtattcctggagaaaaaggccctgcag gtgaacgcggagcaccgggtgagcgtggcgcacctggttttcgtggtcctgcaggcccgaacggtattccgggcg aaaaaggtccagcaggtgaacgtggtgctccgggtgaacgtggtgcacctggatttcgcggtcctgctggtccgaa tggtattccaggtgaaaaaggtccggcaggagagcgtggagcaccgggagaacgtggtgcaccgggctttcgtg gtccggccggtcctaacggtatcccaggtgaaaaaggtccggccggcgagcgtggcgcccctggtgagcgtggt gctcctggttttcgtggtccggctggtccgaacggaattcctggtgagaaaggtccggctggcgaacgtggtgcacc gggtgaacgtggtgcaccgggtttccgtggtccggcgggtcctaatggtatcccgggtgaaaaaggtccggcagg tgaacgtggtgcaccgggtgaacgtggtgcaccgggttttcgcggaccggcaggacctaatggtattccgggaga aaaaggacctgcgggtgaacgtggtgcaccgggtgaacgtggtgcaccgggttttcgtggtccggcaggtcctaat ggaattcctggagagaaaggacctgcaggtgaacgtggtgcaccgggtgaacgtggtgcaccgggttttcgtggtc cggcaggtccaaatggtattccgggtgaaaaaggtccggcaggtgaacgtggtgcaccgggtgaacgtggtgcac cgggttttcgtggtccggcaggtccgaatggcattcctggtgaaaaaggtccggcaggtgaacgtggtgcaccggg tgaacgtggtgcaccgggttttcgtggtccggcaggtccgaatggtattccgggtgaaaaaggtccggcaggtgaa cgtggtgcaccgggcccgcctggtc-cttgttgtggcggcggc, SEQ ID No. 10), Shanghai HuaGen Biotech Co., Ltd. was commissioned to synthesize a gene fragment, and the synthesized T16a gene fragment was inserted into PET32a expression vector *via* restriction sites of *BamH* I (NEB Company, Cat. No.: R0136L) and *Xho* I (NEB Company, Cat. No.: R0146L).

2. The successfully constructed expression plasmid was transformed into *E. coli* competent cell BL21 (DE3) (Merck Company). The specific process is as described above.

3. Monoclonal colonies were picked up from the transformed LB plate and cultured in 10 mL of LB (containing 100 μg/mL ampicillin) medium for 12h-16h, then transferred to 2xYT medium (16g/L peptone, 10g/L yeast extract, and

5g/L sodium chloride) at a ratio of 1:100 for scale-up culture, and cultured at 37°C and 220rpm until the OD600 of the bacterial solution was between 0.4 and 0.6, and 0.5mM IPTG (Sigma Company, Cat. No.: I5502-1G) was added to a final concentration of 0.5 mM for induction expression, and induction conditions were culture at 18°C and 180 rpm for 20 h. Finally, bacteria were collected by centrifugation, and were stored at -20°C or immediately entered into the next step for purification.

4. (1L) The bacterial pellets were resuspended in about 50 mL of phosphate buffer (pH 7.8) (40 mM sodium dihydrogen phosphate, 500 mM sodium chloride), then broken by high-pressure bacteria breaking equipment (Scientz Biotechnology Co.,LTD.), and centrifuged at 13,000 rpm for 30 min to separate soluble proteins from inclusion bodies and collagen colloids.

5. A Ni-NTA (Qiagen Company, Cat. No.: 30210) affinity column was equilibrated with 5 column volumes of binding buffer (40 mM NaH_2PO_3, 500 mM NaCl, pH 7.8). Then, the protein supernatant was added to the column and incubated at 4°C for 0.5-1 h to make the recombinant protein of interest fully bind to the column. Impurity proteins were washed out with 200 mL of washing buffer (10 mM imidazole, 40 mM NaH_2PO_3, 500 mM NaCl, pH 7.8) containing 10 mM imidazole (Sigma Company). Finally, an appropriate amount of His-tagged prescission protease (Ppase, for short) (Sigma, SAE0045) was added, and after incubation at 4°C for 16 h, the flow-through fluid, i.e. the collagen of interest separated from the vector protein was collected. The resulting product was dialyzed overnight and lyophilized to dry powder for use.

6. The resulting T16a protein was measured for purity by SDS-PAGE. The specific procedure was as follows. 40 μL of the purified protein solution was taken, added with 10 μL of 5x protein loading buffer (250 mM Tris-HCl (pH: 6.8), 10% SDS, 0.5% bromophenol blue, 50% glycerol, and 5% β-mercaptoethanol), placed in boiling water at 100°C and boiled for 10 min. The resulting solution was added to SDS-PAGE protein gel at 10 μL per well, and run at 80 V for 2 h. Then the gel was stained with Coomassie Brilliant Blue staining solution (0.1% Coomassie Brilliant Blue R-250, 25% isopropyl alcohol, and 10% glacial acetic acid) for 20min, and then decolorized with protein decolorization solution (10% acetic acid, 5% ethanol). Finally, protein activity was measured as compared to a human native collagen.

Construction and expression of polypeptide T16b containing repeat sequences and linker amino acids

[0039]

1. Human collagen T16b used in Example 1, which differs from the TE16C gene or the HC16 gene in that T16b removes the hinge region amino acid SEQ ID No. 2 from the SEQ ID NO. 6 with linker amino acids, has a full length of 486 aa, and its sequence is:

GERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGI
PGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPG
FRGPAGPNGIPGEKGPAGERGAPRSGERGAPGFRGPAGPNGIPGEKGPAGE
RGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNG
IPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPRSGERG
APGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPA
GERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGP
NGIPGEKGPAGERGAPRSGERGAPGFRGPAGPNGIPGEKGPAGERGAPGE
RGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKG
PAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAP (SEQ ID No. 11)

Its corresponding gene has a full length of 1458 bp. After codon optimization for the codon of *E. coli* (nucleotide sequence: ggcgagcgtggtgcacctggttttcgtggccctgcaggcccgaatggcatcccgggtgaaaaggcccggcagg cgaacgtggcgcccctggtgaacgcggcgcacctggtttccgtggcccggcaggtcctaacggtatcccgggcg

aaaagggtcctgcaggcgagcgtggcgccccgggtgaacgcggtgcccctggctttcgcggtcctgccggccct
aacggcattccgggtgagaaaggtcctgccggtgagcgcggtgcccctggtgagcgcggcgcaccgggctttcg
tggcccggccggtcctaatggtattcctggcgagaagggtccggcaggtgaacgcggtgcacctagatccggcga
gcgtggtgcacctggttttcgtggccctgcaggcccgaatggcatcccgggtgaaaaaggcccggcaggcgaac
gtggcgcccctggtgaacgcggcgcacctggtttccgtggcccggcaggtcctaacggtatcccgggcgaaaag
ggtcctgcaggcgagcgtggcgccccgggtgaacgcggtgcccctggctttcgcggtcctgccggccctaacgg
cattccgggtgagaaaggtcctgccggtgagcgcggtgcccctggtgagcgcggcgcaccgggctttcgtggcc
cggccggtcctaatggtattcctggcgagaagggtccggcaggtgaacgcggtgcacctagatccggcgagcgtg
gtgcacctggttttcgtggccctgcaggcccgaatggcatcccgggtgaaaaaggcccggcaggcgaacgtggc
gcccctggtgaacgcggcgcacctggtttccgtggcccggcaggtcctaacggtatcccgggcgaaaagggtcct
gcaggcgagcgtggcgccccgggtgaacgcggtgcccctggctttcgcggtcctgccggccctaacggcattcc
gggtgagaaaggtcctgccggtgagcgcggtgcccctggtgagcgcggcgcaccgggctttcgtggcccggcc
ggtcctaatggtattcctggcgagaagggtccggcaggtgaacgcggtgcacctagatccggcgagcgtggtgca
cctggttttcgtggccctgcaggcccgaatggcatcccgggtgaaaaaggcccggcaggcgaacgtggcgcccct
ggtgaacgcggcgcacctggtttccgtggcccggcaggtcctaacggtatcccgggcgaaaagggtcctgcagg
cgagcgtggcgccccgggtgaacgcggtgcccctggctttcgcggtcctgccggccctaacggcattccgggtga
gaaaggtcctgccggtgagcgcggtgcccctggtgagcgcggcgcaccgggctttcgtggcccggccggtccta atggtattcctggcgagaagggtccg-gcaggtgaacgcggtgcacct, SEQ ID No. 12), Shanghai HuaGen Biotech Co., Ltd. was commissioned to synthesize a gene fragment, and the synthesized T16a gene fragment was inserted into PET32a expression vector *via* restriction sites of *BamH* I (NEB Company, Cat. No.: R0136L) and *Xho* I (NEB Company, Cat. No.: R0146L).

2. The successfully constructed expression plasmid was transformed into *E. coli* competent cell BL21 (DE3) (Merck Company). The specific process is as described above.

3. Monoclonal colonies were picked up from the transformed LB plate and cultured in 10 mL of LB (containing 100 μg/mL ampicillin) medium for 12h-16h, then transferred to 2xYT medium (16g/L peptone, 10g/L yeast extract, and 5g/L sodium chloride) at a ratio of 1:100 for scale-up culture, and cultured at 37°C and 220rpm until the OD600 of the bacterial solution was between 0.4 and 0.6, and 0.5mM IPTG (Sigma Company, Cat. No.: I5502-1G) was added to a final concentration of 0.5 mM for induction expression, and induction conditions were culture at 18°C and 180 rpm for 20 h. Finally, bacteria were collected by centrifugation, and were stored at -20°C or immediately entered into the next step for purification.

4. (1L) The bacterial pellets were resuspended in about 50 mL of phosphate buffer (pH 7.8) (40 mM sodium dihydrogen phosphate, 500 mM sodium chloride), then broken by high-pressure bacteria breaking equipment (Scientz Biotechnology Co.,LTD.), and centrifuged at 13,000 rpm for 30 min to separate soluble proteins from inclusion bodies.

5. A Ni-NTA (Qiagen Company, Cat. No.: 30210) affinity column was equilibrated with 5 column volumes of binding buffer (40 mM NaH_2PO_3, 500 mM NaCl, pH 7.8). Then, the protein supernatant was added to the column and incubated at 4°C for 0.5-1 h to make the recombinant protein of interest fully bind to the column. Impurity proteins were washed out with 200 mL of washing buffer (10 mM imidazole, 40 mM NaH_2PO_3, 500 mM NaCl, pH 7.8) containing 10 mM imidazole (Sigma Company). Finally, an appropriate amount of His-tagged prescission protease (Ppase, for short) (Sigma, SAE0045) was added, and after incubation at 4°C for 16 h, the flow-through fluid, i.e. the collagen of interest separated the vector protein was collected. The resulting product was dialyzed overnight and lyophilized to dry powder for use.

6. The resulting T16b protein was measured for purity by SDS-PAGE. The specific procedure was as follows. 40 μL of the purified protein solution was taken, added with 10 μL of 5x protein loading buffer (250 mM Tris-HCl (pH: 6.8), 10% SDS, 0.5% bromophenol blue, 50% glycerol, and 5% β-mercaptoethanol), placed in boiling water at 100°C and boiled for 10 min. The resulting solution was added to SDS-PAGE protein gel at 10 μL per well, and run at 80 V for 2 h. Then the gel was stained with Coomassie Brilliant Blue staining solution (0.1% Coomassie Brilliant Blue R-250, 25% isopropyl alcohol, and 10% glacial acetic acid) for 20min, and then decolorized with protein decolorization solution (10% acetic acid, 5% ethanol). Finally, protein activity was measured as compared to a human native collagen.

Example 2: Mass Spectrometric Detection of TE16C Protein

Experimental method

[0040]

Instrument name	Matrix-assisted laser desorption ionization - time-of-flight mass spectrometer MALDI-TOF/TOF Ultraflextreme™, Brucker, Germany		
Matrix	CHCA	**Laser energy**	125
Data retrieval software	Mascot	**Retrieval species**	ALL entries
Retrieval database	NCBIprot		

[0041] The protein sample was subjected to DTT reduction and alkylation of iodoacetamide, and then trypsin was added to digest overnight. The peptide segment obtained after enzymolysis was desalted by C18 ZipTip and then mixed with matrix α-cyano-4-hydroxycinnamic acid (CHCA) for spotting. Finally, the matrix-assisted laser desorption ionization - time-of-flight mass spectrometer MALDI-TOF/TOF Ultraflextreme™, Brucker, Germany was used for analysis (for the technique of peptide fingerprinting, see: Protein J. 2016; 35: 212-7).

[0042] Database retrieval was handled through the MS/MS Ion Search page on the local mascot website. The protein identification results were obtained based on the primary mass spectrum of the peptide segments produced after enzymolysis. Retrieval parameters: trypsin enzymolysis, set two missed restriction sites. The alkylation of cysteine was set to a fixed modification, and the oxidation of methionine was set to a variable modification. The database used for the identification is NCBprot.

[0043] The mass spectrometry peak results are shown in Figure 1.

[0044] Table 1: Mass spectrometry detection of molecular weight and corresponding peptide

Observed value	Mr (Expected value	Peptide
2246.2323	2246.1424	GAPGERGAPGFRGPAGPNGIPGEK
2246.2323	2246.1424	GAPGFRGPAGPNGIPGEKGPAGER
1738.8731	1738.9095	GPAGERGAPGERGAPGFR
1678.9003	1678.8659	GAPGFRGPAGPNGIPGEK
1660.8401	1661.8607	GPAGPNGIPGEKGPAGER
1171.5966	1171.6295	GAPGERGAPGFR
1093.5636	1093.5784	GPAGPNGIPGEK

Coverage of the peptide fragment detected

[0045]

GERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGP
NGIPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPG
ERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGIP
GEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERG
APGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEK

**GPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPG
FRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPA
GERGAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRG
PAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGPNGIPGEKGPAGER
GAPGERGAPGFRGPAGPNGIPGEKGPAGERGAPGERGAPGFRGPAGP
NGIPGEKGPAGER**GAP

[0046] More than 98.8% of the sequence of the protein TE16C to be determined can be detected by mass spectrometry, and the results are very reliable. The mass spectrometric characteristic peaks of the undetermined protein were 2246.2323, 2246.2323, 1738.8731, 1678.9003, 1660.8401, 1171.5966, and 1093.5636, and it was concluded that the TE16C protein was correctly expressed.

Example 3: Detection of Properties of Recombinant Type III Collagens

[0047] Methods for detecting collagen activity can be found in the reference Juming Yao, Satoshi Yanagisawa, Tetsuo Asakura, Design, Expression and Characterization of Collagen-like Proteins Based on the Cell Adhesive and Crosslinking Sequences Derived from Native Collagens, J Biochem. 136, 643-649 (2004). The specific implementation method is as follows:

1. The concentrations of the protein samples to be tested were measured by ultraviolet absorption method, including control human collagen (Sigma, C7774), control old type III collagen HC16, new type III collagen TE16C protein, control T16a protein and T16b protein samples. Specifically, the respective ultraviolet light absorption of the samples at 215 nm and 225 nm was measured, and the respective protein concentration was calculated using the empirical formula C (μg/mL) = 144X (A215-A225), and it should be noted that the detection was performed under the condition of A215<1.5. The principle of the method is to determine the characteristic absorption of a peptide bond under far ultraviolet light, and the method is not affected by chromophore content, has few interfering substances, is easy to operate, and is suitable for detecting human collagen and its analogs which are not colored by Coomassie Brilliant Blue. (Reference is Walker JM., The Protein Protocols Handbook, Second edition, Humana Press, 43-45). After protein concentration detection, all the concentrations of the proteins to be tested were adjusted to 0.5 mg/mL with PBS.
2. 100 μL of each protein solution was added to a 96-well plate and compared with a blank PBS solution, and allowed to stand at room temperature for 60 min.
3. 10^5 well-cultured 3T3 cells (from Teacher Tong Pei, Tsinghua University) were added to each well and incubated at 37 °C for 60 min.
4. Each well was washed 4 times with PBS.
5. The absorbance of OD492nm was measured using an LDH assay kit (Roche, 04944926001). Based on the value of the blank control, the attachment rate of cells can be calculated. The calculation formula is as follows: cell adherence rate = (test well - blank well) x 100% / (positive well - blank well). The adherence rate of the cells can reflect the activity of collagen. The higher the activity of a protein, the faster it can provide a cell with a superior external environment to help the cell adhere.

[0048] See Figure 2 for the results.

[0049] The results in Figure 2 indicate that the two human recombinant collagens (i.e., type III collagen HC16 and type III collagen TE16C) have better adhesion activities than commercial human collagen, and the recombinant type III collagen TE16C of the present invention achieves the most potent cell adhesion activity. The proteins obtained by the two construction methods of T16a and T16b as controls have lower cell adhesion activities than the TE16C protein of the present invention.

Example 4: Expression and Purification of Recombinant Type III Collagens

TE16C protein expression and purification

5 **[0050]**

1. According to steps 1-5 in Example 1, the collagen of interest TE16C separated from the vector protein was obtained and dialyzed into solution A (10 mM Na_2CO_3, pH 10.5, 10 mM NaCl) of an anion exchange column.
2. The anion exchange column (Hitrap Q HP column, 5 mL, GE Healthcare Biosciences) was equilibrated with 5 column volumes of solution A (10 mM Na_2CO_3, pH 10.5, 10 mM NaCl). The dialyzed TE16C protein was then applied to the column, and the protein peak that had passed through the column, i.e. the purified TE16C protein, was collected. The resulting product was dialyzed overnight and lyophilized to dry powder for use.
3. The anion exchange column was eluted with 5 column volumes of solution B (10 mM Na_2CO_3, pH 10.5, 1 M NaCl) to wash out the impurity proteins and endotoxin from the column.

10

15

[0051] The HC16 protein, the polypeptide T16a comprising repeat sequences and a C-terminal stable sequence, and the polypeptide T16b comprising repeat sequences and linker amino acids were treated in the same manner.

[0052] Figure 3 shows the expression of recombinant type III collagens. After the purification of the type III collagen HC16, about 8 mg of pure protein can be obtained from 1 L of the bacterial solution. After the purification of the type III collagen TE16C, about 15 mg or more of pure protein can be obtained from 1 L of the bacterial solution. The expression levels of the proteins obtained by the two construction methods of T16a and T16b as controls were also not as high as that of the TE16C protein of the present invention. It is important to note that for the construction method of T16a, the amount of soluble protein in the supernatant is very low because of the formation of a large amount of colloid during the purification of the protein.

20

25

[0053] Figure 4 shows the purification of recombinant type III collagens. The type III collagen HC16 had a purity of about 95% after Ni column and anion exchange column, while the type III collagen TE16C could be rapidly purified to a purity of more than 99% by a Ni column and an anion exchange column. Though the proteins obtained by the two construction methods of T16a and T16b as controls had high purities, their yields were not high and were much lower than the TE16C protein of the present invention.

30

Example 5: Stabilization of Recombinant Type III Collagens in Aqueous Solutions

Stability study

35 **[0054]**

1. The purified TE16C protein and HC16, T16a and T16b proteins were obtained according to Example 4 and then dialyzed against physiological saline (0.9% NaCl). The dialyzed protein solutions were collected, sterilized by filtration through a 0.22 μM filter membrane, and then allowed to stand for a long time in an environment of 4-8 °C.
2. At different time points, such as one month, two months, three months, half a year, one year, aqueous solutions of TE16c protein and HC16 protein were collected, and were measured for purity by SDS-PAGE (see the method of Example 1).

40

[0055] Figure 5 shows the electrophoresis images of HC16, TE16C, T16a and T16b at the time point of half a year. The type III collagen HC16 was unstable in the aqueous solution, significant degradation occurred after one month, and degradation after six months was very serious. However, the type III collagen TE16C has a stable structure, and after being placed in an aqueous solution for half a year, more than 90% of the protein remained as a complete full-length protein, and only a small amount was degraded. The proteins obtained by the two construction methods of T16a and T16b as controls also had different degrees of degradation, with the degradation of T16b being more serious; their stability is much lower than that of the TE16C protein of the present invention.

45

50

Example 6: Detection of Endotoxin in Recombinant Type III Collagens

Endotoxin detection of TE16c protein, HC16 protein, T16a protein and T16b protein

55

[0056]

1. Purified TE16c protein, HC16 protein, T16a protein and T16b protein were prepared according to Example 4.

2. Reagents including tachypleus amebocyte lysate (Zhanjiang A&C Biological Ltd., λ=0.015EU/ml) and diluent I (Zhanjiang A&C Biological Ltd.) were prepared.
3. The endotoxin standards were diluted to E1, E0.5, E0.25, E0.125, and E0.0625. The samples were diluted 5, 8, 16, 32, 50, 100, 200, 300 and 400 times with diluent I, and the other samples were diluted with water for detection, to dilute the samples 10, 16, 32, 64, 100, 200, 400, 600 and 800 times, respectively.
4. The endotoxin standards and the diluted protein samples were added to the tachypleus amebocyte lysate for observation. The results were compared to both negative and positive controls. The endotoxin concentrations of the samples were calculated.

[0057] Figure 6 shows the residual endotoxin status of recombinant type III collagens. The type III collagen is not uniform in structure and is easy to bind to endotoxin, so the residual endotoxin after Ni column and anion exchange is about 100EU/mL, while the new type III collagen TE16C can be rapidly purified to contain less than 5EU/mL of endotoxin by a Ni column and an anion exchange column. The proteins obtained by the two construction methods of T16a and T16b as controls are also tightly bound to endotoxin, and it is not easy to remove endotoxin.

Example 7: Analysis of Recombinant Type III Collagens

[0058]

1. The sequence of TE16C contains the segment of the human type III collagen Pro488 to Gly510, and several polypeptides (Taihe Biotechnology Co., Ltd., Beijing) were synthesized based on this regional sequence for extensive crystal screening, e.g., GFRGPAGPNGIPGEKGPAGERG polypeptide.
2. The polypeptide was dissolved in water to a solution of 15 mg/mL, and crystals were grown by a hanging drop method in which the formulation of tank solution was 30% (w/v) PEG 400, 0.1 M Na Acetate pH 4.6, and 0.1 M Cadmium Chloride. 1 μL of each of the polypeptide solution and the tank solution were taken and mixed, and then sealed with 1 mL of the tank solution.
3. After about one week, single crystals of the polypeptide gradually grew in the droplet, and was quickly cooled and stored with liquid nitrogen after being taken out.
4. The collected polypeptide crystals were sent to the BL-18U1 Beamline of Shanghai Synchrotron Radiation Facility for X-ray crystal diffraction data collection, and at the same time the data analysis and structure analysis were carried out using the calculator of the line station.

[0059] We obtained the high-resolution three-dimensional structure of the human type III collagen Pro488 to Gly510 region contained in the new type III collagen TE16C by the method of protein crystallography, as shown in Figure 7. It was confirmed that this region formed a very stable trimer structure. Therefore, the structure and function of intact collagen can be possessed without additionally adding the C-terminal sequence of GPPGPCCGGG

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- CN 201210482543 **[0006] [0015] [0037]**

Non-patent literature cited in the description

- *Journal of Biochemistry,* 2004, vol. 136, 643-649 **[0014]**
- Current Protocols in Molecular Biology. John Wiley & Sons, 1989, 6.3.1-6.3.6 **[0034]**
- *Protein J.,* 2016, vol. 35, 212-7 **[0041]**
- **JUMING YAO ; SATOSHI YANAGISAWA ; TETSUO ASAKURA.** Design, Expression and Characterization of Collagen-like Proteins Based on the Cell Adhesive and Crosslinking Sequences Derived from Native Collagens. *J Biochem.,* 2004, vol. 136, 643-649 **[0047]**
- **WALKER JM.** The Protein Protocols Handbook. Humana Press, 43-45 **[0047]**

Patentkrav

1. Polypeptid omfattende 16 gentagelser af en sekvens angivet i SEQ ID No. 1, gentagelserne af sekvensen er direkte forbundet, hvor polypeptidet ikke omfatter
5 en sekvens angivet i SEQ ID No. 2.

2. Polynukleotid, der koder for polypeptidet ifølge krav 1.

3. Ekspressionsvektor omfattende polynukleotidet ifølge krav 2 og eventuelt
10 omfattende en nukleinsyresekvens, der koder for SEQ ID No. 4, hvor nukleinsyresekvensen, der koder for SEQ ID No. 4 er direkte bundet til 5'-enden af nukleinsyresekvensen, der koder for polypeptidet, fortrinsvis omfatter ekspressionsvektoren nukleinsyresekvensen af SEQ ID No. 5.

15 **4.** Værtscelle omfattende ekspressionsvektoren ifølge krav 3, hvor værtscellen fortrinsvis er *Escherichia coli.*

5. Fremgangsmåde til fremstillingen af polypeptidet ifølge krav 1, omfattende:
(1) dyrkning af værtscellen ifølge krav 4 i et fremstillingsmedium og
20 fremstilling af polypeptidet;
(2) høst af polypeptidet og eventuelt opløsning af polypeptidet, fortrinsvis opløsning af polypeptidet med TEV-protease; og
(3) oprensning af polypeptidet med en Ni-søjle og/eller en anionbytterkromatografi;
25 hvor fremgangsmåden til fremstillingen eventuelt ikke omfatter et yderligere trin til fjernelse af endotoksin; hvor det oprensede polypeptid fortrinsvis er i det væsentlige fri for endotoksin eller indeholder mindre end 5 EU/ml endotoksin.

6. Sammensætning omfattende polypeptidet ifølge krav 1, hvor
30 sammensætningen fortrinsvis er en medicinsk anordning, et vævs-teknologisk produkt, et kosmetikprodukt eller et sundhedsplejeprodukt, fortrinsvis er polypeptidet i form af en vandig opløsning af polypeptidet, fortrinsvis er sammensætningen fri for en komponent, der forhindrer nedbrydning af polypeptidet, og fortrinsvis er sammensætningen en sammensætning til

langtidsbrug, hvor langtidsbrugen fortrinsvis er mere end et halvt års brug.

7. Anvendelse af polypeptidet ifølge krav 1 til fremstillingen af en sammensætning, fortrinsvis en medicinsk anordning, et vævs-teknologisk produkt, et kosmetikprodukt eller et sundhedssupplement, hvor polypeptidet fortrinsvis er i form af en vandig opløsning af polypeptidet, fortrinsvis er sammensætningen fri for en komponent, der forhindrer nedbrydning af et polypeptid, og sammensætningen er en sammensætning til langtidsbrug, hvor langtidsbrugen er mere end et halvt års brug.

8. Anvendelse af polypeptidet ifølge krav 1 til at fremme celleadhæsion in vitro.

9. Anvendelse af en aminosyresekvens ifølge SEQ ID No. 4 eller ekspressionsvektoren ifølge krav 3 til fremstillingen af polypeptidet ifølge krav 1.

DRAWINGS

Drawing

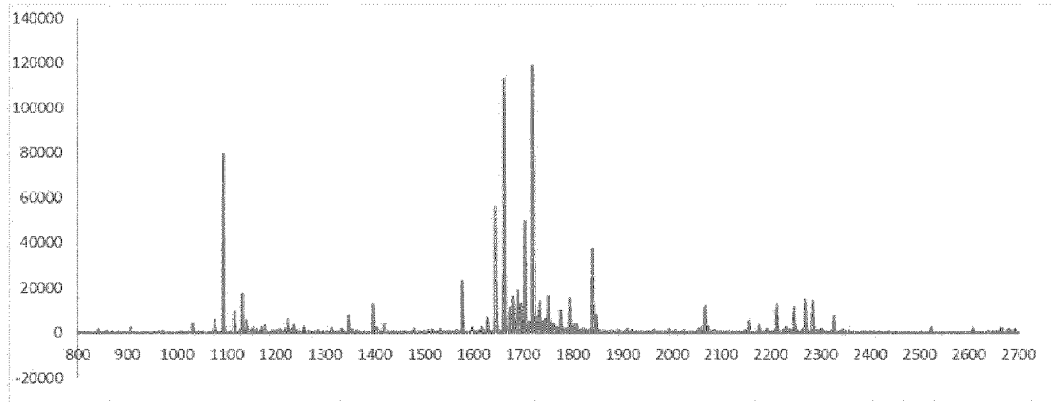

FIG. 1

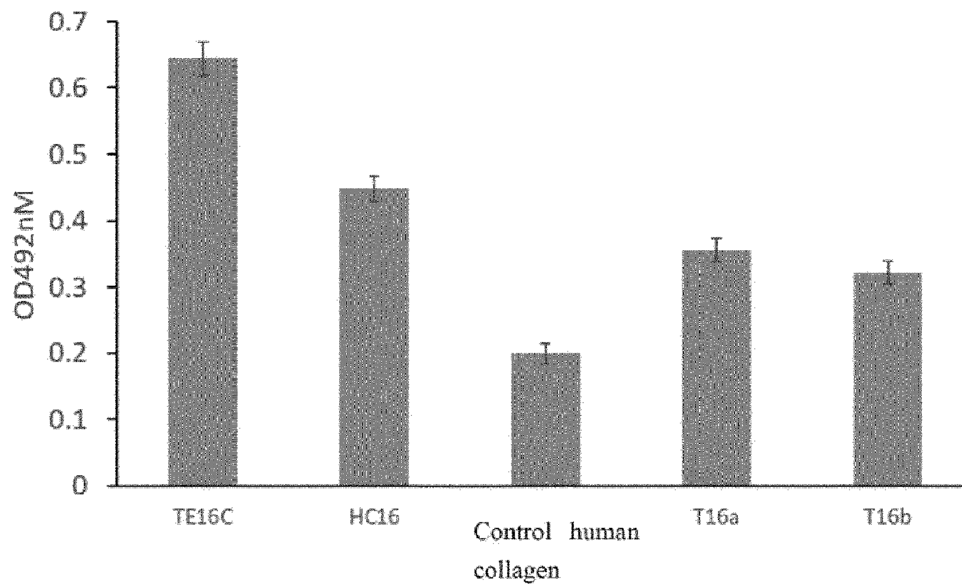

FIG. 2

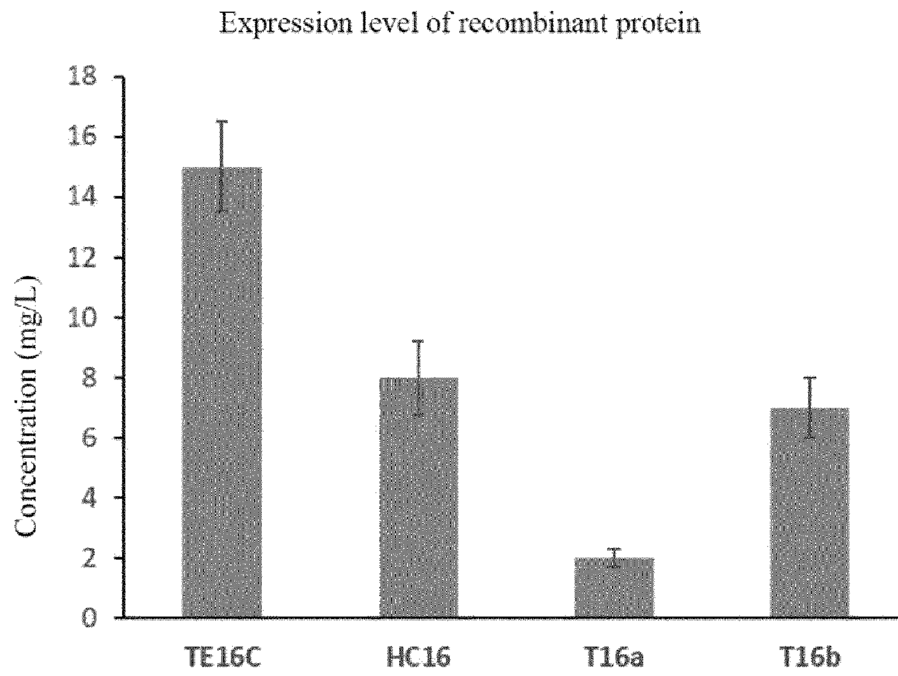

FIG. 3

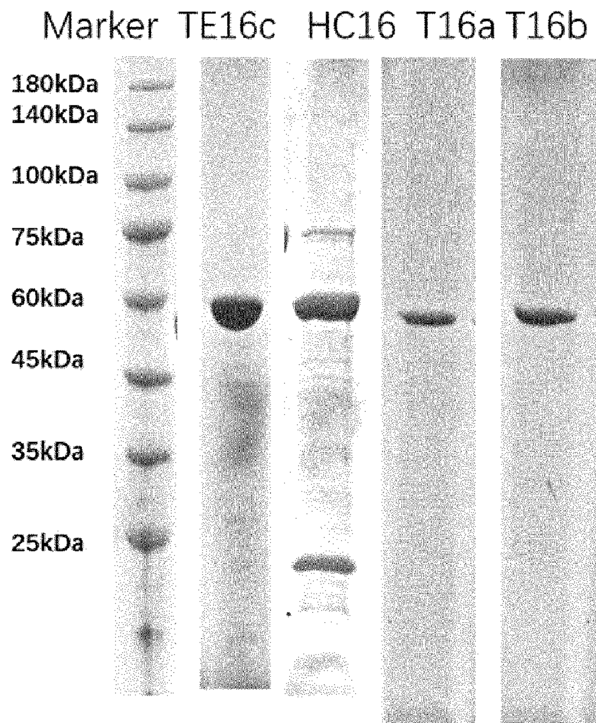

FIG. 4

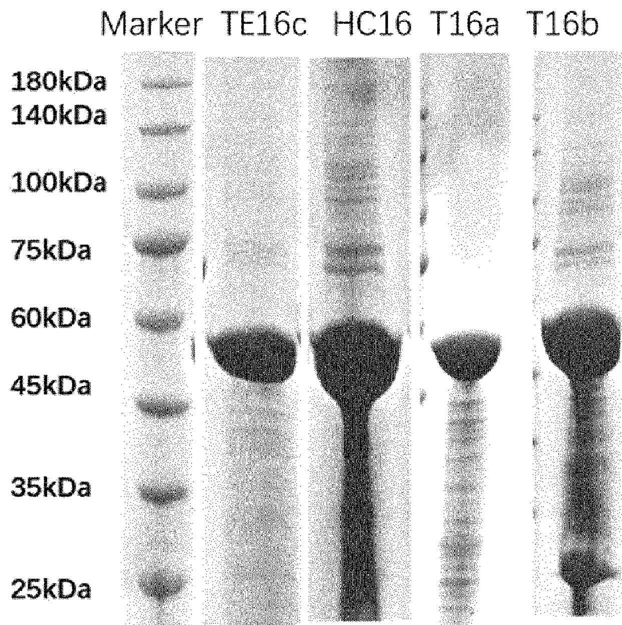

FIG. 5

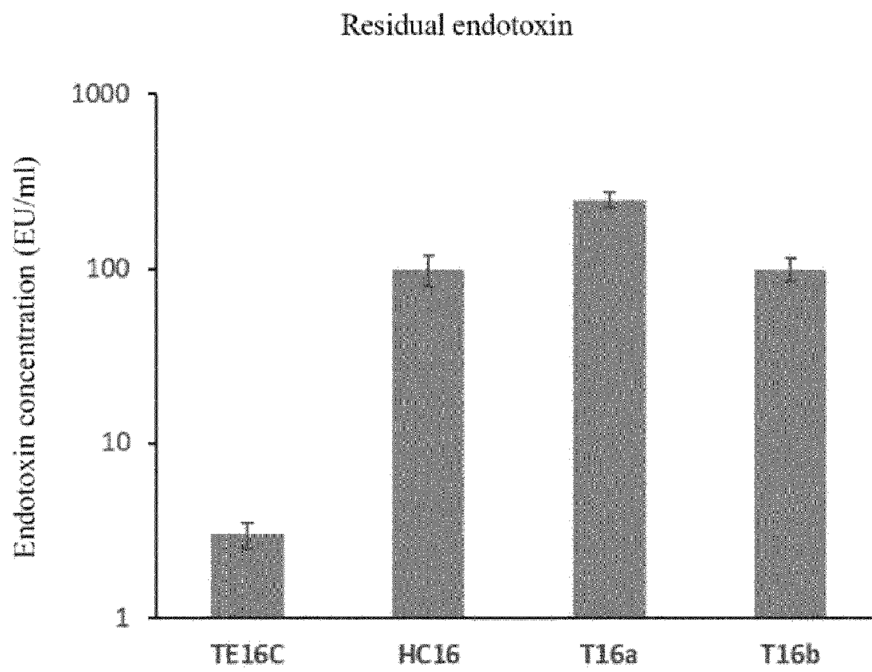

FIG. 6

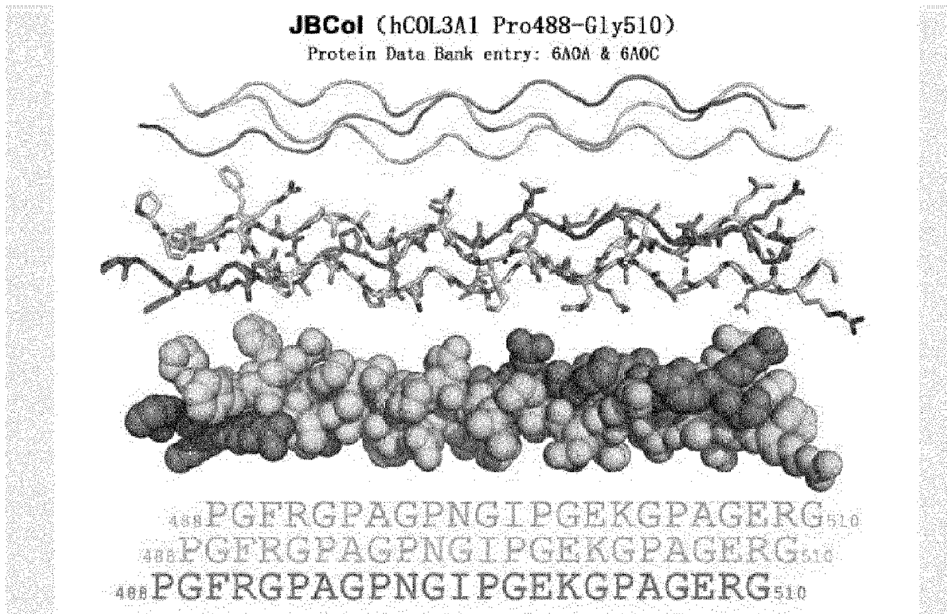

FIG. 7

SEKVENSLISTE

Sekvenslisten er udeladt af skriftet og kan hentes fra det Europæiske Patent Register.

The Sequence Listing was omitted from the document and can be downloaded from the European Patent Register.