

(19) **DANMARK**



Patent- og
Varemærkestyrelsen

(12) Oversættelse af
europæisk patentskrift

(10) **DK/EP 3168229 T3**

-
- (51) Int.Cl.: **C 07 K 14/435 (2006.01)** **C 12 N 15/62 (2006.01)**
- (45) Oversættelsen bekendtgjort den: **2019-09-23**
- (80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2019-06-19**
- (86) Europæisk ansøgning nr.: **16190909.8**
- (86) Europæisk indleveringsdag: **2010-10-27**
- (87) Den europæiske ansøgnings publiceringsdag: **2017-05-17**
- (30) Prioritet: **2010-03-18 EP 10156927** **2010-04-21 WO PCT/SE2010/050439**
- (62) Stamansøgningsnr: **10848078.1**
- (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: **Spiber Technologies AB, KTH Avd Proteinteknologi , Alba Nova Univers Centrum, 106 91 Stockholm, Sverige**
- (72) Opfinder: **Johansson, Jan, Renstiernas Gata 40, 116 31 Stockholm, Sverige**
Rising, Anna, Doppingvägen 12, 756 51 Uppsala, Sverige
Hedhammar, My, Odengatan 40, 113 51 Stockholm, Sverige
NORDLING, Kerstin, Primusgatan 76, 112 67 Stockholm, Sverige
- (74) Fuldmægtig i Danmark: **AWA Denmark A/S, Strandgade 56, 1401 København K, Danmark**
- (54) Benævnelse: **FREMSTILLING AF PROTEINER OG POLYPEPTIDER.**
- (56) Fremdragne publikationer:
WO-A2-2007/078239
LIN ZHI ET AL: "Solution structure of eggcase silk protein and its implications for silk fiber formation.", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA 2 JUN 2009, vol. 106, no. 22, 2 June 2009 (2009-06-02) , pages 8906-8911, XP002716184, ISSN: 1091-6490
MOTRIUK-SMITH DAGMARA ET AL: "Analysis of the conserved N-terminal domains in major ampullate spider silk proteins", BIOMACROMOLECULES, AMERICAN CHEMICAL SOCIETY, US, vol. 6, no. 6, 10 August 2005 (2005-08-10) , pages 3152-3159, XP002453721, ISSN: 1525-7797, DOI: 10.1021/BM050472B
PAN HONG-CHUN ET AL: "[Cloning and prokaryotic expression of major ampullate spidroin gene of spider]", MEDLINE, May 2007 (2007-05), XP002547539,
HEDHAMMAR MY ET AL: "Structural properties of recombinant nonrepetitive and repetitive parts of major ampullate spidroin 1 from Euprosthenops australis: Implications for fiber formation", BIOCHEMISTRY, AMERICAN CHEMICAL SOCIETY, US, vol. 47, no. 11, 1 March 2008 (2008-03-01) , pages 3407-3417, XP002547535, ISSN: 0006-2960, DOI: 10.1021/BI702432Y [retrieved on 2008-02-23]
ASKARIEH GLAREH ET AL: "Self-assembly of spider silk proteins is controlled by a pH-sensitive relay", NATURE, NATURE PUBLISHING GROUP, UNITED KINGDOM, vol. 465, no. 7295, 13 May 2010 (2010-05-13), pages 236-238, XP002593882, ISSN: 0028-0836, DOI: 10.1038/NATURE08962

Fortsættes ...

DESCRIPTION

Technical field of the invention

[0001] The present invention relates to the field of production of proteins and polypeptides, and more specifically to production of proteins and polypeptides selected from the group consisting of (a) polypeptide drugs and (b) growth factors. The present invention provides a method of producing a desired protein/polypeptide selected from the group consisting of (a) polypeptide drugs and (b) growth factors. There is also provided novel fusion protein intermediates for production of the desired proteins and polypeptides selected from the group consisting of (a) polypeptide drugs and (b) growth factors.

Background to the invention

[0002] Production of proteins and polypeptides from DNA can be achieved in various hosts, but a common problem is the formation of insoluble protein/polypeptide aggregates. This may severely impede or even prevent production of a functional protein/polypeptide. One solution to this problem is to express the desired protein or polypeptide as a fusion protein with a protein or polypeptide that provides the required solubility. The fusion protein may be cleaved, and the desired protein isolated.

[0003] The problem is typically aggravated with low-solubility proteins and polypeptides, e.g. membrane-associated proteins and polypeptides. For instance, lung surfactant protein C (SP-C; Table 6) is a transmembrane protein that is produced by alveolar type II cells and is a constituent of surfactant, that is necessary to prevent alveolar collapse at end expiration. Neonatals often suffer from respiratory distress due to insufficient amounts of surfactant. Today, this condition is treated with surfactant preparations extracted from animal lungs. SP-C-33 is a variant of SP-C, where the residues in the transmembrane helix (normally mainly valines) are exchanged for leucines. SP-C-33 retains the function of native SP-C, including proper insertion in membranes, but is less prone to aggregate and therefore feasible to produce in large quantities for development of a synthetic surfactant preparation. Since SP-C-33 so far has not been possible to produce from DNA, it is today manufactured by chemical synthesis.

[0004] Other examples of proteins and polypeptides that pose difficulties when expressed from recombinant DNA are A β -peptide, IAPP, PrP, α -synuclein, calcitonin, prolactin, cystatin, ATF and actin; SP-B, α -defensins and β -defensins; class A-H apolipoproteins; LL-37, SP-C, SP-C33Leu, Brichos, GFP, neuroserpin; hormones, including EPO and GH, and growth factors, including IGF-I and IGF-II; avidin and streptavidin; and protease 3C.

[0005] WO 2007/078239 discloses a miniature major ampullate spidroin protein including a

repetitive part and a C-terminal part derived from *Euprosthrops*, and a method of producing polymers thereof.

[0006] Hedhammar M. et al., *Biochemistry* 47:3407-3417 (2008) discusses the structural properties of recombinant non-repetitive and repetitive parts of major ampullate spidroin 1 from *Euprosthrops australis* and the implications for spider silk fiber formation.

Summary of the invention

[0007] It is an object of the present invention to provide new means and methods for production of proteins and polypeptides, and in particular non-spidroin proteins and polypeptides.

[0008] It is also an object of the present invention to provide new means and methods for production of proteins and polypeptides, and in particular non-spidroin proteins and polypeptides, with low solubility in water, e.g. proteins and polypeptides that are prone to aggregate when produced from recombinant DNA, membrane proteins and polypeptides, and amyloid-forming proteins and polypeptides.

[0009] It is a further object of the present invention to provide alternative means and methods for production of protein or polypeptide drugs and drug targets.

[0010] It is an object of the present invention to provide new means and methods for production of disulphide-containing proteins and polypeptides.

[0011] It is also an object of the present invention to provide new means and methods for production of apolipoproteins.

[0012] For these and other objects that will be evident from the following specification, the present invention provides according to a first aspect a fusion protein that is useful in a method of producing a desired protein or polypeptide selected from the group consisting of (a) polypeptide drugs and (b) growth factors. The fusion protein may be useful as such, or it may be cleaved to obtain the desired protein or polypeptide in isolated form. Fusion proteins according to the invention are comprising (i) at least one solubility-enhancing moiety which is derived from the N-terminal (NT) fragment of a spider silk protein; and (ii) at least one moiety which is a desired protein or polypeptide; wherein each solubility-enhancing moiety is linked directly or indirectly to the desired protein or polypeptide moiety; and wherein each solubility-enhancing moiety has at least 80% identity to any one of SEQ ID NO 6 and SEQ ID NO 50-63 or at least 50% identity to SEQ ID NO 8.

[0013] In certain embodiments of the fusion protein, each solubility-enhancing moiety has at least 80% identity to SEQ ID NO 6 or at least 50% identity to SEQ ID NO 8. In specific embodiments of the fusion protein, each solubility-enhancing moiety contains from 100 to 160

amino acid residues.

[0014] In some embodiments, the desired protein or polypeptide has less than 30% identity to any of SEQ ID NO: 6-10.

[0015] In certain embodiments, the fusion protein is comprising at least two solubility-enhancing moieties, each being derived from the N-terminal (NT) fragment of a spider silk protein. In specific embodiments, the fusion protein is comprising at least two consecutive solubility-enhancing moieties, each being derived from the N-terminal (NT) fragment of a spider silk protein.

[0016] In some embodiments of the fusion protein, at least one solubility-enhancing moiety is linked directly or indirectly to the amino-terminal or the carboxy-terminal end of at least one desired protein or polypeptide moiety. In specific embodiments, at least one solubility-enhancing moiety constitutes the amino-terminal and/or the carboxy-terminal end of the fusion protein.

[0017] In one embodiment, the fusion protein is further comprising (iii) at least one cleavage site arranged between at least one desired protein or polypeptide moiety and at least one solubility-enhancing moiety.

[0018] In certain embodiments of the fusion protein, the desired protein or polypeptide is derived from sponges, comb jellies, jellyfishes, corals, anemones, flatworms, rotifers, roundworms, ribbon worms, clams, snails, octopuses, segmented worms, crustaceans, insects, bryozoans, brachiopods, phoronids, sea stars, sea urchins, tunicates, lancelets, vertebrates, including human, plants, fungi, yeast, bacteria, archaeobacteria or viruses or is an artificial protein or polypeptide. In specific embodiments, the desired protein or polypeptide is derived from molluscs, insects, vertebrates, including human, plants, fungi, yeast, bacteria, archaeobacteria or viruses or is an artificial protein or polypeptide. In further specific embodiments, the desired protein or polypeptide is derived from vertebrates, including human, plants, fungi, yeast, bacteria, archaeobacteria or viruses or is an artificial protein or polypeptide.

[0019] The desired protein or polypeptide is selected from the group consisting of polypeptide drugs and growth factors. In specific embodiments of the fusion protein, the desired protein or polypeptide is selected from the group consisting of LL-37, SP-C, SP-C33, SP-C33Leu, Brichos, GFP, neuroserpin; hormones, including EPO and GH, and growth factors, including IGF-I and IGF-II.

[0020] Preferred embodiments of the fusion protein are selected from the group consisting of SEQ ID NOS 26, 28, 30, 34, 37, 39, 42 and 47 and proteins having at least 80%, preferably at least 90%, more preferably at least 95% identity, to any of these proteins.

[0021] There is also provided isolated polynucleic acids encoding the fusion proteins according to the invention. Certain of the isolated polynucleic acids are selected from the group

consisting of nucleic acids encoding a fusion protein selected from the group consisting of SEQ ID NOS 26, 28, 30, 34, 37, 39, 42 and 47 and proteins having at least 80%, preferably at least 90%, more preferably at least 95% identity, to any of these proteins; and the group of nucleic acids consisting of SEQ ID NOS 27, 29, 31, 38, 40, 43 and 48.

[0022] According to one aspect, the present invention provides a novel use of at least one moiety which is derived from the N-terminal (NT) fragment of a spider silk protein as a solubility enhancing moiety as defined herein in a fusion protein for production of a desired protein or polypeptide selected from the group consisting of (a) polypeptide drugs and (b) growth factors.

[0023] One method of producing a fusion protein is comprising the following steps: a) expressing in a suitable host a fusion protein according to the invention; and b) obtaining a mixture containing the fusion protein, and optionally isolating the fusion protein.

[0024] According to a related aspect, the present invention provides a method of producing a desired protein or polypeptide, comprising the following steps: a) expressing in a suitable host a fusion protein according to the invention; and b) obtaining a mixture containing the fusion protein or polypeptide, and optionally isolating the fusion protein or polypeptide; c) cleaving the fusion protein to provide the desired protein or polypeptide; and d) isolating the desired protein or polypeptide; wherein said fusion protein is comprising: (iii) at least one cleavage site arranged between at least one desired protein or polypeptide moiety and at least one solubility-enhancing moiety.

[0025] In certain embodiments of these methods, step b) further involves purification of the fusion protein on an affinity medium with an immobilized NT moiety and/or on an anion exchange medium. In specific embodiments, the purification of the fusion protein on an affinity medium is carried out with association to an affinity medium with an immobilized NT moiety at a pH of 6.3 or lower, followed by dissociation from the affinity medium with a desired dissociation medium. In further specific embodiments, the dissociation medium has a pH of 6.4 or higher, a pH of 4.1 or lower and/or has a high ionic strength. In some embodiments, purification of the fusion protein on an anion exchange medium is carried out with association to the anion exchange medium at a pH of 6.4 or higher, followed by dissociation from the anion exchange medium with a dissociation medium having a high ionic strength. In some embodiments of these methods, the purification of the fusion protein in step b) occurs in a column, on magnetic beads with functionalized surfaces, or on filters with functionalized surfaces.

List of appended sequences

SEQ ID NO	
1	4Rep
2	4RepCT
3	NT4Rep
4	NT5Rep
5	NT4RepCTHis

SEQ ID NO	
6	NT
7	CT
8	consensus NT sequence
9	consensus CT sequence
10	repetitive sequence from <i>Euprosthynops australis</i> MaSp1
11	consensus G segment sequence 1
12	consensus G segment sequence 2
13	consensus G segment sequence 3
14	NT4Rep (DNA)
15	NT4RepCT (DNA)
16	NT5Rep (DNA)
17	NT4RepCTHis 2
18	NT4RepCTHis 2 (DNA)
19	ZbasicNT4RepCT
20	NT4RepCT
21	HisTrxHisThrNT4RepCT
22	NT4RepCT 2
23	HisNTNT4RepCT
24	HisNTNT4RepCT (DNA)
25	NT8RepCT
26	HisNTMetSP-C33Leu
27	HisNTMetSP-C33Leu (DNA)
28	HisNTNTMetSP-C33Leu
29	HisNTNTMetSP-C33Leu (DNA)
30	HisNTNTMetLL37
31	HisNTNTMetLL37 (DNA)
32	NTHis
33	HisNTNT8RepCT
34	HisNTNTBrichos
35	HisTrxHisSP-C33Leu
36	HisTrxHisSP-C33Leu (DNA)
37	HisTrxNtSP-C33Leu
38	HisTrxNtSP-C33Leu (DNA)
39	2HisNtNtQGBrichos

SEQ ID NO	
40	2HisNtNtQGBrichos (DNA)
41	Brichos
42	2HisNtNtQGGFP
43	2HisNtNtQGGFP (DNA)
44	GFP (Green Fluorescent Protein)
45	ZbGFP
46	HisABPGFP
47	2HisNtNtQGNS
48	2HisNtNtQGNS (DNA)
49	NS (Neuroserpin)
50	NT <i>Euprosthops australis</i> MaSp1
51	NT <i>Latrodectus geometricus</i> MaSp1
52	NT <i>Latrodectus hesperus</i> MaSp1
53	NT <i>Nephila clavipes</i> MaSp1
54	NT <i>Argiope trifasciata</i> MaSp2
55	NT <i>Latrodectus geometricus</i> MaSp2
56	NT <i>Latrodectus hesperus</i> MaSp2
57	NT <i>Nephila inaurata madagascariensis</i> MaSp2
58	NT <i>Nephila clavipes</i> MaSp2
59	NT <i>Argiope bruennichi</i> cylindriform spidroin 1
60	NT <i>Nephila clavata</i> cylindriform spidroin 1
61	NT <i>Latrodectus hesperus</i> tubuliform spidroin
62	NT <i>Nephila clavipes</i> flagelliform silk protein
63	NT <i>Nephila inaurata madagascariensis</i> flagelliform silk protein
64	Linker peptide 1, p. 21
65	Linker peptide 2, p. 21
66	Linker peptide 3, p. 21
67	Linker peptide 4, p. 21
68	CT <i>Euprosthops sp</i> MaSp1
69	CT <i>Euprosthops australis</i> MaSp1
70	CT <i>Argiope trifasciata</i> MaSp1
71	CT <i>Cyrtophora moluccensis</i> Sp1
72	CT <i>Latrodectus geometricus</i> MaSp1
73	CT <i>Latrodectus hesperus</i> MaSp1
74	CT <i>Macrothele holsti</i> Sp1

SEQ ID NO	
75	CT <i>Nephila clavipes</i> MaSp1
76	CT <i>Nephila pilipes</i> MaSp1
77	CT <i>Nephila madagascariensis</i> MaSp1
78	CT <i>Nephila senegalensis</i> MaSp1
79	CT <i>Octonoba varians</i> Sp1
80	CT <i>Psechrus sinensis</i> Sp1
81	CT <i>Tetragnatha kauaiensis</i> MaSp1
82	CT <i>Tetragnatha versicolor</i> MaSp1
83	CT <i>Araneus bicentenarius</i> Sp2
84	CT <i>Argiope amoena</i> MaSp2
85	CT <i>Argiope aurantia</i> MaSp2
86	CT <i>Argiope trifasciata</i> MaSp2
87	CT <i>Gasteracantha mammosa</i> MaSp2
88	CT <i>Latrodectus geometricus</i> MaSp2
89	CT <i>Latrodectus hesperus</i> MaSp2
90	CT <i>Nephila clavipes</i> MaSp2
91	CT <i>Nephila madagascariensis</i> MaSp2
92	CT <i>Nephila senegalensis</i> MaSp2
93	CT <i>Dolomedes tenebrosus</i> Fb1
94	CT <i>Dolomedes tenebrosus</i> Fb2
95	CT <i>Araneus diadematus</i> ADF-1
96	CT <i>Araneus diadematus</i> ADF-2
97	CT <i>Araneus diadematus</i> ADF-3
98	CT <i>Araneus diadematus</i> ADF-4
99	Human SP-B
100	Mouse SP-B
101	Pig SP-B
102	Rabbit SP-B
103	Rat SP-B
104	Mini-B
105	Mini-BLeu
106	Mini-B27
107	1a AA
108	1b AA

SEQ ID NO	
109	1a LL
110	1b LL
111	SP-C
112	SP-C(Leu)
113	SP-C33
114	SP-C30
115	SP-C33(Leu)
116	LL-37
117	KL4
118	0 AAAA
119	0 LLLL
120	protease 3C recognition site

Brief description of the drawings

[0026]

Fig. 1 shows a sequence alignment of spidroin N-terminal domains.

Fig. 2 shows a sequence alignment of spidroin C-terminal domains.

Fig. 3 shows electrophoresis gels of fusion proteins.

Fig. 4 shows an electrophoresis gel of a SP-C33Leu protein obtained from a fusion protein.

Fig. 5 shows *in vitro* surface activity of surfactant suspensions comprising SP-C33Leu obtained from a fusion protein.

Fig. 6 shows electrophoresis gels of SP-C33Leu fusion proteins.

Fig. 7 shows an electrophoresis gel of a Brichos fusion protein.

Fig. 8 shows an electrophoresis gel of a GFP fusion protein and GFP obtained from the fusion protein.

Fig. 9 shows an electrophoresis gel of a neuroserpin fusion protein and neuroserpin obtained from the fusion protein.

Detailed description of the invention

[0027] The present invention is concerned with production of proteins and polypeptides selected from the group consisting of (a) polypeptide drugs and (b) growth factors. Depending on the purpose with this production, the end product may vary. It may for instance be desirable to obtain the protein or polypeptide inserted in a lipid membrane, in solution or associated with other biomolecules. It shall also be realized that it may also be highly desirable to obtain the desired protein or polypeptide as part of a fusion protein, which may provide a suitable handle for purification and detection and/or provide desirable properties, e.g. stability and solubility.

[0028] The present invention is generally based on the insight of the usefulness of the N-terminal (NT) fragment of a spider silk protein as a solubility-enhancing moiety in a fusion protein that is produced from recombinant DNA. Thus, the present invention provides according to a first aspect a fusion protein comprising (i) at least one solubility-enhancing moiety which is derived from the NT fragment of a spider silk protein; and (ii) at least one moiety which is a desired protein or polypeptide selected from the group consisting of (a) polypeptide drugs and (b) growth factors. In a preferred embodiment, the fusion proteins consists of (i) at least one solubility-enhancing moiety which is derived from the NT fragment of a spider silk protein; and (ii) at least one moiety which is a desired protein or polypeptide, optionally including other preferred features disclosed herein, e.g. a linker peptide and/or a cleavage site between the solubility-enhancing moiety and the desired protein or polypeptide. In experiments, surprisingly high yields of different fusion proteins have been achieved in *E. coli*. The fusion protein may be useful as such in isolated form, e.g. for studies of otherwise aggregated or poorly soluble proteins in soluble form, or in crystallization associated with X-ray crystallography. The fusion protein may also be cleaved to release the desired protein.

[0029] The term "fusion protein" implies here a protein that is made by expression from a recombinant nucleic acid, i.e. DNA or RNA that is created artificially by combining two or more nucleic acid sequences that would not normally occur together (genetic engineering). The fusion proteins according to the invention are recombinant proteins, and they are therefore not identical to naturally occurring proteins. In particular, wildtype spidroins are not fusion proteins according to the invention, because they are not expressed from a recombinant nucleic acid as set out above. The combined nucleic acid sequences encode different proteins, partial proteins or polypeptides with certain functional properties. The resulting fusion protein, or recombinant fusion protein, is a single protein with functional properties derived from each of the original proteins, partial proteins or polypeptides.

[0030] In certain embodiments, the fusion protein according to the invention and the corresponding genes are chimeric, i.e. the protein/gene fragments are derived from at least two different species. The solubility-enhancing moiety is derived from the N-terminal fragment of a spider silk protein. According to this aspect, it is preferred that, the desired protein or polypeptide is a non-spidroin protein. This implies that the desired protein or polypeptide is not derived from the C-terminal, repetitive or N-terminal fragment of a spider silk protein.

[0031] The fusion protein according to the invention may also contain one or more linker peptides. The linker peptide(s) may be arranged between the solubility-enhancing moiety and the desired protein or polypeptide moiety, or may be arranged at either end of the solubility-enhancing moiety and the desired protein or polypeptide moiety. If the fusion protein contains two or more solubility-enhancing moieties, the linker peptide(s) may also be arranged in between two solubility-enhancing moieties. The linker(s) may provide a spacer between the functional units of the fusion protein, but may also constitute a handle for identification and purification of the fusion protein, e.g. a His and/or a Trx tag. If the fusion protein contains two or more linker peptides for identification and purification of the fusion protein, it is preferred that they are separated by a spacer sequence, e.g. His₆-spacer-His₆-. The linker may also constitute a signal peptide, such as a signal recognition particle, which directs the fusion protein to the membrane and/or causes secretion of the fusion protein from the host cell into the surrounding medium. The fusion protein may also include a cleavage site in its amino acid sequence, which allows for cleavage and removal of the linker(s) and/or the solubility-enhancing moiety or moieties. Various cleavage sites are known to the person skilled in the art, e.g. cleavage sites for chemical agents, such as CNBr after Met residues and hydroxylamine between Asn-Gly residues, cleavage sites for proteases, such as thrombin or protease 3C. and self-splicing sequences, such as intein self-splicing sequences.

[0032] Each solubility-enhancing moiety is linked directly or indirectly to the desired protein or polypeptide moiety. A direct linkage implies a direct covalent binding between the two moieties without intervening sequences, such as linkers. An indirect linkage also implies that the two moieties are linked by covalent bonds, but that there are intervening sequences, such as linkers and/or one or more further solubility-enhancing moieties.

[0033] The at least one solubility-enhancing moiety may be arranged at either end of the desired protein or polypeptide, i.e. C-terminally arranged or N-terminally arranged. It is preferred that the least one solubility-enhancing moiety is arranged at the N-terminal end of the desired protein or polypeptide. If the fusion protein contains one or more linker peptide(s) for identification and purification of the fusion protein, e.g. a His or Trx tag(s), it is preferred that it is arranged at the N-terminal end of the fusion protein. The at least one solubility-enhancing moiety may also be integrated within the desired protein or polypeptide, for instance between domains or parts of a desired protein. In a preferred embodiment, at least one solubility-enhancing moiety constitutes the N-terminal and/or the C-terminal end of the fusion protein, i.e. no linker peptide or other sequence is present terminal of the solubility-enhancing moiety. A typical fusion protein according to the invention may contain 1-6, such as 1-4, such as 1-2 solubility-enhancing moieties.

[0034] In a preferred embodiment, the fusion protein is comprising at least two solubility-enhancing moieties, each being derived from the N-terminal (NT) fragment of a spider silk protein. The solubility-enhancing moieties, preferably two solubility-enhancing moieties, may be consecutively arranged at either end of the desired protein or polypeptide, i.e. C-terminally arranged or N-terminally arranged. Consecutively arranged solubility-enhancing moieties may also be integrated within the desired protein or polypeptide, for instance between domains or

parts of a desired protein. The solubility-enhancing moieties may also be non-consecutively arranged, either at each end of the desired protein or polypeptide, i.e. C-terminally and N-terminally arranged, or at one end of the desired protein or polypeptide and integrated within the desired protein or polypeptide. A typical preferred fusion protein according to the invention may contain 2-6, such as 2-4 solubility-enhancing moieties.

[0035] In a preferred embodiment, the fusion protein according has at least one cleavage site arranged between at least one desired protein or polypeptide moiety and at least one solubility-enhancing moiety. This allows for cleavage of the fusion protein and purification of the desired protein. It is however noted that it may be desirable to obtain the desired protein or polypeptide as part of a fusion protein, which may provide a suitable handle for purification and detection and/or provide desirable properties, e.g. stability and solubility. In this case, the cleavage site may be omitted, or the cleavage site may be included but the cleavage step omitted.

[0036] A preferred fusion protein has the form of an N-terminally arranged solubility-enhancing moiety, coupled by a linker peptide of 1-30 amino acid residues, such as 1-10 amino acid residues, to a C-terminally arranged desired protein or polypeptide. The linker peptide may contain a cleavage site. Optionally, the fusion protein has an N-terminal or C-terminal linker peptide, which may contain a purification tag, such as a His tag, and a cleavage site.

[0037] Another preferred fusion protein has the form of an N-terminally arranged solubility-enhancing moiety coupled directly to a C-terminally arranged desired protein or polypeptide. Optionally, the fusion protein has an N-terminal or C-terminal linker peptide, which may contain a purification tag, such as a His tag, and a cleavage site.

[0038] One preferred fusion protein has the form of a two consecutive N-terminally arranged solubility-enhancing moieties, coupled by a linker peptide of 1-30 amino acid residues, such as 1-10 amino acid residues, to a C-terminally arranged desired protein or polypeptide. The linker peptide may contain a cleavage site. Optionally, the fusion protein has an N-terminal or C-terminal linker peptide, which may contain a purification tag, such as a His tag, and a cleavage site.

[0039] Another preferred fusion protein has the form of two consecutive N-terminally arranged solubility-enhancing moieties coupled directly to a C-terminally arranged desired protein or polypeptide. Optionally, the fusion protein has an N-terminal or C-terminal linker peptide, which may contain a purification tag, such as a His tag, and a cleavage site.

[0040] The solubility-enhancing moiety is derived from the NT fragment of a spider silk protein, or spidroin. Although the examples by necessity relate to specific NT fragments, in this case proteins derived from major spidroin 1 (MaSp1) from *Euprosthenops australis*, it is considered that the method disclosed herein is applicable to any similar protein moiety. The terms "spidroins" and "spider silk proteins" are used interchangeably throughout the description and encompass all known spider silk proteins, including major ampullate spider silk proteins which

typically are abbreviated "MaSp", or "ADF" in the case of *Araneus diadematus*. These major ampullate spider silk proteins are generally of two types, 1 and 2. These terms furthermore include the new NT protein fragments according to the invention, as defined in the appended claims and itemized embodiments, and other non-natural proteins with a high degree of identity and/or similarity to the known spider silk NT protein fragments.

[0041] The solubility-enhancing moiety has a high degree of similarity to the N-terminal (NT) amino acid sequence of spider silk proteins. As shown in Fig 1, this amino acid sequence is well conserved among various species and spider silk proteins, including MaSp1 and MaSp2. In Fig 1, the following spidroin NT fragments (SEQ ID NO: 50-63) are aligned, denoted with GenBank accession entries where applicable:

TABLE 1 - Spidroin NT fragments

Code	Species and spidroin protein	GenBank acc. no.
Ea MaSp1	<i>Euprosthops australis</i> MaSp 1	AM259067
Lg MaSp1	<i>Latrodectus geometricus</i> MaSp 1	ABY67420
Lh MaSp1	<i>Latrodectus hesperus</i> MaSp 1	ABY67414
Nc MaSp1	<i>Nephila clavipes</i> MaSp 1	ACF19411
At MaSp2	<i>Argiope trifasciata</i> MaSp 2	AAZ15371
Lg MaSp2	<i>Latrodectus geometricus</i> MaSp 2	ABY67417
Lh MaSp2	<i>Latrodectus hesperus</i> MaSp 2	ABR68855
Nim MaSp2	<i>Nephila inaurata madagascariensis</i> MaSp 2	AAZ15322
Nc MaSp2	<i>Nephila clavipes</i> MaSp 2	ACF19413
Ab CySp1	<i>Argiope bruennichi</i> cylindriform spidroin 1	BAE86855
Ncl CySp1	<i>Nephila clavata</i> cylindriform spidroin 1	BAE54451
Lh TuSp1	<i>Latrodectus hesperus</i> tubuliform spidroin	ABD24296
Nc Flag	<i>Nephila clavipes</i> flagelliform silk protein	AF027972
Nim Flag	<i>Nephila inaurata madagascariensis</i> flagelliform silk protein	AF218623 (translated)

[0042] Only the part corresponding to the N-terminal domain is shown for each sequence, omitting the signal peptide. Nc flag and Nlm flag are translated and edited according to Rising A. et al. Biomacromolecules 7, 3120-3124 (2006)).

[0043] It is not critical which specific solubility-enhancing moiety is present in fusion proteins according to the invention, as long as the solubility-enhancing moiety is not entirely missing. Thus, the solubility-enhancing moiety according to the invention can be selected from any of the amino acid sequences shown in Fig 1 or sequences with a high degree of similarity. A wide variety of solubility-enhancing sequences can be used in the fusion protein according to the invention. Based on the homologous sequences of Fig 1, the following sequence constitutes a

consensus solubility-enhancing amino acid sequence:

QANTPWSSPNLADAFINSF(M/L)SA(A/I)SSSGAFSADQLDDMSTIG(D/N/Q)T
 LMSAMD(N/S/K)MGRSG(K/R)STKSKLQALNMAFASSMAEIAAAESGG(G/Q)
 SVGVKTNAISDALSSAFYQTTGSVNPQFV(N/S)EIRSLI(G/N)M(F/L)(A/S)QAS
 ANEV (SEQ ID NO 8).

[0044] According to a preferred embodiment of the invention, the sequence of the solubility-enhancing moiety according to the invention has at least 50% identity, preferably at least 60% identity, to the consensus amino acid sequence SEQ ID NO 8, which is based on the amino acid sequences of Fig 1. In a preferred embodiment, the sequence of the solubility-enhancing moiety according to the invention has at least 65% identity, preferably at least 70% identity, to the consensus amino acid sequence SEQ ID NO 8. In preferred embodiments, the solubility-enhancing moiety according to the invention has furthermore 70%, preferably 80%, similarity to the consensus amino acid sequence SEQ ID NO 8.

[0045] A representative solubility-enhancing moiety according to the invention is the *Euprosthrops australis* sequence SEQ ID NO 6. According to a preferred embodiment of the invention, the solubility-enhancing moiety has at least 80% identity to SEQ ID NO 6 or any individual amino acid sequence in Fig 1, i.e. SEQ ID NO 50-63. In preferred embodiments of the invention, the solubility-enhancing moiety has at least 90%, such as at least 95% identity, to SEQ ID NO 6 or any individual amino acid sequence in Fig 1. In preferred embodiments of the invention, the solubility-enhancing moiety is identical to SEQ ID NO 6 or any individual amino acid sequence in Fig 1, in particular to Ea MaSp1.

[0046] The term "% identity", as used throughout the specification and the appended claims, is calculated as follows. The query sequence is aligned to the target sequence using the CLUSTAL W algorithm (Thompson, J.D., Higgins, D.G. and Gibson, T.J., Nucleic Acids Research, 22: 4673-4680 (1994)). A comparison is made over the window corresponding to the shortest of the aligned sequences. The amino acid residues at each position are compared, and the percentage of positions in the query sequence that have identical correspondences in the target sequence is reported as % identity.

[0047] The term "% similarity", as used throughout the specification and the appended claims, is calculated as described for "% identity", with the exception that the hydrophobic residues Ala, Val, Phe, Pro, Leu, Ile, Trp, Met and Cys are similar; the basic residues Lys, Arg and His are similar; the acidic residues Glu and Asp are similar; and the hydrophilic, uncharged residues Gln, Asn, Ser, Thr and Tyr are similar. The remaining natural amino acid Gly is not similar to any other amino acid in this context.

[0048] Throughout this description, alternative embodiments according to the invention fulfill, instead of the specified percentage of identity, the corresponding percentage of similarity. Other alternative embodiments fulfill the specified percentage of identity as well as another, higher percentage of similarity, selected from the group of preferred percentages of identity for

each sequence. For example, a sequence may be 70% similar to another sequence; or it may be 70% identical to another sequence; or it may be 70% identical and 90% similar to another sequence.

[0049] The solubility-enhancing moiety contains from 100 to 160 amino acid residues. It is preferred that the solubility-enhancing moiety contains at least 100, or more than 110, preferably more than 120, amino acid residues. It is also preferred that the solubility-enhancing moiety contains at most 160, or less than 140 amino acid residues. A typical solubility-enhancing moiety contains approximately 130-140 amino acid residues.

[0050] The desired protein or polypeptide is a non-spidroin protein or polypeptide. The sequence of a desired non-spidroin protein or polypeptide according to the invention preferably has less than 30% identity, such as less than 20% identity, preferably less than 10% identity, to any of the spidroin amino acid sequences disclosed herein, and specifically to any of SEQ ID NO: 6-10.

[0051] In a preferred embodiment, the desired non-spidroin protein or polypeptide is derived from sponges, comb jellies, jellyfishes, corals, anemones, flatworms, rotifers, roundworms, ribbon worms, clams, snails, octopuses, segmented worms, crustaceans, insects, bryozoans, brachiopods, phoronids, sea stars, sea urchins, tunicates, lancelets, vertebrates, including human, plants, fungi, yeast, bacteria, archaeobacteria or viruses or is an artificial protein or polypeptide. By "derived" is meant that the sequence of a desired non-spidroin protein or polypeptide according to the invention has preferably at least 50% identity, preferably at least 60%, preferably at least 70%, more preferably at least 80% identity, or even at least 90% identity, such as 95-100% identity, to a corresponding naturally occurring protein and having a maintained function. In one preferred embodiment, the desired non-spidroin protein or polypeptide is derived from molluscs, insects, vertebrates, including human, plants, fungi, yeast, bacteria, archaeobacteria or viruses or is an artificial protein or polypeptide. In a preferred embodiment, the desired non-spidroin protein or polypeptide is derived from vertebrates, including human, plants, fungi, yeast, bacteria, archaeobacteria or viruses or is an artificial protein or polypeptide.

[0052] The desired non-spidroin protein or polypeptide is selected from the group consisting of polypeptide drugs and growth factors.

[0053] Amyloid-forming proteins and polypeptides include proteins and polypeptides that are associated with disease and functional amyloid. Examples of amyloid-forming proteins and polypeptides include amyloid beta peptide (A β -peptide), islet amyloid polypeptide (amylin or IAPP), prion protein (PrP), α -synuclein, calcitonin, prolactin, cystatin, atrial natriuretic factor (ATF) and actin. Examples of amyloid-forming proteins and polypeptides are listed in Table 3.

TABLE 3 - Amyloid-forming proteins and polypeptides

Protein	Uniprot ID
A β 1-42	P05067

<u>Protein</u>	<u>Uniprot ID</u>
Apolipoprotein SAA	P02735
Cystatin C	P01034
Transthyretin	P02766
Lysozyme	P61626
α -synuclein	P37840
Prion protein	P04156
ODAM	A1E959
Lactadherin	Q08431
Tau	P10636
Gelsolin	P06396
ABri, ADan	Q9Y287
Insulin	P01308
Apolipoprotein A-II	P02652
Apolipoprotein A-IV	P06727
Semenogelin I	P04279
Keratoepithelin	Q15582
Lactotransferrin	P02788
Fibrinogen α -chain	P02671
ANF	P01160
IAPP	P10997
β 2-microglobulin	P61769
Calcitonin	P01258
Prolactin	P01236
Apolipoprotein A-I	P02647
CsgA	P28307
Sup35	C7GN25
Pmel17	P40967
HET-s	A8HR89
Ure2p	Q8NIE6

[0054] Examples of disulphide-containing proteins and polypeptides include surfactant protein B (SP-B) and variants thereof, such as Mini-B, Mini-B27, Mini-BLeu, α -defensins and β -defensins. Without being limited to any specific theory, it is contemplated that the solubility-enhancing moiety promotes the desired formation of intrachain disulphide bonds over interchain disulphide bonds in defensins and other disulphide-containing proteins and

polypeptides. Examples of disulphide-containing proteins and polypeptides are listed in Table 4.

TABLE 4 - Disulphide-containing proteins and polypeptides

Protein	Sequence / Uniprot ID
Human SP-B (SEQ ID NO 99)	FPIPLPYCWLCRALIKRIQAMIPKGALAVAVAQVCRVVPL VAGGICQCLAERYSVILLDTLLGRMLPQLVCRLVLRCSM ^a
Mouse SP-B (SEQ ID NO 100)	LPIPLPFCWLCRTLIKRVQAVIPKGVLA VAVSQVCHVVPL VVGICQCLAERYTVLLLDALLGRVVPQLVCGLVLRCS ^a
Pig SP-B (SEQ ID NO 101)	FPIPLPFCWLCRTLIKRIQAVVPKGVLLKAVAQVCHVVPL PVGGICQCLAERYIVICLNMLLDRTLPQLVCGLVLRCS ^a
Rabbit SP-B (SEQ ID NO 102)	FPIPLPCWLCRTLLKRIQAMIPKGVLAMAVAQVCHVVPL VVGICQCLAERYTVILLEVLLGHVLPQLVCGLVLRCS ^a
Rat SP-B (SEQ ID NO 103)	LPIPLPFCWLCRTLIKRVQAVIPKGVLA VAVSQVCHVVPL VVGICQCLAERYTVLLLDALLGRVVPQLVCGLVLRCS ^a
Mini-B (SEQ ID NO 104)	CWLCRALIKRIQAMIPKGG RMLPQLVCRLVLRCS ^b
Mini-BLeu (SEQ ID NO 105)	CWLCRALIKRIQALIPKGG RLLPQLVCRLVLRCS ^b
Mini-B27 (SEQ ID NO 106)	CLLCRALIKRFNRYLTPQLVCRLVLRCS ^c
1a AA (SEQ ID NO 107)	CWLARALIKRIQALIPKGG RLLPQLVARLVLRCS ^d
1b AA (SEQ ID NO 108)	AWLCRALIKRIQALIPKGG RLLPQLVCRLVLRAS ^e
1a LL (SEQ ID NO 109)	CWLLRALIKRIQALIPKGG RLLPQLVLRVLRCS ^d
1b LL (SEQ ID NO 110)	LWLCRALIKRIQALIPKGG RLLPQLVCRLVLRLS ^e
Proinsulin	P01308
CAR D1 ^f	P78310
Brichos	SEQ ID NO: 41
^a Cys8-Cys77, Cys11-Cys71, Cys35-Cys46 and intermolecular Cys48-Cys48 linkages	

Protein	Sequence / Uniprot ID
^b Cys1-Cys33 and C4-C27 linkages	
^c Cys1-Cys27 and C4-C21 linkages	
^d Cys1-Cys33 linkage	
^e Cys4-Cys27 linkage	
^f Coxsackievirus and adenovirus receptor	

[0055] Examples of apolipoproteins include class A-H apolipoproteins. Examples of apolipoproteins are listed in Table 5.

TABLE 5 - Apolipoproteins

Protein	Sequence / Uniprot ID
Apolipoprotein B-100	P04114
Apolipoprotein C-1	P02654
Apolipoprotein D	P05090
Apolipoprotein E	P02649

[0056] Examples of membrane proteins and polypeptides include membrane-associated receptors, including cytokine receptors, KL4, LL-37, surfactant protein C (SP-C) and variants thereof, such as SP-C(Leu), SP-C33, SP-C30 and SP-C33Leu. Other specific examples include SP-C33Leu fused to Mini-B, Mini-BLeu, 1a AA, 1b AA, 0 AAAA, 1a LL, 1b LL, 0 LLLL or SP-B proteins, optionally via a linker, e.g. Gly_n, Leu_n, Gly-Ala_n or the like. SP-C33Leu may be arranged N-terminal or, preferably, C-terminal to the Mini-B, Mini-BLeu, 1a AA, 1b AA, 0 AAAA, 1a LL, 1b LL, 0 LLLL or SP-B protein. Examples of membrane proteins and polypeptides are listed in Table 6.

TABLE 6 - Membrane proteins and polypeptides

Protein	Sequence
SP-C (SEQ ID NO 111)	FGIPCCPVHLKRLIVVVVVVLIVWIVGALLMGL *
SP-C(Leu) (SEQ ID NO 112)	FGIPSSPVHLKRLKLLLLLLLLLILLLLILGALLMGL
SP-C33 (SEQ ID NO 113)	IPSSPVHLKRLKLLLLLLLLLILLLLILGALLMGL
SP-C30 (SEQ ID NO 114)	IPSSPVHLKRLKLLLLLLLLLILLLLILGALL
SP-C33(Leu) (SEQ ID NO 115)	IPSSPVHLKRLKLLLLLLLLLILLLLILGALLLGL
LL-37 (SEQ ID NO 116)	LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLVPRTES
KL4 (SEQ ID NO 117)	KLLLLKLLLLKLLLLKLLLLK
*Cys-5 and Cys-6 in native SP-C are palmitoylated	

<u>Protein</u>	<u>Uniprot ID</u>
Growth hormone receptor	P10912
G-protein coupled receptor 35	Q9HC97
Insulin receptor,	P06213
Gonadotropin releasing hormone receptor	P30968
Very low density lipoprotein receptor	P98155
TGF-beta receptor, type 1	P36897
Prostaglandin D2 receptor	Q13258
Receptor tyrosine-protein kinase erbB-2 (HER2)	P04626
Receptor tyrosine-protein kinase erbB-4 (HER4)	Q15303
Receptor tyrosine-protein kinase erbB-3 (HER3)	P21860
Aquaporin-1	P29972
Aquaporin-2	P41181
Chloride channel protein CIC-Ka	P51800
Chloride channel protein CIC-Kb	P51801
Integral membrane protein DGCR2/IDD	P98153
Interleukin 9 receptor	Q01113

[0057] Examples of protein and polypeptide drugs and drug targets include hormones that are produced recombinantly, including peptide and protein hormones, such as erythropoietin (EPO) and growth hormone (GH), cytokines, growth factors, such as insulin-like growth factors (IGF-I and IGF-II), KL4, LL-37, surfactant protein C (SP-C) and variants thereof, such as SP-C(Leu), SP-C33, SP-C30 and SP-C33Leu. Other specific examples include SP-C33Leu fused to Mini-B, Mini-BLeu, 1a AA, 1b AA, 0 AAAA, 1a LL, 1b LL, 0 LLLL or SP-B proteins, optionally via a linker, e.g. Gly_n, Leu_n, Gly-Ala_n or the like. SP-C33Leu may be arranged N-terminal or, preferably, C-terminal to the Mini-B, Mini-BLeu, 1a AA, 1b AA, 0 AAAA, 1a LL, 1b LL, 0 LLLL or SP-B protein. Examples of protein and polypeptide drugs and drug targets according to the invention are listed in Table 7.

TABLE 7 - Protein and polypeptide drugs and drug targets

<u>Protein</u>	<u>Sequence / Uniprot ID</u>
Insulin-like growth factor IA	P01243
Insulin like growth factor IB	P05019
Growth hormone 1, variant 1	Q6IYF 1
Growth hormone 1, variant 2	Q6IYF0
Growth hormone 2, variant 2	B1A4H7
Insulin	P01308
Erythropoietin	P01588

<u>Protein</u>	<u>Sequence / Uniprot ID</u>
Coagulation Factor VIII	P00451
Coagulation Factor IX	P00740
Prothrombin	P00734
Serum albumin	P02768
Antithrombin III	P01008
Interferon alfa	P01563
Somatotropin	P01241
Major pollen allergen Bet v 1-A	P15494
OspA (<i>Piscirickettsia salmonis</i>)	Q5BMB7
17 kDa antigen variant of OspA (<i>P. salmonis</i>)	Q9F9K8
Transforming growth factor beta-1	P01137
Transforming growth factor beta-2	P61812
Transforming growth factor beta-3	P10600
Interleukin 1 beta	P01584
Interleukin 1 alfa	P01583
Interleukin 2	P60568
Interleukin 3	P08700
Interleukin 4	P05112
Interleukin 5	P05113
Interleukin 6	P05231
Interleukin 7	P13232
Interleukin 8	P10145
Interleukin 9	P15248
Interleukin 10	P22301
Interleukin 12 subunit alfa	P29459
Interleukin 12 subunit beta	P29460
Interleukin 18	Q14116
Interleukin 21	Q9HBE4
Thymic stromal lymphopoietin	Q969D9
Brichos	SEQ ID NO:41
Neuroserpin	SEQ ID NO:49
<u>Protein</u>	<u>Sequence</u>
SP-C (SEQ ID NO 111)	FGIPCCPVHLKRLIVVVVVVLIVVVIVGALLMGL ^a
SP-C(Leu) (SEQ ID NO 112)	FGIPSSPVHLKRLKLLLLLLLLLILLILGALLMGL

Protein	Sequence
SP-C33 (SEQ ID NO 113)	IPSSPVHLKRLKLLLLLLLLLILLILGALLMGL
SP-C30 (SEQ ID NO 114)	IPSSPVHLKRLKLLLLLLLLLILLILGALL
SP-C33(Leu) (SEQ ID NO 115)	IPSSPVHLKRLKLLLLLLLLLILLILGALLLGL
LL-37 (SEQ ID NO 116)	LLGDFFRKSKEKIGKEFKRIVQRIKDFLRNLPRTES
KL4 (SEQ ID NO 117)	KLLLLKLLLLKLLLLKLLLLK
1a AA (SEQ ID NO 107)	CWLARALIKRIQALIPKGGRLLPQLVARLVLRCs ^b
1b AA (SEQ ID NO 108)	AWLCRALIKRIQALIPKGGRLLPQLVCRLVLRAS ^c
0 AAAA (SEQ ID NO 118)	AWLARALIKRIQALIPKGGRLLPQLVARLVLRAS
1a LL (SEQ ID NO 109)	CWLLRALIKRIQALIPKGGRLLPQLVLRVLRCs ^b
Protein	Sequence / Uniprot ID
1b LL (SEQ ID NO 110)	LWLCRALIKRIQALIPKGGRLLPQLVCRLVLRLS ^c
0 LLLL (SEQ ID NO 119)	LWLLRALIKRIQALIPKGGRLLPQLVLRVLRLS
^a Cys-5 and Cys-6 in native SP-C are palmitoylated	
^b Cys1-Cys33 linkage	
^c Cys4-Cys27 linkage	

[0058] Examples of aggregation-prone proteins and polypeptides include avidin, streptavidin and extracellular, ligand-binding parts of cytokine receptors. Examples of aggregation-prone proteins and polypeptides are listed in Table 8.

TABLE 8 - Aggregation-prone proteins and polypeptides

Protein	UniprotID /other reference
Streptavidin, <i>Streptomyces avidinii</i>	P22629
Streptavidin, <i>Streptomyces lavendulae</i>	B8YQ01
Streptavidin V1, <i>Streptomyces venezuelae</i>	Q53532
Streptavidin V2, <i>Streptomyces venezuelae</i>	Q53533
Putative streptavidin, <i>Burkholderia mallei</i> (strain SAVP1)	A1V7Z0
Putative streptavidin, <i>Burkholderia thailandensis</i>	Q2T1V4
Putative streptavidin, <i>Burkholderia mallei</i>	Q62EP2
Core streptavidin	GenBank: CAA77107.1
M4 (quadruple mutein of streptavidin)	J Biol Chem 280(24): 23225-23231

<u>Protein</u>	<u>UniprotID /other reference</u>
	(2005)
Avidin, <i>Gallus gallus</i>	P02701 GenBank: CAC34569.1
Actin	P68133
Interleukin 6 receptor subunit alfa	P08887
Interleukin 6 receptor subunit beta	P40189
Interleukin 2 receptor subunit alfa	P01589
Interleukin 2 receptor subunit beta	P14784
<u>Protein</u>	<u>Uniprot ID / other reference</u>
Cytokine receptor common subunit gamma	P31785
Green Fluorescent Protein (GFP)	SEQ ID NO: 44

[0059] Examples of proteases include protease 3C from coxsackie virus or human rhinovirus. Further examples of proteases are listed in Table 9.

TABLE 9 - Proteases

<u>Protease</u>	<u>Class</u>	<u>Accession no.</u>
Trypsin (bovine)	serine	P00760
Chymotrypsin (bovine)	serine	P00766
Elastase (porcine)	serine	P00772
Endoproteinase Arg-C (mouse submaxillary gland)	serine	
Endoproteinase Glu-C (V8 protease) (<i>Staphylococcus aureus</i>)	serine	P04188
Acylamino-acid-releasing enzyme (porcine)	serine	P19205
Carboxypeptidase (<i>Penicillium janthinellum</i>)	serine	P43946
Proteinase K (<i>Tritirachium album</i>)	serine	P06873
Subtilisin (<i>Bacillus subtilis</i>)	serine	P04189 P29122
Carboxypeptidase Y (yeast)	serine	P00729
Endoproteinase Lys-C (<i>Lysobacter enzymogenes</i>)	serine	S77957
Enteropeptidase (human)	serine	P98073
Prothrombin	serine	P00734
Factor X	serine	P00742
Pepsin	aspartic	P00791 P00790
Cathepsin D (human)	aspartic	P07339
HIV-1 protease	aspartic	Q9YQ34

Protease	Class	Accession no.
Cathepsin C	cysteine	
Clostripain (endoproteinase-Arg-C) (<i>Clostridium histolyticum</i>)	cysteine	P09870
Papain (<i>Carica papaya</i>)	cysteine	P00784
Protease 3C	cysteine	Q04107
Tobacco Etch virus (TEV)	cysteine	Q0GDU8
Thermolysin (<i>Bacillus thermo-proteolyticus</i>)	metallo	P00800
Endoproteinase Asp-N (<i>Pseudomonas fragi</i>)	metallo	Q9R4J4
Carboxypeptidase A (bovine)	metallo	P00730
Carboxypeptidase B (porcine)	metallo	P00732
IgA protease	metallo	Q97QP7

[0060] The desired non-spidroin protein may be selected from surfactant protein B (SP-B) and variants thereof, such as Mini-B, Mini-B27, Mini-BLeu, KL4, LL-37, and surfactant protein C (SP-C) and variants thereof, such as SP-C(Leu), SP-C33, SP-C30 and SP-C33Leu. Other preferred non-spidroin proteins are neuroserpin, GFP, and the 1a AA, 1b AA, 0 AAAA, 1a LL, 1b LL and 0 LLLL proteins.

[0061] Certain fusion proteins are selected from the group consisting of SEQ ID NOS 26, 28, 30, 34, 37, 39, 42 and 47; and proteins having at least 80%, preferably at least 90%, more preferably at least 95% identity, to any of these proteins.

[0062] There is also provided an isolated polynucleic acid encoding a fusion protein as set out herein, one isolated polynucleic acid is selected from the group consisting of SEQ ID NOS 27, 29, 31, 38, 40, 43 and 48. Another isolated polynucleic acid is selected from the group consisting of SEQ ID NOS 14-16, 18 and 24.

[0063] According to one aspect, the present invention provides a novel use of at least one moiety which is derived from the N-terminal (NT) fragment of a spider silk protein as a solubility enhancing moiety in a fusion protein for production of a desired protein or polypeptide selected from the group consisting of (a) polypeptide drugs and (b) growth factors.

[0064] There is also provided a method of producing a fusion protein. The first step involves expressing in a suitable host a fusion protein according to the invention. Suitable hosts are well known to a person skilled in the art and include e.g. bacteria and eukaryotic cells, such as yeast, insect cell lines and mammalian cell lines. Typically, this step involves expression of a polynucleic acid molecule which encodes the fusion protein in *E. coli*.

[0065] The second method step involves obtaining a mixture containing the fusion protein. The mixture may for instance be obtained by lysing or mechanically disrupting the host cells. The

mixture may also be obtained by collecting the cell culture medium, if the fusion protein is secreted by the host cell. The thus obtained protein can be isolated using standard procedures. If desired, this mixture can be subjected to centrifugation, and the appropriate fraction (precipitate or supernatant) be collected. The mixture containing the fusion protein can also be subjected to gel filtration, chromatography, e.g. anion exchange chromatography, dialysis, phase separation or filtration to cause separation. Optionally, lipopolysaccharides and other pyrogens are actively removed at this stage. If desired, linker peptides may be removed by cleavage in this step.

[0066] In a preferred method, the obtained mixture comprises the fusion protein dissolved in a liquid medium, typically a salt buffer or cell culture medium. In one preferred method, the mixture has a pH below 6.3, and preferably below 6, which promotes assembly of soluble NT domains. In another preferred method, the mixture has a pH above 6.4, and preferably above 7, which prevents or decreases assembly of soluble NT domains. A pH above 6.4, such as above 7, may be particularly useful to improve solubility of fusion proteins wherein the desired protein/polypeptide is an amyloid-forming or aggregation-prone protein/polypeptide.

[0067] According to a related aspect, the present invention provides a method of producing a desired protein or polypeptide selected from the group consisting of (a) polypeptide drugs and (b) growth factors. The first step involves expressing in a suitable host a fusion protein according to the invention. Suitable hosts are well known to a person skilled in the art and include e.g. bacteria and eukaryotic cells, such as yeast, insect cell lines and mammalian cell lines. Typically, this step involves expression of a polynucleic acid molecule which encodes the fusion protein in *E. coli*.

[0068] The second method step involves obtaining a mixture containing the fusion protein. The mixture may for instance be obtained by lysing or mechanically disrupting, e.g. sonicating, the host cells. The mixture may also be obtained by collecting the cell culture medium, if the fusion protein is secreted by the host cell. The thus obtained protein can be isolated using standard procedures. If desired, this mixture can be subjected to centrifugation, and the appropriate fraction (precipitate or supernatant) be collected. The mixture containing the fusion protein can also be subjected to gel filtration, chromatography, e.g. anion exchange chromatography, dialysis, phase separation or filtration to cause separation. Optionally, lipopolysaccharides and other pyrogens are actively removed at this stage. If desired, linker peptides may be removed by cleavage in this step. As set out above, this may be the most suitable form of the desired protein or polypeptide, i.e. as part of a fusion protein. It may provide a suitable handle for purification and detection and/or provide desirable properties, e.g. stability and in particular solubility.

[0069] In a preferred embodiment, the method may also comprise the step of cleaving the fusion protein to provide the desired protein or polypeptide. In this embodiment, the fusion protein is comprising at least one cleavage site arranged between at least one desired protein or polypeptide moiety and at least one solubility-enhancing moiety. In a typical fusion protein, this implies the presence of a single cleavage site between the solubility-enhancing moiety or

moieties and the desired protein or polypeptide. Cleavage may be achieved using standard procedures, for instance cleavage by cyanogen bromide (CNBr) after Met residues, cleavage by hydroxylamine between Asn and Gly residues, cleavage by protease 3C between Gln and Gly residues at -XLETLFQGX- sites (SEQ ID NO 120), and at various other protease sites that are well known to the person skilled in the art.

[0070] The thus obtained desired protein or polypeptide can be isolated using standard procedures. If desired, this mixture can be subjected to centrifugation, and the appropriate fraction (precipitate or supernatant) be collected. The mixture containing the desired protein or polypeptide can also be subjected to gel filtration, chromatography, dialysis, phase separation or filtration to cause separation. Optionally, lipopolysaccharides and other pyrogens are actively removed at this stage. If desired, linker peptides may be removed by cleavage in this step.

[0071] In a preferred embodiment, the obtained mixture comprises the fusion protein dissolved in a liquid medium, typically a salt buffer or cell culture medium. In one preferred embodiment, the mixture has a pH below 6.3, and preferably below 6, such as in the interval 4.2-6.3 or 4.2-6, which promotes assembly of soluble NT domains. In another preferred embodiment, the mixture has a pH above 6.4, and preferably above 7, which prevents or decreases assembly of soluble NT domains. A pH above 6.4, such as above 7, may be particularly useful to improve solubility of fusion proteins wherein the desired protein/polypeptide is an amyloid-forming or aggregation-prone protein/polypeptide.

[0072] Thus, the fusion protein is typically obtained as a solution in a liquid medium. By the terms "soluble" and "in solution" is meant that the fusion protein is not visibly aggregated and does not precipitate from the solvent at 60 000×g. The liquid medium can be any suitable medium, such as an aqueous medium, preferably a physiological medium, typically a buffered aqueous medium, such as a 10-50 mM Tris-HCl buffer or phosphate buffer. The liquid medium preferably has a pH of 6.4 or higher, such as 7 or higher, and/or an ion composition that prevents polymerisation of the solubility-enhancing moiety. That is, the liquid medium typically has either a pH of 6.4 or higher, such as 7 or higher, or an ion composition that prevents polymerisation of the solubility-enhancing moiety, or both.

[0073] Ion compositions that prevent polymerisation of the solubility-enhancing moiety can readily be prepared by the skilled person. A preferred ion composition that prevents polymerisation of the solubility-enhancing moiety has an ionic strength of more than 300 mM. Specific examples of ion compositions that prevent polymerisation of the solubility-enhancing moiety include above 300 mM NaCl, 100 mM phosphate and combinations of these ions having desired preventive effect on the polymerisation of the solubility-enhancing moiety, e.g. a combination of 10 mM phosphate and 300 mM NaCl.

[0074] It has surprisingly been found that the presence of a solubility-enhancing moiety improves the stability of the solution and prevents polymer formation under these conditions. This can be advantageous when immediate polymerisation may be undesirable, e.g. during

protein purification, in preparation of large batches, or when other conditions need to be optimized. It is preferred that the pH of the liquid medium is adjusted to 6.7 or higher, such as 7.0 or higher to achieve high solubility of the fusion protein. It can also be advantageous that the pH of the liquid medium is adjusted to the range of 6.4-6.8, which provides sufficient solubility of the spider silk protein but facilitates subsequent pH adjustment to 6.3 or lower.

[0075] Another aspect of the invention is based on the insight that the NT domain will form large soluble assemblies when the pH is lowered from ca 7 to 6, or more specifically from above 6.4 to below 6.3. This assembly occurs most efficiently at a pH above 4.2, i.e. in the range of 4.2-6.3, such as 4.2-6. This property can be used for affinity purification, e.g. if NT is immobilized on a column. This approach allows release of bound proteins by a shift in pH within a physiologically relevant interval, since the assembly will resolve when pH is elevated from ca 6 to 7.

[0076] In a preferred embodiment of the methods according to the invention, the step of isolating the fusion protein involves purification of the fusion protein on an affinity medium, such as an affinity column, with an immobilized NT moiety and/or on an anion exchange medium, such as an anion exchange column. Purification of the fusion protein on an affinity medium is preferably carried out with association to an affinity medium with an immobilized NT moiety at a pH of 6.3 or lower, preferably in the range of 4.2-6.3, followed by dissociation from the affinity medium with a desired dissociation medium, e.g. having a pH of 6.4 or higher, a pH of 4.1 or lower and/or having a high ionic strength. Purification of the fusion protein on an anion exchange medium is preferably carried out with association to the anion exchange medium at a pH of 6.4 or higher, followed by dissociation from the anion exchange medium with a dissociation medium having a high ionic strength. If desired, purification of the fusion protein on an affinity medium, such as an affinity column, with an immobilized NT moiety can be combined with purification on an anion exchange medium, such as an anion exchange column. A dissociation medium having high ionic strength typically has an ionic strength of more than 300 mM, such as above 300 mM NaCl.

[0077] These two affinity-based procedures utilize the inherent properties of the solubility-enhancing moiety according to the invention. Of particular interest is the strong tendency of spidroin NT protein fragments to associate at a pH below 6.3, in particular in the range of 4.2-6.3. This can advantageously be utilized as a powerful affinity purification tool, allowing one-step purification of fusion proteins according to the invention from complex mixtures. Although chromatography is preferred, other affinity-based purification methods than chromatography can obviously be employed, such as magnetic beads with functionalized surfaces or filters with functionalized surfaces.

[0078] The present invention will in the following be further illustrated by the following non-limiting examples.

Examples

Example 1 - Production of an SP-C33Leu fusion protein

[0079] An expression vector was constructed comprising a gene encoding NT-MetSP-C33Leu as a fusion to His₆ (SEQ ID NOS: 26-27). The vector was used to transform *Escherichia coli* BL21 (DE3) cells (Merck Biosciences) that were grown at 30°C in Luria-Bertani medium containing kanamycin to an OD₆₀₀ of 0.9-1, induced with isopropyl-p-D-thiogalactopyranoside (IPTG), and further incubated for 3 hours at 25°C. The cells were harvested by centrifugation and resuspended in 20 mM Tris-HCl, pH 8.

[0080] Lysozyme was added, and the cells were incubated for 30 min on ice. Tween was added to a final concentration of 0.7%. The cells were disrupted by sonication on ice for 5 min, alternating 2 seconds on and 2 seconds off. The cell lysate was centrifuged at 20 000 × g for 30 min. The supernatant was loaded on a Ni-NTA sepharose column, equilibrated with 20 mM Tris-HCl, pH 8 buffer containing 0.7% Tween. The column was washed with 20 mM Tris-HCl, pH 8 buffer containing 0.7% Tween, and the bound protein was eluted with 20 mM Tris-HCl pH 8, 300 mM imidazole buffer containing 0.7% Tween.

[0081] The eluate was subjected to SDS-PAGE on a 12% Tris-Glycine gel under reducing conditions. A major band corresponding to the fusion protein is indicated by the arrow in Fig. 3A. The yield was determined by mg purified protein from 1 litre shake flask culture grown to an OD₆₀₀ of 1. The yield was 64 mg/l. It is concluded that a fusion protein containing a single NT moiety results in surprisingly high yield in the presence of detergent in the cell lysate.

Example 2 - Production of an SP-C33Leu fusion protein

[0082] An expression vector was constructed comprising a gene encoding NT₂-MetSP-C33Leu (i.e. NTNT-MetSP-C33Leu) as a fusion to His₆ (SEQ ID NOS: 28-29). The vector was used to transform *Escherichia coli* BL21 (DE3) cells (Merck Biosciences) that were grown at 30°C in Luria-Bertani medium containing kanamycin to an OD₆₀₀ of 0.9-1, induced with isopropyl-β-D-thiogalactopyranoside (IPTG), and further incubated for 3 hours at 25°C. The cells were harvested by centrifugation and resuspended in 20 mM Tris-HCl, pH 8.

[0083] Lysozyme was added, and the cells were incubated for 30 min on ice. Tween was either not added or added to a final concentration of 0.7%. The cells were disrupted by sonication on ice for 5 min, alternating 2 seconds on and 2 seconds off. The cell lysate was centrifuged at 20 000 × g for 30 min. The supernatants were loaded on a Ni-NTA sepharose column, equilibrated with 20 mM Tris-HCl, pH 8 buffer ± 0.7% Tween. The column was washed with 20 mM Tris-HCl, pH 8 buffer ± 0.7% Tween, and the bound protein was eluted with 20 mM Tris-HCl pH 8, 300 mM imidazole buffer ± 0.7% Tween.

[0084] The eluate was subjected to SDS-PAGE on a 12% Tris-Glycine gel under reducing conditions. A major band corresponding to the fusion protein is indicated by the arrow in Fig. 3B. The yield was determined by mg purified protein from 1 litre shake flask culture grown to an OD₆₀₀ of 1. The yield was 40 mg/l in the absence of Tween, and 68 mg/l in the presence of 0.7% Tween. It is concluded that a fusion protein containing two consecutive NT moieties results in surprisingly high yield in the absence of detergent in the cell lysate, and an even further increased yield in the presence of detergent in the cell lysate.

Example 3 - Production of SP-C33Leu fusion proteins

[0085] Expression vectors are constructed comprising a gene encoding NT-MetSP-C33Leu, NT₂-MetSP-C33Leu and NT-MetSP-C33Leu-NT, respectively. The vectors are used to transform *Escherichia coli* BL21(DE3) cells (Merck Biosciences) that are grown at 30°C in Luria-Bertani medium containing kanamycin to an OD₆₀₀ of 0.9-1, induced with isopropyl-β-D-thiogalactopyranoside (IPTG), and further incubated for 3 hours at 25°C. The cells are harvested by centrifugation and resuspended in 20 mM Tris-HCl, pH 8.

[0086] Lysozyme is added, and the cells are incubated for 30 min on ice. Tween is either not added or added to a final concentration of 0.7%. The cells are disrupted by sonication on ice for 5 min, alternating 2 seconds on and 2 seconds off. The cell lysates are centrifuged at 20 000 × g for 30 min.

Example 4 - Preparation of NT-Sepharose

[0087] A CysHis₆NT construct is used to transform *Escherichia coli* BL21 (DE3) cells (Merck Biosciences). The cells are grown at 30°C in Luria-Bertani medium containing kanamycin to an OD₆₀₀ of 0.8-1, induced with isopropyl-p-D-thiogalactopyranoside (IPTG), and further incubated for up to 4 hours at room temperature. Thereafter, cells are harvested and resuspended in 20 mM Tris-HCl, pH 8.0, supplemented with lysozyme and DNase I. After complete lysis, the 15000g supernatants are loaded on a column packed with Ni sepharose (GE Healthcare). The column is washed extensively, and then bound proteins are eluted with 100-300 mM imidazole. Fractions containing the target proteins are pooled and dialyzed against 20 mM Tris-HCl, pH 8.0. Purified Cys-His₆-NT protein is coupled to activated thiol Sepharose using standard protocol (GE Healthcare).

Example 5 - Purification of fusion proteins using NT Sepharose

[0088] Cell lysates from Example 3 are loaded on a column packed with NT sepharose, pre-equilibrated with 20 mM NaPi, pH 6. The column is washed extensively with 20 mM NaPi, pH 6

and then bound proteins are eluted with 20 mM NaPi, pH 7. Fractions containing the target proteins are pooled. Protein samples are separated on SDS-PAGE gels and then stained with Coomassie Brilliant Blue R-250. Protein content is determined from absorbance at 280 nm.

Example 6 - Purification of fusion proteins on anion exchanger

[0089] Cell lysates from Example 3 are loaded on a HiTrap Q FF column (GE Healthcare), pre-equilibrated with 20 mM NaP pH 6.5. The column is washed extensively and then bound proteins are eluted with a linear gradient of NaCl up to 1 M. Fractions containing the target proteins are pooled. Protein samples are separated on SDS-PAGE gels and then stained with Coomassie Brilliant Blue R-250. Protein content is determined from absorbance at 280 nm.

Example 7 - Cleavage and isolation of desired protein

[0090] The fusion proteins of Examples 3, 5 and 6 are dissolved in 70% aqueous formic acid, supplemented with 0.1 g/ml CNBr and left at room temp. for 24 hours. Thereafter the mixtures are dried, and separated in the two-phase system chloroform/methanol/water, 8:4:3, by vol. SP-C33Leu is found in the organic phase and can thereafter optionally be further purified with reversed-phase HPLC using a C18 column. The activity of SP-C33Leu mixed with synthetic phospholipids can be tested *in vitro* or *in vivo*, as described in e.g. J. Johansson et al., J. Appl. Physiol. 95, 2055-2063 (2003).

Example 8 - Production of LL-37 fusion protein

[0091] An expression vector was constructed comprising a gene encoding NT₂-LL37 (i.e. NTNT-LL37) as a fusion to His₆ (SEQ ID NOS: 30-31). The vector was used to transform *Escherichia coli* BL21 (DE3) cells (Merck Biosciences) that were grown at 30°C in Luria-Bertani medium containing kanamycin to an OD₆₀₀ of 0.9-1, induced with isopropyl-β-D-thiogalactopyranoside (IPTG), and further incubated for 3 hours at 25°C. The cells were harvested by centrifugation and resuspended in 20 mM Tris-HCl, pH 8.

[0092] Lysozyme was added, and the cells were incubated for 30 min on ice. The cells were disrupted by sonication on ice for 5 min, alternating 2 seconds on and 2 seconds off. The cell lysate was centrifuged at 20 000 × g for 30 min. The supernatants were loaded on a Ni-NTA sepharose column, equilibrated with 20 mM Tris-HCl, pH 8, 250 mM NaCl buffer. The column was washed with 20 mM Tris-HCl, pH 8, 250 mM NaCl buffer, and the bound protein was eluted with 20 mM Tris-HCl pH 8, 300 mM imidazole buffer.

Example 9 - Production of NT-REP₄-CT

[0093] An expression vector was constructed to produce NT-REP₄-CT as an N-terminal fusion to His₆ (SEQ ID NOS 17-18). The vector was used to transform *Escherichia coli* BL21 (DE3) cells (Merck Biosciences) that were grown at 30°C in Luria-Bertani medium containing kanamycin to an OD₆₀₀ of ~1, induced with isopropyl-p-D-thiogalactopyranoside (IPTG), and further incubated for up to 4 hours at room temperature. Thereafter, cells were harvested and resuspended in 20 mM Tris-HCl (pH 8.0) supplemented with lysozyme and DNase I.

[0094] After complete lysis, the 15000g supernatants were loaded onto a column packed with Ni-Sepharose (GE Healthcare, Uppsala, Sweden). The column was washed extensively before bound proteins were eluted with 300 mM imidazole. Fractions containing the target proteins were pooled and dialyzed against 20 mM Tris-HCl (pH 8.0).

[0095] Protein samples were separated via SDS-PAGE and then stained with Coomassie Brilliant Blue R-250. The resulting NT-REP₄-CT protein was concentrated by ultrafiltration using a 5 kDa molecular mass cutoff cellulose filter (Millipore).

Example 10 - Production of NT-REP₄-CT

[0096] An expression vector was constructed to produce NT-REP₄-CT as a C-terminal fusion to Zbasic (SEQ ID NO 19). The vector was used to transform *Escherichia coli* BL21 (DE3) cells (Merck Biosciences) that were grown at 30°C in Luria-Bertani medium containing kanamycin to an OD₆₀₀ of ~1, induced with isopropyl-p-D-thiogalactopyranoside (IPTG), and further incubated for up to 2-4 hours at room temperature. Thereafter, cells were harvested and resuspended in 50 mM Na phosphate (pH 7.5) supplemented with lysozyme and DNase I.

[0097] After complete lysis, the 15000g supernatants were loaded onto cation exchanger (HiTrap S, GE Healthcare, Uppsala, Sweden). The column was washed extensively before bound proteins were eluted with a gradient against 500 mM NaCl. Fractions containing the target proteins were pooled and dialyzed against 50 mM Na phosphate (pH 7.5). The NT-REP₄-CT protein (SEQ ID NO 20) was released from the Zbasic tags by proteolytic cleavage using a protease 3C:fusion protein ratio of 1:50 (w/w) at 4°C over night. To remove the released Zbasic tag, the cleavage mixture was loaded onto a second cation exchanger, and the flowthrough was collected.

Example 11 - Production of NT-REP₄-CT

[0098] An expression vector was constructed to produce NT-REP₄-CT as an C-terminal fusion to HisTrxHis (SEQ ID NO 21). The vector was used to transform *Escherichia coli* BL21 (DE3)

cells (Merck Biosciences) that were grown at 30°C in Luria-Bertani medium containing kanamycin to an OD₆₀₀ of ~1, induced with isopropyl-p-D-thiogalactopyranoside (IPTG), and further incubated for up to 2-4 hours at room temperature. Thereafter, cells were harvested and resuspended in 20 mM Tris-HCl (pH 8.0) supplemented with lysozyme and DNase I.

[0099] After complete lysis, the 15000g supernatants were loaded onto column packed with Ni-Sephrose (GE Healthcare, Uppsala, Sweden). The column was washed extensively before bound proteins were eluted with a gradient against 500 mM NaCl. Fractions containing the target proteins were pooled and dialyzed against 20 mM Tris-HCl (pH 8.0). The NT-REP₄-CT protein (SEQ ID NO 22) was released from the HisTrxHis tags by proteolytic cleavage using a thrombin:fusion protein ratio of 1:1000 (w/w) at 4°C over night. To remove the released HisTrxHis, the cleavage mixture was loaded onto a second Ni-Sephrose column, and the flowthrough was collected.

Example 12 - Production of NT₂-REP₄-CT

[0100] An expression vector was constructed comprising a gene encoding NT₂-REP₄-CT (i.e. NTNT-REP₄-CT) as a fusion to His₆ (SEQ ID NOS: 23-24). The vector was used to transform *Escherichia coli* BL21 (DE3) cells (Merck Biosciences) that were grown at 30°C in Luria-Bertani medium containing kanamycin to an OD₆₀₀ of 0.9-1, induced with isopropyl-p-D-thiogalactopyranoside (IPTG), and further incubated for 3 hours at 25°C. The cells were harvested by centrifugation and resuspended in 20 mM Tris-HCl, pH 8.

[0101] Lysozyme and DNase were added, and the cells were incubated for 30 min on ice. The cell lysate was centrifuged at 20 000 × g for 30 min. The supernatants were loaded on a Ni-NTA sepharose column, equilibrated with 20 mM Tris-HCl, pH 8 buffer. The column was washed with 20 mM Tris-HCl, pH 8 buffer, and the bound protein was eluted with 20 mM Tris-HCl pH 8, 300 mM imidazole buffer.

[0102] The eluate was subjected to SDS-PAGE on a 12% Tris-Glycine gel under reducing conditions. A major band corresponding to the fusion protein is indicated by the arrow in Fig. 3C. The yield was determined by mg purified protein from 1 litre shake flask culture grown to an OD₆₀₀ of 1. The yield was 30 mg/l. It is concluded that spidroin miniature proteins can advantageously be expressed as fusions with two NT moieties.

Example 13 - Production of NT-REP₄-CT, NT₂-REP₄-CT and NT-REP₈-CT

[0103] Expression vectors are constructed comprising a gene encoding NT-REP₄-CT (SEQ ID NOS 20 and 22), NT₂-REP₄-CT (SEQ ID NO 23), and NT-REP₈-CT (SEQ ID NO: 25), respectively. The vectors are used to transform *Escherichia coli* BL21 (DE3) cells (Merck

Biosciences) that are grown at 30°C in Luria-Bertani medium containing kanamycin to an OD₆₀₀ of 0.9-1, induced with isopropyl-p-D-thiogalactopyranoside (IPTG), and further incubated for 3 hours at 25°C. The cells are harvested by centrifugation and resuspended in 20 mM Tris-HCl, pH 8.

[0104] Lysozyme is added, and the cells are incubated for 30 min on ice. Tween is either not added or added to a final concentration of 0.7%. The cell lysates are centrifuged at 20 000 × g for 30 min. One portion of the supernatant is loaded on an anion exchange column in accordance with Example 6.

[0105] An NT affinity medium is prepared as described in Example 4. Another portion of the supernatant is loaded on an NT affinity column in accordance with Example 5.

[0106] Eluates from the anion exchange column and the NT affinity column are subjected to gel electrophoresis.

Example 14 - Production of NTHis, NT₂-REP₈-CT and NT₂-Brichos

A) NTHis

[0107] An expression vector was constructed to produce NT as an N-terminal fusion to His₆ (SEQ ID NO 32). The vector was used to transform *Escherichia coli* BL21(DE3) cells (Merck Biosciences) that were grown at 30°C in Luria-Bertani medium containing kanamycin to an OD₆₀₀ of ~1, induced with isopropyl-p-D-thiogalactopyranoside (IPTG), and further incubated for up to 4 hours at room temperature. Thereafter, cells were harvested and resuspended in 20 mM Tris-HCl (pH 8.0) supplemented with lysozyme and DNase I.

[0108] After complete lysis, the 15000g supernatants were loaded onto a column packed with Ni- Sepharose (GE Healthcare, Uppsala, Sweden). The column was washed extensively before bound proteins were eluted with 300 mM imidazole. Fractions containing the target proteins were pooled and dialyzed against 20 mM Tris-HCl (pH 8.0). Protein samples were separated via SDS-PAGE and then stained with Coomassie Brilliant Blue R-250. The resulting NT protein (SEQ ID NO 32) was concentrated by ultrafiltration using a 5 kDa molecular mass cutoff cellulose filter (Millipore). The yield was 112 mg/litre shake flask grown to an OD₆₀₀ of 1.

B) NT₂-REP₈-CT

[0109] An expression vector was constructed to produce NT₂-REP₈-CT (NTNT8REPCT) as an N-terminal fusion to His₆ (SEQ ID NO 33). The vector were used to transform *Escherichia coli*

BL21(DE3) cells (Merck Biosciences) that were grown at 30°C in Luria-Bertani medium containing kanamycin to an OD₆₀₀ of ~1, induced with isopropyl-p-D-thiogalactopyranoside (IPTG), and further incubated for up to 4 hours at room temperature. Thereafter, cells were harvested and resuspended in 20 mM Tris-HCl (pH 8.0) supplemented with lysozyme and DNase I. Protein samples were separated via SDS-PAGE and then stained with Coomassie Brilliant Blue R-250 to confirm protein expression.

[0110] After complete lysis, the 15000g supernatants are loaded onto a column packed with Ni-Sepharose (GE Healthcare, Uppsala, Sweden). The column is washed extensively before bound proteins are eluted with 300 mM imidazole. Fractions containing the target proteins are pooled and dialyzed against 20 mM Tris-HCl (pH 8.0). Protein samples are separated via SDS-PAGE and then stained with Coomassie Brilliant Blue R-250.

C) NT₂-Brichos

[0111] An expression vector was constructed to produce NT₂-Brichos (NT-NT-Brichos) as an N-terminal fusion to His₆ (SEQ ID NO 34). The vector was used to transform *Escherichia coli* BL21 (DE3) cells (Merck Biosciences) that were grown at 30°C in Luria-Bertani medium containing kanamycin to an OD₆₀₀ of ~1, induced with isopropyl-p-D-thiogalactopyranoside (IPTG), and further incubated for up to 4 hours at room temperature. Thereafter, cells were harvested and resuspended in 20 mM Tris-HCl (pH 8.0) supplemented with lysozyme and DNase I. The cells were further disrupted by sonication on ice for 5 minutes, 2 seconds on and 2 seconds off.

[0112] After complete lysis, the 15000g supernatants were loaded onto a column packed with Ni-Sepharose (GE Healthcare, Uppsala, Sweden). The column was washed extensively before bound proteins were eluted with 300 mM imidazole. Fractions containing the target proteins were pooled and dialyzed against 20 mM Tris-HCl (pH 8.0). Protein samples were separated via SDS-PAGE and then stained with Coomassie Brilliant Blue R-250. The resulting NT₂-Brichos protein (SEQ ID NO 34) was concentrated by ultrafiltration using a 5 kDa molecular mass cutoff cellulose filter (Millipore). The yield was 20 mg/litre shake flask grown to an OD₆₀₀ of 1.

Example 15 - NT for pH-dependent, reversible capture

[0113] Purpose: Use covalently immobilised NT (and NTNT) to reversibly capture NT fusion proteins.

[0114] Strategy: Investigate pH dependent assembly of NT (and NTNT) fusion proteins to fibers (and film) with covalently linked NT (and NTNT). Fibers and films without NT are used as

control.

A) Fibers

[0115] Fibers (-0.5 cm long, -50 µg) of NT-REP₄-CT (SEQ ID NO 20), NT₂-REP₄-CT (SEQ ID NO 23) and REP₄-CT (SEQ ID NO 2, control) were submerged in 100 µl solution of 5 mg/ml soluble NTHis (SEQ ID NO 32) or NT₂-Brichos (SEQ ID NO 34) at pH 8 for 10 minutes. The pH was decreased by addition of 400 µl sodium phosphate buffer (NaP) to pH 6 and incubated for 10 minutes to allow assembly of soluble NT to the fiber. The fibers were transferred to 500 µl of NaP at pH 6, and washed twice. Finally, the fibers were transferred to 500 µl of NaP at pH 7, and incubated 10 minutes to allow release of soluble NT. The same was done in the presence of 300 mM NaCl in all pH 6 NaP buffers. Samples from the different solutions were analysed on SDS-PAGE.

[0116] Using the NT₂-REP₄-CT and NT-REP₄-CT fibers, both NTHis and NT₂-Brichos were captured at pH 6. Upon pH raise to pH 7, both NTHis and NT₂-Brichos were released again and could be detected on SDS-PAGE. The addition of 300 mM NaCl decreased capture at pH 6.

B) Film:

[0117] Films of NT-REP₄-CT (SEQ ID NO 20) and REP₄-CT (SEQ ID NO 2, control) were prepared by casting 50 µl of a protein solution of 3 mg/ml in a plastic well and left to dry over night. The next day, 100 µl solution of 5 mg/ml soluble NTHis (SEQ ID NO 32) at pH 8 was added to wells with film, and left for 10 minutes. The pH was then decreased to 6 by addition of 400 µl NaP and incubated for 10 minutes to allow assembly of soluble NT to the film. The films were then washed twice with 500 µl of NaP at pH 6. For release of soluble NTHis, 500 µl of NaP at pH 7 was added and incubated for 10 minutes. The same was done in presence of 300 mM NaCl in all pH 6 NaP buffers. Samples from the different solutions were analysed on SDS-PAGE.

[0118] Analysis on SDS-PAGE showed that a NT-REP₄-CT film allowed NTHis to be captured at pH 6 and released again upon raise of the pH to 7.

Example 16 - NT for pH-dependent, reversible assembly of fusion proteins

[0119] Purpose: Use NT as a reversible tag that allows analysis of interaction between protein moieties, e.g. analyse the interaction of Brichos with targets with beta sheet structures e.g. surfactant protein C (SP-C).

[0120] NT₂-Brichos (SEQ ID NO 34) is mixed with either NT₂-MetSP-C33Leu (SEQ ID NO 28) or NTHis (SEQ ID NO 32) to a total volume of 100 µl at pH 8. NaP buffer (400 µl) is added to give a final pH of 6, and the mixture is incubated for 20 minutes to allow NT assembly. The pH is then raised again to pH 7 to allow reversal of NT assembly. Samples from the different solutions are analysed on native gel and size exclusion chromatography (SEC).

Example 17 - Cleavage of NT-MetSP-C33Leu and isolation of SP-C33Leu

[0121] About 58 mg of lyophilized HisNT-MetSP-C33Leu (SEQ ID NOS: 26) obtained in Example 1 was dissolved in 3 ml of 70 % aqueous formic acid by vortexing and sonication. To this solution, 200 µl of 5 M CNBr in acetonitrile was added, and the mixture was incubated at room temperature for 24 h. Thereafter solvents were evaporated under a stream of nitrogen, and the residue was washed three times by solubilisation in 70% aqueous formic acid and drying under nitrogen.

[0122] To the dried residue was then added 4.56 ml of chloroform/methanol/ water (8:4:3, by vol), after which the mixture was vortexed and centrifuged. The upper phase was removed, and 1 ml of chloroform/methanol/water (8:4:3, by vol) was added to the lower phase, and vortexing, centrifugation and removal of upper phase were repeated. The two upper phases were combined and dried under vacuum. The lower phase was dried under nitrogen.

[0123] The contents of the lower (left lane) and upper (right lane) phases were analyzed by SDS-PAGE (Fig 4). This showed that the lower phase contains one major band with an estimated molecular mass that agrees well with that of SP-C33Leu. The identity of SP-C33Leu was confirmed by ESI mass spectrometry and amino acid sequencing, which showed a monoisotopic mass of 3594.6 Da (calculated 3594.4 Da) and the expected amino acid sequence.

Example 18 - Analysis of surface activity of SP-C33Leu/phospholipid mixture

[0124] 1,2-Dipalmitoyl-sn-glycero-3-phosphocholine (DPPC)/1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol (POPG) (68:31, w/w) was dissolved in chloroform:methanol (1:1, v/v) and mixed with SP-C33Leu (obtained in Example 17) in the same solvent. The peptide content in the preparations was 2% in relation to the phospholipid weight. The solvents were evaporated with nitrogen, and the preparations were resuspended in saline to a final phospholipid concentration of 10 mg/ml by slow rotation at 37°C.

[0125] Surface tension was measured in triplicates in an alveolus by a captive bubble surfactometer (CBS) (Schurch S et al., J. Appl. Physiol. 67: 2389-2396, 1989). In the CBS, surfactant and an air-bubble representing the lung alveolus are present in an air-tight enclosed chamber. To evaluate surface activity under dynamic circumstances, the chamber is

compressed and surface tension can be calculated by studying the shape and height/width ratio of the bubble.

[0126] In the experiment, 2 μ l of the SP-C33Leu surfactant preparation (10 mg/ml) was inserted into the sucrose-filled test chamber. After insertion, an air-bubble was created and surface tension was measured during five minutes of adsorption. In the following quasi-static cycling experiments, the bubble was compressed stepwise from the initial volume until a surface tension less than 5 mN/m was reached, alternatively to a maximum area compression of 50% and then expanded during five cycles.

[0127] The results are illustrated in Fig. 5, where the first and fifth cycle from one representative example out of three measurements are shown. The surface activity of the SP-C33Leu/DPPC/POPG mixture (Fig. 5) was very similar to that of synthetic SP-C33 in the same phospholipid mixture, see e.g. Johansson et al, J. Appl. Physiol, 95: 2055-2063 (2003).

Example 19- Production of an SP-C33Leu fusion protein

A) without NT (comparative example)

[0128] An expression vector was constructed comprising a gene encoding Thioredoxin (TRX)-SP-C33Leu as a fusion to 2 $\times$ His₆ (SEQ ID NOS: 35-36). The vector was used to transform *Escherichia coli* BL21 (DE3) cells (Merck Biosciences) that were grown at 30°C in Luria-Bertani medium containing kanamycin to an OD₆₀₀ of 0.9-1, induced with IPTG and further incubated for 3 hours at 25°C. The cells were harvested and resuspended in 20 mM Tris-HCl (pH 8.0).

[0129] Lysozyme was added, and the cells were incubated for 30 min on ice. Tween was added to a final concentration of 0.7%. The cells were disrupted by sonication on ice for 5 min, alternating 2 sec on and 2 sec off. The cell lysate was centrifuged at 20 000 $\times$ g for 30 min. The supernatants were loaded onto a column packed with Ni-Sepharose (GE Healthcare, Uppsala, Sweden), equilibrated with 20 mM Tris-HCl (pH 8.0) + 0.7% Tween. The column was washed extensively before bound proteins were eluted with a 300 mM imidazole + 0.7% Tween.

[0130] The target protein was eluted with 300 mM imidazole + 0.7 % Tween and analyzed by SDS-PAGE (Fig. 6A). The eluate contained a small and impure amount of target protein.

B) with NT

[0131] An expression vector was constructed comprising a gene encoding TRX-NT-SP-C33Leu as a fusion to 2 $\times$ His₆ (SEQ ID NOS:37-38). The vector was used to transform *Escherichia coli* BL21 (DE3) cells (Merck Biosciences) that were grown at 30°C in Luria-Bertani

medium containing kanamycin to an OD₆₀₀ of 0.9-1, induced with IPTG, and further incubated for 3 hours at 25°C. The cells were harvested and resuspended in 20 mM Tris-HCl (pH 8.0).

[0132] Lysozyme was added, and the cells were incubated for 30 min on ice. Tween was added to a final concentration of 0.7%. The cells were disrupted by sonication on ice for 5 min, alternating 2 sec on and 2 sec off. The cell lysate was centrifuged at 20 000×g for 30 min. The supernatants were loaded onto a column packed with Ni-Sepharose (GE Healthcare, Uppsala, Sweden), equilibrated with 20 mM Tris-HCl (pH 8.0)+ 0.7% Tween. The column was washed extensively before bound proteins were eluted with a 300 mM imidazole + 0.7% Tween. Fractions containing the fusion proteins were pooled and dialyzed against deionized water.

[0133] The eluate was subjected to SDS-PAGE on a 12 % Tris-Glycine gel under reducing conditions. A major band corresponding to the protein is indicated by the arrow in Fig. 6B. The yield was determined by mg purified protein from 1 litre shake flask culture grown to an OD₆₀₀ of 1. The yield was 30 mg/l.

Example 20 - Production of Brichos

[0134] An expression vector was constructed comprising a gene encoding NT₂-Brichos (i.e. NTNT-Brichos) as a fusion to His₆LinkHis₆ (SEQ ID NOS:39-40). The vector was used to transform *Escherichia coli* BL21 (DE3) cells (Merck Biosciences) that were grown at 30°C in Luria-Bertani medium containing kanamycin to an OD₆₀₀ of 0.9-1, induced with IPTG, and further incubated for 3 hours at 25°C. The cells were harvested and resuspended in 20 mM Tris-HCl (pH 8.0).

[0135] Lysozyme was added, and the cells were incubated for 30 min on ice. The cells were disrupted by sonication on ice for 5 min, alternating 2 sec on and 2 sec off. The cell lysate was centrifuged at 20 000×g for 30 min. The supernatants were loaded onto a column packed with Ni-Sepharose (GE Healthcare, Uppsala, Sweden). The column was washed extensively before bound proteins were eluted with a 300 mM imidazole. Fractions containing the fusion proteins were pooled and dialyzed against 20 mM Tris-HCl (pH 8.0).

[0136] The eluate was subjected to SDS-PAGE on a 12 % Tris-Glycine gel under reducing conditions. A major band corresponding to the fusion protein is indicated by the arrow in Fig. 7. The yield was determined by mg purified protein from 1 litre shake flask culture grown to an OD₆₀₀ of 1. The yield was 28 mg/l of the fusion protein.

[0137] The Brichos protein (SEQ ID NO: 41) is released from the 2His₆NT₂ tags by proteolytic cleavage using a protease 3C:fusion protein ratio of 1:100 (w/w) at 4°C. To remove the released 2His₆NT₂ tag, the cleavage mixture is loaded onto a second Ni-Sepharose, and the flow through is collected.

Example 21- Production of Green Fluorescent Protein (GFP)

[0138] The GFP utilized in this example is a S147P variant, see Kimata, Y et al., Biochem. Biophys. Res. Commun. 232: 69-73 (1997).

A) with NT

[0139] An expression vector was constructed comprising a gene encoding NT₂-GFP (i.e. NTNT-GFP) as a fusion to His₆LinkHis₆ (SEQ ID NOS:42-43). The vector was used to transform *Escherichia coli* BL21 (DE3) cells (Merck Biosciences) that were grown at 30°C in Luria-Bertani medium containing kanamycin to an OD₆₀₀ of 0.9-1, induced with IPTG, and further incubated for 3 hours at 25°C. The cells were harvested and resuspended in 20 mM Tris-HCl (pH 8.0).

[0140] Lysozyme was added, and the cells were incubated for 30 min on ice. The cells were disrupted by sonication on ice for 5 min, alternating 2 sec on and 2 sec off. The cell lysate was centrifuged at 20 000×g for 30 min. The supernatants were loaded onto a column packed with Ni-Sepharose (GE Healthcare, Uppsala, Sweden). The column was washed extensively before bound proteins were eluted with a 300 mM imidazole. Fractions containing the fusion proteins were pooled and dialyzed against 20 mM Tris-HCl (pH 8.0). The GFP protein (SEQ ID NO: 44) was released from the 2His₆NT₂ tags by proteolytic cleavage using a protease 3C:fusion protein ratio of 1:100 (w/w) at 4°C. To remove the released 2His₆NT₂ tag, the cleavage mixture was loaded onto a second Ni-Sepharose, and the flow through was collected.

[0141] The eluates were subjected to SDS-PAGE on a 12 % Tris-Glycine gel under reducing conditions (Fig. 8). Major bands corresponding to the fusion protein (first eluate, left lane) and the target protein (second eluate, right lane) are indicated by the arrows in Fig. 8. The yield was determined by mg purified protein from 1 litre shake flask culture grown to an OD₆₀₀ of 1. The yield was 44 mg/l of the fusion protein and 16 mg/l of the target protein.

[0142] The purified GFP was highly fluorescent, confirming the right fold (beta barrel with linking alpha helix) that is obligate for autocatalytic formation of the chromophore.

B) with other purification tags: Zb and His₆ABP (comparative example)

[0143] BL21 (DE3) cells harboring the vectors (pT7ZbGFP, pT7His₆ABPGFP) were grown over night at 37°C in tryptic soy broth media supplemented with kanamycin. On the following morning, the cultures were inoculated into 100 ml fresh media in 1 litre shake flasks and grown until an OD₆₀₀ of 1 was reached. Protein production was then induced by addition of IPTG to a

final concentration of 1 mM, and production continued for 18 h. The cells were harvested and resuspended in 50 mM sodium phosphate buffer (pH 7.5). The cells were disrupted by sonication on ice for 3 min, alternating 1 sec on and 1 sec off. The cell lysate was centrifuged at 10 000×g for 20 min. The supernatants were loaded onto columns.

[0144] The ZbGFP (SEQ ID NO: 45) fusion protein was purified on 1 ml HiTrap S HP columns in 50 mM sodium phosphate pH 7.5 and eluted with the same buffer supplemented with 160 mM NaCl.

[0145] The His₆ABPGFP (SEQ ID NO: 46) fusion protein was purified on 1 ml Talon columns in 50 mM sodium phosphate pH 8 and eluted with the same buffer supplemented with 30 mM acetic acid and 70 mM sodium acetate, which gives a pH of 5.0.

[0146] The eluates were subjected to SDS-PAGE on a 10-20 % gradient gel under reducing conditions. The yield was determined by mg purified protein/1 litre shake flask culture grown to an OD₆₀₀ of 1. The yield was 10 mg/l for ZbGFP and 7 mg/l for His₆ABPGFP.

Example 22- Production of Neuroserpin

[0147] An expression vector was constructed comprising a gene encoding NT₂-Neuroserpin (i.e. NTNT-Neuroserpin) as a fusion to HiS₆LinkHiS₆ (SEQ ID NOS:47-48). The vector was used to transform *Escherichia coli* BL21 (DE3) cells (Merck Biosciences) that were grown at 30°C in Luria-Bertani medium containing kanamycin to an OD₆₀₀ of 0.9-1, induced with IPTG, and further incubated for 3 hours at 25°C. The cells were harvested and resuspended in 20 mM Tris-HCl (pH 8.0).

[0148] Lysozyme was added, and the cells were incubated for 30 min on ice. The cells were disrupted by sonication on ice for 5 min, alternating 2 sec on and 2 sec off. The cell lysate was centrifuged at 20 000×g for 30 min. The supernatants were loaded onto a column packed with Ni-Sepharose (GE Healthcare, Uppsala, Sweden). The column was washed extensively before bound proteins were eluted with 300 mM imidazole. Fractions containing the fusion proteins were pooled and dialyzed against 20 mM Tris-HCl (pH 8.0).

[0149] The neuroserpin protein (SEQ ID NO: 49) was released from the 2His₆NT₂ tags by proteolytic cleavage using a protease 3C:fusion protein ratio of 1:100 (w/w) at 4°C. To remove the released 2His₆NT₂ tag, the cleavage mixture was loaded onto a second Ni-Sepharose, and the flow through was collected.

[0150] The eluates were subjected to SDS-PAGE on a 12 % Tris-Glycine gel under reducing conditions (Fig. 9). Major bands corresponding to the fusion protein (first eluate, left lane) and the target protein (second eluate, right lane) are indicated by the arrows in Fig. 9. The yield

was determined by mg purified protein from 1 litre shake flask culture grown to an OD₆₀₀ of 1. The yield was 8 mg/l of the fusionprotein and 4 mg/l of the target protein. As a comparison, the expression yield of neuroserpin with His₆ tag was 1.7 mg/l (Belorgey et al. Eur J Biochem. 271(16):3360-3367 (2004).

[0151] The inhibition rate of tPa (tissue plasminogen activator) by the expressed neuroserpin was determined to be the same as published earlier (Belorgey et al. J. Biol. Chem. 277, 17367-17373 (2002).

Example 23 - Production of a protease 3C fusion proteins

[0152] Expression vectors are constructed comprising a gene encoding His₆NT-3C and His₆LinkHis₆NTNT3C, respectively (Gräslund T. et al., Protein Expr Purif 9(1): 125-132 (1997); Cordingley MG. et al., J. Virol. 63(12): 5037-5045 (1989)). The vectors are used to transform *Escherichia coli* BL21(DE3) cells (Merck Biosciences), which are grown at 30°C in Luria-Bertani medium containing kanamycin to an OD₆₀₀ of 0.9-1, induced with IPTG, and further incubated for 3 h at 25°C. The cells are harvested and resuspended in 20 mM Tris-HCl (pH 8.0).

[0153] Lysozyme and DNase are added, and the cells are incubated for 30 min on ice. The cells are further disrupted by sonication on ice for 3 min, alternating 1 sec on and 1 sec off. The cell lysate is centrifuged at 15 000×g for 30 min. The supernatants are loaded onto a column packed with Ni-Sepharose (GE Healthcare, Uppsala, Sweden), equilibrated with 20 mM Tris-HCl (pH 8.0). The column is washed extensively before bound proteins are eluted with 300 mM imidazole. Fractions containing the fusion proteins are pooled and dialyzed against deionized water. The eluate is subjected to SDS-PAGE under reducing conditions. The yield is determined by mg purified protein from 1 litre shake flask culture grown to an OD₆₀₀ of 1.

SEQUENCE LISTING

[0154]

<110> Spiber Technologies AB

<120> PRODUCTION OF PROTEINS AND POLYPEPTIDES

<130> EP-21083749

<150> EP 10156927

<151> 2010-03-18

<150> PCT/SE2010/050439

<151> 2010-04-21

<150> PCT/SE2010/051163

<151> 2010-10-27

<150> EP 10848078

<151> 2010-10-27

<160> 120

<170> PatentIn version 3.5

<210> 1

<211> 149

<212> PRT

<213> Euprosthénops australis

<400> 1

Gly	Ser	Gly	Asn	Ser	Gly	Ile	Gln	Gly	Gln	Gly	Gly	Tyr	Gly	Gly	Leu
1				5					10						15

Gly	Gln	Gly	Gly	Tyr	Gly	Gln	Gly	Ala	Gly	Ser	Ser	Ala	Ala	Ala	Ala
		20						25					30		

Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Gly	Gly	Gln	Gly	Gly	Gln	Gly	Gln
		35					40					45			

Gly	Gly	Tyr	Gly	Gln	Gly	Ser	Gly	Gly	Ser	Ala	Ala	Ala	Ala	Ala	Ala
	50					55					60				

Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Gly	Arg	Gly	Gln	Gly	Gly	Tyr
65					70					75					80

Gly	Gln	Gly	Ser	Gly	Gly	Asn	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala
				85					90					95	

Ala	Ala	Ala	Ala	Ala	Gly	Gln	Gly	Gly	Gln	Gly	Gly	Tyr	Gly	Arg	Gln
			100					105					110		

Ser	Gln	Gly	Ala	Gly	Ser	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala
		115					120					125			

Ala	Ala	Ala	Ala	Gly	Ser	Gly	Gln	Gly	Gly	Tyr	Gly	Gln	Gly	Gln	Gly
		130					135				140				

Gly	Tyr	Gly	Gln	Ser
145				

<210> 2

<211> 265

<212> PRT

<213> Euprosthénops australis

<220>

<221> DOMAIN

<222> (1)..(167)

<223> REP fragment

<220>

<221> DOMAIN

<222> (168)..(265)

<223> CT fragment

<400> 2

Gly Ser Gly Asn Ser Gly Ile Gln Gly Gln Gly Gly Tyr Gly Gly Leu
 1 5 10 15

Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala
 20 25 30

Ala Ala Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln
 35 40 45

Gly Gly Tyr Gly Gln Gly Ser Gly Gly Ser Ala Ala Ala Ala Ala Ala
 50 55 60

Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr
 65 70 75 80

Gly Gln Gly Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala
 85 90 95

Ala Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Arg Gln
 100 105 110

Ser Gln Gly Ala Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 115 120 125

Ala Ala Ala Ala Gly Ser Gly Gln Gly Gly Tyr Gly Gln Gly Gln Gly
 130 135 140

Gly Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala Ser Ala Ala

145 150 155 160

Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro Ser Ala Val
 165 170 175

Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn Gly Gln Val
 180 185 190

Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser Ser Val
 195 200 205

Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile Val Gln Ala
 210 215 220

Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser Ser Ser Ser
 225 230 235 240

Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile Thr Asn Val Val
 245 250 255

Ala Asn Ala Met Ala Gln Val Met Gly
 260 265

<210> 3

<211> 296

<212> PRT

<213> Euprosthénops australis

<220>

<221> DOMAIN

<222> (1)..(137)

<223> NT fragment

<220>

<221> DOMAIN

<222> (138)..(296)

<223> REP fragment

<400> 3

Gly Ser Gly Asn Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala
 1 5 10 15

Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly
 20 25 30

Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met
 35 40 45

Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn
 50 55 60

Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile
 65 70 75 80

Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser
 85 90 95

Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val
 100 105 110

Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala
 115 120 125

Gln Ala Gly Met Asn Asp Val Ser Ala Ser Ala Ser Ala Gly Ala Ser
 130 135 140

Ala Ala Ala Ser Ala Gly Ala Ala Ser Gly Gln Gly Gly Tyr Gly Gly
 145 150 155 160

Leu Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala
165 170 175

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly
180 185 190

Gln Gly Gly Tyr Gly Gln Gly Ser Gly Gly Ser Ala Ala Ala Ala Ala
195 200 205

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly
210 215 220

Tyr Gly Gln Gly Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala
225 230 235 240

Ala Ala Ala Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly
245 250 255

Arg Gln Ser Gln Gly Ala Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala
260 265 270

Ala Ala Ala Ala Ala Ala Gly Ser Gly Gln Gly Gly Tyr Gly Gly Gln
275 280 285

Gly Gln Gly Gly Tyr Gly Gln Ser
290 295

<210> 4

<211> 340

<212> PRT

<213> Euprosthonops australis

<220>

<221> DOMAIN

<222> (1)..(137)

<223> NT fragment

<220>

<221> DOMAIN

<222> (138)..(340)

<223> REP fragment

<400> 4

Gly Ser Gly Asn Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala
1 5 10 15

Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly
20 25 30

Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met
35 40 45

Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn
50 55 60

```

Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile
65          70          75          80

Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser
85          90          95

Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val
100         105         110

Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala
115         120         125

Gln Ala Gly Met Asn Asp Val Ser Ala Ser Ala Ser Ala Gly Ala Ser
130         135         140

Ala Ala Ala Ser Ala Gly Ala Pro Gly Tyr Ser Pro Ala Pro Ser Tyr
145         150         155         160

Ser Ser Gly Gly Tyr Ala Ser Ser Ala Ala Ser Ala Ala Ala Ala Ala
165         170         175

Gly Gln Gly Gly Pro Gly Gly Tyr Gly Pro Ala Pro Asn Gln Gly Ala
180         185         190

Ser Ser Ala Ala Ala Ala Ala Ala Gly Ser Gly Gln Gly Pro Ser Gly
195         200         205

Pro Tyr Gly Thr Ser Tyr Gln Ile Ser Thr Gln Tyr Thr Gln Thr Thr
210         215         220

Thr Ser Gln Gly Gln Gly Tyr Gly Ser Ser Ser Ala Gly Ala Ala Ala
225         230         235         240

Ala Gly Ala Ala Gly Ala Gly Gln Gly Gly Tyr Gly Gly Gln Gly Gln
245         250         255

Gly Gly Tyr Gly Gln Gly Ala Gly Gly Ala Ala Ala Ala Ala Ala Ala
260         265         270

Ala Ala Ala Ala Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Gly
275         280         285

Gly Tyr Gly Gln Gly Gly Gln Gly Gly Gln Gly Gly Gln Gly Gln Gly
290         295         300

Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
305         310         315         320

Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Pro
325         330         335

Gly Ser Gly Gly
340

```

<210> 5
 <211> 424
 <212> PRT
 <213> Euprostenops australis

<220>
 <221> DOMAIN
 <222> (1)..(136)
 <223> NT fragment

<220>
 <221> DOMAIN
 <222> (137)..(313)
 <223> REP fragment

<220>
 <221> DOMAIN
 <222> (314)..(411)
 <223> CT fragment

<220>
 <221> DOMAIN
 <222> (412)..(424)
 <223> His tag

<400> 5
 Met Lys Ala Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala Glu
 1 5 10 15
 Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly Phe
 20 25 30
 Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met Val
 35 40 45
 Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn Lys
 50 55 60
 Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala
 65 70 75 80
 Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser Ile
 85 90 95
 Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val Asn
 100 105 110
 Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala Gln
 115 120 125
 Ala Gly Met Asn Asp Val Ser Ala Ser Ala Ser Ala Gly Ala Ser Ala
 130 135 140

130	135	140
Ala Ala Ser Ala Gly Ala Ala Ser Gly Gln Gly Gly Tyr Gly Gly Leu		
145	150	155
Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala		
	165	170
		175
Ala Ala Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln		
	180	185
		190
Gly Gly Tyr Gly Gln Gly Ser Gly Gly Ser Ala Ala Ala Ala Ala Ala		
	195	200
		205
Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr		
	210	215
		220
Gly Gln Gly Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala		
	225	230
		235
		240
Ala Ala Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Arg		
	245	250
		255
Gln Ser Gln Gly Ala Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala		
	260	265
		270
Ala Ala Ala Ala Ala Gly Ser Gly Gln Gly Gly Tyr Gly Gly Gln Gly		
	275	280
		285
Gln Gly Gly Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala Ala Ser		
	290	295
		300
Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro Ser		
	305	310
		315
		320
Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn Gly		
	325	330
		335
Gln Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser		
	340	345
		350
Ser Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile Val		
	355	360
		365
Gln Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser Ser		
	370	375
		380
Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile Thr Asn		
	385	390
		395
		400
Val Val Ala Asn Ala Met Ala Gln Val Met Gly Lys Leu Ala Ala Ala		
	405	410
		415
Leu Glu His His His His His His		
	420	

<210> 6

<211> 137

<212> PRT

<213> Euprosthénops australis

<220>

<221> VARIANT

<222> (6)..(6)

<223> deletion (deltaHis)

<400> 6

```

Gly Ser Gly Asn Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala
1          5          10          15

Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly
20          25          30

Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met
35          40          45

Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn
50          55          60

Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile
65          70          75          80

Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser
85          90          95

Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val
100         105         110

Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala
115         120         125

Gln Ala Gly Met Asn Asp Val Ser Ala
130         135

```

<210> 7

<211> 98

<212> PRT

<213> Euprosthénops australis

<400> 7

```

Ser Arg Leu Ser Ser Pro Ser Ala Val Ser Arg Val Ser Ser Ala Val
1          5          10          15

Ser Ser Leu Val Ser Asn Gly Gln Val Asn Met Ala Ala Leu Pro Asn
20          25          30

Ile Ile Ser Asn Ile Ser Ser Ser Val Ser Ala Ser Ala Pro Gly Ala
35          40          45

```

Ser Gly Cys Glu Val Ile Val Gln Ala Leu Leu Glu Val Ile Thr Ala
 50 55 60

Leu Val Gln Ile Val Ser Ser Ser Ser Val Gly Tyr Ile Asn Pro Ser
 65 70 75 80

Ala Val Asn Gln Ile Thr Asn Val Val Ala Asn Ala Met Ala Gln Val
 85 90 95

Met Gly

<210> 8

<211> 131

<212> PRT

<213> Artificial Sequence

<220>

<223> Consensus sequence derived from spidroin NT fragments

<220>

<221> VARIANT

<222> (20)..(20)

<223> Leu

<220>

<221> VARIANT

<222> (42)..(42)

<223> Asn

<220>

<221> VARIANT

<222> (42)..(42)

<223> Gln

<220>

<221> VARIANT

<222> (50)..(50)

<223> Ser

<220>

<221> VARIANT

<222> (50)..(50)

<223> Lys

<220>

<221> VARIANT

<222> (56)..(56)

<223> Arg

<220>

<221> VARIANT

<222> (84)..(84)

<223> Leu

<220>

<221> VARIANT

<222> (114)..(114)

<223> Ser

<220>

<221> VARIANT

<222> (121)..(121)

<223> Asn

<220>

<221> VARIANT

<222> (123)..(123)

<223> Leu

<220>

<221> VARIANT

<222> (124)..(124)

<223> Ser

<400> 8

Gln	Ala	Asn	Thr	Pro	Trp	Ser	Ser	Pro	Asn	Leu	Ala	Asp	Ala	Phe	Ile
1			5					10						15	

Asn	Ser	Phe	Met	Ser	Ala	Ala	Ser	Ser	Ser	Gly	Ala	Phe	Ser	Ala	Asp
		20					25					30			

Gln	Leu	Asp	Asp	Met	Ser	Thr	Ile	Gly	Asp	Thr	Leu	Met	Ser	Ala	Met
	35						40				45				

Asp	Asn	Met	Gly	Arg	Ser	Gly	Lys	Ser	Thr	Lys	Ser	Lys	Leu	Gln	Ala
50						55				60					

Leu	Asn	Met	Ala	Phe	Ala	Ser	Ser	Met	Ala	Glu	Ile	Ala	Ala	Ala	Glu
65					70					75					80

Ser	Gly	Gly	Gly	Ser	Val	Gly	Val	Lys	Thr	Asn	Ala	Ile	Ser	Asp	Ala
				85					90					95	

Leu	Ser	Ser	Ala	Phe	Tyr	Gln	Thr	Thr	Gly	Ser	Val	Asn	Pro	Gln	Phe
			100					105					110		

Val	Asn	Glu	Ile	Arg	Ser	Leu	Ile	Gly	Met	Phe	Ala	Gln	Ala	Ser	Ala
	115						120					125			

Asn	Glu	Val
-----	-----	-----

130

<210> 9

<211> 100

<212> PRT

<213> Artificial Sequence

<220>

<223> Consensus sequence derived from known MaSp1 and MaSp2 proteins

<220>

<221> MISC_FEATURE

<222> (1)..(71)

<223> Sequence length present in known species variants

<220>

<221> VARIANT

<222> (7)..(7)

<223> Glu

<400> 9

Ser	Arg	Leu	Ser	Ser	Pro	Gln	Ala	Ser	Ser	Arg	Val	Ser	Ser	Ala	Val
1				5					10					15	

Ser	Asn	Leu	Val	Ser	Ser	Gly	Pro	Thr	Asn	Ser	Ala	Ala	Leu	Ser	Asn
		20						25					30		

Thr	Ile	Ser	Asn	Val	Val	Ser	Gln	Ile	Ser	Ala	Ser	Asn	Pro	Gly	Leu
		35					40					45			

Ser	Gly	Cys	Asp	Val	Leu	Val	Gln	Ala	Leu	Leu	Glu	Val	Val	Ser	Ala
	50					55					60				

Leu	Val	His	Ile	Leu	Gly	Ser	Ser	Ser	Ile	Gly	Gln	Val	Asn	Tyr	Gly
65					70					75					80

Ser	Ala	Gly	Gln	Ala	Thr	Gln	Ile	Val	Gly	Gln	Ser	Val	Ala	Gln	Ala
				85					90					95	

Leu	Gly	Glu	Phe
			100

<210> 10

<211> 1110

<212> PRT

<213> Euprosthonops australis

<220>

<221> REPEAT

<222> (7)..(19)

<220>

<221> REPEAT
<222> (20)..(42)

<220>
<221> REPEAT
<222> (43)..(56)

<220>
<221> REPEAT
<222> (57)..(70)

<220>
<221> REPEAT
<222> (71)..(83)

<220>
<221> REPEAT
<222> (84)..(106)

<220>
<221> REPEAT
<222> (107)..(120)

<220>
<221> REPEAT
<222> (121)..(134)

<220>
<221> REPEAT
<222> (135)..(147)

<220>
<221> REPEAT
<222> (148)..(170)

<220>
<221> REPEAT
<222> (171)..(183)

<220>
<221> REPEAT
<222> (184)..(197)

<220>
<221> REPEAT
<222> (198)..(211)

<220>

<221> REPEAT
<222> (212)..(234)

<220>
<221> REPEAT
<222> (235)..(248)

<220>
<221> REPEAT
<222> (249)..(265)

<220>
<221> REPEAT
<222> (266)..(279)

<220>
<221> REPEAT
<222> (280)..(293)

<220>
<221> REPEAT
<222> (294)..(306)

<220>
<221> REPEAT
<222> (307)..(329)

<220>
<221> REPEAT
<222> (330)..(342)

<220>
<221> REPEAT
<222> (343)..(356)

<220>
<221> REPEAT
<222> (357)..(370)

<220>
<221> REPEAT
<222> (371)..(393)

<220>
<221> REPEAT
<222> (394)..(406)

<220>

<221> REPEAT
<222> (407)..(420)

<220>
<221> REPEAT
<222> (421)..(434)

<220>
<221> REPEAT
<222> (435)..(457)

<220>
<221> REPEAT
<222> (458)..(470)

<220>
<221> REPEAT
<222> (471)..(488)

<220>
<221> REPEAT
<222> (489)..(502)

<220>
<221> REPEAT
<222> (503)..(516)

<220>
<221> REPEAT
<222> (517)..(529)

<220>
<221> REPEAT
<222> (530)..(552)

<220>
<221> REPEAT
<222> (553)..(566)

<220>
<221> REPEAT
<222> (567)..(580)

<220>
<221> REPEAT
<222> (581)..(594)

<220>

<221> REPEAT
<222> (595)..(617)

<220>
<221> REPEAT
<222> (618)..(630)

<220>
<221> REPEAT
<222> (631)..(647)

<220>
<221> REPEAT
<222> (648)..(661)

<220>
<221> REPEAT
<222> (662)..(675)

<220>
<221> REPEAT
<222> (676)..(688)

<220>
<221> REPEAT
<222> (689)..(711)

<220>
<221> REPEAT
<222> (712)..(725)

<220>
<221> REPEAT
<222> (726)..(739)

<220>
<221> REPEAT
<222> (740)..(752)

<220>
<221> REPEAT
<222> (753)..(775)

<220>
<221> REPEAT
<222> (776)..(789)

<220>

<221> REPEAT
<222> (790)..(803)

<220>
<221> REPEAT
<222> (804)..(816)

<220>
<221> REPEAT
<222> (817)..(839)

<220>
<221> REPEAT
<222> (840)..(853)

<220>
<221> REPEAT
<222> (854)..(867)

<220>
<221> REPEAT
<222> (868)..(880)

<220>
<221> REPEAT
<222> (881)..(903)

<220>
<221> REPEAT
<222> (904)..(917)

<220>
<221> REPEAT
<222> (918)..(931)

<220>
<221> REPEAT
<222> (932)..(945)

<220>
<221> REPEAT
<222> (946)..(968)

<220>
<221> REPEAT
<222> (969)..(981)

<220>

<221> REPEAT

<222> (982)..(998)

<220>

<221> REPEAT

<222> (999)..(1013)

<220>

<221> REPEAT

<222> (1014)..(1027)

<220>

<221> REPEAT

<222> (1028)..(1042)

<220>

<221> REPEAT

<222> (1043)..(1059)

<220>

<221> REPEAT

<222> (1060)..(1073)

<220>

<221> REPEAT

<222> (1074)..(1092)

<400> 10

Gln Gly Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 1 5 10 15

Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln
 20 25 30

Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala
 35 40 45

Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly
 50 55 60

Gln Gly Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 65 70 75 80

Ala Ala Ser Gly Gln Gly Gly Gln Gly Gly Gln Gly Gly Gln Gly Gln
 85 90 95

Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala
 100 105 110

Ala Ala Ala Ala Ala Ala Ala Ala Gly Gln Gly Gln Gly Arg Tyr Gly
 115 120 125

Gln Gly Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala
 130 135 140

Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Gln Gly Gly Leu Gly Gln
 145 150 155 160

Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala
 165 170 175

Ser Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln
 180 185 190

Gly Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 195 200 205

Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln
 210 215 220

Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala
 225 230 235 240

Ala Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln Gly
 245 250 255

Arg Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
 260 265 270

Ala Ala Ala Ala Ala Ala Ala Gly Gln Gly Gln Gly Gly Tyr Gly Gln
 275 280 285

Gly Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 290 295 300

Ala Ala Gly Gln Gly Gly Gln Gly Gly Gln Gly Gly Leu Gly Gln Gly
 305 310 315 320

Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
 325 330 335

Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly
 340 345 350

Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Glu Ala Ala
 355 360 365

Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln Gly
 370 375 380

Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
 385 390 395 400

Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly
 405 410 415

Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 420 425 430

Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln Gly

Ala Ala Gly Gln Gly Gly Gln Gly Tyr Gly Gly Leu Gly Gln Gly
 435 440 445
 Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala
 450 455 460
 Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln Gly Arg
 465 470 475 480
 Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
 485 490 495
 Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly
 500 505 510
 Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 515 520 525
 Ser Gly Gln Gly Ser Gln Gly Gly Gln Gly Gly Gln Gly Gln Gly Gly
 530 535 540
 Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
 545 550 555 560
 Ala Ala Ala Ala Ala Ser Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly
 565 570 575
 Ala Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 580 585 590
 Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln Gly
 595 600 605
 Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala
 610 615 620
 Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln Gly Gly Tyr
 625 630 635 640
 Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala
 645 650 655
 Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Ser
 660 665 670
 Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ser
 675 680 685
 Gly Gln Gly Gly Gln Gly Gly Gln Gly Gly Gln Gly Gln Gly Gly Tyr
 690 695 700
 Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala
 705 710 715 720
 Ala Ala Ala Ala Ala Gly Gln Gly Gln Gly Gly Tyr Gly Gln Gly Ala
 725 730 735

Gly	Gly	Asn	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala
			740						745					750	
Gly	Gln	Gly	Gly	Gln	Gly	Gly	Gln	Gly	Gly	Leu	Gly	Gln	Gly	Gly	Tyr
			755					760					765		
Gly	Gln	Gly	Ala	Gly	Ser	Ser	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala
		770				775						780			
Ala	Ala	Ala	Ala	Ala	Gly	Arg	Gly	Gln	Gly	Gly	Tyr	Gly	Gln	Gly	Val
					790					795					800
Gly	Gly	Asn	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala
				805					810					815	
Gly	Gln	Gly	Gly	Gln	Gly	Gly	Gln	Gly	Gly	Leu	Gly	Gln	Gly	Gly	Tyr
			820					825					830		
Gly	Gln	Gly	Ala	Gly	Ser	Ser	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala
			835				840					845			
Ala	Ala	Ala	Ala	Ala	Gly	Arg	Gly	Gln	Gly	Gly	Tyr	Gly	Gln	Gly	Ser
			850			855					860				
Gly	Gly	Asn	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ser
					870					875					880
Gly	Gln	Gly	Ser	Gln	Gly	Gly	Gln	Gly	Gly	Gln	Gly	Gln	Gly	Gly	Tyr
				885					890					895	
Gly	Gln	Gly	Ala	Gly	Ser	Ser	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala
			900					905					910		
Ala	Ala	Ala	Ala	Ser	Gly	Arg	Gly	Gln	Gly	Gly	Tyr	Gly	Gln	Gly	Ala
			915				920					925			
Gly	Gly	Asn	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala
			930				935					940			
Ala	Gly	Gln	Gly	Gly	Gln	Gly	Gly	Tyr	Gly	Gly	Leu	Gly	Gln	Gly	Gly
					950					955					960
Tyr	Gly	Gln	Gly	Ala	Gly	Ser	Ser	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala
				965					970						975
Ala	Ala	Ala	Ala	Gly	Gly	Gln	Gly	Gly	Gln	Gly	Gln	Gly	Gly	Tyr	Gly
				980				985					990		
Gln	Gly	Ser	Gly	Gly	Ser	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala
			995				1000						1005		
Ala	Ala	Ala	Ala	Ala	Gly	Arg	Gly	Gln	Gly	Gly	Tyr	Gly	Gln	Gly	
						1015						1020			
Ser	Gly	Gly	Asn	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala
							1030						1035		

```

1035
1040
1045
1050
1055
1060
1065
1070
1075
1080
1085
1090
1095
1100
1105
1110

```

Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Arg Gln
 1040 1045 1050
 Ser Gln Gly Ala Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala
 1055 1060 1065
 Ala Ala Ala Ala Ala Gly Ser Gly Gln Gly Gly Tyr Gly Gly Gln
 1070 1075 1080
 Gly Gln Gly Gly Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala
 1085 1090 1095
 Ala Ser Ala Ala Ser Thr Val Ala Asn Ser Val Ser
 1100 1105 1110

<210> 11

<211> 23

<212> PRT

<213> Artificial Sequence

<220>

<223> Consensus sequence derived from internal repeats of Euprostenops australis MaSp1

<220>

<221> VARIANT

<222> (4)..(4)

<223> Ser

<220>

<221> VARIANT

<222> (8)..(8)

<223> Tyr

<220>

<221> VARIANT

<222> (11)..(11)

<223> Gln

<400> 11

Gly Gln Gly Gly Gln Gly Gly Gln Gly Gly Leu Gly Gln Gly Gly Tyr
 1 5 10 15

Gly Gln Gly Ala Gly Ser Ser
 20

<210> 12

<211> 17

<212> PRT

<213> Artificial Sequence

<220>

<223> Consensus sequence derived from internal repeats of Euprostenops australis MaSp1

<220>

<221> VARIANT

<222> (9)..(9)

<223> Arg

<220>

<221> VARIANT

<222> (14)..(14)

<223> Ser

<220>

<221> VARIANT

<222> (16)..(16)

<223> Gly

<400> 12

Gly	Gln	Gly	Gly	Gln	Gly	Gln	Gly	Gly	Tyr	Gly	Gln	Gly	Ala	Gly	Ser
1				5					10					15	

Ser

<210> 13

<211> 14

<212> PRT

<213> Artificial Sequence

<220>

<223> Consensus sequence derived from internal repeats of Euprostenops australis MaSp1

<220>

<221> VARIANT

<222> (2)..(2)

<223> Gln

<220>

<221> VARIANT

<222> (6)..(6)

<223> Arg

<220>

<221> VARIANT

<222> (11)..(11)

<223> Ser

<220>

<221> VARIANT

<222> (11)..(11)

<223> Val

<400> 13

Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Gly Asn
 1 5 10

<210> 14

<211> 888

<212> DNA

<213> Euprostenops australis

<400> 14

```

ggttctggga attcacacac tacaccatgg acaaaccag gactcgcaga aaacttcatg      60
aacagtttca tgcaaggcct gagctcgatg ccaggtttca cggcaagcca attggatgat      120
atgtcaacca tcgcacaatc catggtacag tcaatacaat ccttggcggc acaaggcagg      180
acatcaccca ataagctgca ggcccttaac atggcttttg catcttcgat ggcagaaatc      240
gcggcatccg aagaaggagg gggaagcctt tccacaaaa ctagctctat agccagtgca      300
atgtccaacg cgtttctgca aacaactgga gtggtaaacc aaccgttcat aaatgaaata      360
actcagctcg ttagcatggt tgctcaagca ggtatgaatg atgtcagtgc ttccgcatca      420
gcaggagcat ccgcagcagc atccgcagga gcggctagcg gtcaagggtg gatgggtgga      480
ctagggtcaag gaggatatgg acaagggtgca ggaagttctg cagccgctgc cgccgccgca      540
gcagccgccg cagcaggtgg acaagggtgga caagggtcaag gaggatatgg acaagggttca      600
ggaggttctg cagccgccgc cgccgccgca gcagcagcag cagctgcagc agctggacga      660
ggtcaaggag gatattgtca aggttctgga ggtaatgtg ctgccgcagc cgctgccgcc      720
gccgccgccg ctgcagcagc cggacaggga ggtcaagggt gatattggtag acaaagccaa      780
ggtgctggtt ccgctgctgc tgctgctgct gctgctgccg ctgctgctgc tgcaggatct      840
ggacaagggt gatacggtag acaagggtcaa ggaggttatg gtcagagt      888

```

<210> 15

<211> 1272

<212> DNA

<213> Euprostenops australis

<400> 15

```

atgaaagcat cacacactac accatggaca aaccaggac tcgcagaaaa cttcatgaac      60
agtttcatgc aaggcctgag ctcgatgccg ggtttcacgg caagccaatt ggatgatatg      120
tcaaccatcg cacaatccat ggtacagtca atacaatcct tggcggcaca aggcaggaca      180
tcaccgaata agctgcaggc cottaacatg gcttttgcac cttcgatggc agaaatcgcg      240
gcatccgaag aaggaggggg aagcctttcc accaaaacta gctctatagc cagtgcfaat      300
tccaacgcgt ttctgcaaac aactggagtg gtaaaccaac cgttcataaa tgaaataact      360
cagctcggtt gcatgtttgc tcaagcaggt atgaatgatg tcagtgcctc cgcatcagca      420

```

```

ggagcatccg cagcagcatc cgcaggagcg gctagcggtc aaggtggata tggtaggacta 480
ggtcaaggag gatattggaca aggtgcagga agttctgcag ccgctgccgc cgcgcgagca 540
gccgccgcag caggtggaca aggtggacaa ggtcaaggag gatattggaca aggttcagga 600
ggttctgcag ccgccgccgc cgcgcgagca gcagcagcag ctgcagcagc tggacgaggt 660
caaggaggat atggtcaagg ttctggaggt aatgctgctg ccgcagccgc tgcgcgccgc 720
gccgccgctg cagcagccgg acagggaggt caaggtggat atggtagaca aagccaaggt 780
gctggttccg ctgctgctgc tgetgctgct gctgccgctg ctgctgctgc aggatctgga 840
caaggtggat acggtggaca aggtcaagga ggttatggtc agagtagtgc ttctgcttca 900
gctgctgctg cagctgctag tactgtagct aattcggtag gtgcctctc atgccttcc 960
gcagtatctc gagtttcttc agcagtttct agcttggttt caaatgggtca agtgaatatg 1020
gcagcggtac ctaatatcat ttccaacatt tcttcttctg tcagtgcctc tgctcctggt 1080
gcttctggat gtgaggtcat agtgcaagct ctactgaag tcatcactgc tcttggtcaa 1140
atcgtagtt cttctagtgt tggatatatt aatccatctg ctgtgaacca aattactaat 1200
gttggtgcta atgccatggc tcaagtaatg ggcaagcttg cggccgcact cgagcaccac 1260
caccaccacc ac 1272

```

<210> 16

<211> 1020

<212> DNA

<213> Euprostenops australis

<400> 16

```

ggttctggga attcacacac tacaccatgg acaaaccag gactcgcaga aaacttcatg 60
aacagtttca tgcaaggcct gagctcgatg ccaggtttca cggcaagcca attggatgat 120
atgtcaacca tcgcacaatc catggtacag tcaatacaat ccttggcggc acaaggcagg 180
acatcacoga ataagctgca ggcccttaac atggcttttg catcttcgat ggcagaaatc 240
ggggcatccg aagaaggagg gggaagcctt tccacaaaa ctagctctat agccagtgca 300
atgtccaacg cgtttctgca aacaactgga gtggtaaacc aaccgttcat aaatgaaata 360
actcagctcg ttagcatggt tgctcaagca ggtatgaatg atgtcagtgc ttccgcatca 420
gcaggagcat ccgcagcagc atccgcagga gcgccagggt acagtccctgc accaagctac 480
agttcgggag gttatgcttc aagtgctgcc tcagcagccg ctgcagcagg acaaggagga 540
cctgggggat acggtccagc acctaaccaa ggagcttcat ctgccgctgc tgcagccgca 600
ggatcaggac aaggaccatc aggaccgtac ggtacatctt accagataag tacacaatat 660
actcaaacia cgacttcaca gggacaagga tatgggtcaa gtagcgctgg agccgcagct 720
gcaggcgctg caggtgctgg acaagggggc tacggagggtc aaggtcaagg aggatatggt 780
caaggagccg gaggtgctgc cgcagcagcc gccgctgccg cagccgctgc cgcgcagcc 840
ggacaagggt gacaagggtg aggaggatat ggacaaggag gacaaggagg acaaggagga 900
caaggtcaag gaggatatgg acaaggtgca ggaagtctg cagccgccgc cgcgcagca 960
ggggggggg gggggggggg gggggggggt gggggggggt atggtggggg ttctgggggt 1020

```

gcagcagccg ccgcagcagc aggcagaggt caaggaggat atggccagc cctcggaggt 1020

<210> 17

<211> 422

<212> PRT

<213> Euprostenops australis

<400> 17

Met Lys Ala Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala Glu
1 5 10 15

Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly Phe
20 25 30

Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met Val
35 40 45

Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn Lys
50 55 60

Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala
65 70 75 80

Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser Ile
85 90 95

Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val Asn
100 105 110

Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala Gln
115 120 125

Ala Gly Met Asn Asp Val Ser Ala Ser Ala Ser Ala Gly Ala Ser Ala
130 135 140

Ala Ala Ser Ala Gly Ala Ala Ser Gly Gln Gly Gly Tyr Gly Gly Leu
145 150 155 160

Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala
165 170 175

Ala Ala Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln
180 185 190

Gly Gly Tyr Gly Gln Gly Ser Gly Gly Ser Ala Ala Ala Ala Ala
195 200 205

Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr
210 215 220

Gly Gln Gly Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala
225 230 235 240

Ala Ala Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Arg

245 250 255
 Gln Ser Gln Gly Ala Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala
 260 265 270
 Ala Ala Ala Ala Ala Gly Ser Gly Gln Gly Gly Tyr Gly Gly Gln Gly
 275 280 285
 Gln Gly Gly Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala Ala Ser
 290 295 300
 Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro Ser
 305 310 315 320
 Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn Gly
 325 330 335
 Gln Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser
 340 345 350
 Ser Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile Val
 355 360 365
 Gln Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser Ser
 370 375 380
 Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile Thr Asn
 385 390 395 400
 Val Val Ala Asn Ala Met Ala Gln Val Met Gly Ala Ala Ala Leu Glu
 405 410 415
 His His His His His His
 420

<210> 18

<211> 1266

<212> DNA

<213> Euprostenops australis

<400> 18

atgaaagcat cacacactac accatggaca aaccaggac tcgcagaaaa cttcatgaac	60
agtttcatgc aaggcctgag ctcgatgccca ggtttcacgg caagccaatt ggatgatatg	120
tcaaccatcg cacaatccat ggtacagtca atacaatcct tggcggcaca aggcaggaca	180
tcaccgaata agctgcaggc ccttaacatg gcttttgcac cttcgatggc agaaatcgcg	240
gcatccgaag aaggaggggg aagcctttcc accaaaacta gctctatagc cagtgcgaatg	300
tccaacgcgt ttctgcgaac aactggagtg gtaaaccaac cgttcataaa tgaaataact	360
cagctcgtaa gcatgtttgc tcaagcaggt atgaatgatg tcagtgcctc cgcacagca	420
ggagcatccg cagcagcatc cgcaggagcg gctagcggtc aaggaggata tggaggacta	480
ggtcaaggag gatattggaca aggtgcagga agttctgcag ccgctgccgc cgcgcagca	540
gccgccgcag caggaggaca aggtggacaa ggtcaaggag gatattggaca aggttcagga	600

```

ggttctgcag ccgccgccgc cgccgcagca gcagcagcag ctgcagcagc tggacgaggt      660
caaggaggat atggtcaagg ttctggaggt aatgctgctg ccgcagccgc tgccgccgcc      720
gccgccgctg cagcagccgg acagggaggt caaggaggat atggtagaca aagccaaggt      780
gctggttccg ctgctgctgc tgetgctgct gctgccgctg ctgctgctgc aggatctgga      840
caaggaggat acggtggaca aggtcaagga ggttatggtc agagtagtgc ttctgcttca      900
gctgctgctg cagctgctag tactgtagct aattcgggtga gtgcctctc atgccttcc      960
gcagtatctc gagtttcttc agcagtttct agcttggttt caaatgggtca agtgaatatg     1020
gcagcggtac ctaatatcat ttccaacatt tcttctctctg tcagtgcctc tgctcctggt     1080
gcttctggat gtgaggtcat agtgcaagct ctactcgaag tcatcactgc tcttggtcaa     1140
atcgtagttt cttctagtgt tggatatatt aatccatctg ctgtgaacca aattactaat     1200
gttggtgcta atgcatggc tcaagtaatg ggcgcgccgc cactcgagca ccaccaccac     1260
caccac                                                                    1266

```

<210> 19

<211> 491

<212> PRT

<213> Euprostenops australis

<400> 19

```

Met Ala Ser Met Thr Gly Gly Gln Gln Met Gly Arg Gly Ser Val Asp
1           5           10          15

```

```

Asn Lys Phe Asn Lys Glu Arg Arg Arg Ala Arg Arg Glu Ile Arg His

```

20

25

30

```

Leu Pro Asn Leu Asn Arg Glu Gln Arg Arg Ala Phe Ile Arg Ser Leu
35           40           45

```

```

Arg Asp Asp Pro Ser Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys Lys
50           55           60

```

```

Leu Asn Asp Ala Gln Ala Pro Lys Pro Asn Leu Glu Ala Leu Phe Gln
65           70           75           80

```

```

Gly Pro Asn Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala Glu
85           90           95

```

```

Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly Phe
100          105          110

```

```

Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met Val
115          120          125

```

```

Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn Lys
130          135          140

```

```

Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala
145          150          155          160

```

Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser Ile
 165 170 175
 Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val Asn
 180 185 190
 Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala Gln
 195 200 205
 Ala Gly Met Asn Asp Val Ser Ala Ser Ala Ser Ala Gly Ala Ser Ala
 210 215 220
 Ala Ala Ser Ala Gly Ala Ala Ser Gly Gln Gly Gly Tyr Gly Gly Leu
 225 230 235 240
 Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala
 245 250 255
 Ala Ala Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln
 260 265 270
 Gly Gly Tyr Gly Gln Gly Ser Gly Gly Ser Ala Ala Ala Ala Ala Ala
 275 280 285
 Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr
 290 295 300
 Gly Gln Gly Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala
 305 310 315 320
 Ala Ala Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Arg
 325 330 335
 Gln Ser Gln Gly Ala Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala
 340 345 350
 Ala Ala Ala Ala Ala Gly Ser Gly Gln Gly Gly Tyr Gly Gly Gln Gly
 355 360 365
 Gln Gly Gly Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala Ala Ser
 370 375 380
 Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro Ser
 385 390 395 400
 Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn Gly
 405 410 415
 Gln Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser
 420 425 430
 Ser Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile Val
 435 440 445
 Gln Ala Val Val Glu Val Ile Phe Ala Val Val Glu Val Val Ser Ser

Gln Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser Ser
450 455 460

Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile Thr Asn
465 470 475 480

Val Val Ala Asn Ala Met Ala Gln Val Met Gly
485 490

<210> 20

<211> 411

<212> PRT

<213> Euprosthénops australis

<400> 20

Gly Pro Asn Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala Glu
1 5 10 15

Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly Phe
20 25 30

Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met Val
35 40 45

Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn Lys
50 55 60

Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala
65 70 75 80

Ala Ser Glu Glu Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser Ile
85 90 95

Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val Asn
100 105 110

Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala Gln
115 120 125

Ala Gly Met Asn Asp Val Ser Ala Ser Ala Ser Ala Gly Ala Ser Ala
130 135 140

Ala Ala Ser Ala Gly Ala Ala Ser Gly Gln Gly Gly Tyr Gly Gly Leu
145 150 155 160

Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala
165 170 175

Ala Ala Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln
180 185 190

Gly Gly Tyr Gly Gln Gly Ser Gly Gly Ser Ala Ala Ala Ala Ala Ala
195 200 205

Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr
210 215 220

Gly Gln Gly Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala
 225 230 235 240
 Ala Ala Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Arg
 245 250 255
 Gln Ser Gln Gly Ala Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala
 260 265 270
 Ala Ala Ala Ala Ala Gly Ser Gly Gln Gly Gly Tyr Gly Gly Gln Gly
 275 280 285
 Gln Gly Gly Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala Ala Ser
 290 295 300
 Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro Ser
 305 310 315 320
 Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn Gly
 325 330 335
 Gln Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser
 340 345 350
 Ser Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile Val
 355 360 365
 Gln Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser Ser
 370 375 380
 Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile Thr Asn
 385 390 395 400
 Val Val Ala Asn Ala Met Ala Gln Val Met Gly
 405 410

<210> 21

<211> 551

<212> PRT

<213> Euprosthénops australis

<400> 21

Met Gly His His His His His His Met Ala Ser Ser Asp Lys Ile Ile
 1 5 10 15
 His Leu Thr Asp Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly
 20 25 30
 Ala Ile Leu Val Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met
 35 40 45
 Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp Glu Tyr Gln Gly Lys Leu
 50 55 60
 Thr Val Ala Lys Leu Asn Ile Asp Gln Asn Pro Gly Thr Ala Pro Lys
 65 70 75 80

Tyr	Gly	Ile	Arg	Gly	Ile	Pro	Thr	Leu	Leu	Phe	Lys	Asn	Gly	Glu	
				85				90					95		
Val	Ala	Ala	Thr	Lys	Val	Gly	Ala	Leu	Ser	Lys	Gly	Gln	Leu	Lys	Glu
			100					105					110		
Phe	Leu	Asp	Ala	Asn	Leu	Ala	Gly	Ser	Gly	Ser	Gly	His	Met	His	His
		115					120					125			
His	His	His	His	Ser	Ser	Gly	Leu	Val	Pro	Arg	Gly	Ser	Gly	Asn	Ser
	130					135					140				
His	Thr	Thr	Pro	Trp	Thr	Asn	Pro	Gly	Leu	Ala	Glu	Asn	Phe	Met	Asn
145					150					155					160
Ser	Phe	Met	Gln	Gly	Leu	Ser	Ser	Met	Pro	Gly	Phe	Thr	Ala	Ser	Gln
				165					170					175	
Leu	Asp	Asp	Met	Ser	Thr	Ile	Ala	Gln	Ser	Met	Val	Gln	Ser	Ile	Gln
			180					185					190		
Ser	Leu	Ala	Ala	Gln	Gly	Arg	Thr	Ser	Pro	Asn	Lys	Leu	Gln	Ala	Leu
		195					200					205			
Asn	Met	Ala	Phe	Ala	Ser	Ser	Met	Ala	Glu	Ile	Ala	Ala	Ser	Glu	Glu
	210					215					220				
Gly	Gly	Gly	Ser	Leu	Ser	Thr	Lys	Thr	Ser	Ser	Ile	Ala	Ser	Ala	Met
225					230					235					240
Ser	Asn	Ala	Phe	Leu	Gln	Thr	Thr	Gly	Val	Val	Asn	Gln	Pro	Phe	Ile
				245					250					255	
Asn	Glu	Ile	Thr	Gln	Leu	Val	Ser	Met	Phe	Ala	Gln	Ala	Gly	Met	Asn
			260					265					270		
Asp	Val	Ser	Ala	Ser	Ala	Ser	Ala	Gly	Ala	Ser	Ala	Ala	Ala	Ser	Ala
		275					280					285			
Gly	Ala	Ala	Ser	Gly	Gln	Gly	Gly	Tyr	Gly	Gly	Leu	Gly	Gln	Gly	Gly
	290					295					300				
Tyr	Gly	Gln	Gly	Ala	Gly	Ser	Ser	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala
305					310					315					320
Ala	Ala	Ala	Ala	Gly	Gly	Gln	Gly	Gly	Gln	Gly	Gln	Gly	Gly	Tyr	Gly
				325					330					335	
Gln	Gly	Ser	Gly	Gly	Ser	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala
			340					345					350		
Ala	Ala	Ala	Ala	Ala	Gly	Arg	Gly	Gln	Gly	Gly	Tyr	Gly	Gln	Gly	Ser
		355					360					365			
Gly	Gly	Asn	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala	Ala
	370					375					380				

Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Arg Gln Ser Gln Gly
385 390 395 400

Ala Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
405 410 415

Ala Gly Ser Gly Gln Gly Gly Tyr Gly Gly Gln Gly Gln Gly Gly Tyr
420 425 430

Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala Ala Ser Ala Ala Ser Thr
435 440 445

Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro Ser Ala Val Ser Arg
450 455 460

Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn Gly Gln Val Asn Met
465 470 475 480

Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser Ser Val Ser Ala
485 490 495

Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile Val Gln Ala Leu Leu
500 505 510

Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser Ser Ser Ser Val Gly
515 520 525

Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile Thr Asn Val Val Ala Asn
530 535 540

Ala Met Ala Gln Val Met Gly
545 550

<210> 22

<211> 412

<212> PRT

<213> Euprostenops australis

<400> 22

Gly Ser Gly Asn Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala
1 5 10 15

Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly
20 25 30

Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met
35 40 45

Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn
50 55 60

Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile
65 70 75 80

Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser
 85 90 95
 Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val
 100 105 110
 Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala
 115 120 125
 Gln Ala Gly Met Asn Asp Val Ser Ala Ser Ala Ser Ala Gly Ala Ser
 130 135 140
 Ala Ala Ala Ser Ala Gly Ala Ala Ser Gly Gln Gly Gly Tyr Gly Gly
 145 150 155 160
 Leu Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala
 165 170 175
 Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly
 180 185 190
 Gln Gly Gly Tyr Gly Gln Gly Ser Gly Gly Ser Ala Ala Ala Ala Ala
 195 200 205
 Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly
 210 215 220
 Tyr Gly Gln Gly Ser Gly Gly Asn Ala Ala Ala Ala Ala Ala Ala
 225 230 235 240
 Ala Ala Ala Ala Ala Ala Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly
 245 250 255
 Arg Gln Ser Gln Gly Ala Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala
 260 265 270
 Ala Ala Ala Ala Ala Ala Gly Ser Gly Gln Gly Gly Tyr Gly Gly Gln
 275 280 285
 Gly Gln Gly Gly Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala Ala
 290 295 300
 Ser Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro
 305 310 315 320
 Ser Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn
 325 330 335
 Gly Gln Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser
 340 345 350
 Ser Ser Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile
 355 360 365

Val Gln Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser
370 375 380

Ser Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile Thr
385 390 395 400

Asn Val Val Ala Asn Ala Met Ala Gln Val Met Gly
405 410

<210> 23

<211> 534

<212> PRT

<213> Euprostenops australis

<400> 23

Met Gly His His His His His Met Ser His Thr Thr Pro Trp Thr
1 5 10 15

Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu
20 25 30

Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr
35 40 45

Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly
50 55 60

Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser
65 70 75 80

Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser
85 90 95

Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln
100 105 110

Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu
115 120 125

Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Gly Gly Gly Thr Pro
130 135 140

Trp Thr Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln
145 150 155 160

Gly Leu Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met
165 170 175

Ser Thr Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala
180 185 190

Gln Gly Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe
195 200 205

Ala Ser Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser
 210 215 220
 Leu Ser Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe
 225 230 235 240
 Leu Gln Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr
 245 250 255
 Gln Leu Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Val Ser Ala
 260 265 270
 Gly Asn Ser Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln Gly Gly Tyr
 275 280 285
 Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala
 290 295 300
 Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln Gly Tyr Gly Gln
 305 310 315 320
 Gly Ser Gly Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 325 330 335
 Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Ser Gly
 340 345 350
 Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 355 360 365
 Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Arg Gln Ser Gln Gly Ala
 370 375 380
 Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 385 390 395 400
 Gly Ser Gly Gln Gly Gly Tyr Gly Gly Gln Gly Gln Gly Gly Tyr Gly
 405 410 415
 Gln Ser Ser Ala Ser Ala Ser Ala Ala Ala Ser Ala Ala Ser Thr Val
 420 425 430
 Ala Asn Ser Val Ser Arg Leu Ser Ser Pro Ser Ala Val Ser Arg Val
 435 440 445
 Ser Ser Ala Val Ser Ser Leu Val Ser Asn Gly Gln Val Asn Met Ala
 450 455 460
 Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser Ser Val Ser Ala Ser
 465 470 475 480
 Ala Pro Gly Ala Ser Gly Cys Glu Val Ile Val Gln Ala Leu Leu Glu
 485 490 495
 Val Ile Thr Ala Leu Val Gln Ile Val Ser Ser Ser Ser Val Gly Tyr
 500 505 510

Ile Asn Pro Ser Ala Val Asn Gln Ile Thr Asn Val Val Ala Asn Ala
 515 520 525

Met Ala Gln Val Met Gly
 530

<210> 24

<211> 1602

<212> DNA

<213> Euprosthénops australis

<400> 24

atgggcatc atcatcatca tcatatgagc cataccaccc cgtggaccaa cccgggcctg	60
gcggaaaact ttatgaacag ctttatgcag ggcctgagca gcatgccggg ctttaccgag	120
agccagctgg atgatatgag caccattgag cagagcatgg tgcagagcat tcagagcctg	180
gcggcgaggg gccgtaccag cccgaacaaa ctgcaggcgc tgaacatggc gtttgcgagc	240
agcatggcgg aaattgcggc gagcgaagaa ggcggcggca gcctgagcac caaaaccagc	300
agcattgcga gcgcgatgag caacgcgttt ctgcagacca ccggcggtgt gaaccagccg	360
tttattaacg aaattaccca gctggtgagc atgtttgcgc aggcgggtat gaatgacggc	420
ggcgacaccc cgtggaccaa cccgggcctg gcggaaaact ttatgaacag ctttatgcag	480
ggcctgagca gcatgccggg ctttaccgag agccagctgg atgatatgag caccattgag	540
cagagcatgg tgcagagcat tcagagcctg gcggcgaggg gccgtaccag cccgaacaaa	600
ctgcaggcgc tgaacatggc gtttgcgagc agcatggcgg aaattgcggc gagcgaagaa	660
ggcggcggca gcctgagcac caaaaccagc agcattgcga gcgcgatgag caacgcgttt	720
ctgcagacca ccggcggtgt gaaccagccg tttattaacg aaattaccca gctggtgagc	780
atgtttgcgc aggcgggcat gaacgatgtg agcgcgggga attcaggtca aggtggatat	840
ggtggactag gtcaaggagg atatggacaa ggtgcaggaa gttctgcagc cgctgccgcc	900
gccgcagcag ccgccgcagc aggtggacaa ggtggacaag gtcaaggagg atatggacaa	960
ggttcaggag gttctgcagc cgccgccgcc gccgcagcag cagcagcagc tgcagcagct	1020
ggacgaggtc aaggaggata tggtaagggt tctggaggta atgctgctgc cgcagccgct	1080
gccgccgccg ccgccgctgc agcagccgga caggagggtc aaggaggata tggtagacaa	1140
agccaagggtg ctggttcgcg tgcctgctgt gctgctgctg ctgccgctgc tgcctgctga	1200
ggatctggac aaggaggata cggtggacaa ggtcaaggag gttatggtca gagtagtgct	1260
tctgcttcag ctgctgcgct agctgctagt actgtagcta attcggtgag tcgcctctca	1320
tgcctctccg cagtatctcg agtttcttca gcagtttcta gcttggtttc aaatggtcaa	1380
gtgaatatgg cagcgttacc taatatcatt tccaacattt cttcttctgt cagtgcattc	1440
gctcctggtg cttctggatg tgaggcata gtgcaagctc tactogaagt catcactgct	1500
cttgttcaaa tcgttagttc ttctagtgtt ggatatatta atccatctgc tgtgaaccaa	1560
attactaatg ttgttgctaa tgccatggct caagtaatgg gc	1602

<210> 25

<211> 514

<212> PRT

<213> Euprosthénops australis

<400> 25

Gly Ser Gly Asn Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala
1 5 10 15

Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly
20 25 30

Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met
35 40 45

Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn
50 55 60

Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile
65 70 75 80

Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser
85 90 95

Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val
100 105 110

Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala
115 120 125

Gln Ala Gly Met Asn Asp Val Ser Ala Gly Tyr Gly Gln Gly Ala Gly
130 135 140

Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
145 150 155 160

Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Ser Gly Gly Asn Ala Ala
165 170 175

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ser Gly Gln Gly Ser Gln
180 185 190

Gly Gly Gln Gly Gly Gln Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly
195 200 205

Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ser
210 215 220

Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly Gly Asn Ala Ala

225 230 235 240

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Gln Gly Gly
245 250 255

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala

Gln Gly Gly Tyr Gly Gly Leu Gly Gln Gly Gly Tyr Gly Gln Gly Ala
260 265 270

Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly
275 280 285

Gly Gln Gly Gly Gln Gly Gln Gly Gly Tyr Gly Gln Gly Ser Gly Gly
290 295 300

Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
305 310 315 320

Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Ser Gly Gly Asn Ala Ala
325 330 335

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Gln Gly
340 345 350

Gly Gln Gly Gly Tyr Gly Arg Gln Ser Gln Gly Ala Gly Ser Ala Ala
355 360 365

Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gly Ser Gly Gln
370 375 380

Gly Gly Tyr Gly Gly Gln Gly Gln Gly Gly Tyr Gly Gln Ser Ser Ala
385 390 395 400

Ser Ala Ser Ala Ala Ala Ser Ala Ala Ser Thr Val Ala Asn Ser Val
405 410 415

Ser Arg Leu Ser Ser Pro Ser Ala Val Ser Arg Val Ser Ser Ala Val
420 425 430

Ser Ser Leu Val Ser Asn Gly Gln Val Asn Met Ala Ala Leu Pro Asn
435 440 445

Ile Ile Ser Asn Ile Ser Ser Ser Val Ser Ala Ser Ala Pro Gly Ala
450 455 460

Ser Gly Cys Glu Val Ile Val Gln Ala Leu Leu Glu Val Ile Thr Ala
465 470 475 480

Leu Val Gln Ile Val Ser Ser Ser Ser Val Gly Tyr Ile Asn Pro Ser
485 490 495

Ala Val Asn Gln Ile Thr Asn Val Val Ala Asn Ala Met Ala Gln Val
500 505 510

Met Gly

<210> 26

<211> 176

<212> PRT

<213> Artificial Sequence

<220>

<223> Fusion protein

<400> 26

Met Gly His His His His His His Met Ser His Thr Thr Pro Trp Thr
1 5 10 15

Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu
20 25 30

Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr
35 40 45

Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly
50 55 60

Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser
65 70 75 80

Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser
85 90 95

Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln
100 105 110

Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu
115 120 125

Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Val Ser Ala Met Ile
130 135 140

Pro Ser Ser Pro Val His Leu Lys Arg Leu Lys Leu Leu Leu Leu Leu
145 150 155 160

Leu Leu Leu Ile Leu Leu Leu Ile Leu Gly Ala Leu Leu Leu Gly Leu

165 170 175

<210> 27

<211> 528

<212> DNA

<213> Artificial Sequence

<220>

<223> Fusion protein

<400> 27

atgggccatc atcatcatca tcatatgagc cataccaccc cgtggaccaa cccgggcctg	60
gcggaaaact ttatgaacag ctttatgcag ggcctgagca gcatgccggg ctttaccgcg	120
agccagctgg atgatatgag caccattgcg cagagcatgg tgcagagcat tcagagcctg	180
gcggcgcagg gccgtaccag cccgaacaaa ctgcaggcgc tgaacatggc gtttgcgagc	240
agcatggcgg aaattgcggc gagcgaagaa ggcggcgcca gcctgagcac caaaaccagc	300

```

- - - - -
agcattgcga ggcgatgag caacgcgttt ctgcagacca ccggcgtggt gaaccagccg      360
tttattaacg aaattaccca gctggtgagc atgtttgcgc aggcgggcat gaacgatgtg      420
agecgatga ttccgagcag cccggtgcat ctgaaacgcc tgaaactgct gctgctgctg      480
ctgctgctga ttctgctgct gattctgggc gcgctgctgc tgggcctg                    528

```

<210> 28

<211> 309

<212> PRT

<213> Artificial Sequence

<220>

<223> Fusion protein

<400> 28

```

Met Gly His His His His His His Met Ser His Thr Thr Pro Trp Thr
1          5          10          15

Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu
20          25          30

Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr
35          40          45

Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly
50          55          60

Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser
65          70          75          80

Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser
85          90          95

Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln
100         105         110

Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu
115         120         125

Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Gly Gly Gly Thr Pro
130         135         140

Trp Thr Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln
145         150         155         160

Gly Leu Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met
165         170         175

Ser Thr Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala
180         185         190

Gln Gly Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe
195         200         205

```

Ala Ser Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser
210 215 220

Leu Ser Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe
225 230 235 240

Leu Gln Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr
245 250 255

Gln Leu Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Val Ser Ala
260 265 270

Gly Asn Ser Met Ile Pro Ser Ser Pro Val His Leu Lys Arg Leu Lys
275 280 285

Leu Leu Leu Leu Leu Leu Leu Leu Ile Leu Leu Leu Ile Leu Gly Ala
290 295 300

Leu Leu Leu Gly Leu
305

<210> 29

<211> 927

<212> DNA

<213> Artificial Sequence

<220>

<223> Fusion protein

<400> 29

atgggccatc atcatcatca tcatatgagc cataccaccc cgtggaccaa cccgggcctg	60
gcggaaaact ttatgaacag ctttatgcag ggcctgagca gcatgccggg ctttaccgcg	120
agccagctgg atgatatgag caccattgcg cagagcatgg tgcagagcat tcagagcctg	180
gcggcgcagg gccgtaccag cccgaacaaa ctgcaggcgc tgaacatggc gtttgcgagc	240
agcatggcgg aaattgcggc gagcgaagaa ggcggcgcca gcctgagcac caaaaccagc	300
agcattgcga gcgcgatgag caacgcgttt ctgcagacca ccggcggtgt gaaccagccg	360
tttattaacg aaattaccca gctggtgagc atgtttgcgc aggcgggtat gaatgacggc	420
ggcggcaccc cgtggaccaa cccgggcctg gcggaact ttatgaacag ctttatgcag	480
ggcctgagca gcatgccggg ctttaccgcg agccagctgg atgatatgag caccattgcg	540
cagagcatgg tgcagagcat tcagagcctg gcggcgcagg gcgtaccag cccgaacaaa	600
ctgcaggcgc tgaacatggc gtttgcgagc agcatggcgg aaattgcggc gagcgaagaa	660
ggcggcgcca gcctgagcac caaaaccagc agcattgcga gcgcgatgag caacgcgttt	720
ctgcagacca ccggcggtgt gaaccagccg tttattaacg aaattaccca gctggtgagc	780
atgtttgcgc aggcgggcat gaacgatgtg agcgcgggga attctatgat tccgagcagc	840
ccggtgcac tgaaacgcct gaaactgctg ctgctgctgc tgctgctgat tetgctgctg	900
attctgggcg cgctgctgct gggcctg	927

<210> 30

<211> 313

<212> PRT

<213> Artificial Sequence

<220>

<223> Fusion protein

<400> 30

Met Gly His His His His His His Met Ser His Thr Thr Pro Trp Thr
1 5 10 15

Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu
20 25 30

Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr
35 40 45

Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly
50 55 60

Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser
65 70 75 80

Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser
85 90 95

Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln
100 105 110

Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu
115 120 125

Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Gly Gly Gly Thr Pro
130 135 140

Trp Thr Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln
145 150 155 160

Gly Leu Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met
165 170 175

Ser Thr Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala
180 185 190

Gln Gly Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe
195 200 205

Ala Ser Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser
210 215 220

Leu Ser Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe
225 230 235 240

Leu Gln Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr
245 250 255 260 265 270 275 280 285 290 295 300

245 250 255
 Gln Leu Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Val Ser Ala
 260 265 270
 Gly Asn Ser Met Leu Leu Gly Asp Phe Phe Arg Lys Ser Lys Glu Lys
 275 280 285
 Ile Gly Lys Glu Phe Lys Arg Ile Val Gln Arg Ile Lys Asp Phe Leu
 290 295 300
 Arg Asn Leu Val Pro Arg Thr Glu Ser
 305 310

<210> 31

<211> 939

<212> DNA

<213> Artificial Sequence

<220>

<223> Fusion protein

<400> 31

```

atggggccatc atcatcatca tcatatgagc cataccaccc cgtggaccaa cccgggcctg      60
gcggaaaact ttatgaacag ctttatgcag ggcctgagca gcatgccggg ctttaccgcg      120
agccagctgg atgatatgag caccattgcg cagagcatgg tgcagagcat tcagagcctg      180
gcggcgcagg gccgtaccag cccgaacaaa ctgcaggcgc tgaacatggc gtttgcgagc      240
agcatggcgg aaattgcggc gagcgaagaa ggcggcggca gcctgagcac caaaaccagc      300
agcattgcga gcgcgatgag caacgcgttt ctgcagacca ccggcgtggt gaaccagccg      360
tttattaacg aaattaccca gctggtgagc atgtttgcgc aggcoggtat gaatgacggc      420
ggcggcaccc cgtggaccaa cccgggcctg gcggaaaact ttatgaacag ctttatgcag      480
ggcctgagca gcatgccggg ctttaccgcg agccagctgg atgatatgag caccattgcg      540
cagagcatgg tgcagagcat tcagagcctg gcggcgcagg gccgtaccag cccgaacaaa      600
ctgcaggcgc tgaacatggc gtttgcgagc agcatggcgg aaattgcggc gagcgaagaa      660
ggcggcggca gcctgagcac caaaaccagc agcattgcga gcgcgatgag caacgcgttt      720
ctgcagacca ccggcgtggt gaaccagccg tttattaacg aaattaccca gctggtgagc      780
atgtttgcgc aggcgggcat gaacgatgtg agcgcgggga attctatgct gctgggtgat      840
ttcttccgca aatctaaaga gaagattggc aaagagttta aaagaattgt ccagagaatc      900
aaggattttt tgcggaatct tgtaccagc agagagtcc      939

```

<210> 32

<211> 158

<212> PRT

<213> Euprostenops australis

<400> 32

Met Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala Glu Asn Phe

Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly
 100 105 110
 Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met
 115 120 125
 Phe Ala Gln Ala Gly Met Asn Asp Gly Gly Gly Thr Pro Trp Thr Asn
 130 135 140
 Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser
 145 150 155 160
 Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile
 165 170 175
 Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg
 180 185 190
 Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser
 195 200 205
 Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr
 210 215 220
 Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr
 225 230 235 240
 Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val
 245 250 255
 Ser Met Phe Ala Gln Ala Gly Met Asn Asp Val Ser Ala Gly Tyr Gly
 260 265 270
 Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 275 280 285
 Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Ser Gly
 290 295 300
 Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ser Gly
 305 310 315 320
 Gln Gly Ser Gln Gly Gly Gln Gly Gly Gln Gly Gln Gly Gly Tyr Gly
 325 330 335
 Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 340 345 350
 Ala Ala Ala Ser Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Ala Gly
 355 360 365
 Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 370 375 380
 Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln Gly Gly Tyr
 385 390 395 400

Gly Gln Gly Ala Gly Ser Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala
 405 410 415
 Ala Ala Ala Gly Gly Gln Gly Gly Gln Gly Gln Gly Gly Tyr Gly Gln
 420 425 430
 Gly Ser Gly Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 435 440 445
 Ala Ala Ala Ala Gly Arg Gly Gln Gly Gly Tyr Gly Gln Gly Ser Gly
 450 455 460
 Gly Asn Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 465 470 475 480
 Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Arg Gln Ser Gln Gly Ala
 485 490 495
 Gly Ser Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala
 500 505 510
 Gly Ser Gly Gln Gly Gly Tyr Gly Gly Gln Gly Gln Gly Gly Tyr Gly
 515 520 525
 Gln Ser Ser Ala Ser Ala Ser Ala Ala Ala Ser Ala Ala Ser Thr Val
 530 535 540
 Ala Asn Ser Val Ser Arg Leu Ser Ser Pro Ser Ala Val Ser Arg Val
 545 550 555 560
 Ser Ser Ala Val Ser Ser Leu Val Ser Asn Gly Gln Val Asn Met Ala
 565 570 575
 Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser Ser Val Ser Ala Ser
 580 585 590
 Ala Pro Gly Ala Ser Gly Cys Glu Val Ile Val Gln Ala Leu Leu Glu
 595 600 605
 Val Ile Thr Ala Leu Val Gln Ile Val Ser Ser Ser Ser Val Gly Tyr
 610 615 620
 Ile Asn Pro Ser Ala Val Asn Gln Ile Thr Asn Val Val Ala Asn Ala
 625 630 635 640
 Met Ala Gln Val Met Gly
 645

<210> 34

<211> 387

<212> PRT

<213> Artificial Sequence

<220>

<223> Fusion protein

<400> 34

Met Gly His His His His His His Met Ser His Thr Thr Pro Trp Thr
1 5 10 15

Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu
20 25 30

Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr
35 40 45

Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly
50 55 60

Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser
65 70 75 80

Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser
85 90 95

Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln
100 105 110

Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu
115 120 125

Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Gly Gly Gly Thr Pro
130 135 140

Trp Thr Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln
145 150 155 160

Gly Leu Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met
165 170 175

Ser Thr Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala
180 185 190

Gln Gly Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe
195 200 205

Ala Ser Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser
210 215 220

Leu Ser Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe
225 230 235 240

Leu Gln Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr
245 250 255

Gln Leu Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Val Ser Ala
260 265 270

Gly Asn Ser Glu His Leu Val Thr Thr Ala Thr Phe Ser Ile Gly Ser
275 280 285

Thr Gly Leu Val Val Tyr Asp Tyr Gln Gln Leu Leu Ile Ala Tyr Lys
 290 295 300

Pro Ala Pro Gly Thr Cys Cys Tyr Ile Met Lys Ile Ala Pro Glu Ser
 305 310 315 320

Ile Pro Ser Leu Glu Ala Leu Thr Arg Lys Val His Asn Phe Gln Met
 325 330 335

Glu Cys Ser Leu Gln Ala Lys Pro Ala Val Pro Thr Ser Lys Leu Gly
 340 345 350

Gln Ala Glu Gly Arg Asp Ala Gly Ser Ala Pro Ser Gly Gly Asp Pro
 355 360 365

Ala Phe Leu Gly Met Ala Val Ser Thr Leu Cys Gly Glu Val Pro Leu
 370 375 380

Tyr Tyr Ile
 385

<210> 35

<211> 178

<212> PRT

<213> Artificial Sequence

<220>

<223> Fusion protein

<400> 35

Met Gly His His His His His His Met Ala Ser Ser Asp Lys Ile Ile
 1 5 10 15

His Leu Thr Asp Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly
 20 25 30

Ala Ile Leu Val Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met
 35 40 45

Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp Glu Tyr Gln Gly Lys Leu
 50 55 60

Thr Val Ala Lys Leu Asn Ile Asp Gln Asn Pro Gly Thr Ala Pro Lys
 65 70 75 80

Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu
 85 90 95

Val Ala Ala Thr Lys Val Gly Ala Leu Ser Lys Gly Gln Leu Lys Glu
 100 105 110

Phe Leu Asp Ala Asn Leu Ala Gly Ser Gly Ser Gly His Met His His
 115 120 125

His His His His Ser Ser Gly Leu Val Pro Arg Gly Ser Gly Asn Ser
130 135 140

Met Ile Pro Ser Ser Pro Val His Leu Lys Arg Leu Lys Leu Leu Leu
145 150 155 160

Leu Leu Leu Leu Leu Ile Leu Leu Leu Ile Leu Gly Ala Leu Leu Leu
165 170 175

Gly Leu

<210> 36

<211> 537

<212> DNA

<213> Artificial Sequence

<220>

<223> Fusion protein

<400> 36

atgggccatc atcatcatca tcatatggct agcagcgata aaattattca cctgactgac	60
gacagttttg acacggatgt actcaaagcg gacggggcga tcctcgtcga tttctgggca	120
gagtgggtcg gtccgtgcaa aatgatcgcc ccgattctgg atgaaatcgc tgacgaatat	180
cagggcaaac tgaccgttgc aaaactgaac atcgatcaaa accctggcac tgcgccgaaa	240
tatggcatcc gtggtatccc gactctgctg ctgttcaaaa acggtgaagt ggcggaacc	300
aaagtgggtg cactgtctaa aggtcagttg aaagagttcc tcgacgctaa cctggccggt	360
tctggttctg gccatatgca ccatcatcat catcattctt ctggtctggt gccacgcggt	420
tctgggaatt ctatgattcc gagcagcccg gtgcatctga aacgcctgaa actgctgctg	480
ctgctgctgc tgctgattct gctgctgatt ctgggcgcgc tgctgctggg cctgtaa	537

<210> 37

<211> 313

<212> PRT

<213> Artificial Sequence

<220>

<223> Fusion protein

<400> 37

Met Gly His His His His His His Met Ala Ser Ser Asp Lys Ile Ile
1 5 10 15

His Leu Thr Asp Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly
20 25 30

Ala Ile Leu Val Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met
35 40 45

Ile Ala Pro Ile Leu Asp Glu Ile Ala Asp Glu Trp Gln Glv Lys Leu

```

-----
50          55          60          65          70          75          80          85          90          95          100          105          110          115          120          125          130          135          140          145          150          155          160
Thr Val Ala Lys Leu Asn Ile Asp Gln Asn Pro Gly Thr Ala Pro Lys
65          70          75          80

Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu Leu Phe Lys Asn Gly Glu
85          90          95

Val Ala Ala Thr Lys Val Gly Ala Leu Ser Lys Gly Gln Leu Lys Glu
100          105          110

Phe Leu Asp Ala Asn Leu Ala Gly Ser Gly Ser Gly His Met His His
115          120          125

His His His His Ser Ser Gly Leu Val Pro Arg Gly Ser Gly Asn Ser
130          135          140

His Met Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala Glu Asn
145          150          155          160

Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly Phe Thr
165          170          175

Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met Val Gln
180          185          190

Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn Lys Leu
195          200          205

Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala Ala
210          215          220

Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser Ile Ala
225          230          235          240

Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val Asn Gln
245          250          255

Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala Gln Ala
260          265          270

Gly Met Asn Asp Val Ser Ala Met Ile Pro Ser Ser Pro Val His Leu
275          280          285

Lys Arg Leu Lys Leu Leu Leu Leu Leu Leu Leu Ile Leu Leu Leu
290          295          300

Ile Leu Gly Ala Leu Leu Leu Gly Leu
305          310

```

<210> 38

<211> 939

<212> DNA

<213> Artificial Sequence

<220>

<223> Fusion protein

<400> 38

```

atgggccatc atcatcatca tcatatggct agcagcgata aaattattca cctgactgac      60
gacagttttg acacggatgt actcaaagcg gacggggcga tcctcgtcga tttctgggca     120
gagtgggtgcg gtccgtgcaa aatgatcgcc ccgattctgg atgaaatcgc tgacgaatat     180
cagggcaaac tgaccgttgc aaaactgaac atcgatcaaa accctggcac tgcgccgaaa     240
tatggcatcc gtggtatccc gactctgctg ctgttcaaaa acggtgaagt ggcggaacc      300
aaagtgggtg cactgtctaa aggtcagttg aaagagttcc tcgacgctaa cctggccggt     360

tctggttctg gccatatgca ccatcatcat catcattctt ctggtctggt gccacgcggt     420
tctgggaatt cccatatgag ccataccacc ccgtggacca acccgggcct ggcgaaaac      480
tttatgaaca gctttatgca gggcctgagc agcatgccgg gctttaccgc gagccagctg     540
gatgatatga gcaccattgc gcagagcatg gtgcagagca ttcagagcct ggcggcgcag     600
ggcgcacca gccgaacaa actgcaggcg ctgaacatgg cgtttgcgag cagcatggcg     660
gaaattgcgg cgagcgaaga aggcggcggc agcctgagca ccaaaccag cagcattgcg     720
agcgcgatga gcaacgcgtt tctgcagacc accggcgtgg tgaaccagcc gtttattaac     780
gaaattaccc agctggtgag catgtttgcg caggcgggca tgaacgatgt gagcgcgatg     840
attccgagca gcccggtgca tctgaaacgc ctgaaattgc tgctgctgtt actgctgctg     900
attctgcttc tgattctggg cgcgctgctg ctgggcctg      939

```

<210> 39

<211> 410

<212> PRT

<213> Artificial Sequence

<220>

<223> Fusion protein

<400> 39

```

Met Gly His His His His His Gly Gly Gly Gly Ser Gly Gly Gly
1          5          10         15

Gly Ser His His His His His Met Ser His Thr Thr Pro Trp Thr
20          25          30

Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu
35          40          45

Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr
50          55          60

Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly
65          70          75          80

```

Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser
 85 90 95

Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser
 100 105 110

Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln
 115 120 125

Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu
 130 135 140

Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Gly Gly Gly Thr Pro
 145 150 155 160

Trp Thr Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln
 165 170 175

Gly Leu Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met
 180 185 190

Ser Thr Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala
 195 200 205

Gln Gly Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe
 210 215 220

Ala Ser Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser
 225 230 235 240

Leu Ser Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe
 245 250 255

Leu Gln Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr
 260 265 270

Gln Leu Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Val Ser Ala
 275 280 285

Leu Glu Ala Leu Phe Gln Gly Pro Asn Ser Glu His Leu Val Thr Thr
 290 295 300

Ala Thr Phe Ser Ile Gly Ser Thr Gly Leu Val Val Tyr Asp Tyr Gln
 305 310 315 320

Gln Leu Leu Ile Ala Tyr Lys Pro Ala Pro Gly Thr Cys Cys Tyr Ile
 325 330 335

Met Lys Ile Ala Pro Glu Ser Ile Pro Ser Leu Glu Ala Leu Thr Arg
 340 345 350

Lys Val His Asn Phe Gln Met Glu Cys Ser Leu Gln Ala Lys Pro Ala
 355 360 365

Val Pro Thr Ser Lys Leu Gly Gln Ala Glu Gly Arg Asp Ala Gly Ser
 370 375 380

Ala Pro Ser Gly Gly Asp Pro Ala Phe Leu Gly Met Ala Val Ser Thr
 385 390 395 400

Leu Cys Gly Glu Val Pro Leu Tyr Tyr Ile
 405 410

<210> 40

<211> 1233

<212> DNA

<213> Artificial Sequence

<220>

<223> Fusion protein

<400> 40

```

atgggccatc atcatcatca tcatggcggc ggtggtagcg gtggcgggtg cagccatcat      60
catcatcatc atatgagcca taccaccccg tggaccaacc cgggcctggc ggaaaacttt      120
atgaacagct ttatgcaggg cctgagcagc atgccgggct ttaccgcgag ccagctggat      180
gatatgagca ccattgcgca gagcatggtg cagagcattc agagcctggc gggcagggc      240
cgtaccagcc cgaacaaact gcaggcgctg aacatggcgt ttgcgagcag catggcggaa      300
attgcggcga gcgaagaagg cggcggcagc ctgagcacca aaaccagcag cattgcgagc      360
gcgatgagca acgcgtttct gcagaccacc ggcgtggtga accagccgtt tattaacgaa      420
attaccagc tggtagcat gtttgcgag gcgggcatga acgatggcg cggcaccccg      480
tggaccaacc cgggcctggc ggaaaacttt atgaacagct ttatgcaggg cctgagcagc      540
atgccgggct ttaccgcgag ccagctggat gatatgagca ccattgcgca gagcatggtg      600
cagagcattc agagcctggc ggcgcagggc cgtaccagcc cgaacaaact gcaggcgctg      660
aacatggcgt ttgcgagcag catggcggaa attgcggcga gcgaagaagg cggcggcagc      720
ctgagcacca aaaccagcag cattgcgagc gcgatgagca acgcgtttct gcagaccacc      780
ggcgtggtga accagccgtt tattaacgaa attaccagc tggtagcat gtttgcgag      840
gcgggcatga acgatgtgag gcgcctggaa gcgctgttcc agggcccgaa ttcagagcac      900
ctggttacca ctgccacctt ctccatcggc tccactggcc tcgtggtgta tgactaccag      960
cagctgctga tcgcctacaa gccagccctt ggcacctgct gctacatcat gaagatagct     1020
ccagagagca tcccagctct tgaggctctc actagaaaag tccacaactt ccagatggaa     1080
tgctctctgc aggccaaagg ccgagtgctt acgtctaagc tgggccaggc agaggggcga     1140
gatgcaggct cagcacccct cggaggggac ccggccttcc tgggcatggc cgtgagcacc     1200
ctgtgtggcg aggtgccgct ctactacatc tag                                     1233

```

<210> 41

<211> 116

<212> PRT

<213> Homo sapiens

<400> 41

Gly Pro Asn Ser Glu His Leu Val Thr Thr Ala Thr Phe Ser Ile Gly
1 5 10 15

Ser Thr Gly Leu Val Val Tyr Asp Tyr Gln Gln Leu Leu Ile Ala Tyr
20 25 30

Lys Pro Ala Pro Gly Thr Cys Cys Tyr Ile Met Lys Ile Ala Pro Glu
35 40 45

Ser Ile Pro Ser Leu Glu Ala Leu Thr Arg Lys Val His Asn Phe Gln
50 55 60

Met Glu Cys Ser Leu Gln Ala Lys Pro Ala Val Pro Thr Ser Lys Leu
65 70 75 80

Gly Gln Ala Glu Gly Arg Asp Ala Gly Ser Ala Pro Ser Gly Gly Asp
85 90 95

Pro Ala Phe Leu Gly Met Ala Val Ser Thr Leu Cys Gly Glu Val Pro
100 105 110

Leu Tyr Tyr Ile
115

<210> 42

<211> 537

<212> PRT

<213> Artificial Sequence

<220>

<223> Fusion protein

<400> 42

Met Gly His His His His His His Gly Gly Gly Gly Ser Gly Gly Gly
1 5 10 15

Gly Ser His His His His His His Met Ser His Thr Thr Pro Trp Thr
20 25 30

Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu
35 40 45

Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr
50 55 60

Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly
65 70 75 80

Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser
85 90 95

Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser
100 105 110

Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln
115 120 125

```

110
120
130
Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu
130 135 140

Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Gly Gly Gly Thr Pro
145 150 155 160

Trp Thr Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln
165 170 175

Gly Leu Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met
180 185 190

Ser Thr Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala
195 200 205

Gln Gly Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe
210 215 220

Ala Ser Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser
225 230 235 240

Leu Ser Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe
245 250 255

Leu Gln Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr
260 265 270

Gln Leu Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Val Ser Ala
275 280 285

Leu Glu Ala Leu Phe Gln Gly Pro Asn Ser Lys Gly Glu Glu Leu Phe
290 295 300

Thr Gly Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly
305 310 315 320

His Lys Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly
325 330 335

Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro
340 345 350

Trp Pro Thr Leu Val Thr Thr Leu Thr Tyr Gly Val Gln Cys Phe Ser
355 360 365

Arg Tyr Pro Asp His Met Lys Gln His Asp Phe Phe Lys Ser Ala Met
370 375 380

Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly
385 390 395 400

Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val
405 410 415

Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile
---
```

420

425

430

Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile
 435 440 445

Met Ala Asp Lys Gln Lys Asn Gly Ile Lys Val Asn Phe Lys Ile Arg
 450 455 460

His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln Gln
 465 470 475 480

Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr
 485 490 495

Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp
 500 505 510

His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr Leu Gly
 515 520 525

Met Asp Glu Leu Tyr Lys Leu Ile Asn
 530 535

<210> 43

<211> 1620

<212> DNA

<213> Artificial Sequence

<220>

<223> Fusion protein

<400> 43

atgggccatc atcatcatca tcatggcggc ggtggtagcg gtggcggtgg cagccatcat	60
catcatcatc atatgagcca taccaccccg tggaccaacc cgggcctggc ggaaaacttt	120
atgaacagct ttatgcaggg cctgagcagc atgccgggct ttaccgcgag ccagctggat	180
gatatgagca ccattgcgca gagcatggtg cagagcattc agagcctggc ggcgcagggc	240
cgtaccagcc cgaacaaact gcaggcgctg aacatggcgt ttgcgagcag catggcggaa	300
attgcggcga gcgaagaagg cggcggcagc ctgagcacca aaaccagcag cattgcgagc	360
gcgatgagca acgcgtttct gcagaccacc ggcgtggtga accagccgtt tattaacgaa	420
attaccagc tggtagcat gtttgcgcag gcgggcatga acgatggcgg cggcaccccg	480
tggaccaacc cgggcctggc ggaaaacttt atgaacagct ttatgcaggg cctgagcagc	540
atgccgggct ttaccgcgag ccagctggat gatatgagca ccattgcgca gagcatggtg	600
cagagcattc agagcctggc ggcgcagggc cgtaccagcc cgaacaaact gcaggcgctg	660
aacatggcgt ttgcgagcag catggcggaa attgcggcga gcgaagaagg cggcggcagc	720
ctgagcacca aaaccagcag cattgcgagc gcgatgagca acgcgtttct gcagaccacc	780
ggcgtggtga accagccgtt tattaacgaa attaccagc tggtagcat gtttgcgcag	840
gcgggcatga acgatgtgag gcgcgtggaa gcgctgttcc agggcccga ttcaaagggc	900

gaggagctgt tcaccgggggt ggtgcccattc ctggctcgagc tggacggcga cgtaaaccgc 960
 cacaagttca gcgtgtccgg cgagggcgag ggcgatgcca cctacggcaa gctgaccctg 1020
 aagttcatct gcaccaccgg caagctgccc gtgccctggc ccaccctcgt gaccaccctg 1080
 acctacggcg tgcagtgtt cagccgtac cccgaccaca tgaagcagca cgacttcttc 1140
 aagtcgccca tgcccgaagg ctacgtccag gagcgaccca tcttcttcaa ggacgacggc 1200
 aactacaaga cccgcgccga ggtgaagttc gagggcgaca ccctggtgaa ccgcatcgag 1260
 ctgaagggca tcgacttcaa ggaggacggc aacatcctgg ggacaaagct ggagtacaac 1320
 tacaacagcc acaacgtcta tatcatggcc gacaagcaga agaacggcat caaggtgaac 1380
 ttcaagatcc gccacaacat cgaggacggc agcgtgcagc tcgccgacca ctaccagcag 1440
 aacacccccca tcggcgacgg ccccgctgct ctgcccgcga accactacct gagcaccagg 1500
 tccgccctga gcaaagaccc caacgagaag cgcgatcaca tggctctgct ggagttcgtg 1560
 accgccgcgc ggatcactct cggcatggac gagctgtaca agttaattaa ctaatgataa 1620

<210> 44

<211> 243

<212> PRT

<213> Aequorea victoria

<400> 44

Gly Pro Asn Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile
 1 5 10 15

Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser
 20 25 30

Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe
 35 40 45

Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr
 50 55 60

Thr Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met
 65 70 75 80

Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln
 85 90 95

Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala
 100 105 110

Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys
 115 120 125

Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu
 130 135 140

Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys
 145 150 155 160

Asn Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly
165 170 175

Ser Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp
180 185 190

Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala
195 200 205

Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu
210 215 220

Phe Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys
225 230 235 240

Leu Ile Asn

<210> 45

<211> 307

<212> PRT

<213> Artificial Sequence

<220>

<223> Fusion protein

<400> 45

Val Asp Asn Lys Phe Asn Lys Glu Gln Gln Asn Ala Phe Tyr Glu Ile
1 5 10 15

Leu His Leu Pro Asn Leu Asn Glu Glu Gln Arg Asn Ala Phe Ile Gln
20 25 30

Ser Leu Lys Asp Asp Pro Ser Gln Ser Ala Asn Leu Leu Ala Glu Ala
35 40 45

Lys Lys Leu Asn Asp Ala Gln Ala Pro Lys Leu Glu Ala Leu Phe Gln
50 55 60

Gly Pro Asn Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile
65 70 75 80

Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser
85 90 95

Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe
100 105 110

Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr
115 120 125

Thr Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met
130 135 140

Lys Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln
145 150 155 160

Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala
 165 170 175

Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys
 180 185 190

Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu
 195 200 205

Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys
 210 215 220

Asn Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly
 225 230 235 240

Ser Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp
 245 250 255

Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala
 260 265 270

Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu
 275 280 285

Phe Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys
 290 295 300

Leu Ile Asn
 305

<210> 46

<211> 368

<212> PRT

<213> Artificial Sequence

<220>

<223> Fusion protein

<400> 46

His His His His His His Leu Ala Glu Ala Lys Val Leu Ala Asn Arg
 1 5 10 15

Glu Leu Asp Lys Tyr Gly Val Ser Asp Tyr His Lys Asn Leu Ile Asn
 20 25 30

Asn Ala Lys Thr Val Glu Gly Val Lys Asp Leu Gln Ala Gln Val Val
 35 40 45

Glu Ser Ala Lys Lys Ala Arg Ile Ser Glu Ala Thr Asp Gly Leu Ser
 50 55 60

Asp Phe Leu Lys Ser Gln Thr Pro Ala Glu Asp Thr Val Lys Ser Ile

65	70	75	80
Glu Leu Ala Glu Ala Lys Val Leu Ala Asn Arg Glu Leu Asp Lys Tyr	85	90	95
Gly Val Ser Asp Tyr Tyr Lys Asn Leu Ile Asn Asn Ala Lys Thr Val	100	105	110
Glu Gly Val Lys Ala Leu Ile Asp Glu Ile Leu Ala Ala Leu Pro Asn	115	120	125
Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu Val Glu	130	135	140
Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly Glu Gly	145	150	155
Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile Cys Thr	165	170	175
Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr Leu Thr	180	185	190
Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys Gln His	195	200	205
Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu Arg Thr	210	215	220
Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu Val Lys	225	230	235
Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly Ile Asp	245	250	255
Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr Asn Tyr	260	265	270
Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn Gly Ile	275	280	285
Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser Val Gln	290	295	300
Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly Pro Val	305	310	315
Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu Ser Lys	325	330	335
Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe Val Thr	340	345	350
Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys Leu Ile Asn	355	360	365

<210> 47

<211> 692

<212> PRT

<213> Artificial Sequence

<220>

<223> Fusion protein

<400> 47

Met Gly His His His His His His Gly Gly Gly Gly Ser Gly Gly Gly
 1 5 10 15

Gly Ser His His His His His His Met Ser His Thr Thr Pro Trp Thr
 20 25 30

Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu
 35 40 45

Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr
 50 55 60

Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly
 65 70 75 80

Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser
 85 90 95

Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser
 100 105 110

Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln
 115 120 125

Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu
 130 135 140

Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Gly Gly Gly Thr Pro
 145 150 155 160

Trp Thr Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln
 165 170 175

Gly Leu Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met
 180 185 190

Ser Thr Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala
 195 200 205

Gln Gly Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe
 210 215 220

Ala Ser Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser
 225 230 235 240

Leu Ser Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe
245 250 255

Leu Gln Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr
260 265 270

Gln Leu Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp Val Ser Ala
275 280 285

Leu Glu Ala Leu Phe Gln Gly Pro Asn Ser Thr Gly Ala Thr Ser Pro
290 295 300

Glu Glu Ala Ile Ala Asp Leu Ser Val Asn Met Tyr Asn Arg Leu Arg
305 310 315 320

Ala Thr Gly Glu Asp Glu Asn Ile Leu Phe Ser Pro Leu Ser Ile Ala
325 330 335

Leu Ala Met Gly Met Met Glu Leu Gly Ala Gln Gly Ser Thr Gln Lys
340 345 350

Glu Ile Arg His Ser Met Gly Tyr Asp Ser Leu Lys Asn Gly Glu Glu
355 360 365

Phe Ser Phe Leu Lys Glu Phe Ser Asn Met Val Thr Ala Lys Glu Ser
370 375 380

Gln Tyr Val Met Lys Ile Ala Asn Ser Leu Phe Val Gln Asn Gly Phe
385 390 395 400

His Val Asn Glu Glu Phe Leu Gln Met Met Lys Lys Tyr Phe Asn Ala
405 410 415

Ala Val Asn His Val Asp Phe Ser Gln Asn Val Ala Val Ala Asn Tyr
420 425 430

Ile Asn Lys Trp Val Glu Asn Asn Thr Asn Asn Leu Val Lys Asp Leu
435 440 445

Val Ser Pro Arg Asp Phe Asp Ala Ala Thr Tyr Leu Ala Leu Ile Asn
450 455 460

Ala Val Tyr Phe Lys Gly Asn Trp Lys Ser Gln Phe Arg Pro Glu Asn
465 470 475 480

Thr Arg Thr Phe Ser Phe Thr Lys Asp Asp Glu Ser Glu Val Gln Ile
485 490 495

Pro Met Met Tyr Gln Gln Gly Glu Phe Tyr Tyr Gly Glu Phe Ser Asp
500 505 510

Gly Ser Asn Glu Ala Gly Gly Ile Tyr Gln Val Leu Glu Ile Pro Tyr
515 520 525

Glu Gly Asp Glu Ile Ser Met Met Leu Val Leu Ser Arg Gln Glu Val

530 535 540
 Pro Leu Ala Thr Leu Glu Pro Leu Val Lys Ala Gln Leu Val Glu Glu
 545 550 555 560
 Trp Ala Asn Ser Val Lys Lys Gln Lys Val Glu Val Tyr Leu Pro Arg
 565 570 575
 Phe Thr Val Glu Gln Glu Ile Asp Leu Lys Asp Val Leu Lys Ala Leu
 580 585 590
 Gly Ile Thr Glu Ile Phe Ile Lys Asp Ala Asn Leu Thr Gly Leu Ser
 595 600 605
 Asp Asn Lys Glu Ile Phe Leu Ser Lys Ala Ile His Lys Ser Phe Leu
 610 615 620
 Glu Val Asn Glu Glu Gly Ser Glu Ala Ala Ala Val Ser Gly Met Ile
 625 630 635 640
 Ala Ile Ser Arg Met Ala Val Leu Tyr Pro Gln Val Ile Val Asp His
 645 650 655
 Pro Phe Phe Phe Leu Ile Arg Asn Arg Arg Thr Gly Thr Ile Leu Phe
 660 665 670
 Met Gly Arg Val Met His Pro Glu Thr Met Asn Thr Ser Gly His Asp
 675 680 685
 Phe Glu Glu Leu
 690

<210> 48

<211> 2085

<212> DNA

<213> Artificial Sequence

<220>

<223> Fusion protein

<400> 48

```

atgggccatc atcatcatca tcatggcggc ggtggtagcg gtggcggtgg cagccatcat      60
catcatcatc atatgagcca taccaccccg tggaccaacc cgggcctggc ggaaaacttt      120
atgaacagct ttatgcaggg cctgagcagc atgccgggct ttaccgcgag ccagctggat      180
gatatgagca ccattgcgca gagcatggtg cagagcattc agagcctggc ggcgcagggc      240
cgtaccagcc cgaacaaact gcaggcgctg aacatggcgt ttgcgagcag catggcggaa      300
attgcggcga gcgaagaagg cgggcggcagc ctgagcacca aaaccagcag cattgcgagc      360
gcgatgagca acgcgtttct gcagaccacc ggcgtggtga accagccgtt tattaacgaa      420
attaccagc  tggtagcat  gtttgcgcag gcgggcatga acgatggcgg cggcaccccg      480
tggaccaacc cgggcctggc ggaaaacttt atgaacagct ttatgcaggg cctgagcagc      540
atcccaact  ttaccacaa  caactaat  aatatgaca ccattacaa  aacataata      600

```

Asn Ile Leu Phe Ser Pro Leu Ser Ile Ala Leu Ala Met Gly Met Met
35 40 45

Glu Leu Gly Ala Gln Gly Ser Thr Gln Lys Glu Ile Arg His Ser Met
 50 55 60
 Gly Tyr Asp Ser Leu Lys Asn Gly Glu Glu Phe Ser Phe Leu Lys Glu
 65 70 75 80
 Phe Ser Asn Met Val Thr Ala Lys Glu Ser Gln Tyr Val Met Lys Ile
 85 90 95
 Ala Asn Ser Leu Phe Val Gln Asn Gly Phe His Val Asn Glu Glu Phe
 100 105 110
 Leu Gln Met Met Lys Lys Tyr Phe Asn Ala Ala Val Asn His Val Asp
 115 120 125
 Phe Ser Gln Asn Val Ala Val Ala Asn Tyr Ile Asn Lys Trp Val Glu
 130 135 140
 Asn Asn Thr Asn Asn Leu Val Lys Asp Leu Val Ser Pro Arg Asp Phe
 145 150 155 160
 Asp Ala Ala Thr Tyr Leu Ala Leu Ile Asn Ala Val Tyr Phe Lys Gly
 165 170 175
 Asn Trp Lys Ser Gln Phe Arg Pro Glu Asn Thr Arg Thr Phe Ser Phe
 180 185 190
 Thr Lys Asp Asp Glu Ser Glu Val Gln Ile Pro Met Met Tyr Gln Gln
 195 200 205
 Gly Glu Phe Tyr Tyr Gly Glu Phe Ser Asp Gly Ser Asn Glu Ala Gly
 210 215 220
 Gly Ile Tyr Gln Val Leu Glu Ile Pro Tyr Glu Gly Asp Glu Ile Ser
 225 230 235 240
 Met Met Leu Val Leu Ser Arg Gln Glu Val Pro Leu Ala Thr Leu Glu
 245 250 255
 Pro Leu Val Lys Ala Gln Leu Val Glu Glu Trp Ala Asn Ser Val Lys
 260 265 270
 Lys Gln Lys Val Glu Val Tyr Leu Pro Arg Phe Thr Val Glu Gln Glu
 275 280 285
 Ile Asp Leu Lys Asp Val Leu Lys Ala Leu Gly Ile Thr Glu Ile Phe
 290 295 300
 Ile Lys Asp Ala Asn Leu Thr Gly Leu Ser Asp Asn Lys Glu Ile Phe
 305 310 315 320
 Leu Ser Lys Ala Ile His Lys Ser Phe Leu Glu Val Asn Glu Glu Gly
 325 330 335
 Ser Glu Ala Ala Ala Val Ser Gly Met Ile Ala Ile Ser Arg Met Ala
 340 345 350

Val Leu Tyr Pro Gln Val Ile Val Asp His Pro Phe Phe Phe Leu Ile
 355 360 365

Arg Asn Arg Arg Thr Gly Thr Ile Leu Phe Met Gly Arg Val Met His
 370 375 380

Pro Glu Thr Met Asn Thr Ser Gly His Asp Phe Glu Glu Leu
 385 390 395

<210> 50

<211> 131

<212> PRT

<213> Euprosthénops australis

<400> 50

Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala Glu Asn Phe Met
 1 5 10 15

Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly Phe Thr Ala Ser
 20 25 30

Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met Val Gln Ser Ile
 35 40 45

Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn Lys Leu Gln Ala
 50 55 60

Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala Ala Ser Glu
 65 70 75 80

Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser Ile Ala Ser Ala
 85 90 95

Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val Asn Gln Pro Phe
 100 105 110

Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala Gln Ala Gly Met
 115 120 125

Asn Asp Val
 130

<210> 51

<211> 127

<212> PRT

<213> Latrodectus geometricus

<400> 51

Gln Ala Asn Thr Pro Trp Ser Ser Lys Gln Asn Ala Asp Ala Phe Ile
 1 5 10 15

Ser Ala Phe Met Thr Ala Ala Ser Gln Ser Gly Ala Phe Ser Ser Asp
 20 25 30

Gln Ile Asn Asn Met Ser Val Ile Ser Asn Thr Leu Met Ala Ala Met

Gln Ile Asp Asp Met Ser Val Ile Ser Asn Thr Leu Met Ala Ala Met
35 40 45

Asp Asn Met Gly Gly Arg Ile Thr Pro Ser Lys Leu Gln Ala Leu Asp
50 55 60

Met Ala Phe Ala Ser Ser Val Ala Glu Ile Ala Ala Val Glu Gly Gln
65 70 75 80

Asn Ile Gly Val Thr Thr Asn Ala Ile Ser Asp Ala Leu Thr Ser Ala
85 90 95

Phe Tyr Gln Thr Thr Gly Val Val Asn Asn Lys Phe Ile Ser Glu Ile
100 105 110

Arg Ser Leu Ile Asn Met Phe Ala Gln Ala Ser Ala Asn Asp Val
115 120 125

<210> 52

<211> 127

<212> PRT

<213> Latrodectus hesperus

<400> 52

Gln Ala Asn Thr Pro Trp Ser Ser Lys Ala Asn Ala Asp Ala Phe Ile
1 5 10 15

Asn Ser Phe Ile Ser Ala Ala Ser Asn Thr Gly Ser Phe Ser Gln Asp
20 25 30

Gln Met Glu Asp Met Ser Leu Ile Gly Asn Thr Leu Met Ala Ala Met
35 40 45

Asp Asn Met Gly Gly Arg Ile Thr Pro Ser Lys Leu Gln Ala Leu Asp
50 55 60

Met Ala Phe Ala Ser Ser Val Ala Glu Ile Ala Ala Ser Glu Gly Gly
65 70 75 80

Asp Leu Gly Val Thr Thr Asn Ala Ile Ala Asp Ala Leu Thr Ser Ala
85 90 95

Phe Tyr Gln Thr Thr Gly Val Val Asn Ser Arg Phe Ile Ser Glu Ile
100 105 110

Arg Ser Leu Ile Gly Met Phe Ala Gln Ala Ser Ala Asn Asp Val
115 120 125

<210> 53

<211> 130

<212> PRT

<213> Nephila clavipes

<400> 53

Gln Asn Thr Pro Trp Ser Ser Thr Glu Leu Ala Asp Ala Phe Ile Asn
1 5 10 15

Ala Phe Met Asn Glu Ala Gly Arg Thr Gly Ala Phe Thr Ala Asp Gln
20 25 30

Leu Asp Asp Met Ser Thr Ile Gly Asp Thr Ile Lys Thr Ala Met Asp
35 40 45

Lys Met Ala Arg Ser Asn Lys Ser Ser Lys Gly Lys Leu Gln Ala Leu
50 55 60

Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala Ala Val Glu Gln
65 70 75 80

Gly Gly Leu Ser Val Asp Ala Lys Thr Asn Ala Ile Ala Asp Ser Leu
85 90 95

Asn Ser Ala Phe Tyr Gln Thr Thr Gly Ala Ala Asn Pro Gln Phe Val
100 105 110

Asn Glu Ile Arg Ser Leu Ile Asn Met Phe Ala Gln Ser Ser Ala Asn
115 120 125

Glu Val
130

<210> 54

<211> 131

<212> PRT

<213> Argiope trifasciata

<400> 54

Gln Gly Ala Thr Pro Trp Glu Asn Ser Gln Leu Ala Glu Ser Phe Ile
1 5 10 15

Ser Arg Phe Leu Arg Phe Ile Gly Gln Ser Gly Ala Phe Ser Pro Asn
20 25 30

Gln Leu Asp Asp Met Ser Ser Ile Gly Asp Thr Leu Lys Thr Ala Ile
35 40 45

Glu Lys Met Ala Gln Ser Arg Lys Ser Ser Lys Ser Lys Leu Gln Ala
50 55 60

Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala Val Ala Glu
65 70 75 80

Gln Gly Gly Leu Ser Leu Glu Ala Lys Thr Asn Ala Ile Ala Ser Ala
85 90 95

Leu Ser Ala Ala Phe Leu Glu Thr Thr Gly Tyr Val Asn Gln Gln Phe
100 105 110

Val Asn Glu Ile Lys Thr Leu Ile Phe Met Ile Ala Gln Ala Ser Ser
115 120 125

Asn Glu Ile
130

<210> 55

<211> 124

<212> PRT

<213> Latrodectus geometricus

<400> 55

Leu Arg Trp Ser Ser Lys Asp Asn Ala Asp Arg Phe Ile Asn Ala Phe
1 5 10 15

Leu Gln Ala Ala Ser Asn Ser Gly Ala Phe Ser Ser Asp Gln Val Asp
20 25 30

Asp Met Ser Val Ile Gly Asn Thr Leu Met Thr Ala Met Asp Asn Met
35 40 45

Gly Gly Arg Ile Thr Pro Ser Lys Leu Gln Ala Leu Asp Met Ala Phe
50 55 60

Ala Ser Ser Val Ala Glu Ile Ala Val Ala Asp Gly Gln Asn Val Gly
65 70 75 80

Gly Ala Thr Asn Ala Ile Ser Asn Ala Leu Arg Ser Ala Phe Tyr Gln
85 90 95

Thr Thr Gly Val Val Asn Asn Gln Phe Ile Ser Glu Ile Ser Asn Leu
100 105 110

Ile Asn Met Phe Ala Gln Val Ser Ala Asn Glu Val
115 120

<210> 56

<211> 127

<212> PRT

<213> Latrodectus hesperus

<400> 56

Gln Ala Asn Thr Pro Trp Ser Ser Lys Glu Asn Ala Asp Ala Phe Ile
1 5 10 15

Gly Ala Phe Met Asn Ala Ala Ser Gln Ser Gly Ala Phe Ser Ser Asp
20 25 30

Gln Ile Asp Asp Met Ser Val Ile Ser Asn Thr Leu Met Ala Ala Met
35 40 45

Asp Asn Met Gly Gly Arg Ile Thr Gln Ser Lys Leu Gln Ala Leu Asp
50 55 60

Met Ala Phe Ala Ser Ser Val Ala Glu Ile Ala Val Ala Asp Gly Gln
65 70 75 80

Asn Val Gly Ala Ala Thr Asn Ala Ile Ser Asp Ala Leu Arg Ser Ala
85 90 95

Phe Tyr Gln Thr Thr Gly Val Val Asn Asn Gln Phe Ile Thr Gly Ile
100 105 110

Ser Ser Leu Ile Gly Met Phe Ala Gln Val Ser Gly Asn Glu Val
115 120 125

<210> 57

<211> 131

<212> PRT

<213> *Nephila inaurata madagascariensis*

<400> 57

Gln Ala Asn Thr Pro Trp Ser Asp Thr Ala Thr Ala Asp Ala Phe Ile
1 5 10 15

Gln Asn Phe Leu Gly Ala Val Ser Gly Ser Gly Ala Phe Thr Pro Asp
20 25 30

Gln Leu Asp Asp Met Ser Thr Val Gly Asp Thr Ile Met Ser Ala Met
35 40 45

Asp Lys Met Ala Arg Ser Asn Lys Ser Ser Lys Ser Lys Leu Gln Ala
50 55 60

Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala Ala Val Glu
65 70 75 80

Gln Gly Gly Gln Ser Met Asp Val Lys Thr Asn Ala Ile Ala Asn Ala
85 90 95

Leu Asp Ser Ala Phe Tyr Met Thr Thr Gly Ser Thr Asn Gln Gln Phe
100 105 110

Val Asn Glu Met Arg Ser Leu Ile Asn Met Leu Ser Ala Ala Ala Val
115 120 125

Asn Glu Val
130

<210> 58

<211> 131

<212> PRT

<213> *Nephila clavipes*

<400> 58

Gln Ala Arg Ser Pro Trp Ser Asp Thr Ala Thr Ala Asp Ala Phe Ile
1 5 10 15

Gln Asn Phe Leu Ala Ala Val Ser Gly Ser Gly Ala Phe Thr Ser Asp
20 25 30

Gln Leu Asp Asp Met Ser Thr Ile Gly Asp Thr Ile Met Ser Ala Met
 35 40 45

Asp Lys Met Ala Arg Ser Asn Lys Ser Ser Gln His Lys Leu Gln Ala
 50 55 60

Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala Ala Val Glu
 65 70 75 80

Gln Gly Gly Met Ser Met Ala Val Lys Thr Asn Ala Ile Val Asp Gly
 85 90 95

Leu Asn Ser Ala Phe Tyr Met Thr Thr Gly Ala Ala Asn Pro Gln Phe
 100 105 110

Val Asn Glu Met Arg Ser Leu Ile Ser Met Ile Ser Ala Ala Ser Ala
 115 120 125

Asn Glu Val
 130

<210> 59

<211> 129

<212> PRT

<213> Argiope bruennichi

<400> 59

Ala Val Pro Ser Val Phe Ser Ser Pro Asn Leu Ala Ser Gly Phe Leu
 1 5 10 15

Gln Cys Leu Thr Phe Gly Ile Gly Asn Ser Pro Ala Phe Pro Thr Gln
 20 25 30

Glu Gln Gln Asp Leu Asp Ala Ile Ala Gln Val Ile Leu Asn Ala Val
 35 40 45

Ser Ser Asn Thr Gly Ala Thr Ala Ser Ala Arg Ala Gln Ala Leu Ser
 50 55 60

Thr Ala Leu Ala Ser Ser Leu Thr Asp Leu Leu Ile Ala Glu Ser Ala
 65 70 75 80

Glu Ser Asn Tyr Ser Asn Gln Leu Ser Glu Leu Thr Gly Ile Leu Ser
 85 90 95

Asp Cys Phe Ile Gln Thr Thr Gly Ser Asp Asn Pro Ala Phe Val Ser
 100 105 110

Arg Ile Gln Ser Leu Ile Ser Val Leu Ser Gln Asn Ala Asp Thr Asn
 115 120 125

Ile

<210> 60

<211> 129

<212> PRT

<213> *Nephila clavata*

<400> 60

Pro Val Pro Ser Val Phe Ser Ser Pro Ser Leu Ala Ser Gly Phe Leu
 1 5 10 15

Gly Cys Leu Thr Thr Gly Ile Gly Leu Ser Pro Ala Phe Pro Phe Gln
 20 25 30

Glu Gln Gln Asp Leu Asp Asp Leu Ala Lys Val Ile Leu Ser Ala Val
 35 40 45

Thr Ser Asn Thr Asp Thr Ser Lys Ser Ala Arg Ala Gln Ala Leu Ser
 50 55 60

Thr Ala Leu Ala Ser Ser Leu Ala Asp Leu Leu Ile Ser Glu Ser Ser
 65 70 75 80

Gly Ser Ser Tyr Gln Thr Gln Ile Ser Ala Leu Thr Asn Ile Leu Ser
 85 90 95

Asp Cys Phe Val Thr Thr Thr Gly Ser Asn Asn Pro Ala Phe Val Ser
 100 105 110

Arg Val Gln Thr Leu Ile Gly Val Leu Ser Gln Ser Ser Ser Asn Ala
 115 120 125

Ile

<210> 61

<211> 128

<212> PRT

<213> *Latrodectus hesperus*

<400> 61

Ala Ser Val Asn Ile Phe Asn Ser Pro Asn Ala Ala Thr Ser Phe Leu
 1 5 10 15

Asn Cys Leu Arg Ser Asn Ile Glu Ser Ser Pro Ala Phe Pro Phe Gln
 20 25 30

Glu Gln Ala Asp Leu Asp Ser Ile Ala Glu Val Ile Leu Ser Asp Val
 35 40 45

Ser Ser Val Asn Thr Ala Ser Ser Ala Thr Ser Leu Ala Leu Ser Thr
 50 55 60

Ala Leu Ala Ser Ser Leu Ala Glu Leu Leu Val Thr Glu Ser Ala Glu
 65 70 75 80

Glu Asp Ile Asp Asn Gln Val Val Ala Leu Ser Thr Ile Leu Ser Gln
 85 90 95

Cys Phe Val Glu Thr Thr Gly Ser Pro Asn Pro Ala Phe Val Ala Ser
 100 105 110

Val Lys Ser Leu Leu Gly Val Leu Ser Gln Ser Ala Ser Asn Tyr Glu
 115 120 125

<210> 62

<211> 130

<212> PRT

<213> *Nephila clavipes*

<400> 62

Ile Ala Asn Ser Pro Phe Ser Asn Pro Asn Thr Ala Glu Ala Phe Ala
 1 5 10 15

Arg Ser Phe Val Ser Asn Ile Val Ser Ser Gly Glu Phe Gly Ala Gln
 20 25 30

Gly Ala Glu Asp Phe Asp Asp Ile Ile Gln Ser Leu Ile Gln Ala Gln
 35 40 45

Ser Met Gly Lys Gly Arg His Asp Thr Lys Ala Lys Ala Lys Ala Met
 50 55 60

Gln Val Ala Leu Ala Ser Ser Ile Ala Glu Leu Val Ile Ala Glu Ser
 65 70 75 80

Ser Gly Gly Asp Val Gln Arg Lys Thr Asn Val Ile Ser Asn Ala Leu
 85 90 95

Arg Asn Ala Leu Met Ser Thr Thr Gly Ser Pro Asn Glu Glu Phe Val
 100 105 110

His Glu Val Gln Asp Leu Ile Gln Met Leu Ser Gln Glu Gln Ile Asn
 115 120 125

Glu Val
 130

<210> 63

<211> 130

<212> PRT

<213> *Nephila inaurata madagascariensis*

<400> 63

Ile Val Asn Ser Pro Phe Ser Asn Pro Asn Thr Ala Glu Ala Phe Ala
 1 5 10 15

Arg Ser Phe Val Ser Asn Val Val Ser Ser Gly Glu Phe Gly Ala Gln
 20 25 30

Gly Ala Glu Asp Phe Asp Asp Ile Ile Gln Ser Leu Ile Gln Ala Gln
 35 40 45

Ser Met Gly Lys Gly Arg His Asp Thr Lys Ala Lys Ala Lys Ala Met
50 55 60

Gln Val Ala Leu Ala Ser Ser Ile Ala Glu Leu Val Ile Ala Glu Ser
65 70 75 80

Ser Gly Gly Asp Val Gln Arg Lys Thr Asn Val Ile Ser Asn Ala Leu
85 90 95

Arg Asn Ala Leu Met Ser Thr Thr Gly Ser Pro Asn Glu Glu Phe Val
100 105 110

His Glu Val Gln Asp Leu Ile Gln Met Leu Ser Gln Glu Gln Ile Asn
115 120 125

Glu Val
130

<210> 64

<211> 18

<212> PRT

<213> Euprostenops australis

<400> 64

Ala Ser Ala Ser Ala Ala Ala Ser Ala Ala Ser Thr Val Ala Asn Ser
1 5 10 15

Val Ser

<210> 65

<211> 8

<212> PRT

<213> Euprostenops australis

<400> 65

Ala Ser Ala Ala Ser Ala Ala Ala
1 5

<210> 66

<211> 8

<212> PRT

<213> Euprostenops australis

<400> 66

Gly Ser Ala Met Gly Gln Gly Ser
1 5

<210> 67

<211> 5

<212> PRT

<213> Euprostenops australis

<400> 67

Ser Ala Ser Ala Glu

Ser Ala Ser Ala Gly
1 5

<210> 68

<211> 100

<212> PRT

<213> Euprosthénops sp

<400> 68

Ser Arg Leu Ser Ser Pro Glu Ala Ser Ser Arg Val Ser Ser Ala Val
1 5 10 15

Ser Asn Leu Val Ser Ser Gly Pro Thr Asn Ser Ala Ala Leu Ser Ser
20 25 30

Thr Ile Ser Asn Val Val Ser Gln Ile Gly Ala Ser Asn Pro Gly Leu
35 40 45

Ser Gly Cys Asp Val Leu Val Gln Ala Leu Leu Glu Val Val Ser Ala
50 55 60

Leu Ile His Ile Leu Gly Ser Ser Ser Ile Gly Gln Val Asn Tyr Gly
65 70 75 80

Ser Ala Gly Gln Ala Thr Gln Leu Val Gly Gln Ser Val Tyr Gln Ala
85 90 95

Leu Gly Glu Phe
100

<210> 69

<211> 98

<212> PRT

<213> Euprosthénops australis

<400> 69

Ser Arg Leu Ser Ser Pro Ser Ala Val Ser Arg Val Ser Ser Ala Val
1 5 10 15

Ser Ser Leu Val Ser Asn Gly Gln Val Asn Met Ala Ala Leu Pro Asn
20 25 30

Ile Ile Ser Asn Ile Ser Ser Ser Val Ser Ala Ser Ala Pro Gly Ala
35 40 45

Ser Gly Cys Glu Val Ile Val Gln Ala Leu Leu Glu Val Ile Thr Ala
50 55 60

Leu Val Gln Ile Val Ser Ser Ser Ser Val Gly Tyr Ile Asn Pro Ser
65 70 75 80

Ala Val Asn Gln Ile Thr Asn Val Val Ala Asn Ala Met Ala Gln Val
85 90 95

Met Gly

<210> 70

<211> 99

<212> PRT

<213> Argiope trifasciata

<400> 70

Ser Arg Leu Ser Ser Pro Gly Ala Ala Ser Arg Val Ser Ser Ala Val
1 5 10 15

Thr Ser Leu Val Ser Ser Gly Gly Pro Thr Asn Ser Ala Ala Leu Ser
20 25 30

Asn Thr Ile Ser Asn Val Val Ser Gln Ile Ser Ser Ser Asn Pro Gly
35 40 45

Leu Ser Gly Cys Asp Val Leu Val Gln Ala Leu Leu Glu Ile Val Ser
50 55 60

Ala Leu Val His Ile Leu Gly Ser Ala Asn Ile Gly Gln Val Asn Ser
65 70 75 80

Ser Gly Val Gly Arg Ser Ala Ser Ile Val Gly Gln Ser Ile Asn Gln
85 90 95

Ala Phe Ser

<210> 71

<211> 89

<212> PRT

<213> Cyrtophora moluccensis

<400> 71

Ser His Leu Ser Ser Pro Glu Ala Ser Ser Arg Val Ser Ser Ala Val
1 5 10 15

Ser Asn Leu Val Ser Ser Gly Ser Thr Asn Ser Ala Ala Leu Pro Asn
20 25 30

Thr Ile Ser Asn Val Val Ser Gln Ile Ser Ser Ser Asn Pro Gly Leu
35 40 45

Ser Gly Cys Asp Val Leu Val Gln Ala Leu Leu Glu Val Val Ser Ala
50 55 60

Leu Ile His Ile Leu Gly Ser Ser Ser Ile Gly Gln Val Asn Tyr Gly
65 70 75 80

Ser Ala Gly Gln Ala Thr Gln Ile Val
85

<210> 72

<211> 98

<212> PRT

<213> Latrodectus geometricus

<400> 72

Ser Ala Leu Ala Ala Pro Ala Thr Ser Ala Arg Ile Ser Ser His Ala
1 5 10 15

Ser Thr Leu Leu Ser Asn Gly Pro Thr Asn Pro Ala Ser Ile Ser Asn
20 25 30

Val Ile Ser Asn Ala Val Ser Gln Ile Ser Ser Ser Asn Pro Gly Ala
35 40 45

Ser Ser Cys Asp Val Leu Val Gln Ala Leu Leu Glu Leu Val Thr Ala
50 55 60

Leu Leu Thr Ile Ile Gly Ser Ser Asn Val Gly Asn Val Asn Tyr Asp
65 70 75 80

Ser Ser Gly Gln Tyr Ala Gln Val Val Ser Gln Ser Val Gln Asn Ala
85 90 95

Phe Val

<210> 73

<211> 98

<212> PRT

<213> Latrodectus hesperus

<400> 73

Ser Ala Leu Ser Ala Pro Ala Thr Ser Ala Arg Ile Ser Ser His Ala
1 5 10 15

Ser Ala Leu Leu Ser Ser Gly Pro Thr Asn Pro Ala Ser Ile Ser Asn
20 25 30

Val Ile Ser Asn Ala Val Ser Gln Ile Ser Ser Ser Asn Pro Gly Ala
35 40 45

Ser Ala Cys Asp Val Leu Val Gln Ala Leu Leu Glu Leu Val Thr Ala
50 55 60

Leu Leu Thr Ile Ile Gly Ser Ser Asn Ile Gly Ser Val Asn Tyr Asp
65 70 75 80

Ser Ser Gly Gln Tyr Ala Gln Val Val Thr Gln Ser Val Gln Asn Val
85 90 95

Phe Gly

<210> 74

<211> 93

<212> PRT

<213> Macrothele holsti

<400> 74

Ser His Leu Ser Ser Pro Glu Ala Ser Ser Arg Val Ser Ser Ala Val
 1 5 10 15

Ser Asn Leu Val Ser Gly Gly Ser Thr Asn Ser Ala Ala Leu Pro Asn
 20 25 30

Thr Ile Ser Asn Val Val Ser Gln Ile Ser Ser Ser Asn Pro Gly Leu
 35 40 45

Ser Gly Cys Asp Val Leu Val Gln Ala Leu Leu Glu Val Val Ser Ala
 50 55 60

Leu Ile His Ile Leu Gly Ser Ser Ser Ile Gly Gln Val Asp Tyr Gly
 65 70 75 80

Ser Ala Gly Gln Ala Thr Gln Ile Val Gly Gln Ser Ala
 85 90

<210> 75

<211> 98

<212> PRT

<213> Nephila clavipes

<400> 75

Ser Arg Leu Ser Ser Pro Gln Ala Ser Ser Arg Val Ser Ser Ala Val
 1 5 10 15

Ser Asn Leu Val Ala Ser Gly Pro Thr Asn Ser Ala Ala Leu Ser Ser
 20 25 30

Thr Ile Ser Asn Val Val Ser Gln Ile Gly Ala Ser Asn Pro Gly Leu
 35 40 45

Ser Gly Cys Asp Val Leu Ile Gln Ala Leu Leu Glu Val Val Ser Ala
 50 55 60

Leu Ile Gln Ile Leu Gly Ser Ser Ser Ile Gly Gln Val Asn Tyr Gly
 65 70 75 80

Ser Ala Gly Gln Ala Thr Gln Ile Val Gly Gln Ser Val Tyr Gln Ala
 85 90 95

Leu Gly

<210> 76

<211> 89

<212> PRT

<213> Nephila pilipes

<400> 76

Ser Arg Leu Ser Ser Pro Glu Ala Ser Ser Arg Val Ser Ser Ala Val
 1 5 10 15

Ser Asn Leu Val Ser Ser Gly Pro Thr Asn Ser Ala Ala Leu Ser Asn
20 25 30

Thr Ile Ser Asn Val Val Ser Gln Ile Ser Ser Ser Asn Pro Gly Leu
35 40 45

Ser Gly Cys Asp Val Leu Val Gln Ala Leu Leu Glu Val Val Ser Ala
50 55 60

Leu Ile His Ile Leu Gly Ser Ser Ser Ile Gly Gln Val Asn Tyr Gly
65 70 75 80

Ser Ala Gly Gln Ala Thr Gln Ile Val
85

<210> 77

<211> 87

<212> PRT

<213> *Nephila madagascariensis*

<400> 77

Ser Arg Leu Ser Ser Pro Gln Ala Ser Ser Arg Val Ser Ser Ala Val
1 5 10 15

Ser Asn Leu Val Ala Ser Gly Pro Thr Asn Ser Ala Ala Leu Ser Ser
20 25 30

Thr Ile Ser Asn Ala Val Ser Gln Ile Gly Ala Ser Asn Pro Gly Leu
35 40 45

Ser Gly Cys Asp Val Leu Ile Gln Ala Leu Leu Glu Val Val Ser Ala
50 55 60

Leu Ile His Ile Leu Gly Ser Ser Ser Ile Gly Gln Val Asn Tyr Gly
65 70 75 80

Ser Ala Gly Gln Ala Thr Gln
85

<210> 78

<211> 87

<212> PRT

<213> *Nephila senegalensis*

<400> 78

Ser Arg Leu Ser Ser Pro Glu Ala Ser Ser Arg Val Ser Ser Ala Val
1 5 10 15

Ser Asn Leu Val Ser Ser Gly Pro Thr Asn Ser Ala Ala Leu Ser Ser
20 25 30

Thr Ile Ser Asn Val Val Ser Gln Ile Gly Ala Ser Asn Pro Gly Leu
35 40 45

Ser Gly Cys Asp Val Leu Ile Gln Ala Leu Leu Glu Val Val Ser Ala
50 55 60

Leu Val His Ile Leu Gly Ser Ser Ser Ile Gly Gln Val Asn Tyr Gly
65 70 75 80

Ser Ala Gly Gln Ala Thr Gln
85

<210> 79

<211> 89

<212> PRT

<213> Octonoba varians

<400> 79

Ser Arg Leu Ser Ser Pro Glu Ala Ser Ser Arg Val Ser Ser Ala Val
1 5 10 15

Ser Asn Leu Val Ser Ser Gly Pro Thr Asn Ser Ala Ala Leu Ser Asn
20 25 30

Thr Ile Ser Asn Val Val Ser Gln Ile Ser Ser Ser Asn Pro Gly Leu
35 40 45

Ser Gly Cys Asp Val Leu Val Gln Ala Leu Leu Glu Val Val Ser Ala
50 55 60

Pro Ile His Ile Leu Gly Ser Ser Ser Ile Gly Gln Val Asn Tyr Gly
65 70 75 80

Ser Ala Gly Gln Ala Thr Gln Ile Val
85

<210> 80

<211> 89

<212> PRT

<213> Psechrus sinensis

<400> 80

Ser Arg Leu Ser Ser Pro Glu Ala Ser Ser Arg Val Ser Ser Ala Val
1 5 10 15

Ser Asn Leu Val Ser Ser Gly Pro Thr Asn Ser Ala Ala Leu Pro Asn
20 25 30

Thr Ile Ser Asn Val Val Ser Gln Ile Ser Ser Ser Asn Pro Gly Leu
35 40 45

Ser Gly Cys Asp Val Leu Val Gln Ala Leu Leu Glu Val Val Ser Ala
50 55 60

Leu Ile His Ile Leu Gly Ser Ser Ser Ile Gly Gln Val Asn Tyr Gly
65 70 75 80

Ser Ala Gly Gln Ala Thr Gln Ile Val
85

<210> 81

<211> 88

<212> PRT

<213> *Tetragnatha kauaiensis*

<400> 81

Ser Leu Leu Ser Ser Pro Ala Ser Asn Ala Arg Ile Ser Ser Ala Val
1 5 10 15

Ser Ala Leu Ala Ser Gly Ala Ala Ser Gly Pro Gly Tyr Leu Ser Ser

20

25

30

Val Ile Ser Asn Val Val Ser Gln Val Ser Ser Asn Ser Gly Gly Leu
35 40 45

Val Gly Cys Asp Thr Leu Val Gln Ala Leu Leu Glu Ala Ala Ala Ala
50 55 60

Leu Val His Val Leu Ala Ser Ser Ser Gly Gly Gln Val Asn Leu Asn
65 70 75 80

Thr Ala Gly Tyr Thr Ser Gln Leu
85

<210> 82

<211> 88

<212> PRT

<213> *Tetragnatha versicolor*

<400> 82

Ser Arg Leu Ser Ser Pro Ala Ser Asn Ala Arg Ile Ser Ser Ala Val
1 5 10 15

Ser Ala Leu Ala Ser Gly Gly Ala Ser Ser Pro Gly Tyr Leu Ser Ser
20 25 30

Ile Ile Ser Asn Val Val Ser Gln Val Ser Ser Asn Asn Asp Gly Leu
35 40 45

Ser Gly Cys Asp Thr Val Val Gln Ala Leu Leu Glu Val Ala Ala Ala
50 55 60

Leu Val His Val Leu Ala Ser Ser Asn Ile Gly Gln Val Asn Leu Asn
65 70 75 80

Thr Ala Gly Tyr Thr Ser Gln Leu
85

<210> 83

<211> 89

<212> PRT

<213> Araneus bicentenarius

<400> 83

Ser Arg Leu Ser Ser Ser Ala Ala Ser Ser Arg Val Ser Ser Ala Val
 1 5 10 15

Ser Ser Leu Val Ser Ser Gly Pro Thr Thr Pro Ala Ala Leu Ser Asn
 20 25 30

Thr Ile Ser Ser Ala Val Ser Gln Ile Ser Ala Ser Asn Pro Gly Leu
 35 40 45

Ser Gly Cys Asp Val Leu Val Gln Ala Leu Leu Glu Val Val Ser Ala
 50 55 60

Leu Val His Ile Leu Gly Ser Ser Ser Val Gly Gln Ile Asn Tyr Gly
 65 70 75 80

Ala Ser Ala Gln Tyr Ala Gln Met Val
 85

<210> 84

<211> 97

<212> PRT

<213> Argiope amoena

<400> 84

Arg Leu Ser Ser Pro Gln Ala Ser Ser Arg Val Ser Ser Ala Val Ser
 1 5 10 15

Thr Leu Val Ser Ser Gly Pro Thr Asn Pro Ala Ser Leu Ser Asn Ala
 20 25 30

Ile Gly Ser Val Val Ser Gln Val Ser Ala Ser Asn Pro Gly Leu Pro
 35 40 45

Ser Cys Asp Val Leu Val Gln Ala Leu Leu Glu Ile Val Ser Ala Leu
 50 55 60

Val His Ile Leu Gly Ser Ser Ser Ile Gly Gln Ile Asn Tyr Ser Ala
 65 70 75 80

Ser Ser Gln Tyr Ala Arg Leu Val Gly Gln Ser Ile Ala Gln Ala Leu
 85 90 95

Gly

<210> 85

<211> 82

<212> PRT

<213> Argiope aurantia

<400> 85

Ser Arg Leu Ser Ser Pro Gln Ala Ser Ser Arg Val Ser Ser Ala Val
1 5 10 15

Ser Thr Leu Val Ser Ser Gly Pro Thr Asn Pro Ala Ala Leu Ser Asn
20 25 30

Ala Ile Ser Ser Val Val Ser Gln Val Ser Ala Ser Asn Pro Gly Leu
35 40 45

Ser Gly Cys Asp Val Leu Val Gln Ala Leu Leu Glu Leu Val Ser Ala
50 55 60

Leu Val His Ile Leu Gly Ser Ser Ser Ile Gly Gln Ile Asn Tyr Ala
65 70 75 80

Ala Ser

<210> 86

<211> 98

<212> PRT

<213> Argiope trifasciata

<400> 86

Ser Arg Leu Ser Ser Pro Gln Ala Ser Ser Arg Val Ser Ser Ala Val
1 5 10 15

Ser Thr Leu Val Ser Ser Gly Pro Thr Asn Pro Ala Ser Leu Ser Asn
20 25 30

Ala Ile Ser Ser Val Val Ser Gln Val Ser Ser Ser Asn Pro Gly Leu
35 40 45

Ser Gly Cys Asp Val Leu Val Gln Ala Leu Leu Glu Ile Val Ser Ala
50 55 60

Leu Val His Ile Leu Gly Ser Ser Ser Ile Gly Gln Ile Asn Tyr Ala
65 70 75 80

Ala Ser Ser Gln Tyr Ala Gln Leu Val Gly Gln Ser Leu Thr Gln Ala
85 90 95

Leu Gly

<210> 87

<211> 89

<212> PRT

<213> Gasteracantha mammosa

<400> 87

Ser Arg Leu Ser Ser Pro Gln Ala Gly Ala Arg Val Ser Ser Ala Val
1 5 10 15

Ser Ala Leu Val Ala Ser Gly Pro Thr Ser Pro Ala Ala Val Ser Ser
20 25 30

Ala Ile Ser Asn Val Ala Ser Gln Ile Ser Ala Ser Asn Pro Gly Leu
35 40 45

Ser Gly Cys Asp Val Leu Val Gln Ala Leu Leu Glu Ile Val Ser Ala
50 55 60

Leu Val Ser Ile Leu Ser Ser Ala Ser Ile Gly Gln Ile Asn Tyr Gly
65 70 75 80

Ala Ser Gly Gln Tyr Ala Ala Met Ile
85

<210> 88

<211> 90

<212> PRT

<213> Latrodectus geometricus

<400> 88

Ser Ala Leu Ser Ser Pro Thr Thr His Ala Arg Ile Ser Ser His Ala
1 5 10 15

Ser Thr Leu Leu Ser Ser Gly Pro Thr Asn Ser Ala Ala Ile Ser Asn
20 25 30

Val Ile Ser Asn Ala Val Ser Gln Val Ser Ala Ser Asn Pro Gly Ser
35 40 45

Ser Ser Cys Asp Val Leu Val Gln Ala Leu Leu Glu Leu Ile Thr Ala
50 55 60

Leu Ile Ser Ile Val Asp Ser Ser Asn Ile Gly Gln Val Asn Tyr Gly
65 70 75 80

Ser Ser Gly Gln Tyr Ala Gln Met Val Gly
85 90

<210> 89

<211> 98

<212> PRT

<213> Latrodectus hesperus

<400> 89

Ser Ala Leu Ser Ser Pro Thr Thr His Ala Arg Ile Ser Ser His Ala

1 5 10 15

Ser Thr Leu Leu Ser Ser Gly Pro Thr Asn Ala Ala Ala Leu Ser Asn
20 25 30

Val Ile Ser Asn Ala Val Ser Gln Val Ser Ala Ser Asn Pro Gly Ser
35 40 45

Ser Ser Cys Asp Val Leu Val Gln Ala Leu Leu Glu Ile Ile Thr Ala
50 55 60

Ju Ju Ju
 Leu Ile Ser Ile Leu Asp Ser Ser Ser Val Gly Gln Val Asn Tyr Gly
 65 70 75 80

Ser Ser Gly Gln Tyr Ala Gln Ile Val Gly Gln Ser Met Gln Gln Ala
 85 90 95

Met Gly

<210> 90

<211> 97

<212> PRT

<213> *Nephila clavipes*

<400> 90

Ser Arg Leu Ala Ser Pro Asp Ser Gly Ala Arg Val Ala Ser Ala Val
 1 5 10 15

Ser Asn Leu Val Ser Ser Gly Pro Thr Ser Ser Ala Ala Leu Ser Ser
 20 25 30

Val Ile Ser Asn Ala Val Ser Gln Ile Gly Ala Ser Asn Pro Gly Leu
 35 40 45

Ser Gly Cys Asp Val Leu Ile Gln Ala Leu Leu Glu Ile Val Ser Ala
 50 55 60

Cys Val Thr Ile Leu Ser Ser Ser Ser Ile Gly Gln Val Asn Tyr Gly
 65 70 75 80

Ala Ala Ser Gln Phe Ala Gln Val Val Gly Gln Ser Val Leu Ser Ala
 85 90 95

Phe

<210> 91

<211> 82

<212> PRT

<213> *Nephila madagascariensis*

<400> 91

Ser Arg Leu Ala Ser Pro Asp Ser Gly Ala Arg Val Ala Ser Ala Val
 1 5 10 15

Ser Asn Leu Val Ser Ser Gly Pro Thr Ser Ser Ala Ala Leu Ser Ser
 20 25 30

Val Ile Ser Asn Ala Val Ser Gln Ile Gly Ala Ser Asn Pro Gly Leu
 35 40 45

Ser Gly Cys Asp Val Leu Ile Gln Ala Leu Leu Glu Ile Val Ser Ala
 50 55 60

Cys Val Thr Ile Leu Ser Ser Ser Ile Gly Gln Val Asn Tyr Gly
65 70 75 80

Ala Ala

<210> 92

<211> 82

<212> PRT

<213> *Nephila senegalensis*

<220>

<221> misc_feature

<222> (35)..(35)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc_feature

<222> (56)..(56)

<223> Xaa can be any naturally occurring amino acid

<400> 92

Ser Arg Leu Ala Ser Pro Asp Ser Gly Ala Arg Val Ala Ser Ala Val
1 5 10 15

Ser Asn Leu Val Ser Ser Gly Pro Thr Ser Ser Ala Ala Leu Ser Ser
20 25 30

Val Ile Xaa Asn Ala Val Ser Gln Ile Gly Ala Ser Asn Pro Gly Leu
35 40 45

Ser Gly Cys Asp Val Leu Ile Xaa Ala Leu Leu Glu Ile Val Ser Ala
50 55 60

Cys Val Thr Ile Leu Ser Ser Ser Ile Gly Gln Val Asn Tyr Gly
65 70 75 80

Ala Ala

<210> 93

<211> 71

<212> PRT

<213> *Dolomedes tenebrosus*

<400> 93

Ser Arg Leu Ser Ser Pro Glu Ala Ala Ser Arg Val Ser Ser Ala Val
1 5 10 15

Ser Ser Leu Val Ser Asn Gly Gln Val Asn Val Asp Ala Leu Pro Ser
20 25 30

Ile Ile Ser Asn Leu Ser Ser Ser Ile Ser Ala Ser Ala Thr Thr Ala
35 40 45

Ser Asp Cys Glu Val Leu Val Gln Val Leu Leu Glu Val Val Ser Ala
50 55 60

Leu Val Gln Ile Val Cys Ser
65 70

<210> 94

<211> 97

<212> PRT

<213> Dolomedes tenebrosus

<400> 94

Ser Arg Leu Ser Ser Pro Gln Ala Ala Ser Arg Val Ser Ser Ala Val
1 5 10 15

Ser Ser Leu Val Ser Asn Gly Gln Val Asn Val Ala Ala Leu Pro Ser
20 25 30

Ile Ile Ser Ser Leu Ser Ser Ser Ile Ser Ala Ser Ser Thr Ala Ala
35 40 45

Ser Asp Cys Glu Val Leu Val Gln Val Leu Leu Glu Ile Val Ser Ala
50 55 60

Leu Val Gln Ile Val Ser Ser Ala Asn Val Gly Tyr Ile Asn Pro Glu
65 70 75 80

Ala Ser Gly Ser Leu Asn Ala Val Gly Ser Ala Leu Ala Ala Ala Met
85 90 95

Gly

<210> 95

<211> 93

<212> PRT

<213> Araneus diadematus

<400> 95

Asn Arg Leu Ser Ser Ala Gly Ala Ala Ser Arg Val Ser Ser Asn Val
1 5 10 15

Ala Ala Ile Ala Ser Ala Gly Ala Ala Ala Leu Pro Asn Val Ile Ser
20 25 30

Asn Ile Tyr Ser Gly Val Leu Ser Ser Gly Val Ser Ser Ser Glu Ala
35 40 45

Leu Ile Gln Ala Leu Leu Glu Val Ile Ser Ala Leu Ile His Val Leu
50 55 60

Gly Ser Ala Ser Ile Gly Asn Val Ser Ser Val Gly Val Asn Ser Ala
65 70 75 80

Leu Asn Ala Val Gln Asn Ala Val Gly Ala Tyr Ala Gly
85 90

85

90

<210> 96

<211> 98

<212> PRT

<213> Araneus diadematus

<400> 96

Ser Arg Leu Ser Ser Pro Ser Ala Ala Ala Arg Val Ser Ser Ala Val
1 5 10 15

Ser Leu Val Ser Asn Gly Gly Pro Thr Ser Pro Ala Ala Leu Ser Ser
20 25 30

Ser Ile Ser Asn Val Val Ser Gln Ile Ser Ala Ser Asn Pro Gly Leu
35 40 45

Ser Gly Cys Asp Ile Leu Val Gln Ala Leu Leu Glu Ile Ile Ser Ala
50 55 60

Leu Val His Ile Leu Gly Ser Ala Asn Ile Gly Pro Val Asn Ser Ser

65 70 75 80

Ser Ala Gly Gln Ser Ala Ser Ile Val Gly Gln Ser Val Tyr Arg Ala
85 90 95

Leu Ser

<210> 97

<211> 98

<212> PRT

<213> Araneus diadematus

<400> 97

Ser Arg Leu Ser Ser Pro Ala Ala Ser Ser Arg Val Ser Ser Ala Val
1 5 10 15

Ser Ser Leu Val Ser Ser Gly Pro Thr Lys His Ala Ala Leu Ser Asn
20 25 30

Thr Ile Ser Ser Val Val Ser Gln Val Ser Ala Ser Asn Pro Gly Leu
35 40 45

Ser Gly Cys Asp Val Leu Val Gln Ala Leu Leu Glu Val Val Ser Ala
50 55 60

Leu Val Ser Ile Leu Gly Ser Ser Ser Ile Gly Gln Ile Asn Tyr Gly
65 70 75 80

Ala Ser Ala Gln Tyr Thr Gln Met Val Gly Gln Ser Val Ala Gln Ala
85 90 95

Leu Ala

<210> 98

<211> 94

<212> PRT

<213> Araneus diadematus

<400> 98

Ser Val Tyr Leu Arg Leu Gln Pro Arg Leu Glu Val Ser Ser Ala Val
 1 5 10 15

Ser Ser Leu Val Ser Ser Gly Pro Thr Asn Gly Ala Ala Val Ser Gly
 20 25 30

Ala Leu Asn Ser Leu Val Ser Gln Ile Ser Ala Ser Asn Pro Gly Leu
 35 40 45

Ser Gly Cys Asp Ala Leu Val Gln Ala Leu Leu Glu Leu Val Ser Ala
 50 55 60

Leu Val Ala Ile Leu Ser Ser Ala Ser Ile Gly Gln Val Asn Val Ser
 65 70 75 80

Ser Val Ser Gln Ser Thr Gln Met Ile Ser Gln Ala Leu Ser
 85 90

<210> 99

<211> 79

<212> PRT

<213> Homo sapiens

<220>

<221> DISULFID

<222> (8)..(77)

<220>

<221> DISULFID

<222> (11)..(71)

<220>

<221> DISULFID

<222> (35)..(46)

<220>

<221> DISULFID

<222> (48)..(48)

<223> Intermolecular Cys48-Cys48 linkage

<400> 99

Phe Pro Ile Pro Leu Pro Tyr Cys Trp Leu Cys Arg Ala Leu Ile Lys
 1 5 10 15

Arg Ile Gln Ala Met Ile Pro Lys Gly Ala Leu Ala Val Ala Val Ala

20 25 30

Gln Val Cys Arg Val Val Pro Leu Val Ala Gly Gly Ile Cys Gln Cys
35 40 45

Leu Ala Glu Arg Tyr Ser Val Ile Leu Leu Asp Thr Leu Leu Gly Arg
50 55 60

Met Leu Pro Gln Leu Val Cys Arg Leu Val Leu Arg Cys Ser Met
65 70 75

<210> 100

<211> 79

<212> PRT

<213> Mus musculus

<220>

<221> DISULFID

<222> (8)..(77)

<220>

<221> DISULFID

<222> (11)..(71)

<220>

<221> DISULFID

<222> (35)..(46)

<220>

<221> DISULFID

<222> (48)..(48)

<223> Intermolecular Cys48-Cys48 linkage

<400> 100

Leu Pro Ile Pro Leu Pro Phe Cys Trp Leu Cys Arg Thr Leu Ile Lys
1 5 10 15

Arg Val Gln Ala Val Ile Pro Lys Gly Val Leu Ala Val Ala Val Ser
20 25 30

Gln Val Cys His Val Val Pro Leu Val Val Gly Gly Ile Cys Gln Cys
35 40 45

Leu Ala Glu Arg Tyr Thr Val Leu Leu Leu Asp Ala Leu Leu Gly Arg
50 55 60

Val Val Pro Gln Leu Val Cys Gly Leu Val Leu Arg Cys Ser Thr
65 70 75

<210> 101

<211> 79

<212> PRT

<213> Sus scrofa

<220>

<221> DISULFID

<222> (8)..(77)

<220>

<221> DISULFID

<222> (11)..(71)

<220>

<221> DISULFID

<222> (35)..(46)

<220>

<221> DISULFID

<222> (48)..(48)

<223> Intermolecular Cys48-Cys48 linkage

<400> 101

Phe	Pro	Ile	Pro	Leu	Pro	Phe	Cys	Trp	Leu	Cys	Arg	Thr	Leu	Ile	Lys
1				5					10					15	

Arg	Ile	Gln	Ala	Val	Val	Pro	Lys	Gly	Val	Leu	Leu	Lys	Ala	Val	Ala
		20						25					30		

Gln	Val	Cys	His	Val	Val	Pro	Leu	Pro	Val	Gly	Gly	Ile	Cys	Gln	Cys
		35					40					45			

Leu	Ala	Glu	Arg	Tyr	Ile	Val	Ile	Cys	Leu	Asn	Met	Leu	Leu	Asp	Arg
	50					55					60				

Thr	Leu	Pro	Gln	Leu	Val	Cys	Gly	Leu	Val	Leu	Arg	Cys	Ser	Ser	
65					70					75					

<210> 102

<211> 79

<212> PRT

<213> Oryctolagus cuniculus

<220>

<221> DISULFID

<222> (8)..(77)

<220>

<221> DISULFID

<222> (11)..(71)

<220>

<221> DISULFID

<222> (35)..(46)

<220>

<221> DISULFID

<222> (48)..(48)

<223> Intermolecular Cys48-Cys48 linkage

<400> 102

Phe Pro Ile Pro Leu Pro Leu Cys Trp Leu Cys Arg Thr Leu Leu Lys
1 5 10 15

Arg Ile Gln Ala Met Ile Pro Lys Gly Val Leu Ala Met Ala Val Ala
20 25 30

Gln Val Cys His Val Val Pro Leu Val Val Gly Gly Ile Cys Gln Cys
35 40 45

Leu Ala Glu Arg Tyr Thr Val Ile Leu Leu Glu Val Leu Leu Gly His
50 55 60

Val Leu Pro Gln Leu Val Cys Gly Leu Val Leu Arg Cys Ser Ser
65 70 75

<210> 103

<211> 79

<212> PRT

<213> Rattus norvegicus

<220>

<221> DISULFID

<222> (8)..(77)

<220>

<221> DISULFID

<222> (11)..(71)

<220>

<221> DISULFID

<222> (35)..(46)

<220>

<221> DISULFID

<222> (48)..(48)

<223> Intermolecular Cys48-Cys48 linkage

<400> 103

Leu Pro Ile Pro Leu Pro Phe Cys Trp Leu Cys Arg Thr Leu Ile Lys
1 5 10 15

Arg Val Gln Ala Val Ile Pro Lys Gly Val Leu Ala Val Ala Val Ser
20 25 30

Gln Val Cys His Val Val Pro Leu Val Val Gly Gly Ile Cys Gln Cys
 35 40 45

Leu Ala Glu Arg Tyr Thr Val Leu Leu Leu Asp Ala Leu Leu Gly Arg
 50 55 60

Val Val Pro Gln Leu Val Cys Gly Leu Val Leu Arg Cys Ser Thr
 65 70 75

<210> 104

<211> 34

<212> PRT

<213> Artificial Sequence

<220>

<223> Construct derived from SP-B

<220>

<221> DISULFID

<222> (1)..(33)

<220>

<221> DISULFID

<222> (4)..(27)

<400> 104

Cys Trp Leu Cys Arg Ala Leu Ile Lys Arg Ile Gln Ala Met Ile Pro
 1 5 10 15

Lys Gly Gly Arg Met Leu Pro Gln Leu Val Cys Arg Leu Val Leu Arg
 20 25 30

Cys Ser

<210> 105

<211> 34

<212> PRT

<213> Artificial Sequence

<220>

<223> Construct derived from SP-B

<220>

<221> DISULFID

<222> (1)..(33)

<220>

<221> DISULFID

<222> (4)..(27)

<400> 105

Cys Trp Leu Cys Arg Ala Leu Ile Lys Arg Ile Gln Ala Leu Ile Pro
 1 5 10 15

Lys Gly Gly Arg Leu Leu Pro Gln Leu Val Cys Arg Leu Val Leu Arg
 20 25 30

Cys Ser

<210> 106

<211> 27

<212> PRT

<213> Artificial Sequence

<220>

<223> Construct derived from SP-B

<220>

<221> DISULFID

<222> (1)..(27)

<220>

<221> DISULFID

<222> (4)..(21)

<400> 106

Cys Leu Leu Cys Arg Ala Leu Ile Lys Arg Phe Asn Arg Tyr Leu Thr
 1 5 10 15

Pro Gln Leu Val Cys Arg Leu Val Leu Arg Cys
 20 25

<210> 107

<211> 34

<212> PRT

<213> Artificial Sequence

<220>

<223> Construct derived from SP-B

<220>

<221> DISULFID

<222> (1)..(33)

<400> 107

Cys Trp Leu Ala Arg Ala Leu Ile Lys Arg Ile Gln Ala Leu Ile Pro
 1 5 10 15

Lys Gly Gly Arg Leu Leu Pro Gln Leu Val Ala Arg Leu Val Leu Arg
 20 25 30

Cys Ser

<210> 108

<211> 34

<212> PRT

<213> Artificial Sequence

<220>

<223> Construct derived from SP-B

<220>

<221> DISULFID

<222> (4)..(27)

<400> 108

Ala	Trp	Leu	Cys	Arg	Ala	Leu	Ile	Lys	Arg	Ile	Gln	Ala	Leu	Ile	Pro
1				5					10					15	

Lys	Gly	Gly	Arg	Leu	Leu	Pro	Gln	Leu	Val	Cys	Arg	Leu	Val	Leu	Arg
			20					25					30		

Ala Ser

<210> 109

<211> 34

<212> PRT

<213> Artificial Sequence

<220>

<223> Construct derived from SP-B

<220>

<221> DISULFID

<222> (1)..(33)

<400> 109

Cys	Trp	Leu	Leu	Arg	Ala	Leu	Ile	Lys	Arg	Ile	Gln	Ala	Leu	Ile	Pro
1				5					10					15	

Lys	Gly	Gly	Arg	Leu	Leu	Pro	Gln	Leu	Val	Leu	Arg	Leu	Val	Leu	Arg
			20					25					30		

Cys Ser

<210> 110

<211> 34

<212> PRT

<213> Artificial Sequence

<220>

<223> Construct derived from SP-B

<220>

<221> DISULFID

<222> (4)..(27)

<400> 110

Leu	Trp	Leu	Cys	Arg	Ala	Leu	Ile	Lys	Arg	Ile	Gln	Ala	Leu	Ile	Pro
1			5					10					15		

Lys	Gly	Gly	Arg	Leu	Leu	Pro	Gln	Leu	Val	Cys	Arg	Leu	Val	Leu	Arg
			20				25					30			

Leu Ser

<210> 111

<211> 35

<212> PRT

<213> Homo sapiens

<400> 111

Phe	Gly	Ile	Pro	Cys	Cys	Pro	Val	His	Leu	Lys	Arg	Leu	Leu	Ile	Val
1				5				10					15		

Val	Val	Val	Val	Val	Leu	Ile	Val	Val	Val	Ile	Val	Gly	Ala	Leu	Leu
			20				25					30			

Met Gly Leu
35

<210> 112

<211> 35

<212> PRT

<213> Artificial Sequence

<220>

<223> Construct derived from SP-C

<400> 112

Phe	Gly	Ile	Pro	Ser	Ser	Pro	Val	His	Leu	Lys	Arg	Leu	Lys	Leu	Leu
1				5				10					15		

Leu	Leu	Leu	Leu	Leu	Leu	Ile	Leu	Leu	Leu	Ile	Leu	Gly	Ala	Leu	Leu
			20				25					30			

Met Gly Leu
35

<210> 113

<211> 33

<212> PRT

<213> Artificial Sequence

<220>

<223> Construct derived from SP-C

<400> 113

```
Ile Pro Ser Ser Pro Val His Leu Lys Arg Leu Lys Leu Leu Leu Leu
1          5          10          15
```

```
Leu Leu Leu Leu Ile Leu Leu Leu Ile Leu Gly Ala Leu Leu Met Gly
20          25          30
```

Leu

<210> 114

<211> 30

<212> PRT

<213> Artificial Sequence

<220>

<223> Construct derived from SP-C

<400> 114

```
Ile Pro Ser Ser Pro Val His Leu Lys Arg Leu Lys Leu Leu Leu Leu
1          5          10          15
```

```
Leu Leu Leu Leu Ile Leu Leu Leu Ile Leu Gly Ala Leu Leu
20          25          30
```

<210> 115

<211> 33

<212> PRT

<213> Artificial Sequence

<220>

<223> Construct derived from SP-C

<400> 115

```
Ile Pro Ser Ser Pro Val His Leu Lys Arg Leu Lys Leu Leu Leu Leu
1          5          10          15
```

```
Leu Leu Leu Leu Ile Leu Leu Leu Ile Leu Gly Ala Leu Leu Leu Gly
20          25          30
```

Leu

<210> 116

<211> 37

<212> PRT

<213> Homo sapiens

<400> 116

```
Leu Leu Gly Asp Phe Phe Arg Lys Ser Lys Glu Lys Ile Gly Lys Glu
1          5          10          15
```

```
Phe Lys Arg Ile Val Gln Arg Ile Lys Asp Phe Leu Arg Asn Leu Val
```

20 25 30

Pro Arg Thr Glu Ser
35

<210> 117

<211> 21

<212> PRT

<213> Artificial Sequence

<220>

<223> Construct derived from SP-B

<400> 117

Lys Leu Leu Leu Leu Lys Leu Leu Leu Leu Lys Leu Leu Leu Lys
1 5 10 15

Leu Leu Leu Leu Lys
20

<210> 118

<211> 34

<212> PRT

<213> Artificial Sequence

<220>

<223> Construct derived from SP-B

<400> 118

Ala Trp Leu Ala Arg Ala Leu Ile Lys Arg Ile Gln Ala Leu Ile Pro
1 5 10 15

Lys Gly Gly Arg Leu Leu Pro Gln Leu Val Ala Arg Leu Val Leu Arg
20 25 30

Ala Ser

<210> 119

<211> 34

<212> PRT

<213> Artificial Sequence

<220>

<223> Construct derived from SP-B

<400> 119

Leu Trp Leu Leu Arg Ala Leu Ile Lys Arg Ile Gln Ala Leu Ile Pro
1 5 10 15

Lys Gly Gly Arg Leu Leu Pro Gln Leu Val Leu Arg Leu Val Leu Arg
20 25 30

Leu Ser

Leu Ser

<210> 120

<211> 9

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<222> (1)..(1)

<223> Xaa can be any naturally occurring amino acid

<220>

<221> misc_feature

<222> (9)..(9)

<223> Xaa can be any naturally occurring amino acid

<400> 120

Xaa	Leu	Glu	Thr	Leu	Phe	Gln	Gly	Xaa
1				5				

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

- [WO2007078239A \[0005\]](#)
- [EP21083749A \[0154\]](#)
- [EP10156927A \[0154\]](#)
- [SE2010050439W \[0154\]](#)
- [SE2010051163W \[0154\]](#)
- [EP10848078A \[0154\]](#)

Non-patent literature cited in the description

- **HEDHAMMAR M. et al.**Biochemistry, 2008, vol. 47, 3407-3417 [\[0006\]](#)
- **RISING A. et al.**Biomacromolecules, 2006, vol. 7, 3120-3124 [\[0042\]](#)
- **THOMPSON, J.D.HIGGINS, D.G.GIBSON, T.J.**Nucleic Acids Research, 1994, vol. 22, 4673-4680 [\[0046\]](#)
- J Biol Chem, 2005, vol. 280, 2423225-23231 [\[0058\]](#)
- **J. JOHANSSON et al.**J. Appl. Physiol., 2003, vol. 95, 2055-2063 [\[0090\]](#)
- **SCHURCH S et al.**J. Appl. Physiol., 1989, vol. 67, 2389-2396 [\[0125\]](#)
- **JOHANSSON et al.**J. Appl. Physiol, 2003, vol. 95, 2055-2063 [\[0127\]](#)
- **KIMATA, Y et al.**Biochem. Biophys. Res. Commun., 1997, vol. 232, 69-73 [\[0138\]](#)
- **BELOGEY et al.**Eur J Biochem., 2004, vol. 271, 163360-3367 [\[0150\]](#)
- **BELOGEY et al.**J. Biol. Chem., 2002, vol. 277, 17367-17373 [\[0151\]](#)
- **GRÄSLUND T. et al.**Protein Expr Purif, 1997, vol. 9, 1125-132 [\[0152\]](#)
- **CORDINGLEY MG et al.**J. Virol., 1989, vol. 63, 125037-5045 [\[0152\]](#)

P A T E N T K R A V

1. Fusionsprotein omfattende

- (i) mindst én opløselighedsfremmende gruppe som er afledt fra det N-terminale (NT) fragment af et edderkoppesilkeprotein; og
- 5 (ii) mindst én gruppe som er et ønsket protein eller polypeptid valgt fra gruppen bestående af (a) polypeptidmedikamenter og (b) vækstfaktorer;
- hvor hver opløselighedsfremmende gruppe er forbundet direkte eller indirekte til den ønskede protein- eller polypeptidgruppe; og
- hvor hver opløselighedsfremmende gruppe har mindst 80% identitet med en hvilken som helst af SEQ ID NO 6 og SEQ ID NO 50-63 eller mindst 50% identitet med SEQ ID NO 8.
- 10

2. Fusionsprotein ifølge krav 1 hvor hver opløselighedsfremmende gruppe har mindst 80% identitet med SEQ ID NO 6 eller mindst 50% identitet med SEQ ID NO 8.

3. Fusionsprotein ifølge et hvilket som helst af de foregående krav hvor hver opløselighedsfremmende gruppe indeholder fra 100 til 160 aminosyrerester.

- 15 **4.** Fusionsprotein ifølge et hvilket som helst af de foregående krav hvor fusionsproteinet omfatter mindst to, eventuelt sammenhængende, opløselighedsfremmende grupper hver især afledt fra det N-terminale (NT) fragment af et edderkoppesilkeprotein.

5. Fusionsprotein ifølge et hvilket som helst af de foregående krav hvor mindst én opløselighedsfremmende gruppe er forbundet direkte eller indirekte til den aminoterminalende eller den carboxyterminale ende af mindst én ønsket protein- eller polypeptidgruppe.

20

6. Fusionsprotein ifølge krav 5 hvor mindst én opløselighedsfremmende gruppe udgør den aminoterminalende og/eller den carboxylterminale ende af fusionsproteinet.

7. Fusionsprotein ifølge et hvilket som helst af de foregående krav der yderligere omfatter (iii) mindst ét spaltningssted anbragt mellem mindst én ønsket protein- eller polypeptidgruppe og mindst én opløselighedsfremmende gruppe.

25

8. Fusionsprotein ifølge et hvilket som helst af de foregående krav hvor det ønskede protein eller polypeptid er en vækstfaktor.

9. Fusionsprotein ifølge et hvilket som helst af kravene 1-7 hvor det ønskede protein eller polypeptid er et polypeptidmedikament.

30 **10.** Anvendelse af mindst én gruppe som er afledt af det N-terminale (NT) fragment af et edderkoppesilkeprotein som en opløselighedsfremmende gruppe i et fusionsprotein til fremstilling af et ønsket protein eller polypeptid valgt fra gruppen bestående af (a) polypeptidmedikamenter og (b) vækstfaktorer; og hvor hver opløselighedsfremmende gruppe er som defineret ifølge et hvilket som helst af kravene 1-3.

35 **11.** Anvendelse ifølge krav 10, hvor fusionsproteinet yderligere omfatter mindst ét spaltningssted anbragt mellem mindst én ønsket protein- eller polypeptidgruppe og mindst én opløselighedsfremmende gruppe.

12. Anvendelse ifølge et hvilket som helst af kravene 10-11 hvor det ønskede protein eller polypeptid er en vækstfaktor.

13. Anvendelse ifølge et hvilket som helst af kravene 10-11 hvor det ønskede protein eller polypeptid er et polypeptidmedikament.

14. Fremgangsmåde til fremstilling af et ønsket protein eller polypeptid omfattende de følgende trin:

- 5 a) at eksprimere et fusionsprotein ifølge et hvilket som helst af kravene 7-9 i en egnet vært;
 - b) at opnå en blanding indeholdende fusionsproteinet eller polypeptidet, og valgfrit at isolere fusionsproteinet eller polypeptidet;
 - c) at spalte fusionsproteinet til fremskaffelse af det ønskede protein eller polypeptid; og
 - d) at isolere det ønskede protein eller polypeptid.
- 10 **15.** Fremgangsmåde ifølge krav 14 hvor trin b) yderligere involverer oprensning af fusionsproteinet på et affinitetsmedium med en immobiliseret NT gruppe og/eller på et anionbyttermidmedium.

DRAWINGS

Ea	MaSp1	SHTPTWNPGLAENFMNSFMQGLSSMPGFTASQLDDMSTIAQSMVQSIQSLAAQGRTPNKLQALNNMAFA
Lg	MaSp1	QANTPWSSKANADAFINSFISSAQNTGSFSQDQMDDMSLIGNTMTAMDNMG--GRITPSKLQALDNMAFA
Lh	MaSp1	QANTPWSSKANADAFINSFISAAASNTGSFSQDQMEDMSLIGNTMAAMDNMG--GRITPSKLQALDNMAFA
Nc	MaSp1	-QNTPWSSTELADAFINAFMEAGRTGAFTADQLDDMSTIGDTIKTAMDKMARSNKSSQKGLQALNNMAFA
At	MaSp2	QGATPWENSQLAESFISRFLRFIQSGCAFSPNQLDDMSSIGDTLTKTAIEKMAQSRKSSKSKLQALNNMAFA
Lg	MaSp2	---LRWSSKDNADRFINAFQAASNSGAFSSDQVDDMSVIGNTMTAMDNMG--GRITPSKLQALDNMAFA
Lh	MaSp2	QANTPWSSKENADAFIGAFMAAASQSGAFSSDQIDDMSVISNTLMAAMDNMG--GRITQSKLQALDNMAFA
Nim	MaSp2	QANTPWSDTATADAFIQNELGAVSGGAFTPDQLDDMSTVGDTIMSAMDKMARSNKSSQKGLQALNNMAFA
Nc	MaSp2	QARSPWSDTATADAFIQNFLAAVSGGAFSTDQLDDMSTIGDTIMSAMDKMARSNKSSQKGLQALNNMAFA
Ab	CySp1	AVPSVFSSPNLASCFLOCLTFGIGNSPAFTQEQQDLDAIAQVILNAVSSNTGATASAR--AQALSTALA
Ncl	CySp1	PVPSVFSSPSLASGFLGCLTTGIGLSPAFPQEQQDLDDAKVILSAVTSNTDTSKSAR--AQALSTALA
Lh	TuSp1	ASVNIENSPNAATSFLNCLRSNIESSPAFFQEQAADLDSIAEVILSDVSS-VNTASSAT--SLALSTALA
Nc	flag	IANSPPFSNPNTAEAFARSFVSNIVSSGEGFQAQGAEDFDDIIQSLIQAQ-SMGKGRHDTKAKAKAMQVVALA
Nlm	flag	IVNSPPFSNPNTAEAFARSFVSNVVSSGEGFQAQGAEDFDDIIQSLIQAQ-SMGKGRHDTKAKAKAMQVVALA
Ea	MaSp1	SSMAEIAASEEGGSLSTKTSIIASAMSN AFLQTGTVNQPFINEITQLVSMFAQAQGMNDV
Lg	MaSp1	SSVAEIAASEG--GDLGVTTNAIADALTSAFYQTTGVVNNRFISEIRSLISMFAQAASANDV
Lh	MaSp1	SSVAEIAASEG--GDLGVTTNAIADALTSAFYQTTGVVNSRFISEIRSLIGMFAQAASANDV
Nc	MaSp1	SSMAEIAAVEQGLSVDAKTNAIADSLNSAFYQTTGAANPQFVNEIRSLINMFAQSSANEV
At	MaSp2	SSMAEIAVAEQGLSLEAKTNAIASALSAFLETGTVNQQFVNEIKTLIFMIAQAASSNEI
Lg	MaSp2	SSVAEIAVADG--QNVGGATNAISNALRSAFYQTTGVVNNQFISEISNLINMFAQVVSANEV
Lh	MaSp2	SSVAEIAVADG--QNVGAATNAISDALRSAFYQTTGVVNNQFITGISSLIGMFAQVSGNEV
Nim	MaSp2	SSMAEIAAVEQGGQSMVDVKTNAIANALDSAFYMTTGSTNQQFVNMERSLINMLSAAAVNEV
Nc	MaSp2	SSMAEIAAVEQGGMSMAVKTNAIVDGLNSAFYMTTGAANPQFVNMERSLISMSIAASANEV
Ab	CySp1	SSLTDLLIAESAESNYSNLSELTGILSDCFIQTTGSDNPAFVSRIQSLISVLSQNADTNI
Ncl	CySp1	SSLADLLISESSGSSYQTQISALTNILSDCFVTTGSDNPAFVSQVQTLIGVLSQSSSNAI
Lh	TuSp1	SSLAEELLVTESAEEDIDNQVVALSTILSQCFVETTGSPNPAFVASVKSLLGVLQSASNYE
Nc	flag	SSIAELVIAESSGGDVQRKTNVISNALRNALMSTTGSPNEEFVHEVQDLIQMLSQEQINEV
Nlm	flag	SSIAELVIAESSGGDVQRKTNVISNALRNALMSTTGSPNEEFVHEVQDLIQMLSQEQINEV

Fig 1

CThyb_Esp	SRLSSPEASS	RVSSAVSNLV	SSG-PTNSAA	LSSTISNVVS	QIGASNPGLS
CTnat_Eau	SRLSSPSAVS	RVSSAVSSLV	SNG-QVNMAA	LPNIISNISS	SVSASAPGAS
AF350266_At1	SRLSSPGAAS	RVSSAVTSLV	SSGGPTNSAA	LSNTISNVVS	QISSSNPGLS
AY666062_Cm1	SHLSSPEASS	RVSSAVSNLV	SSG-STNSAA	LPNTISNVVS	QISSSNPGLS
AF350273_Lg1	SALAAPATSA	RISSHASTLL	SNG-PTNPAS	ISNVISNAVS	QISSSNPGAS
AY953074_Lh1	SALSAPATSA	RISSHASALL	SSG-PTNPAS	ISNVISNAVS	QISSSNPGAS
AY666068_Mh1	SHLSSPEASS	RVSSAVSNLV	SSG-STNSAA	LPNTISNVVS	QISSSNPGLS
U20329_Nc1	SRLSSPQASS	RVSSAVSNLV	ASG-PTNSAA	LSSTISNVVS	QIGASNPGLS
AY666076_Np1	SRLSSPEASS	RVSSAVSNLV	SSG-PTNSAA	LSNTISNVVS	QISSSNPGLS
AF350277_Nm1	SRLSSPQASS	RVSSAVSNLV	ASG-PTNSAA	LSSTISNAVS	QIGASNPGLS
AF350279_Ns1	SRLSSPEASS	RVSSAVSNLV	SSG-PTNSAA	LSSTISNVVS	QIGASNPGLS
AY666057_Ov1	SRLSSPEASS	RVSSAVSNLV	SSG-PTNSAA	LSNTISNVVS	QISSSNPGLS
AY666064_Ps1	SRLSSPEASS	RVSSAVSNLV	SSG-PTNSAA	LPNTISNVVS	QISSSNPGLS
AF350285_Tk1	SLLSSPASNA	RISSAVSALA	SGA-ASGPGY	LSSVISNVVS	QVSSNSGGLV
AF350286_Tv1	SRLSSPASNA	RISSAVSALA	SGG-ASSPGY	LSSIISNVVS	QVSSNNDGLS
ABU20328_Ab2	SRLSSSAASS	RVSSAVSSLV	SSG-PTTPAA	LSNTISSAVS	QISASNPGLS
AY365016_Aam2	-RLSSPQASS	RVSSAVSTLV	SSG-PTNPAS	LSNAIGSVVS	QVSASNPGLP
AF350263_Aau2	SRLSSPQASS	RVSSAVSTLV	SSG-PTNPAA	LSNAISSVVS	QVSASNPGLS
AF350267_At2	SRLSSPQASS	RVSSAVSTLV	SSG-PTNPAS	LSNAISSVVS	QVSSSNPGLS
AF350272_Gm2	SRLSSPQAGA	RVSSAVSALV	ASG-PTSPAA	VSSAISNVAS	QISASNPGLS
AF350275_Lg2	SALSSPTTHA	RISSHASTLL	SSG-PTNSAA	ISNVISNAVS	QVSASNPGLS
AY953075_Lh2	SALSSPTTHA	RISSHASTLL	SSG-PTNAAA	LSNVISNAVS	QVSASNPGLS
AY654293_Nc2	SRLASPDGSA	RVASAVSNLV	SSG-PTSSAA	LSSVISNAVS	QIGASNPGLS
AF350278_Nm2	SRLASPDGSA	RVASAVSNLV	SSG-PTSSAA	LSSVISNAVS	QIGASNPGLS
AF350280_Ns2	SRLASPDGSA	RVASAVSNLV	SSG-PTSSAA	LSSVIXNAVS	QIGASNPGLS
AF350269_DtFb1	SRLSSPEAAS	RVSSAVSSLV	SNG-QVNVDA	LPSIISNLSS	SISASATTAS
AF350270_DtFb2	SRLSSPQAAS	RVSSAVSSLV	SNG-QVNVAA	LPSIISSLSS	SISASSTAAS
U47853_ADF1	NRLSSAGAAS	RVSSNVAAIA	SAG-----AAA	LPNVISNIYS	GVLSS--GVS
U47854_ADF2	SRLSSPSAAA	RVSSAVS-LV	SNGGPTSPAA	LSSSISNVVS	QISASNPGLS
U47855_ADF3	SRLSSPAASS	RVSSAVSSLV	SSG-PTKHAA	LSNTISSVVS	QVSASNPGLS
U47856_ADF4	SVYLRLQPRL	EVSSAVSSLV	SSG-PTNGAA	VSGALNSLVS	QISASNPGLS
Consensus	SRLSSPQASS	RVSSAVSNLV	SSG-PTNSAA	LSNTISNVVS	QISASNPGLS

Fig 2

CThyb_Esp	GCDVLVQALL	EVVSALIHIL	GSSSIGQVNY	GSAGQATQLV	GQSVYQALGE	F
CTnat_Eau	GCEVIVQALL	EVITALVQIV	SSSSVGYNP	SAVNQITNVV	ANAMAQVMG-	-
AF350266_At1	GCDVLVQALL	EIVSALVHIL	GSANIGQVNS	SGVGRSASIV	GQSINQAFS-	-
AY666062_Cm1	GCDVLVQALL	EVVSALIHIL	GSSSIGQVNY	GSAGQATQIV	-----	-
AF350273_Lg1	SCDVLVQALL	ELVTALHTII	GSSNVGNVNY	DSSGQYAQVV	SQSVQNAFV-	-
AY953074_Lh1	ACDVLVQALL	ELVTALHTII	GSSNIGSVNY	DSSGQYAQVV	TQSVQNVFG-	-
AY666068_Mh1	GCDVLVQALL	EVVSALIHIL	GSSSIGQVDY	GSAGQATQIV	GQSA-----	-
U20329_Nc1	GCDVLIQALL	EVVSALIQIL	GSSSIGQVNY	GSAGQATQIV	GQSVYQALG-	-
AY666076_Np1	GCDVLVQALL	EVVSALIHIL	GSSSIGQVNY	GSAGQATQIV	-----	-
AF350277_Nm1	GCDVLIQALL	EVVSALIHIL	GSSSIGQVNY	GSAGQATQ--	-----	-
AF350279_Ns1	GCDVLIQALL	EVVSALVHIL	GSSSIGQVNY	GSAGQATQ--	-----	-
AY666057_Ov1	GCDVLVQALL	EVVSAPIHIL	GSSSIGQVNY	GSAGQATQIV	-----	-
AY666064_Ps1	GCDVLVQALL	EVVSALIHIL	GSSSIGQVNY	GSAGQATQIV	-----	-
AF350285_Tk1	GCDTLVQALL	EAAAALVHVL	ASSSGGQVNL	NTAGYTSQL-	-----	-
AF350286_Tv1	GCDTVVQALL	EVAAALVHVL	ASSNIGQVNL	NTAGYTSQL-	-----	-
ABU20328_Ab2	GCDVLVQALL	EVVSALVHIL	GSSSVGQINY	GASAQYAQMV	-----	-
AY365016_Aam2	SCDVLVQALL	EIVSALVHIL	GSSSIGQINY	SASSQYARLV	GQSIAQALG-	-
AF350263_Aau2	GCDVLVQALL	ELVSALVHIL	GSSSIGQINY	AAS-----	-----	-
AF350267_At2	GCDVLVQALL	EIVSALVHIL	GSSSIGQINY	AASSQYAQLV	GQSLTQALG-	-
AF350272_Gm2	GCDVLVQALL	EIVSALVSIL	SSASIGQINY	GASGQYAAMI	-----	-
AF350275_Lg2	SCDVLVQALL	ELITALISIV	DSSNIGQVNY	GSSGQYAQMV	G-----	-
AY953075_Lh2	SCDVLVQALL	EIITALISIL	DSSSVGQVNY	GSSGQYAQIV	GQSMQQAMG-	-
AY654293_Nc2	GCDVLIQALL	EIVSACVTIL	SSSSIGQVNY	GAASQFAQVV	GQSVLSAF--	-
AF350278_Nm2	GCDVLIQALL	EIVSACVTIL	SSSSIGQVNY	GAA-----	-----	-
AF350280_Ns2	GCDVLIXALL	EIVSACVTIL	SSSSIGQVNY	GAA-----	-----	-
AF350269_DtFb1	DCEVLVQVLL	EVVSALVQIV	CS-----	-----	-----	-
AF350270_DtFb2	DCEVLVQVLL	EIVSALVQIV	SSANVGYNP	EASGSLN-AV	GSALAAAMG-	-
U47853_ADF1	SSEALIQALL	EVISALIHVL	GSASIGNVSS	VGVSALNAV	QNAVAYAG-	-
U47854_ADF2	GCDILVQALL	EIISALVHIL	GSANIGPVNS	SSAGQSASIV	GQSVYRALS-	-
U47855_ADF3	GCDVLVQALL	EVVSALVSIL	GSSSIGQINY	GASAQYTQMV	GQSVQAALA-	-
U47856_ADF4	GCDALVQALL	ELVSALVAIL	SSASIGQVNV	SSVSQSTQMI	SQALS-----	-
Consensus	GCDVLVQALL	EVVSALVHIL	GSSSIGQVNY	GSAGQATQIV	GQSVQAALGE	F

Fig 2 (continued)

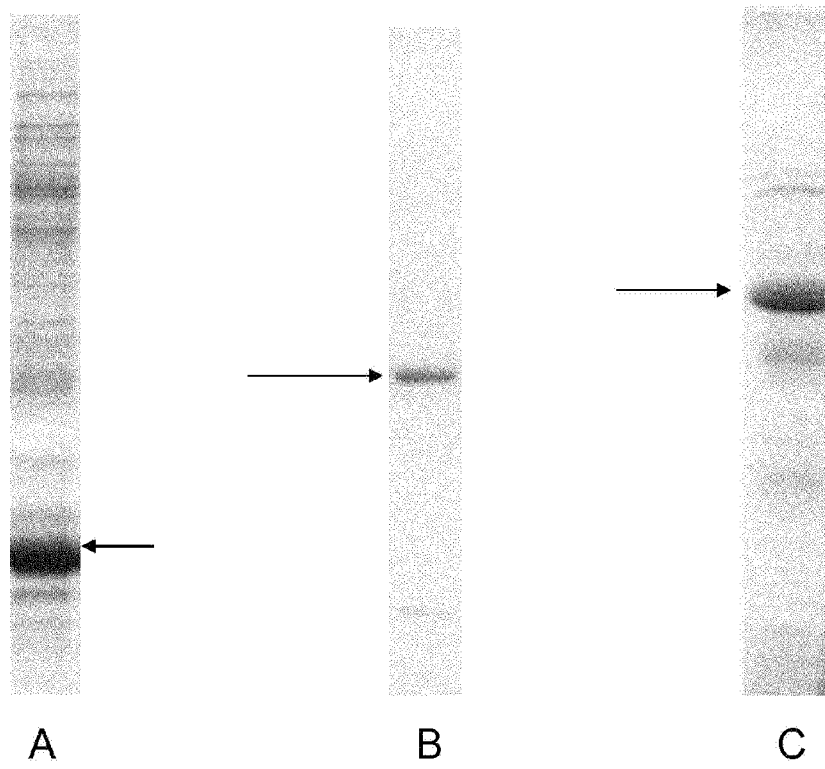


Fig. 3

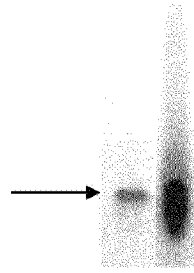


Fig. 4

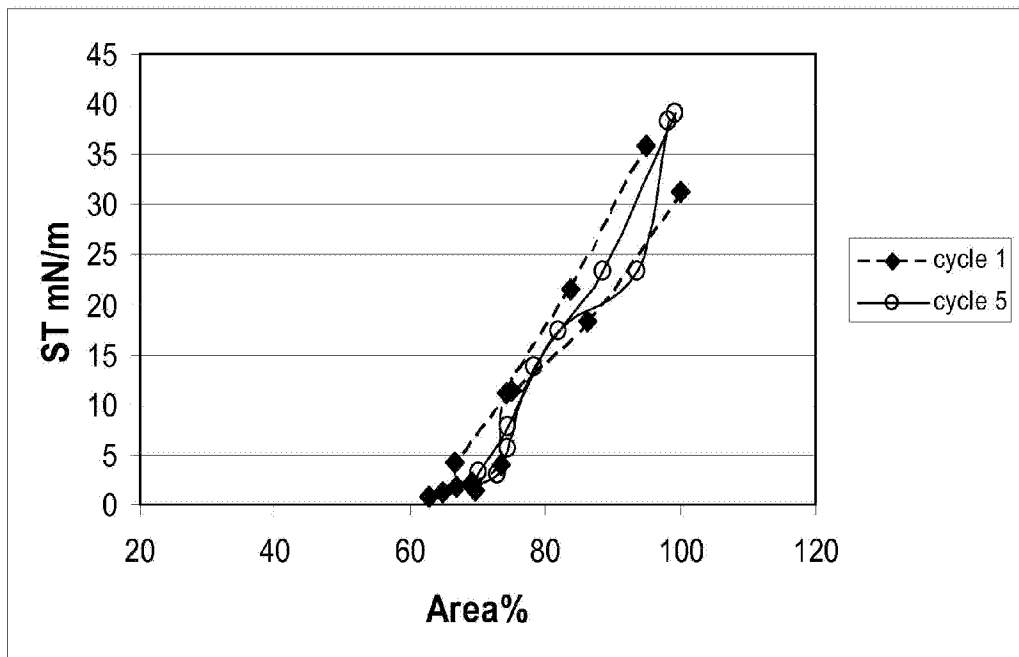


Fig. 5

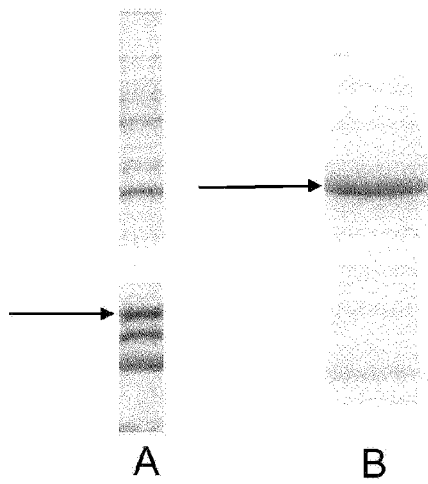


Fig. 6

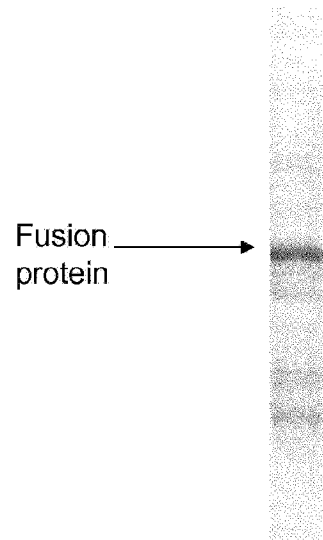


Fig. 7

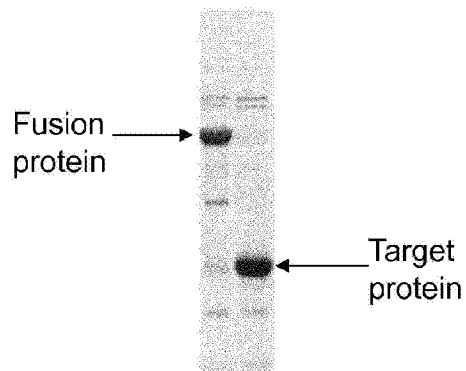


Fig. 8

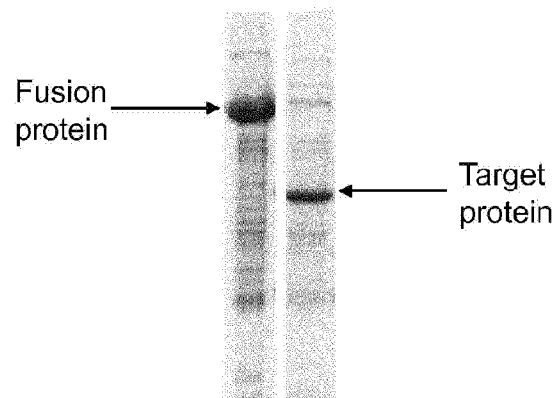


Fig. 9