(12) STANDARD PATENT (11) Application No. **AU 2013255815 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Spider silk fusion protein structures without repetitive fragment for binding to an organic target

(51) International Patent Classification(s)
B01D 15/38 (2006.01) ***C12N 15/62*** (2006.01)
C07K 14/435 (2006.01)

(21) Application No: **2013255815** (22) Date of Filing: **2013.05.02**

(87) WIPO No: **WO13/164405**

(30) Priority Data

(31) Number	(32) Date	(33) Country
12166393.4	**2012.05.02**	**EP**

(43) Publication Date: **2013.11.07**
(44) Accepted Journal Date: **2018.03.29**

(71) Applicant(s)
Spiber Technologies AB

(72) Inventor(s)
Hedhammar, My

(74) Agent / Attorney
Spruson & Ferguson, GPO Box 3898, Sydney, NSW, 2001, AU

(56) Related Art
MORGAN ET AL: "Characterization and optimization of RGD-containing silk blends to support osteoblastic differentiation", BIOMATERIALS, vol. 29, no. 16, 5 March 2008, pages 2556-2563
ANNA RISING ET AL: "Spider silk proteins: recent advances in recombinant production, structure-function relationships and biomedical applications", CMLS CELLULAR AND MOLECULAR LIFE SCIENCES, vol. 68, no. 2, 29 July 2010, pages 169-184
D. Huemmerich et al. "Primary Structure Elements of Spider Dragline Silks and Their Contribution to Protein Solubility", Biochemistry, Vol 43, pages 13604-13612, (2004)
J. Nilsson et al. "Affinity Fusion Strategies for Detection, Purification, and Immobilization of Recombinant Proteins", Protein Expression and Purification, Vol 11, No 1, pages 1-16, (October 1997)

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 November 2013 (07.11.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/164405 A1

(51) International Patent Classification:
B01D 15/38 (2006.01) *C12N 15/62* (2006.01)
C07K 14/435 (2006.01)

(21) International Application Number:
PCT/EP2013/059146

(22) International Filing Date:
2 May 2013 (02.05.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
12166393.4 2 May 2012 (02.05.2012) EP

(71) Applicant: **SPIBER TECHNOLOGIES AB** [SE/SE]; Husargatan 3, Box 575, S-751 23 Uppsala (SE).

(72) Inventor: **HEDHAMMAR, My**; Odengatan 40, S-113 51 Stockholm (SE).

(74) Agent: **HENRIKSSON, Mikael**; Awapatent AB, P.O. Box 45086, S-104 30 Stockholm (SE).

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

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

Published:

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

— *with sequence listing part of description (Rule 5.2(a))*

WO 2013/164405 A1

(54) Title: SPIDER SILK FUSION PROTEIN STRUCTURES WITHOUT REPETITIVE FRAGMENT FOR BINDING TO AN ORGANIC TARGET

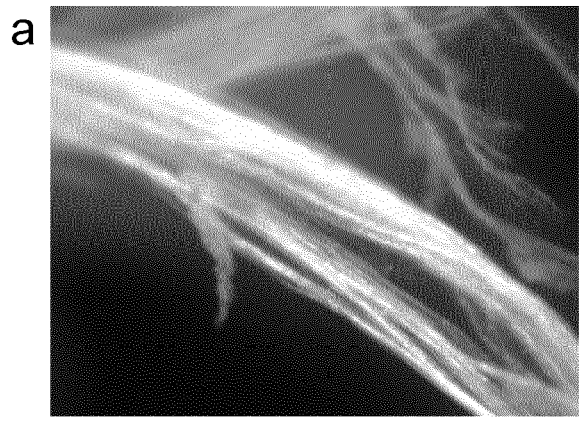

Fig 3

(57) Abstract: A recombinant fusion protein comprising the moieties Band CTis provided. B is a non-spidroin moiety which provides the capacity of selective interaction with an organic target. CTis a moiety of from 70 to 120 amino acid residues and is derived from the C-terminal fragment of a spider silk protein. The fusion protein is not comprising any moiety derived from the repetitive fragment of a spider silk protein.

SPIDER SILK FUSION PROTEIN STRUCTURES WITHOUT REPETITIVE FRAGMENT FOR BINDING TO AN ORGANIC TARGET

Technical field of the invention

The present invention relates to the field of recombinant fusion proteins, and more specifically to novel fusion proteins comprising moieties derived from spider silk proteins (spidroins). The present invention provides methods for providing a protein structure which is a polymer comprising a recombinant fusion protein, which is comprising moieties derived from spidroins. There is also provided novel protein structures for binding to an organic target.

Background to the invention

In applied protein chemistry, it is a common problem how to formulate or present a biologically active peptide or protein to the relevant site of activity, typically an organic target, such as a nucleic acid, a protein, a complex of proteins, or a complex of a protein(s) and/or lipids and/or carbohydrates and/or a nucleic acid(s). The simplest solution is simply to provide an aqueous solution of the biologically active peptide or protein. Many applications do however require some further means to achieve the desired goal. For instance, the peptides/proteins may be associated with a lipid mixture or chemically immobilized to a support structure.

Applications for peptides/proteins immobilized to a support structure include preparative and analytical separation procedures, such as bioprocesses, chromatography, cell capture and culture, active filters, and diagnostics. Structures based on extracellular matrix proteins, e.g. collagen, are disclosed in EP 704 532 and EP 985 732.

It has also been suggested to use spider silk proteins in a supporting structure. Spider silks are nature´s high-performance polymers, obtaining extraordinary toughness and extensibility due to a combination of strength and elasticity. Spiders have up to seven different glands which produce a

variety of silk types with different mechanical properties and functions. Dragline silk, produced by the major ampullate gland, is the toughest fiber. It consists of two main polypeptides, mostly referred to as major ampullate spidroin (MaSp) 1 and 2, but e.g. as ADF-3 and ADF-4 in *Araneus diadematus*. These proteins have molecular masses in the range of 200-720 kDa. Spider dragline silk proteins, or MaSps, have a tripartite composition; a non-repetitive N-terminal domain, a central repetitive region comprised of many iterated poly-Ala/Gly segments, and a non-repetitive C-terminal domain. It is generally believed that the repetitive region forms intermolecular contacts in the silk fibers, while the precise functions of the terminal domains are less clear. It is also believed that in association with fiber formation, the repetitive region undergoes a structural conversion from random coil and α-helical conformation to β-sheet structure. The C-terminal region of spidroins is generally conserved between spider species and silk types.

W0 07/078239 and Stark, M. *et al.*, Biomacromolecules 8: 1695-1701, (2007) disclose a miniature spider silk protein consisting of a repetitive fragment with a high content of Ala and Gly and a C-terminal fragment of a protein, as well as soluble fusion proteins comprising the spider silk protein. Fibers of the spider silk protein are obtained spontaneously upon liberation of the spider silk protein from its fusion partner.

Rising, A. *et al.*, CMLS 68(2): 169-184 (2010) reviews advances in the production of spider silk proteins. Fibres are created from a miniaturized spidroin construct consisting of four repeats of a segment derived from the repetitive region and the non-repetitive C-terminal domain of a spider silk protein.

US 2009/0263430 discloses chemical coupling of the enzyme β-galactosidase to films of a miniature spider silk protein. However, chemical coupling may require conditions which are unfavourable for protein stability and/or function. Proteins containing multiple repeats of a segment derived from the repetitive region of spider silk proteins have been designed to include a RGD cell-binding segment (Bini, E *et al.*, Biomacromolecules 7:3139-3145 (2006); Morgan *et al.*, Biomaterials 29(16): 2556-2563 (2008)) and/or a R5 peptide (Wong Po Foo, C *et al.*, Proc Natl Acad Sci 103 (25):

9428-9433 (2006)) or other protein segments involved in mineralization (Huang, J *et al.*, Biomaterials 28: 2358-2367 (2007); WO 2006/076711). In these prior art documents, films are formed by solubilizing the fusion proteins in the denaturing organic solvent hexafluoroisopropanol (HFIP) and drying.

US 2005/261479 A1 discloses a method of for purification of recombinant silk proteins consisting of a repetitive fragment and an affinity tag, involving magnetic affinity separation of individual silk proteins from complex mixtures without formation of silk protein fibers or other polymer structures.

Known supporting structures and associated techniques have certain drawbacks with regard to e.g. economy, efficiency, stability, regenerating capacity, bioactivity and biocompatibility.

Summary of the Invention

According to one aspect of the present invention provides a recombinant fusion protein comprising the moieties B and CT, wherein: B is a non-spidroin moiety which provides the capacity of selective interaction with an organic target, wherein the B moiety is a protein or polypeptide fragment comprising more than 30 amino acid residues; and wherein the B moiety has less than 30% identity to any of SEQ ID NOS: 6-10; and CT is a moiety of from 70 to 120 amino acid residues and is derived from the C-terminal fragment of a spider silk proteinhas at least 80% identity to any one of SEQ ID NO: 7 and SEQ ID NOS: 14-44; wherein the CT moiety provides the capacity of forming a polymer; with the proviso that the fusion protein is not comprising any moiety derived from the repetitive fragment of a spider silk protein.

It is an aspect of the present invention to provide novel recombinant fusion proteins. In particular, it is an aspect of the present invention to provide novel recombinant fusion proteins which are capable of selective interaction with an organic target.

It is an aspect of the present invention to provide a novel protein

structure that is capable of selective interaction with an organic target.

It is also an aspect of the present invention to provide a protein structure that is capable of selective interaction with an organic target,

wherein the structure is formed without use of harsh solvents which may have

2013255815 07 Mar 2018

an unpredictable effect on the secondary structure or activity of the protein and/or remain in the protein structure.

It is one aspect of the present invention to provide a stable protein structure that is capable of selective interaction with an organic target, which protein structure can readily be regenerated after use, e.g. with chemical treatment.

It is another aspect of the present invention to provide a stable protein structure that is biocompatible and suitable for cell culture and as an implant.

It is yet another aspect of the invention to provide a protein structure with a high density of evenly spaced functionalities that are capable of selective interaction with an organic target.

It is a further aspect of the invention to provide a protein structure which

maintains its selective binding ability upon storage at +4°C or at room temperature for months.

It is also an aspect of the invention to provide a protein structure which is autoclavable, i.e. maintains its selective binding ability after sterilizing heat treatment.

For these and other aspects that will be evident from the following disclosure, the present invention provides according to a first aspect a fusion protein and a protein structure consisting of polymers comprising as a repeating structural unit the fusion protein as set out in the claims.

According to a related aspect, the present invention provides an

isolated nucleic acid encoding the fusion protein and a method of producing the fusion protein as set out in the claims.

The present invention provides according to another aspect a method

for providing a protein structure as set out in the claims.

The present invention provides according to a further aspect an affinity

medium as set out in the claims.

The present invention provides according to one aspect a cell scaffold material as set

out in the claims. According to a related aspect, the present invention also provides a combination of cells and a cell scaffold material according to the claims.

The present invention provides according to an aspect novel uses of a protein structure and a fusion protein as set out in the claims.

The present invention provides according to another aspect a method for separation of an organic target from a sample as set out in the claims.

The present invention provides according to a further aspect a method

for immobilization and optionally cultivation of cells as set out in the claims.

Brief description of the drawings

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

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

Fig. 3 shows microscopic pictures of silk fused ABD in fiber format.

Fig. 4 shows SDS-PAGE gels of fractions of human blood plasma eluted from ABD-NTCT and ABD-CT films.

Fig. 5 shows microscopic pictures of silk fused M4 fragments in fiber and foam format.

Fig. 6 shows fluorescence microscopy pictures of Atto-565-biotin bound to M4-NTCT and M4-CT films.

Fig. 7 shows microscopic pictures of silk fused antibody fragments in foam format.

Fig. 8 shows an antigen binding analysis of pure and silk fused antibody fragments.

Fig. 9 shows a macroscopic foam of silk fused xylanase.

List of appended sequences

SEQ ID NO	
1	4Rep
2	4RepCT
3	NT4Rep
4	NT5Rep
5	NT4RepCTHis
6	NT
7	CT
8	consensus NT sequence
9	consensus CT sequence
10	repetitive sequence from *Euprosthenops australis* MaSp1
11	consensus G segment sequence 1
12	consensus G segment sequence 2
13	consensus G segment sequence 3
14	CT Euprosthenops sp MaSp1
15	CT Euprosthenops australis MaSp1

SEQ ID NO	
16	CT Argiope trifasciata MaSp1
17	CT Cyrtophora moluccensis Sp1
18	CT Latrodectus geometricus MaSp1
19	CT Latrodectus hesperus MaSp1
20	CT Macrothele holsti Sp1
21	CT Nephila clavipes MaSp1
22	CT Nephila pilipes MaSp1
23	CT Nephila madagascariensis MaSp1
24	CT Nephila senegalensis MaSp1
25	CT Octonoba varians Sp1
26	CT Psechrus sinensis Sp1
27	CT Tetragnatha kauaiensis MaSp1
28	CT Tetragnatha versicolor MaSp1
29	CT Araneus bicentenarius Sp2
30	CT Argiope amoena MaSp2
31	CT Argiope aurantia MaSp2
32	CT Argiope trifasciata MaSp2
33	CT Gasteracantha mammosa MaSp2
34	CT Latrodectus geometricus MaSp2
35	CT Latrodectus hesperus MaSp2
36	CT Nephila clavipes MaSp2
37	CT Nephila madagascariensis MaSp2
38	CT Nephila senegalensis MaSp2
39	CT Dolomedes tenebrosus Fb1
40	CT Dolomedes tenebrosus Fb2
41	CT Araneus diadematus ADF-1
42	CT Araneus diadematus ADF-2
43	CT Araneus diadematus ADF-3
44	CT Araneus diadematus ADF-4
45	NT Euprosthenops australis MaSp1
46	NT Latrodectus geometricus MaSp1
47	NT Latrodectus hesperus MaSp1

SEQ ID NO	
48	NT Nephila clavipes MaSp1
49	NT Argiope trifasciata MaSp2
50	NT Latrodectus geometricus MaSp2
51	NT Latrodectus hesperus MaSp2
52	NT Nephila inaurata madagascariensis MaSp2
53	NT Nephila clavipes MaSp2
54	NT *Argiope bruennichi* cylindriform spidroin 1
55	NT *Nephila clavata* cylindriform spidroin 1
56	NT *Latrodectus hesperus* tubuliform spidroin
57	NT *Nephila clavipes* flagelliform silk protein
58	NT Nephila inaurata madagascariensis flagelliform silk protein
59	His_6NT-CT
60	His_6NTNT-CT
61	His_6Z-CT
62	His_6CT-Z
63	His_6Z-NTCT
64	His_6NTCT-Z
65	His_6Z-NTNTCT
66	His_6NTNTCT-Z
67	His_6-ABD-NTCT (DNA)
68	His_6-ABD-NTCT
69	His_6-ABD-CT (DNA)
70	His_6-ABD-CT
71	His_6-M4-NTCT (DNA)
72	His_6-M4-NTCT
73	His_6-M4-CT (DNA)
74	His_6-M4-CT
75	His_6-scFv1-NTCT (DNA)
76	His_6-scFv1-NTCT
77	His_6-scFv1-CT (DNA)
78	His_6-scFv1-CT
79	His_6Xyl-NTCT (DNA)

SEQ ID NO

80	His_6Xyl-NTCT
81	His_6Xyl-CT (DNA)
82	His_6Xyl-CT
83	His_6EGF-NTCT (DNA)
84	His_6EGF-NTCT
85	His_6EGF-CT (DNA)
86	His_6EGF-CT

Detailed description of the invention

The present invention is generally based on the insight that solid protein structures capable of selective interaction with an organic target can be prepared in the form of polymers of a recombinant fusion protein as a repeating structural unit. The fusion protein is comprising at least one non-spidroin moiety, preferably of more than 30 amino acid residues, that is capable of selective interaction with the organic target, and moieties corresponding to at least the C-terminal fragment of a spider silk protein, with the proviso that the fusion protein is not comprising any moiety derived from the repetitive fragment of a spider silk protein.

Surprisingly, the moiety derived from the spider silk protein can be induced to rearrange structurally and as a result form polymeric, solid structures, while the non-spidroin moiety is not structurally rearranged but maintains its desirable structure and function, i.e. capability of selective interaction with the organic target. The protein structures can be obtained without a chemical coupling step or a denaturing method step, which facilitates the procedure and improves the chances of obtaining a fusion protein with maintained functionality of its moieties, in particular when the functions are dependent on the secondary structure of the moieties. The formation of these fusion protein polymers can be tightly controlled, and this insight has been developed into further novel protein structures, methods of producing the protein structures and uses of the protein structures in various applications and methods.

The fusion protein according to the invention thus harbors both the desired selective interaction activity and an internal solid support activity that is employed in the protein structure under physiological conditions. It must be considered as surprising that the binding activity of the fusion protein is maintained although the non-spidroin moiety is covalently attached to the spidroin moiety when the latter is structurally rearranged to form polymeric, solid structures. In fact, the heat and/or chemical stability and/or binding activity of the moiety providing the selective interaction activity may be increased when integrated in a fusion protein structure according to the invention. The protein structure also provides a high and predictable density of the selective interaction activity towards an organic target. Losses of valuable protein moieties with selective interaction activity are minimized, since all expressed protein moieties are associated with the solid support.

The polymers which are formed from the fusion proteins according to the invention are solid structures and are useful for their physical properties, especially the useful combination of high strength, elasticity and light weight. A particularly useful feature is that the spidroin-derived moieties of the fusion protein are biochemically robust and suitable for regeneration, e.g. with acid, base or chaotropic agents, and suitable for heat sterilization, e.g. autoclaving at 120°C for 20 min. The polymers are also useful for their ability to support cell adherence and growth. The properties derived from dragline silk are attractive in development of new materials for medical or technical purposes. In particular, protein structures according to the invention are useful in preparative and analytical separation procedures, such as chromatography, cell capture, selection and culture, active filters, and diagnostics. Protein structures according to the invention are also useful in medical devices, such as implants and medical products, such as wound closure systems, band-aids, sutures, wound dressings, and scaffolds for cell immobilization, cell culture, tissue engineering and guided cell regeneration.

The present invention provides a recombinant fusion protein that is capable of selective interaction with an organic target, which fusion protein is comprising the moieties **B** and **CT**, and optionally **NT**. The present invention also provides a protein structure that is capable of selective interaction with

an organic target, wherein said protein structure is a polymer comprising, and optionally consisting of, the recombinant fusion protein according to the invention, i.e. comprising, and optionally consisting of, the moieties **B** and **CT**, and optionally **NT**. The fusion protein according to the invention is not comprising any moiety derived from the repetitive (**REP**) fragment of a spider silk protein.

Although the **CT** and the optional **NT** moieties of the fusion proteins in the examples by necessity relate to specific proteins, e.g. proteins derived from major spidroin 1 (MaSp1) from *Euprosthenops australis*, it is considered that the present disclosure is applicable to any structurally similar moieties for the purpose of producing fusion protein structures according to the invention. Furthermore, although the **B** moiety which provides the selective interaction activity of the fusion proteins in the examples by necessity relate to specific protein moieties, e.g. moieties derived from protein A, protein G and streptavidin, it is considered that the present disclosure is applicable to any structurally and/or functionally similar **B** moiety for the purpose of producing fusion protein structures according to the invention, capable of selective interaction with an organic target.

Specific fusion proteins according to the invention are defined by the formulas $\mathbf{B_x\text{-}CT\text{-}B_z}$,, wherein x and z are integers from 0 to 5; and $x+z \geq 1$, optionally further containing one or more **NT** moieties, e.g. 1-2 **NT** moieties, at either end of the fusion protein or between any two protein moieties in the fusion protein, e.g. $\mathbf{NT\text{-}B_x\text{-}CT\text{-}B_z}$, $\mathbf{B_x\text{-}NT\text{-}CT\text{-}B_z}$, $\mathbf{B_x\text{-}CT\text{-}NT\text{-}B_z}$, $\mathbf{B_x\text{-}CT\text{-}B_z\text{-}NT}$, $\mathbf{NT\text{-}NT\text{-}B_x\text{-}CT\text{-}B_z}$ or $\mathbf{B_x\text{-}CT\text{-}B_z\text{-}NT\text{-}NT}$,. If $x+z > 1$, i.e. if there are two or more **B** moieties, they may be identical or different. The two or more **B** moieties may have capacity of selective interaction with the same organic target or with different organic targets. It is preferred that the two or more **B** moieties are substantially identical, each having capacity of selective interaction with the same organic target. Alternatively, it is preferred that the two or more **B** moieties are not identical, and that they together provide the capacity of selective interaction with the organic target.

In preferred fusion proteins according to the invention, x and z are integers from 0 to 2, preferably from 0 to 1. In certain preferred fusion proteins

according to the invention, either x or z are 0, i.e. the fusion proteins are defined by the formulas **B_x-CT**, and **CT-B_z**, wherein either x or z is an integer from 1 to 5, and optionally containing 1-2 **NT** moieties.. In preferred fusion proteins according to the invention, x and z are integers from 0 to 1; and x+z = 1. Thus, certain preferred fusion proteins according to the invention are defined by the formulas **B-CT**, and **CT- B**, optionally containing 1-2 **NT** moieties. In preferred fusion proteins according to the invention, the optional **NT** moiety is missing.

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. Furthermore, the fusion protein according to the invention and the corresponding genes are chimeric, i.e. the protein/gene moieties are derived from at least two different species. The **CT** moiety and the optional **NT** moiety are derived from a spider silk protein. For avoidance of doubt, the **B** moiety according to the invention is a non-spidroin protein or polypeptide, i.e. it is not derived from a spider silk protein. In particular, the **B** moiety according to the invention is not derived from the C-terminal, repetitive or N-terminal fragments of a spider silk protein.

The fusion protein typically consists of from 170 to 2000 amino acid residues, such as from 170 to 1000 amino acid residues, such as from 170 to 600 amino acid residues, preferably from 170 to 500 amino acid residues, such as from 170 to 400 amino acid residues. The small size is advantageous because longer proteins containing spider silk protein fragments may form amorphous aggregates, which require use of harsh solvents for solubilisation

and polymerisation. The recombinant fusion protein may contain more than 2000 residues, in particular in cases where the spider silk protein more than one **B** moiety and/or when it contains a **NT** moiety, e.g. 1-2 **NT** moieties.

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 non-natural proteins with a high degree of identity and/or similarity to the known spider silk proteins.

Consequently, the term "non-spidroin" implies proteins that are not derived from a spider silk protein, i.e. with a low (or no) degree of identity and/or similarity to spider silk proteins.

The protein structure according to the invention is capable of selective interaction with an organic target. This capacity resides in the fusion protein according to the invention, and more specifically in the **B** moiety of the fusion protein. Any interactions of the **CT** moiety and the optional **NT** moiety with organic molecules are not encompassed by the term "capable of selective interaction with an organic target". For avoidance of doubt, the term "capable of selective interaction with an organic target" does not encompass dimerization, oligomerization or polymerization of the fusion proteins according to the invention that rely on interactions involving the **CT** moiety and/or the optional **NT** moiety.

The term "organic target" encompasses all chemical molecules containing carbon with the exception of what is traditionally considered inorganic molecules by the skilled person, e.g. carbonates, simple oxides of carbon, cyanides, diamond and graphite. For avoidance of doubt, inorganic molecules, salts and ions, such as silica and calcium chloride, are not organic. The organic target may be a complex containing or consisting of organic molecules, e.g. a receptor complex on a cell surface. The organic target may be a monomer, dimer, oligomer or polymer of one or more organic molecule types, which may be held together by covalent bonds or other types of association. It may of course also simply be a single organic molecule.

Preferred organic targets according to the invention include, but are not limited to, nucleic acids, proteins and polypeptides, lipids and carbohydrates, as well as combinations thereof. Further preferred organic targets according to the invention include, but are not limited to, immunoglobulins, molecules comprising immunoglobulin or derivatives thereof, albumin, molecules comprising albumin or derivatives thereof, biotin, molecules comprising biotin or derivatives or analogues thereof, and biological disease markers, e.g. from blood, serum, urine, saliva or other samples from body tissues.

In the context of the present invention, "specific" or "selective" interaction of a ligand, e.g. a **B** moiety of the fusion protein according to the invention with its target means that the interaction is such that a distinction between specific and non-specific, or between selective and non-selective, interaction becomes meaningful. The interaction between two proteins is sometimes measured by the dissociation constant. The dissociation constant describes the strength of binding (or affinity) between two molecules. Typically the dissociation constant between an antibody and its antigen is from 10^{-7} to 10^{-11} M. However, high specificity does not necessarily require high affinity. Molecules with low affinity (in the molar range) for its counterpart have been shown to be as specific as molecules with much higher affinity. In the case of the present invention, a specific or selective interaction refers to the extent to which a particular method can be used to determine the presence and/or amount of a specific protein, the target protein or a fragment thereof, under given conditions in the presence of other proteins in a sample of a naturally occurring or processed biological or biochemical fluid. In other words, specificity or selectivity is the capacity to distinguish between related proteins. Specific and selective are sometimes used interchangeably in the present description.

The fusion protein according to the invention may also contain one or more linker peptides. The linker peptide(s) may be arranged between any moieties of the fusion protein, e.g. between the **CT** and **NT** moieties, between two **B** moieties, and between **B** and **CT** moieties, or may be arranged at either terminal end of the fusion protein. If the fusion protein contains two or more **B** moieties, the linker peptide(s) may also be arranged in between two

B 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. His_6-spacer-His_6-. 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 other relevant moieties, typically the **B** 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.

The **CT** and **B** moieties are linked directly or indirectly to one another. A direct linkage implies a direct covalent binding between the moieties without intervening sequences, such as linkers. An indirect linkage also implies that the moieties are linked by covalent bonds, but that there are intervening sequences, such as linkers and/or one or more further moieties, e.g. 1-2 **NT** moieties.

The **B** moiety or moieties may be arranged internally or at either end of the fusion protein, i.e. C-terminally arranged or N-terminally arranged. It is preferred that the **B** moiety or moieties are arranged at the N-terminal end of the fusion protein. 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.

A preferred fusion protein has the form of an N-terminally arranged **B** moiety, coupled by a linker peptide of 1-30 amino acid residues, such as 1-10 amino acid residues, to a C-terminally arranged **CT** moiety. 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.

Another preferred fusion protein has the form of an N-terminally arranged **B** moiety coupled directly to a C-terminally arranged **CT** moiety. 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.

The protein structure according to the invention is a polymer comprising as a repeating structural unit recombinant fusion proteins according to the invention, which implies that it contains an ordered plurality of fusion proteins according to the invention, typically well above 100 fusion protein units, e.g. 1000 fusion protein units or more. Optionally, the polymer may comprise as a further repeating structural unit complementary proteins without a **B** moiety, preferably proteins derived from spider silk. This may be advantageous if the **B** moiety of the fusion protein is large and/or bulky. These complementary proteins typically comprise, and may even consist of, a **CT** moiety, and optionally a further **NT** moiety, e.g. 1-2 **NT** moieties. Preferred complementary proteins according to the invention can have any of the structures set out herein with a deleted **B** moiety. It is preferred that the complementary fusion protein is substantially identical to the fusion protein with a deleted **B** moiety. However, it is preferred that the protein structure according to the invention is a polymer consisting of recombinant fusion proteins according to the invention as a repeating structural unit, i.e. that the protein structure according to the invention is a polymer of the recombinant fusion protein according to the invention.

The magnitude of fusion units in the polymer implies that the protein structure obtains a significant size. In a preferred embodiment, the protein structure has a size of at least 0.1 µm in at least two dimensions. Thus, the term "protein structure" as used herein relates to fusion protein polymers having a thickness of at least 0.1 µm, preferably macroscopic polymers that are visible to the human eye, i.e. having a thickness of at least 1 µm. The term "protein structure" does not encompass unstructured aggregates or precipitates. While monomers of the fusion protein are water soluble, it is understood that the protein structures according to the invention are solid

structures, i.e. not soluble in water. The protein structures are polymers comprising as a repeating structural unit monomers of the recombinant fusion proteins according to the invention.

It is preferable that the protein structure according to the invention is in a physical form selected from the group consisting of fiber, film, foam, net, mesh, sphere and capsule.

It is preferable that the protein structure according to the invention is a fiber or film with a thickness of at least 1 nm or at least 0.1 µm, preferably at least 1 µm. It is preferred that the fiber or film has a thickness in the range of 1 nm -400 m, such as 1-400 µm, and preferably 60-120 µm. It is preferred that fibers have a length in the range of 0.5-300 cm, preferably 1-100 cm. Other preferred ranges are 0.5-30 cm and 1-20 cm. The fiber has the capacity to remain intact during physical manipulation, i.e. can be used for spinning, weaving, twisting, crocheting and similar procedures. The film is advantageous in that it is coherent and adheres to solid structures, e.g. the plastics in microtiter plates. This property of the film facilitates washing and regeneration procedures and is very useful for separation purposes. A particularly useful protein structure is a film or a fiber wherein the **B** moiety is the Z domain derived from staphylococcal protein A or a protein fragment having at least 70% identity thereto.

It is also preferred that the protein structure according to the invention has a tensile strength above 1 MPa, preferably above 2 MPa, more preferably 10 MPa or higher. It is preferred that the protein structure according to the invention has a tensile strength above 100 MPa, more preferably 200 MPa or higher.

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.

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.

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.

The **CT** moiety is a protein fragment containing from 70 to 120 amino acid residues and is derived from the C-terminal fragment of a spider silk protein. The expression "derived from" implies in the context of the **CT** moiety according to the invention that it has a high degree of similarity to the C-terminal 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. A consensus sequence of the C-terminal regions of MaSp1 and MaSp2 is provided as SEQ ID NO: 9. In Fig 1, the following MaSp proteins are aligned, denoted with GenBank accession entries where applicable (SEQ ID NOS: 14-44):

TABLE 1 - Spidroin **CT** moieties

Species and spidroin protein	Entry
Euprosthenops sp MaSp1 (Pouchkina-Stantcheva, NN & McQueen-Mason, SJ, *ibid*)	Cthyb_Esp
Euprosthenops australis MaSp1	CTnat_Eau
Argiope trifasciata MaSp1	AF350266_At1

Species and spidroin protein	Entry
Cyrtophora moluccensis Sp1	AY666062_Cm1
Latrodectus geometricus MaSp1	AF350273_Lg1
Latrodectus hesperus MaSp1	AY953074_Lh1
Macrothele holsti Sp1	AY666068_Mh1
Nephila clavipes MaSp1	U20329_Nc1
Nephila pilipes MaSp1	AY666076_Np1
Nephila madagascariensis MaSp1	AF350277_Nm1
Nephila senegalensis MaSp1	AF350279_Ns1
Octonoba varians Sp1	AY666057_Ov1
Psechrus sinensis Sp1	AY666064_Ps1
Tetragnatha kauaiensis MaSp1	AF350285_Tk1
Tetragnatha versicolor MaSp1	AF350286_Tv1
Araneus bicentenarius Sp2	ABU20328_Ab2
Argiope amoena MaSp2	AY365016_Aam2
Argiope aurantia MaSp2	AF350263_Aau2
Argiope trifasciata MaSp2	AF350267_At2
Gasteracantha mammosa MaSp2	AF350272_Gm2
Latrodectus geometricus MaSp2	AF350275_Lg2
Latrodectus hesperus MaSp2	AY953075_Lh2
Nephila clavipes MaSp2	AY654293_Nc2
Nephila madagascariensis MaSp2	AF350278_Nm2
Nephila senegalensis MaSp2	AF350280_Ns2
Dolomedes tenebrosus Fb1	AF350269_DtFb1
Dolomedes tenebrosus Fb2	AF350270_DtFb2
Araneus diadematus ADF-1	U47853_ADF1
Araneus diadematus ADF-2	U47854_ADF2
Araneus diadematus ADF-3	U47855_ADF3
Araneus diadematus ADF-4	U47856_ADF4

It is not critical which specific **CT** moiety is present in spider silk proteins according to the invention, as long as the **CT** moiety is not entirely missing. Thus, the **CT** moiety according to the invention can be selected from any of the amino acid sequences shown in Fig 1 and Table 1 (SEQ ID NOS:

14-44) or sequences with a high degree of similarity. A wide variety of C-terminal sequences can be used in the spider silk protein according to the invention.

The sequence of the **CT** moiety according to the invention has at least 50% identity, preferably at least 60%, more preferably at least 65% identity, or even at least 70% identity, to the consensus amino acid sequence SEQ ID NO: 9, which is based on the amino acid sequences of Fig 1 (SEQ ID NOS: 14-44).

A representative **CT** moiety according to the invention is the *Euprosthenops australis* sequence SEQ ID NO: 7, Thus, according to a preferred aspect of the invention, the **CT** moiety has at least 80%, preferably at least 90%, such as at least 95%, identity to SEQ ID NO: 7 or any individual amino acid sequence of Fig 1 and Table 1 (SEQ ID NOS: 14-44). In preferred aspects of the invention, the **CT** moiety is identical to SEQ ID NO: 7 or any individual amino acid sequence of Fig 1 and Table 1.

The **CT** moiety typically consists of from 70 to 120 amino acid residues. It is preferred that the **CT** moiety contains at least 70, or more than 80, preferably more than 90, amino acid residues. It is also preferred that the **CT** moiety contains at most 120, or less than 110 amino acid residues. A typical **CT** moiety contains approximately 100 amino acid residues.

The optional **NT** moiety is a protein fragment containing from 100 to 160 amino acid residues and is derived from the N-terminal fragment of a spider silk protein. The expression "derived from" implies in the context of the **NT** moiety according to the invention that it has a high degree of similarity to the N-terminal amino acid sequence of spider silk proteins. As shown in Fig 2, this amino acid sequence is well conserved among various species and spider silk proteins, including MaSp1 and MaSp2. In Fig 2, the following spidroin **NT** moieties are aligned, denoted with GenBank accession entries where applicable (SEQ ID NOS: 45-58):

TABLE 2 - Spidroin **NT** moieties

Code	Species and spidroin protein	GenBank acc. no.
Ea MaSp1	Euprosthenops australis MaSp 1	AM259067
Lg MaSp1	Latrodectus geometricus MaSp 1	ABY67420

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

Only the part corresponding to the N-terminal moiety 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)).

It is not critical which specific **NT** moiety is present in spider silk proteins according to the invention. Thus, the **NT** moiety according to the invention can be selected from any of the amino acid sequences shown in Fig 2 and Table 2 (SEQ ID NOS: 45-58) or sequences with a high degree of similarity. A wide variety of N-terminal sequences can be used in the spider silk protein according to the invention. Based on the homologous sequences of Fig 2, the following sequence constitutes a consensus **NT** 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).

The sequence of the **NT** 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 2. In a preferred embodiment, the sequence of the **NT** 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 **NT** moiety according to the invention has furthermore 70%, preferably 80%, similarity to the consensus amino acid sequence SEQ ID NO: 8.

A representative **NT** moiety according to the invention is the *Euprosthenops australis* sequence SEQ ID NO: 6. According to a preferred embodiment of the invention, the **NT** moiety has at least 80% identity to SEQ ID NO: 6 or any individual amino acid sequence in Fig 2 (SEQ ID NOS: 45-58). In preferred embodiments of the invention, the **NT** moiety has at least 90%, such as at least 95% identity, to SEQ ID NO: 6 or any individual amino acid sequence in Fig 2. In preferred embodiments of the invention, the **NT** moiety is identical to SEQ ID NO: 6 or any individual amino acid sequence in Fig 2, in particular to Ea MaSp1 (SEQ ID NO: 45).

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

The fusion protein is not comprising any moiety derived from the repetitive fragment of a spider silk protein. A typical moiety that is derived from the repetitive fragment of a spider silk protein and thus void in the present fusion protein is a **REP** moiety, i.e. a protein fragment containing from 70 to 300 amino acid residues that is derived from the repetitive fragment of a spider silk protein. In fusion proteins lacking a REP moiety, non-specific binding of the B moiety to other molecules than its antigen target has advantageously been observed to decrease even further. It is also particularly surprising that solid structures are formed spontaneously from fusion proteins lacking a REP moiety.

The **REP** moiety has a repetitive character, alternating between alanine-rich stretches and glycine-rich stretches. The **REP** moiety generally contains more than 70, such as more than 140, and less than 300, preferably less than 240, such as less than 200, amino acid residues, and can itself be divided into several **L** (linker) segments, **A** (alanine-rich) segments and **G** (glycine-rich) segments, as will be explained in more detail below. Typically, said linker segments, which are optional, are located at the **REP** moiety terminals, while the remaining segments are in turn alanine-rich and glycine-

rich. Thus, the **REP** moiety can generally have either of the following structures, wherein n is an integer:

$L(AG)_nL$, such as $LA_1G_1A_2G_2A_3G_3A_4G_4A_5G_5L$;

$L(AG)_nAL$, such as $LA_1G_1A_2G_2A_3G_3A_4G_4A_5G_5A_6L$;

$L(GA)_nL$, such as $LG_1A_1G_2A_2G_3A_3G_4A_4G_5A_5L$; or

$L(GA)_nGL$, such as $LG_1A_1G_2A_2G_3A_3G_4A_4G_5A_5G_6L$.

It follows that it is not critical whether an alanine-rich or a glycine-rich segment is adjacent to the N-terminal or C-terminal linker segments. It is preferred that n is an integer from 2 to 10, preferably from 2 to 8, preferably from 4 to 8, more preferred from 4 to 6, i.e. n=4, n=5 or n=6.

The alanine content of the **REP** moiety is typically above 20%, preferably above 25%, more preferably above 30%, and below 50%, preferably below 40%, more preferably below 35%.

Now turning to the segments that constitute the **REP** moiety, it shall be emphasized that each segment is individual, i.e. any two **A** segments, any two **G** segments or any two **L** segments of a specific **REP** moiety may be identical or may not be identical. Thus, it is not a general feature that each type of segment is identical within a specific **REP** moiety. Rather, the following disclosure provides the skilled person with guidelines how to identify a **REP** moiety which is thereby considered to be derived from the repetitive fragment of a spider silk protein, and which does not constitute a part of a functional fusion protein according to the invention.

Each individual **A** segment is an amino acid sequence having from 8 to 18 amino acid residues. A vast majority of these amino acid residues are alanine residues. More specifically, from 0 to 3 of the amino acid residues are not alanine residues, and the remaining amino acid residues are alanine residues. Thus, all amino acid residues in each individual **A** segment are alanine residues, with no exception or the exception of one, two or three amino acid residues, which can be any amino acid. The alanine-replacing amino acid(s) is (are) natural amino acids, preferably individually selected from the group of serine, glutamic acid, cysteine and glycine, more preferably serine. Of course, it is possible that one or more of the **A** segments are all-alanine segments, while the remaining **A** segments contain 1-3 non-alanine residues, such as serine, glutamic acid, cysteine or glycine.

Each individual **A** segment typically has at least 80%, preferably at least 90%, more preferably 95%, most preferably 100% identity to an amino acid sequence selected from the group of amino acid residues 7-19, 43-56, 71-83, 107-120, 135-147, 171-183, 198-211, 235-248, 266-279, 294-306, 330-342, 357-370, 394-406, 421-434, 458-470, 489-502, 517-529, 553-566, 581-594, 618-630, 648-661, 676-688, 712-725, 740-752, 776-789, 804-816, 840-853, 868-880, 904-917, 932-945, 969-981, 999-1013, 1028-1042 and 1060-1073 of SEQ ID NO: 10. Each sequence of this group corresponds to a segment of the naturally occurring sequence of *Euprosthenops australis* MaSp1 protein, which is deduced from cloning of the corresponding cDNA, see WO 2007/078239. Alternatively, each individual **A** segment has at least 80%, preferably at least 90%, more preferably 95%, most preferably 100% identity to an amino acid sequence selected from the group of amino acid residues 143-152, 174-186, 204-218, 233-247 and 265-278 of SEQ ID NO: 3.

Furthermore, it has been concluded from experimental data that each individual **G** segment is an amino acid sequence of from 12 to 30 amino acid residues. It is preferred that each individual **G** segment consists of from 14 to 23 amino acid residues. At least 40% of the amino acid residues of each **G** segment are glycine residues. Typically the glycine content of each individual **G** segment is in the range of 40-60%.

Each individual **G** segment typically has at least 80%, preferably at least 90%, more preferably 95%, most preferably 100% identity to an amino acid sequence selected from the group of amino acid residues 20-42, 57-70, 84-106, 121-134, 148-170, 184-197, 212-234, 249-265, 280-293, 307-329, 343-356, 371-393, 407-420, 435-457, 471-488, 503-516, 530-552, 567-580, 595-617, 631-647, 662-675, 689-711, 726-739, 753-775, 790-803, 817-839, 854-867, 881-903, 918-931, 946-968, 982-998, 1014-1027, 1043-1059 and 1074-1092 of SEQ ID NO: 10. Each sequence of this group corresponds to a segment of the naturally occurring sequence of *Euprosthenops australis* MaSp1 protein, which is deduced from cloning of the corresponding cDNA, see WO 2007/078239. Alternatively, each individual **G** segment has at least 80%, preferably at least 90%, more preferably 95%, most preferably 100% identity to an amino acid sequence selected from the group of amino acid

residues 153-173, 187-203, 219-232, 248-264 and 279-296 of SEQ ID NO: 3. Each sequence of this group corresponds to a segment of expressed, non-natural spider silk proteins, which proteins have capacity to form silk structures under appropriate conditions. Thus, each individual **G** segment may be identical to an amino acid sequence selected from the above-mentioned amino acid segments.

There are the three subtypes of the **G** segment. This classification is based upon careful analysis of the *Euprosthenops australis* MaSp1 protein sequence (WO 2007/078239), and the information has been employed and verified in the construction of novel, non-natural spider silk proteins.

The first subtype of the **G** segment is represented by the amino acid one letter consensus sequence GQG(G/S)QGG(Q/Y)GG (L/Q)GQGGYGQGA GSS (SEQ ID NO: 11). This first, and generally the longest, **G** segment subtype typically contains 23 amino acid residues, but may contain as little as 17 amino acid residues, and lacks charged residues or contain one charged residue. Thus, this first **G** segment subtype typically contains 17-23 amino acid residues, but it is contemplated that it may contain as few as 12 or as many as 30 amino acid residues. Representative **G** segments of this first subtype are amino acid residues 20-42, 84-106, 148-170, 212-234, 307-329, 371-393, 435-457, 530-552, 595-617, 689-711, 753-775, 817-839, 881-903, 946-968, 1043-1059 and 1074-1092 of SEQ ID NO: 10. In certain embodiments, the first two amino acid residues of each **G** segment of this first subtype are not -Gln-Gln-.

The second subtype of the **G** segment is represented by the amino acid one letter consensus sequence GQGGQGQG(G/R)Y GQG(A/S)G(S/G)S (SEQ ID NO: 12). This second, generally mid-sized, **G** segment subtype typically contains 17 amino acid residues and lacks charged residues or contain one charged residue. This second **G** segment subtype typically contains 14-20 amino acid residues, but it is contemplated that it may contain as few as 12 or as many as 30 amino acid residues. Representative **G** segments of this second subtype are amino acid residues 249-265, 471-488, 631-647 and 982-998 of SEQ ID NO: 10; and amino acid residues 187-203 of SEQ ID NO: 3.

The third subtype of the **G** segment is represented by the amino acid one letter consensus sequence G(R/Q)GQG(G/R)YGQG (A/S/V)GGN (SEQ ID NO: 13). This third **G** segment subtype typically contains 14 amino acid residues, and is generally the shortest of the **G** segment subtypes. This third **G** segment subtype typically contains 12-17 amino acid residues, but it is contemplated that it may contain as many as 23 amino acid residues. Representative **G** segments of this third subtype are amino acid residues 57-70, 121-134, 184-197, 280-293, 343-356, 407-420, 503-516, 567-580, 662-675, 726-739, 790-803, 854-867, 918-931, 1014-1027 of SEQ ID NO: 10; and amino acid residues 219-232 of SEQ ID NO: 3.

Thus, in preferred embodiments, each individual G segment has at least 80%, preferably 90%, more preferably 95%, identity to an amino acid sequence selected from SEQ ID NO: 11, SEQ ID NO: 12 and SEQ ID NO: 13.

In a preferred embodiment of the alternating sequence of **A** and **G** segments of the **REP** moiety, every second **G** segment is of the first subtype, while the remaining **G** segments are of the third subtype, e.g. ...$\mathbf{A}_1\mathbf{G}_{short}\mathbf{A}_2\mathbf{G}_{long}\mathbf{A}_3\mathbf{G}_{short}\mathbf{A}_4\mathbf{G}_{long}\mathbf{A}_5\mathbf{G}_{short}$... In another preferred embodiment of the **REP** moiety, one **G** segment of the second subtype interrupts the **G** segment regularity *via* an insertion, e.g. ...$\mathbf{A}_1\mathbf{G}_{short}\mathbf{A}_2\mathbf{G}_{long}\mathbf{A}_3\mathbf{G}_{mid}\mathbf{A}_4\mathbf{G}_{short}\mathbf{A}_5\mathbf{G}_{long}$...

Each individual **L** segment represents an optional linker amino acid sequence, which may contain from 0 to 20 amino acid residues, such as from 0 to 10 amino acid residues. There are also linker amino acid sequences present in the repetitive part (SEQ ID NO: 10) of the deduced amino acid sequence of the MaSp1 protein from *Euprosthenops australis*. In particular, the amino acid sequence of a linker segment may resemble any of the described **A** or **G** segments, but usually not sufficiently to meet their criteria as defined herein.

Representative **L** segments are amino acid residues 1-6 and 1093-1110 of SEQ ID NO: 10; and amino acid residues 138-142 of SEQ ID NO: 3, but the skilled person in the art will readily recognize that there are many suitable alternative amino acid sequences for these segments. In one embodiment of the **REP** moiety, one of the **L** segments contains 0 amino

acids, i.e. one of the **L** segments is void. In another embodiment of the **REP** moiety, both **L** segments contain 0 amino acids, i.e. both **L** segments are void. Thus, these embodiments of the **REP** moieties may be schematically represented as follows: $(\mathbf{AG})_n\mathbf{L}$, $(\mathbf{AG})_n\mathbf{AL}$, $(\mathbf{GA})_n\mathbf{L}$, $(\mathbf{GA})_n\mathbf{GL}$; $\mathbf{L}(\mathbf{AG})_n$, $\mathbf{L}(\mathbf{AG})_n\mathbf{A}$, $\mathbf{L}(\mathbf{GA})_n$, $\mathbf{L}(\mathbf{GA})_n\mathbf{G}$; and $(\mathbf{AG})_n$, $(\mathbf{AG})_n\mathbf{A}$, $(\mathbf{GA})_n$, $(\mathbf{GA})_n\mathbf{G}$.

The fusion protein is not comprising any moiety derived from the repetitive fragment of a spider silk protein. i.e. it has a low (or no) degree of identity and/or similarity to repetitive spider silk protein fragments. The sequence of the fusion protein according to the invention preferably has less than 30% identity, such as less than 20% identity, preferably less than10% identity, to any of the repetitive spidroin amino acid sequences disclosed herein, and specifically to any of SEQ ID NO: 10-13.

The **B** moiety is a protein or polypeptide fragment comprising more than 30 amino acid residues. The **B** moiety is preferably comprising more than 50 amino acid residues, such as more than 100 amino acid residues. The **B** moiety is preferably comprising less than 1000 amino acid residues, such as less than 400 amino acid residues, more preferably less than 300 amino acid residues. It is capable of selective interaction with the organic target, and it is the **B** moiety in the fusion protein which provides the capacity of selective interaction with the organic target.

The **B** moiety is a non-spidroin moiety. This implies that it is not derived from a spider silk protein, i.e. it has a low (or no) degree of identity and/or similarity to spider silk proteins. The sequence of the **B** moiety according to the invention preferably has less than 30% identity, such as less than 20% identity, preferably less than10% identity, to any of the spidroin amino acid sequences disclosed herein, and specifically to any of SEQ ID NO: 6-10.

It is regarded as within the capabilities of those of ordinary skill in the art to select the **B** moiety. Nevertheless, examples of affinity ligands that may prove useful as **B** moieties, as well as examples of formats and conditions for detection and/or quantification, are given below for the sake of illustration.

The biomolecular diversity needed for selection of affinity ligands may be generated by combinatorial engineering of one of a plurality of possible scaffold molecules, and specific and/or selective affinity ligands are then selected using a suitable selection platform. Non-limiting examples of such structures, useful for generating affinity ligands against the organic target, are staphylococcal protein A and domains thereof and derivatives of these domains, such as the Z domain (Nord K *et al.* (1997) Nat. Biotechnol. 15:772-777); lipocalins (Beste G *et al.* (1999) Proc. Natl. Acad. Sci. U.S.A. 96:1898-1903); ankyrin repeat domains (Binz HK *et al.* (2003) J. Mol. Biol. 332:489-503); cellulose binding domains (CBD) (Smith GP *et al.* (1998) J. Mol. Biol. 277:317-332; Lehtiö J *et al.* (2000) Proteins 41:316-322); γ crystallines (Fiedler U and Rudolph R, WO01/04144); green fluorescent protein (GFP) (Peelle B *et al.* (2001) Chem. Biol. 8:521-534); human cytotoxic T lymphocyte-associated antigen 4 (CTLA-4) (Hufton SE *et al.* (2000) FEBS Lett. 475:225-231; Irving RA *et al.* (2001) J. Immunol. Meth. 248:31-45); protease inhibitors, such as Knottin proteins (Wentzel A *et al.* (2001) J. Bacteriol. 183:7273-7284; Baggio R *et al.* (2002) J. Mol. Recognit. 15:126-134) and Kunitz domains (Roberts BL *et al.* (1992) Gene 121:9-15; Dennis MS and Lazarus RA (1994) J. Biol. Chem. 269:22137-22144); PDZ domains (Schneider S *et al.* (1999) Nat. Biotechnol. 17:170-175); peptide aptamers, such as thioredoxin (Lu Z *et al.* (1995) Biotechnology 13:366-372; Klevenz B *et al.* (2002) Cell. Mol. Life Sci. 59:1993-1998); staphylococcal nuclease (Norman TC *et al.* (1999) Science 285:591-595); tendamistats (McConell SJ and Hoess RH (1995) J. Mol. Biol. 250:460-479; Li R *et al.* (2003) Protein Eng. 16:65-72); trinectins based on the fibronectin type III domain (Koide A *et al.* (1998) J. Mol. Biol. 284:1141-1151; Xu L *et al.* (2002) Chem. Biol. 9:933-942); zinc fingers (Bianchi E *et al.* (1995) J. Mol. Biol. 247:154-160; Klug A (1999) J. Mol. Biol. 293:215-218; Segal DJ *et al.* (2003) Biochemistry 42:2137-2148); adnectin; anticalin; DARPin; affilin and avimer.

The above-mentioned examples include scaffold proteins presenting a single randomized loop used for the generation of novel binding specificities, protein scaffolds with a rigid secondary structure where side chains protruding from the protein surface are randomized for the generation of novel binding

specificities, and scaffolds exhibiting a non-contiguous hyper-variable loop region used for the generation of novel binding specificities.

Oligonucleotides may also be used as affinity ligands. Single stranded nucleic acids, called aptamers or decoys, fold into well-defined three-dimensional structures and bind to their target with high affinity and specificity. (Ellington AD and Szostak JW (1990) Nature 346:818-822; Brody EN and Gold L (2000) J. Biotechnol. 74:5-13; Mayer G and Jenne A (2004) BioDrugs 18:351-359). The oligonucleotide ligands can be either RNA or DNA and can bind to a wide range of target molecule classes.

For selection of the desired affinity ligand from a pool of variants of any of the scaffold structures mentioned above, a number of selection platforms are available for the isolation of a specific novel ligand against a target protein of choice. Selection platforms include, but are not limited to, phage display (Smith GP (1985) Science 228:1315-1317), ribosome display (Hanes J and Plückthun A (1997) Proc. Natl. Acad. Sci. U.S.A. 94:4937-4942), yeast two-hybrid system (Fields S and Song O (1989) Nature 340:245-246), yeast display (Gai SA and Wittrup KD (2007) Curr Opin Struct Biol 17:467-473), mRNA display (Roberts RW and Szostak JW (1997) Proc. Natl. Acad. Sci. U.S.A. 94:12297-12302), bacterial display (Daugherty PS (2007) Curr Opin Struct Biol 17:474-480, Kronqvist N *et al.* (2008) Protein Eng Des Sel 1-9, Harvey BR *et al.* (2004) PNAS 101(25):913-9198), microbead display (Nord O *et al.* (2003) J Biotechnol 106:1-13, WO01/05808), SELEX (System Evolution of Ligands by Exponential Enrichment) (Tuerk C and Gold L (1990) Science 249:505-510) and protein fragment complementation assays (PCA) (Remy I and Michnick SW (1999) Proc. Natl. Acad. Sci. U.S.A. 96:5394-5399). A preferred group of **B** moieties with affinity for immunoglobulins, albumin or other organic targets are bacterial receptin domains or derivatives thereof.

A group of preferred **B** moieties are capable of selective interaction with immunoglobulins and molecules comprising immunoglobulin or derivatives thereof, e.g. the fragment crystallisable (Fc) region of IgG. A preferred group of immunoglobulin subclasses are the subclasses that are recognized by the Z domain derived from staphylococcal protein A, i.e. IgG1, IgG2, IgG4, IgA and IgM from human, all Ig subclasess from rabbit and cow,

IgG1 and IgG2 from guinea pig, and IgG1, IgG2a, IgG2b, IgG3 and IgM from mouse (see Hober, S. *et al.*, J. Chromatogr B. 848:40-47 (2007)), more preferably the immunoglobulin subclasses IgG1, IgG2, IgG4, IgA and IgM from human. The Z domain is an engineered version of the immunoglobulin G (IgG) binding domain B of staphylococcal protein A, and is a 58 amino acid long triple-helix motif that binds the Fc region of IgG. Another preferred group of immunoglobulin subclasses are the subclasses that are recognized by the C2 domain streptococcal protein G; i.e. all human subclasses of IgG, including IgG3, and IgG from several animals, including mouse, rabbit and sheep.

One group of preferred **B** moieties are selected from the group consisting of the Z domain derived from staphylococcal protein A, staphylococcal protein A and domains thereof, preferably the E, D, A, B and C domains, streptococcal protein G and domains thereof, preferably the C1, C2 and C3 domains; and protein fragments having at least 70% identity, such as at least 80% identity, or at least 90% identity, to any of these amino acid sequences. Preferably, the **B** moiety is selected from the group consisting of the Z domain derived from staphylococcal protein A, the B domain of staphylococcal protein A, and the C2 domain of streptococcal protein G; and protein fragments having at least 70% identity, such as at least 80% identity, or at least 90% identity, to any of these amino acid sequences. Preferably, the **B** moiety is selected from the group consisting of the Z domain derived from staphylococcal protein A and protein fragments having at least 70% identity, such as at least 80% identity, or at least 90% identity, to this amino acid sequence. It is preferred that the **B** moiety is selected from the group consisting of the Z domain derived from staphylococcal protein A and the C2 domain of streptococcal protein G. A preferred group of **B** moieties with affinity for immunoglobulins are bacterial receptin domains or derivatives thereof.

Another group of preferred **B** moieties are capable of selective interaction with albumin and molecules comprising albumin or derivatives thereof. A preferred group of **B** moieties with affinity for albumin are bacterial receptin domains or derivatives thereof. Preferred **B** moieties are selected

from streptococcal protein G, the albumin-binding domain of streptococcal protein G, GA modules from *Finegoldia magna*; and protein fragments having at least 70% identity, such as at least 80% identity, or at least 90% identity, to any of these amino acid sequences. Preferably, the **B** moiety is selected from the albumin-binding domain of streptococcal protein G and protein fragments having at least 70% identity, such as at least 80% identity, or at least 90% identity, thereto. It is preferred that the **B** moiety is the albumin-binding domain of streptococcal protein G.

A further group of preferred **B** moieties are capable of selective interaction with biotin and molecules comprising biotin or derivatives or analogues thereof. Preferred **B** moieties are selected from the group consisting of streptavidin, monomeric streptavidin (M4); and protein fragments having at least 70% identity, such as at least 80% identity, or at least 90% identity to any of these amino acid sequences. It is preferred that the **B** moiety is monomeric streptavidin (M4).

Another group of preferred **B** moities are enzymes, capable of selective interaction with substrates for an enzymatically catalyzed reaction. Preferred enzyme **B** moieties include xylanase and lysozyme.

A further group of preferred **B** moities are growth factors, capable of stimulating cell growth. Preferred growth factor **B** moities include epidermal growth factor (EGF), in particular human EGF, fibroblast growth factor 2 (FGF2), nerve growth factor 1 (NGF1) and stromal cell-derived factor 1 (SDF1). Specific fusion proteins and protein structures according to the invention are provided in the Examples. These preferred fusion proteins form the group consisting of SEQ ID NOS 61-66, 68, 70, 72, 74, 76, 80, 82, 84 and 86. Further preferred fusion proteins are having at least 80%, preferably at least 90%, more preferably at least 95%, identity to any of these sequences.

The present invention further provides isolated nucleic acids encoding a fusion protein according to the invention. In particular, specific nucleic acids are provided in the Examples and the appended sequence listing. Further preferred nucleic acids encode fusion proteins having at least 80%, preferably at least 90%, more preferably at least 95%, identity to any of SEQ ID NOS 61-66, 68, 70, 72, 74, 76, 80, 82, 84 and 86.

The nucleic acids according to the invention are useful for producing the fusion proteins according to the invention. The present invention provides 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 nucleic acid molecule which encodes the fusion protein in *E. coli*.

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.

Proteins structures, or formats, according to the invention are assembled spontaneously from the fusion proteins according to the invention under suitable conditions, and the assembly into polymers is promoted by the presence of shearing forces and/or an interface between two different phases e.g. between a solid and a liquid phase, between air and a liquid phase or at a hydrophobic/hydrophilic interface, e.g. a mineral oil-water interface. The presence of the resulting interface stimulates polymerization at the interface or in the region surrounding the interface, which region extends into the liquid medium, such that said polymerizing initiates at said interface or in said interface region. Various protein structures can be produced by adapting the conditions during the assembly. For instance, if the assembly is allowed to occur in a container that is gently wagged from side to side, a fiber is formed at the air-water interface. If the mixture is allowed to stand still, a film is

formed at the air-water interface. If the mixture is evaporated, a film is formed at the bottom of the container. If oil is added on top of the aqueous mixture, a film is formed at the oil-water interface, either if allowed to stand still or if wagged. If the mixture is foamed, e.g. by bubbling of air or whipping, the foam is stable and solidifies if allowed to dry.

The present invention thus provides a method for providing a protein structure displaying a binding activity towards an organic target. In the first method step, there is provided a recombinant fusion protein according to the invention. The fusion protein may e.g. be provided by expressing it in a suitable host from a nucleic acid according to the invention. In the second method step, the fusion protein is subjected to conditions to achieve formation of a polymer comprising the recombinant fusion protein. Notably, although the spontaneously assembled protein structures can be solubilized in hexafluoroisopropanol, the solubilized fusion proteins are then not able to spontaneously reassemble into e.g. fibers.

The protein structure is useful as part of an affinity medium for immobilization of an organic target, wherein the **B** moiety is capable of selective interaction with the organic target. A sample, e.g. a biological sample, may be applied to a fusion protein or a protein structure according to the invention which is capable of binding to an organic target present in the biological sample, and the fusion protein or protein structure is then useful for separation of the organic target from the sample. A biological sample, such as blood, serum or plasma which has been removed from a subject may be subjected to detection, separation and/or quantification of the organic target.

The present invention thus provides a method for separation of an organic target from a sample. A sample, e.g. a biological sample such as blood, serum or plasma, containing the organic target is provided. The biological sample may be an earlier obtained sample. If using an earlier obtained sample in a method, no steps of the method are practiced on the human or animal body.

An affinity medium according to the invention is provided, comprising a fusion protein or a protein structure according to the invention. In certain embodiments, the affinity medium is consisting of the fusion protein or protein

structure according to the invention. The affinity medium is capable of selective interaction with the organic target by means of the **B** moiety in the fusion protein according to the invention. The affinity medium is contacted with the sample under suitable conditions to achieve binding between the affinity medium and the organic target. Non-bound sample is removed under suitable conditions to maintain selective binding between the affinity medium and the organic target. This method results in an organic target immobilized to the affinity medium, and specifically to the fusion protein, according to the invention.

In a preferred method according to the invention, the fusion protein in the affinity medium is present as a protein structure according to the invention when contacting the affinity medium with the sample to achieve binding between the affinity medium and the organic target.

A particularly useful protein structure in this respect is a film or a fiber wherein the **B** moiety is the Z domain derived from staphylococcal protein A or a protein fragment having at least 70% identity, such as at least 80% identity, or at least 90% identity, thereto. The film is advantageous in that it adheres to solid supports, e.g. the plastics in microtiter plates. This property of the film facilitates washing and regeneration procedures and is very useful for separation purposes.

It has surprisingly been observed that the alkali stability of the Z domain may even be enhanced when being part of a fusion protein according to the invention in a protein structure according to the invention. This property may be very useful for washing and regeneration purposes, e.g. allowing for high concentrations of NaOH, such as 0.1 M, 0.5 M, 1 M or even above 1 M, e.g. 2 M, and/or for high concentrations of urea, e.g. 6-8 M. The chemical stability may also be useful to allow for repeated cycles of use of the Z domain for affinity purification. This alkali stability may be further increased by utilizing a stabilized mutant of the Z domain. Furthermore, it has advantageously been shown that the fusion proteins according to the invention, including the Z domain, are heat stable. This allows for sterilization by heat with maintained solid protein format/structure as well as binding ability.

A known problem with traditional affinity matrices with Z domains is leakage of the Z domain from the affinity matrix. Due to the stable incorporation of the Z domain by a peptide bond into the fusion protein of the invention, it is contemplated that the undesirable leakage of the Z domain from the protein structures according to the invention is low or absent. Another advantage of the fusion proteins according to the invention is that the resulting protein structure has a high density of Z domains (or other **B** moieties). It is contemplated that this high density provides a high binding capacity. Altogether, these properties of the fusions proteins are very attractive for various **B** moieties, and in particular for affinity purification using protein Z with good production economy. These properties are also useful in other formats than in traditional gel bead affinity columns, e.g. in filter-like formats.

The immobilized organic target is capable of selective interaction with a second organic target. The method is then further comprising the step of contacting said affinity medium and the immobilized organic target with a second organic target, which is capable of selective interaction with the first organic target, under suitable conditions to achieve binding between the first and second organic targets.

The immobilized organic target is detectable and/or quantifiable. The detection and/or quantification of the organic target may be accomplished in any way known to the skilled person for detection and/or quantification of binding reagents in assays based on various biological or non-biological interactions. The organic targets may be labeled themselves with various markers or may in turn be detected by secondary, labeled affinity ligands to allow detection, visualization and/or quantification. This can be accomplished using any one or more of a multitude of labels, which can be conjugated to the organic target or to any secondary affinity ligand, using any one or more of a multitude of techniques known to the skilled person, and not as such involving any undue experimentation. Non-limiting examples of labels that can be conjugated to organic targets and/or secondary affinity ligands include fluorescent dyes or metals (e.g., fluorescein, rhodamine, phycoerythrin, fluorescamine), chromophoric dyes (e.g., rhodopsin), chemiluminescent

compounds (e.g., luminal, imidazole) and bioluminescent proteins (e.g., luciferin, luciferase), haptens (e.g., biotin). A variety of other useful fluorophores and chromophores are described in Stryer L (1968) Science 162:526-533 and Brand L and Gohlke JR (1972) Annu. Rev. Biochem. 41:843-868. Organic targets and/or secondary affinity ligands can also be labeled with enzymes (e.g., horseradish peroxidase, alkaline phosphatase, beta-lactamase), radioisotopes (e.g., ^{3}H, ^{14}C, ^{32}P, ^{35}S or ^{125}I) and particles (e.g., gold). In the context of the present disclosure, "particles" refer to particles, such as metal particles, suitable for labeling of molecules. Further, the affinity ligands may also be labeled with fluorescent semiconductor nanocrystals (quantum dots). Quantum dots have superior quantum yield and are more photostable compared to organic fluorophores and are therefore more easily detected (Chan *et al.* (2002) *Curr Opi Biotech*. 13: 40-46). The different types of labels can be conjugated to an organic target or a secondary affinity ligand using various chemistries, e.g., the amine reaction or the thiol reaction. However, other reactive groups than amines and thiols can be used, e.g., aldehydes, carboxylic acids and glutamine.

If the detection and/or quantification involves exposure to a second organic target or secondary affinity ligand, the affinity medium is washed once again with buffers to remove unbound secondary affinity ligands. As an example, the secondary affinity ligand may be an antibody or a fragment or a derivative thereof. Thereafter, organic targets may be detected and/or quantified with conventional methods. The binding properties for a secondary affinity ligand may vary, but those skilled in the art should be able to determine operative and optimal assay conditions for each determination by routine experimentation.

The detection, localization and/or quantification of a labeled molecule may involve visualizing techniques, such as light microscopy or immunofluoresence microscopy. Other methods may involve the detection via flow cytometry or luminometry. The method of visualization of labels may include, but is not restricted to, fluorometric, luminometric and/or enzymatic techniques. Fluorescence is detected and/or quantified by exposing fluorescent labels to light of a specific wavelength and thereafter detecting

and/or quantifying the emitted light in a specific wavelength region. The presence of a luminescently tagged molecule may be detected and/or quantified by luminescence developed during a chemical reaction. Detection of an enzymatic reaction is due to a color shift in the sample arising from chemical reaction. Those of skill in the art are aware that a variety of different protocols can be modified in order for proper detection and/or quantification.

One available method for detection and/or quantification of the organic target is by linking it or the secondary affinity ligand to an enzyme that can then later be detected and/or quantified in an enzyme immunoassay (such as an EIA or ELISA). Such techniques are well established, and their realization does not present any undue difficulties to the skilled person. In such methods, the biological sample is brought into contact with a protein structure according to the invention which binds to the organic target, which is then detected and/or quantified with an enzymatically labeled secondary affinity ligand. Following this, an appropriate substrate is brought to react in appropriate buffers with the enzymatic label to produce a chemical moiety, which for example is detected and/or quantified using a spectrophotometer, fluorometer, luminometer or by visual means.

The organic target or the secondary affinity ligands can be labeled with radioisotopes to enable detection and/or quantification. Non-limiting examples of appropriate radiolabels in the present disclosure are ^{3}H, ^{14}C, ^{32}P, ^{35}S or ^{125}I. The specific activity of the labeled affinity ligand is dependent upon the half-life of the radiolabel, isotopic purity, and how the label has been incorporated into the affinity ligand. Affinity ligands are preferably labeled using well-known techniques (Wensel TG and Meares CF (1983) in: *Radioimmunoimaging and Radioimmunotherapy* (Burchiel SW and Rhodes BA eds.) Elsevier, New York, pp 185-196). A thus radiolabeled affinity ligand can be used to visualize the organic target by detection of radioactivity. Radionuclear scanning can be performed with e.g. a gamma camera, magnetic resonance spectroscopy, emission tomography, gamma/beta counters, scintillation counters and radiographies.

Thus, the sample may be applied to the protein structure for detection, separation and/or quantification of the organic target. This procedure enables

not only detection of the organic target, but may in addition show the distribution and relative level of expression thereof. Optionally, the organic target may be released from the affinity medium and collected. Thus, the use may comprise affinity purification on an affinity medium onto which the organic target has been immobilized. The protein structure may for example be arranged in a column or in well plates (such as 96 well plates), or on magnetic beads, agarose beads or sepharose beads. Further, the use may comprise use of the protein structures on a soluble matrix, for example using a dextran matrix, or use in a surface plasmon resonance instrument, such as a Biacore™ instrument, wherein the analysis may for example comprise monitoring the affinity for the immobilized organic target or a number of potential affinity ligands.

The protein structures according to the invention can be washed and regenerated with various cleaning agents, including acid, base and chaotropic agents. Particularly useful cleaning agents include NaOH, such as 0.1, 0.5 or 1 M NaOH, and urea, such as 6-8 M urea, Since the protein structures according to the invention are surprisingly resistant to chemical treatment and/or sterilizing heat treatment, the methods according to the invention involving use of the protein structures may comprise a final step of regenerating the protein structure. The methods preferably comprise a final step of regenerating the affinity medium by chemical treatment and/or sterilizing heat treatment. It is preferred that the chemical treatment comprises treatment with NaOH, such as 0.1, 0.5 or 1 M NaOH, and/or urea, such as 6-8 M urea,

Fusion proteins according to the invention can be also be allowed to bind to an organic target in solution, i.e. prior to allowing the fusion protein to polymerize and form a protein structure, such as a film, a foam or a fibre. Both the spidroin-derived moieties (e.g. **CT**) as such and the corresponding fusion proteins incorporating a **B** moiety polymerise into solid structures even in the presence of contaminating proteins, without appreciable incorporation of contaminants into the material, and the functional (**B**) moieties retain their expected binding properties. It is therefore contemplated that the binding properties of the **B** moiety can be used to capture compounds or cells from

the surrounding solution and incorporate the captured compounds or cells into or on a protein structure according to the invention.

Thus, in another preferred method according to the invention, the fusion protein in the affinity medium is present in solution when contacting the affinity medium with the sample to achieve binding between the affinity medium and the organic target. The complex of fusion protein bound to the organic target is then allowed to form a fusion protein structure according to the invention.

This method may be particularly useful when the purpose is to "fish out" specific molecules or cells from a solution, e.g. to obtain target molecules from the media in large scale eukaryotic cell production systems when the target proteins are secreted. Since the binding of target molecules and formation of solid structures by the spidroin-derived moieties can take place at physiological conditions and since the spidroin-derived moieties are cytocompatible, the method can be applied repeatedly to an ongoing production process.

The protein structure according to the invention is also useful in separation, immobilization and/or cultivation of cells. A particularly useful protein structure in this respect is a film, a fiber or a foam. The film is advantageous in that it adheres to solid structures, e.g. the plastics in microtiter plates. This property of the film facilitates washing and regeneration procedures and is very useful for selective detection and separation purposes.

The present invention thus provides a cell scaffold material for cultivation of cells having an organic target that is present on the cell surface. The cell scaffold material is comprising a protein structure according to the invention. In certain embodiments, the cell scaffold material is consisting of the protein structure according to the invention.

It has been found by the present inventors that a cell scaffold material comprising a polymer comprising, and optionally consisting of, the fusion protein according to the invention provides a beneficial environment for the cultivation of cells, and preferably eukaryotic cells, in a variety of different settings. Furthermore, this environment enables the establishment of cultures

of cells that are otherwise very difficult, very costly or even impossible to culture in a laboratory, and for the establishment of cell-containing materials useful for tissue engineering and/or transplantation.

The invention also provides a combination of cells, preferably eukaryotic cells, and the cell scaffold material according to the invention. Such a combination according to the invention may be presented in a variety of different formats, and tailored to suit the needs of a specific situation. It is contemplated, for example, that the inventive combination may be useful as a cell-containing implant for the replacement of cells in damaged or diseased tissue.

The cell scaffold material can be utilized to capture cells either directly or indirectly. In direct capture, the **B** moiety is capable of selective interaction with an organic target that is present on the cell surface. Alternatively, the **B** moiety is capable of selective interaction with and is bound to an intermediate organic target, and that intermediate organic target is capable of selective interaction with an organic target that is present on the cell surface. Thus, in indirect capture, the cell scaffold material is further comprising an intermediate organic target, and the **B** moiety is capable of selective interaction with and is bound to said intermediate organic target. The intermediate organic target, in turn, is capable of selective interaction with the organic target that is present on the cell surface.

In one embodiment of the cell scaffold materials as disclosed herein, the fusion protein is further comprises an oligopeptide cell-binding motif. In connection with the cultivation of certain cells in certain situations, the presence of oligopeptide cell-binding motifs has been observed to improve or maintain cell viability, and the inclusion of such a motif into the cell scaffold material as a part of the spider silk protein is thought to provide additional benefits. The cell-binding motif is an oligopeptide coupled to the rest of the fusion protein via at least one peptide bond. For example, it may be coupled to the N-terminal or the C-terminal of the rest of the fusion protein, or at any position within the amino acid sequence of the rest of the spider silk protein. With regard to the selection of oligopeptidic cell-binding motifs, the skilled person is aware of several alternatives. The coupling of an oligopeptide cell-

binding motif to the rest of the spider silk protein is readily accomplished by the skilled person using standard genetic engineering or chemical coupling techniques. Thus, in some embodiments, the cell-binding motif is introduced via genetic engineering, i.e. forming part of a genetic fusion between a nucleic acid encoding a fusion protein and the cell-binding motif. As an additional beneficial characteristic of such embodiments, the cell-binding motif will be present in a 1:1 ratio to the monomers of fusion protein in the polymer making up the cell scaffold material.

The polymer in the cell scaffold material used in the methods or combination described herein may adopt a variety of physical forms, and use of a specific physical form may offer additional advantages in different specific situations. For example, in an embodiment of the methods or combination, said cell scaffold material is in a physical form selected from the group consisting of film, foam, capsules, fiber and fiber-mesh.

The present invention accordingly provides a method for immobilization of cells. A sample e.g. a biological sample such as blood, comprising cells of interest is provided. The biological sample may be an earlier obtained sample. If using an earlier obtained sample in a method, no steps of the method are practiced on the human or animal body.

The sample is applied to a cell scaffold material according to the invention under suitable conditions to allow selective interaction between the cell scaffold material and an organic target that is present on the surface of the cells of interest. The cells are allowed to immobilize to said cell scaffold material by binding between the organic target on the cell surface and said cell scaffold material. Non-bound sample is removed under suitable conditions to maintain selective binding between the cell scaffold material and the organic target. This method results in cells exhibiting the organic target being immobilized to the cell scaffold material, and specifically to the protein structure, according to the invention.

As set out above, the cell scaffold material can be utilized to capture cells either directly or indirectly. In direct capture, the **B** moiety is capable of selective interaction with an organic target that is present on the cell surface. Alternatively, the **B** moiety is capable of selective interaction with and is

bound to an intermediate organic target, and that intermediate organic target is capable of selective interaction with an organic target that is present on the cell surface. Thus, in indirect capture, the cell scaffold material is further comprising an intermediate organic target, and the **B** moiety is capable of selective interaction with and is bound to said intermediate organic target. The intermediate organic target, in turn, is capable of selective interaction with the organic target that is present on the cell surface.

Regardless of capture method, the captured cells may be released from the fusion protein by cleavage of the fusion protein to release the moiety involved in cell capture from the cell scaffold material. As mentioned hereinabove, the fusion protein may include a cleavage site in its amino acid sequence, which allows for cleavage and removal of the relevant moiety, typically the **B** moiety or a cell-binding motif. 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.

The present invention also provides a method for cultivation of cells. Cells of interest are immobilized to the cell scaffold material using the method disclosed hereinabove. The combination of the cell scaffold material and the immobilized cells are maintained under conditions suitable for cell culture.

In the context of the present invention, the terms “cultivation” of cells, “cell culture” etc are to be interpreted broadly, such that they encompass for example situations in which cells divide and/or proliferate, situations in which cells are maintained in a differentiated state with retention of at least one functional characteristic exhibited by the cell type when present in its natural environment, and situations in which stem cells are maintained in an undifferentiated state.

According to another aspect, the present invention provides a novel recombinant protein comprising a **CT** moiety and at least one **NT** moiety, with the proviso that the protein is not comprising any moiety derived from the repetitive fragment of a spider silk protein. In a preferred embodiment, the protein is comprising 1-2 NT moieties.

In one preferred embodiment, the recombinant protein is consisting of a **CT** moiety and at least one **NT** moiety, such as 1-2 **NT** moieties. The protein may schematically be written as **NT**-**CT**, **CT**-**NT**, **NTNT**-**CT**, **CT**-**NTNT** or **NT**-**CT**-**NT**, and preferably **NT**-**CT** or **NTNT**-**CT**.

A preferred recombinant protein according to the invention is selected from the group consisting of SEQ ID NOS: 59-60 ; and proteins having at least 80%, preferably at least 90%, more preferably at least 95% identity to any of these sequences. The present invention further provides an isolated nucleic acid encoding a recombinant protein according to the invention.

This protein is void of the repetitive fragment of a spider silk protein, and it is therefore surprising that it is still capable of forming solid protein structures, e.g. fibers and films. An advantage with the present recombinant protein is that it can be produced with higher yield than CT alone or the corresponding proteins containing REP but lacking NT, see e.g. Examples 1-2.

The present invention provides a protein structure which is a polymer comprising as a repeating structural unit a recombinant protein according to this aspect. The protein structure preferably has a size of at least 0.1 µm in at least two dimensions. The protein structure is preferably in a physical form selected from the group consisting of fiber, film, foam, net, mesh, sphere and capsule.

Since this recombinant protein is not comprising any moiety derived from the repetitive fragment of a spider silk protein, it has a low (or no) degree of identity and/or similarity to repetitive spider silk protein fragments. The sequence of the protein according to the invention preferably has less than 30% identity, such as less than 20% identity, preferably less than10% identity, to any of the repetitive spidroin amino acid sequences disclosed herein, and specifically to any of SEQ ID NO: 10-13.

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

Examples

Example 1 - Cloning, expression and fiber formation of NT-CT

To investigate if CT covalently linked to NT can form fibers, a NT-CT fusion protein (a **NT** moiety and a **CT** moiety) was produced and purified.

Cloning

Genes encoding the His_6NT-CT fusion protein (SEQ ID NO: 59) were constructed. The vectors were transformed into chemocompetent *Escherichia coli* (*E. coli*) BL21 (DE3) cells that were allowed to grow onto agar plates supplemented with kanamycin (70 µg/ml). Colonies were thereafter picked and PCR screened for correct insert and subsequently also sequenced to confirm the DNA sequence.

Production

E. coli BL21 (DE3) cells possessing the pT7His_6NTCT vector were grown in Luria-Bertani medium (6 litre in total) supplemented with kanamycin (70 µg/ml) to an OD_{600} value of 1-1.5 in 30°C, followed by induction of expression with 300 µM IPTG (isopropyl β-D-1-thiogalactopyranoside) and further incubation in 20ºC for approximately 2 h. Next, the cells were harvested by a 20 min centrifugation at 4 700 rpm, and the resulting cell pellets were dissolved in 20 mM Tris (pH 8.0).

Purification

Cell pellets dissolved in 20 mM Tris (pH 8.0) were supplemented with lysozyme and DNase I in order to lyse the bacterial cells, whereupon the cell lysates were recovered after 15 000 rpm of centrifugation for 30 min. Next, the recovered cell lysates were divided and loaded onto a total of four Chelating Sepharose Fast Flow Zn^{2+} columns, keeping the protein bound to the column matrix via the His_6 tag. After washing, bound proteins were eluted with 20 mM Tris/300 mM imidazole (pH 8.0). The yield of NTCT was typically higher than that of CT or Rep4CT. Next, the pooled eluate liquid was dialysed against 5 litres of 20 mM Tris (pH 8.0) over night, concentrated to 1 mg/ml

and finally allowed to form fibers or films. Fibers were faster formed if pH was decreased to below pH 6.4.

The fact that macroscopic fibers of His_6NTCT could be obtained demonstrates that CT retains its fiber forming properties when fused to NT.

Example 2 - Cloning, expression and fiber formation of NTNT-CT

To investigate if CT can form fibers although covalently linked to NTNT, a NTNT-CT fusion protein (a **NTNT** moiety and a **CT** moiety) was produced and purified.

Cloning

Genes encoding the His_6NTNT-CT fusion protein (SEQ ID NO: 60) were constructed. The vectors were transformed into chemocompetent *E. coli* BL21 (DE3) cells that were allowed to grow onto agar plates supplemented with kanamycin (70 µg/ml). Colonies were thereafter picked and PCR screened for correct insert and subsequently also sequenced to confirm the DNA sequence.

Production

E. coli BL21 (DE3) cells possessing the pT7His_6NTNTCT vector were grown in Luria-Bertani medium (6 litre in total) supplemented with kanamycin (70 µg/ml) to an OD_{600} value of 1-1.5 in 30°C, followed by induction of expression with 300 µM IPTG and further incubation in 20ºC for approximately 2 h. Next, the cells were harvested by a 20 min centrifugation at 4 700 rpm, and the resulting cell pellets were dissolved in 20 mM Tris (pH 8.0).

Purification

Cell pellets dissolved in 20 mM Tris (pH 8.0) were supplemented with lysozyme and DNase I in order to lyse the bacterial cells, whereupon the cell lysates were recovered after 15 000 rpm of centrifugation for 30 min. Next, the recovered cell lysates were divided and loaded onto a total of four

Chelating Sepharose Fast Flow Zn^{2+} columns, keeping the protein bound to the column matrix via the His_6 tag. After washing, bound proteins were eluted with 20 mM Tris/300 mM imidazole (pH 8.0). The yield of NTNTCT was typically higher than that of CT or Rep4CT. Next, the pooled eluate liquid was dialysed against 5 litres of 20 mM Tris (pH 8.0) over night, concentrated to 1 mg/ml and finally allowed to form fibers or films. Fibers were faster formed if pH was decreased to below pH 6.4.

The fact that macroscopic fibers of His_6NTNTCT could be obtained demonstrates that CT retains its fiber forming properties when fused to NTNT.

Example 3 - Cloning, expression and fiber formation of an IgG-binding CT fusion protein

To prove the fusion protein concept, a CT protein (a **CT** moiety) is produced in fusion with the Z protein domain (a **B** moiety). The Z domain is an engineered version of the immunoglobulin G (IgG) binding domain B of staphylococcal protein A, and is a 58 amino acid long triple-helix motif that binds the fragment crystallisable (F_c) region of IgG. The aim is to investigate whether it is possible to produce structures, such as fibers, films and membranes, from a fusion protein consisting of the Z domain fused to CT and still retain the IgG-binding ability of domain Z, as well as the structure forming properties of CT. In order to do so a fusion protein consisting of the Z domain N-terminally or C-terminally to CT is cloned.

Cloning

Genes encoding the His_6ZCT and His_6CTZ fusion proteins (SEQ ID NOS: 61-62) are constructed and transformed into chemocompetent *E. coli* BL21 (DE3) cells that are allowed to grow onto agar plates supplemented with kanamycin (70 µg/ml). Colonies are thereafter picked and PCR screened for correct insert and subsequently also sequenced to confirm the correct DNA sequence.

Production

E. coli BL21 (DE3) cells possessing the pT7His_6ZCT or pT7His_6CTZ vector are grown in Luria-Bertani medium (6 litre in total) supplemented with kanamycin (70 µg/ml) to an OD_{600} value of 1-1.5 in 30°C, followed by induction of His_6ZCT or His_6CTZ expression with 300 µM IPTG and further incubation in 20ºC for approximately 2 h. Next, the cells are harvested by a 20 min centrifugation at 4 700 rpm, and the resulting cell pellets are dissolved in 20 mM Tris (pH 8.0).

Purification

Cell pellets dissolved in 20 mM Tris (pH 8.0) are supplemented with lysozyme and DNase I in order to lyse the bacterial cells, whereupon the cell lysates are recovered after 15 000 rpm of centrifugation for 30 min. Next, the recovered cell lysates are divided and loaded onto a total of four Chelating Sepharose Fast Flow Zn^{2+} columns, keeping the His_6ZCT protein bound to the column matrix via the His_6 tag. After washing, bound proteins are eluted with 20 mM Tris/300 mM imidazole (pH 8.0). The pooled eluate fractions are dialysed against 5 litres of 20 mM Tris (pH 8.0) over night, concentrated to 1 mg/ml and finally allowed to form fibers.

Analysis

To explore the capacity of the **B** moiety in a fusion protein structure of selective interaction with an organic target, the ability of domain Z in the fusion protein to bind IgG is studied. Fibers and films of this fusion protein are used for binding of purified IgG and IgG from serum, followed by elution and subsequent analysis on SDS-PAGE, where IgG under non-reducing conditions appears as a ~146 kDa band.

<u>Example 4 - Cloning, expression and fiber formation of an IgG-binding NTCT fusion protein</u>

To prove the fusion protein concept, a NT-CT protein (a **NT** and a **CT** moiety) is produced in fusion with the Z protein domain (a **B** moiety). The aim is to investigate whether it is possible to produce structures, such as fibers,

films and membranes, from a fusion protein consisting of the Z domain fused to NTCT and still retain the IgG-binding ability of domain Z, as well as the structure forming properties of CT. In order to do so a fusion protein consisting of the Z domain N-terminally and C-terminally to NTCT is cloned.

Cloning

Genes encoding the His_6ZNTCT and His_6NTCTZ fusion proteins (SEQ ID NOS: 63-64) are constructed and transformed into chemocompetent *E. coli* BL21 (DE3) cells that are allowed to grow onto agar plates supplemented with kanamycin (70 µg/ml). Colonies are thereafter picked and PCR screened for correct insert and subsequently also sequenced to confirm the correct DNA sequence.

Production

E. coli BL21 (DE3) cells possessing the pT7His_6ZNTCT and pT7His_6NTCTZ vector are grown in Luria-Bertani medium (6 litre in total) supplemented with kanamycin (70 µg/ml) to an OD_{600} value of 1-1.5 in 30°C, followed by induction of expression with 300 µM IPTG and further incubation in 20ºC for approximately 2 h. Next, the cells are harvested by a 20 min centrifugation at 4 700 rpm, and the resulting cell pellets are dissolved in 20 mM Tris (pH 8.0).

Purification

Cell pellets dissolved in 20 mM Tris (pH 8.0) are supplemented with lysozyme and DNase I in order to lyse the bacterial cells, whereupon the cell lysates are recovered after 15 000 rpm of centrifugation for 30 min. Next, the recovered cell lysates are divided and loaded onto a total of four Chelating Sepharose Fast Flow Zn^{2+} columns, keeping the His_6ZNTCT or His_6NTCTZ protein bound to the column matrix via the His_6 tag. After washing, bound proteins are eluted with 20 mM Tris/300 mM imidazole (pH 8.0). The pooled eluate fractions are dialysed against 5 litres of 20 mM Tris (pH 8.0) over night, concentrated to 1 mg/ml and finally allowed to form fibers.

Analysis

To explore the capacity of the **B** moiety in a fusion protein structure of selective interaction with an organic target, the ability of domain Z in the fusion protein to bind IgG is studied. Fibers and films of this fusion protein are used for binding of purified IgG and IgG from serum, followed by elution and subsequent analysis on SDS-PAGE, where IgG under non-reducing conditions appears as a ~146 kDa band.

Example 5 - Cloning, expression and fiber formation of an IgG-binding NTNTCT fusion protein

To prove the fusion protein concept, a NTNT-CT protein (a **NTNT** and a **CT** moiety) is produced in fusion with the Z protein domain (a **B** moiety). The aim is to investigate whether it is possible to produce structures, such as fibers, films and membranes, from a fusion protein consisting of the Z domain fused to NTNTCT and still retain the IgG-binding ability of domain Z, as well as the structure forming properties of CT. In order to do so a fusion protein consisting of the Z domain N-terminally and C-terminally to NTNTCT is cloned.

Cloning

Genes encoding the His_6ZNTNTCT and His_6NTNTCTZ fusion proteins (SEQ ID NOS: 65-66) are constructed and transformed into chemocompetent *E. coli* BL21 (DE3) cells that are allowed to grow onto agar plates supplemented with kanamycin (70 µg/ml). Colonies are thereafter picked and PCR screened for correct insert and subsequently also sequenced to confirm the correct DNA sequence.

Production

E. coli BL21 (DE3) cells possessing the pT7His_6ZNTNTCT and pT7His_6NTNTCTZ vector are grown in Luria-Bertani medium (6 litre in total) supplemented with kanamycin (70 µg/ml) to an OD_{600} value of 1-1.5 in 30°C, followed by induction of expression with 300 µM IPTG and further incubation in 20ºC for approximately 2 h. Next, the cells are harvested by a 20 min

centrifugation at 4 700 rpm, and the resulting cell pellets are dissolved in 20 mM Tris (pH 8.0).

Purification

Cell pellets dissolved in 20 mM Tris (pH 8.0) are supplemented with lysozyme and DNase I in order to lyse the bacterial cells, whereupon the cell lysates are recovered after 15 000 rpm of centrifugation for 30 min. Next, the recovered cell lysates are divided and loaded onto a total of four Chelating Sepharose Fast Flow Zn^{2+} columns, keeping the His_6ZNTNTCT or His_6NTNTCTZ protein bound to the column matrix via the His_6 tag. After washing, bound proteins are eluted with 20 mM Tris/300 mM imidazole (pH 8.0). The pooled eluate fractions are dialysed against 5 litres of 20 mM Tris (pH 8.0) over night, concentrated to 1 mg/ml and finally allowed to form fibers.

Analysis

To explore the capacity of the **B** moiety in a fusion protein structure of selective interaction with an organic target, the ability of domain Z in the fusion protein to bind IgG is studied. Fibers and films of this fusion protein are used for binding of purified IgG and IgG from serum, followed by elution and subsequent analysis on SDS-PAGE, where IgG under non-reducing conditions appears as a ~146 kDa band.

<u>Example 6 - Cloning, expression and formation of solid structures of ABD-NTCT and ABD-CT fusion proteins</u>

NTCT and CT were produced in fusion with the albumin binding domain (ABD) derived from streptococcal protein G. ABD is a 5 kDa triple-helix motif that binds to albumin. The aim was to investigate whether it is possible to produce structures, such as fibers and films, from the fusion proteins consisting of the ABD domain fused to NTCT (denoted His_6-ABD-NTCT, SEQ ID NO: 68) and to CT (denoted His_6-ABD-CT, SEQ ID NO: 70) respectively, and still retain the albumin binding ability of ABD domain as well as the structure forming properties of NTCT and CT. In order to do so, two

fusion proteins consisting of the ABD domain fused N-terminally to NTCT and to CT were cloned.

Cloning

A gene (SEQ ID NO: 69) encoding the His_6-ABD-CT fusion protein (SEQ ID NO: 70) was constructed as follows. Primers were designed in order to generate PCR fragments of domain ABD from a vector containing such an ABD sequence. Also, the primers contained recognition sites for the restriction endonucleases *NdeI* and *EcoRI*. The resulting PCR products were then treated with the restriction endonucleases *NdeI* and *EcoRI*, as was the target vector (denoted pAff8His_6TrxHis_6CT, harbouring a kanamycin resistance gene). Upon restriction cleavage of the target vector, the His_6TrxHis_6 part was cleaved off. Cleaved PCR fragments and target vector were joined together with the aid of a T4 DNA Ligase, whereupon the resulting correctly ligated vector (pT7His_6-ABD-CT) was transformed into chemocompetent *E. coli* BL21 (DE3) cells that were allowed to grow onto agar plates supplemented with kanamycin (50 µg /ml). Colonies were thereafter picked and screened for correct insert and subsequently sequenced to confirm the DNA sequence of the inserted ABD into the target vector.

Cloning of a gene (SEQ ID NO: 67) encoding the His_6-ABD-NTCT fusion protein (SEQ ID NO: 68) was constructed in the same way as described for His_6-ABD-CT, but the target vector here was denoted by pT7His_6scFv1-NTCT, where the pT7His_6scFv1 part was cleaved off upon treatment with *NdeI* and *EcoRI*. The correctly ligated vector is denoted as pT7His_6ABD-NTCT.

Production

E. coli BL21 (DE3) cells possessing the pT7His_6-ABD-CT vector were grown in Luria-Bertani medium (3 liters in total) supplemented with kanamycin (50 µg/ml) to an OD_{600} value of 1-1.5 in 30°C, followed by induction of pT7His_6-ABD-CT expression with 300 µM IPTG and further incubation at14ºC for approximately 17 h. Next, the cells were harvested by a 20 min

centrifugation at 4 700 rpm, and the resulting cell pellet was dissolved in 20 mM Tris (pH 8.0).

Production of His_6ABD-NTCT was performed in the same way as described for His_6ABD-CT.

Purification

The cell pellet dissolved in 20 mM Tris (pH 8.0) was supplemented with lysozyme and DNase I in order to lyse the bacterial cells, followed by the addition of NaCl and imidazole to a final concentration of 200 mM and 10 mM, respectively. After 30 min of centrifugation (15 000 rpm) the cell lysate was recovered. Next, the recovered cell lysate was loaded onto a Chelating Sepharose Fast Flow Zn^{2+} column, keeping the His_6-ABD-CT protein bound to the column matrix via the His_6 tag. After washing, bound proteins were eluted with 20 mM Tris/200 mM imidazole (pH 8.0)/300 mM NaCl. The eluate contained 28.8 mg of His_6-ABD-CT protein according to an A_{280} measurement. Next, the eluted protein was dialyzed against 3 liters of 20 mM Tris (pH 8.0) over night and thereafter concentrated to 1.48 mg/ml, yielding a final amount of 6.216 mg His_6-ABD-CT fusion protein.

The same purification procedure was carried out for His_6-ABD-NTCT. The eluate concentration of His_6ABD-NTCT was 1.76 mg/ml, and a final amount of 35.2 mg of fusion protein was obtained.

Film and fiber formation

Films of His_6-ABD-CT were casted in 96-well plates (Tissue culture plate, Suspension cells, 83.1835.500, Sarstedt) from 15 µl of 1 mg/ml soluble fusion protein per film. The films were then allowed to solidify over night (20°C, 35% relative humidity). The same procedure was followed for casting films of His_6-ABD-NTCT from 15 µl of 1 mg/ml protein solution.

Fibers were also made for both His_6ABD-CT and His_6ABD-NTCT from 1.76 and 1.06 mg/ml of soluble fusion protein, respectively. Fig. 3a shows a microscopic fiber picture of a His_6-ABD-NTCT fusion protein (SEQ ID NO: 68), while Fig 3b shows a microscopic fiber picture of His_6-ABD-CT fusion protein (SEQ ID NO: 70). The fact that macroscopic fibers of both His_6ABD-

CT and His_6ABD-NTCT could be obtained although CT or NTCT has been fused to another protein, i.e. the 46 amino acid long ABD domain, demonstrates that CT and NTCT retain their structural forming properties despite being fused to the ABD domain.

Analysis

To evaluate the ability of ABD-NTCT films to bind albumin, human blood plasma was used as albumin source. Four films of ABD-NTCT and ABD-CT were pre-wetted with 150 µl of 1×PBS followed by incubation of 100 µl of human blood plasma (1:5 dilution) for 30 min at room temperature. After washing three times with 200 µl 1×PBS, bound albumin was eluted in 50 µl by lowering the pH to approximately 2.7 with elution buffer (i.e. 0.5 M acetic acid, 1 M urea, 100 mM NaCl), after which the eluted fractions were analyzed by non-reducing SDS-PAGE. Films of NTCT and CT were used as control material, and were treated in the same way.

Fig. 4 shows a non-reducing SDS-PAGE gel of eluted fractions after binding of albumin from human blood plasma to ABD-NTCT and ABD-CT films. The gel was loaded according to:

(1-4) Quadruplicates of ABD-NTCT films, 14 µl loaded;
(5-8) Quadruplicates of ABD-CT films, 14 µl loaded;
(9) Protein ladder;
(10-13) Quadruplicates of NTCT (control) films;
(14-16) Triplicates of CT (control) films;
(17) Human blood plasma (1:50), 8 µl loaded.

All films of ABD-NTCT and ABD-CT have bound albumin from human blood plasma (Fig. 4, lanes 1-8). As only one single albumin band (~65 kDa) appears in the eluted fraction of the ABD-NTCT and ABD-CT films, they seem to not bind anything else unspecifically from the human blood plasma. Control films of NTCT and CT do not show any albumin in the eluted fractions (Fig. 4, lanes 10-16). It is concluded that the ABD domain is functional in the macroscopic solid structures of fusion proteins with NTCT and CT.

Example 7 - Cloning, expression and formation of solid structures of M4-NTCT and M4-CT fusion proteins

Monomeric streptavidin (M4) domain was produced in fusion with NTCT and CT protein, respectively. The M4 domain is a mutated version of the tetrameric streptavidin, and is a 159 amino acid long protein domain that binds non-covalently to biotin. Our aim was to investigate whether it is possible to produce structures, such as films, foams and fibers from the fusion proteins consisting of the M4 domain fused to NTCT (denoted His_6-M4-NTCT, SEQ ID NO: 72) and to CT (denoted His_6-M4-CT, SEQ ID NO: 74), respectively, and still retain the biotin-binding ability of M4 domain as well as the structure forming properties of NTCT and CT. In order to do so, two fusion proteins consisting of the M4 domain fused N-terminally to NTCT and to CT were cloned.

Cloning

A gene (SEQ ID NO: 73) encoding the His_6-M4-CT fusion protein (SEQ ID NO: 74) was constructed as set out in Example 6, but the primers were designed in order to generate PCR fragments of domain M4 from a vector containing such a M4 sequence. The target vector was denoted pAff8His_6TrxHis_6CT, where the His_6TrxHis_6 part was cleaved off upon treatment with *NdeI* and *EcoRI*. The correctly ligated vector is denoted pT7His_6M4-CT.

Cloning of a gene (SEQ ID NO: 71) encoding the His_6-M4-NTCT fusion protein (SEQ ID NO: 72) was constructed in the same way as described for His_6-M4-CT, but the target vector here was denoted by pT7His_6scFv1-NTCT, where the pT7His_6scFv1 part was cleaved off upon treatment with *NdeI* and *EcoRI*. The correctly ligated vector is denoted as pT7His_6M4-NTCT.

Production

Production of His_6M4-CT and His_6M4-NTCT was performed in the same way as described in Example 6.

Purification

Purification of His_6M4-CT and His_6M4-NTCT was performed in the same way as described in Example 6.

The eluate contained 3.6 mg of His_6M4-CT protein. After protein concentration to 1.39 mg/ml, a final amount of 0.834 mg His_6M4-CT fusion protein was obtained.

The eluate content of His_6M4-NTCT protein was 3.2 mg. After protein concentration to 1.14 mg/ml, a final amount of 1.368 mg His_6M4-NTCT fusion protein was obtained.

Film, foam and fiber formation

Films of His_6M4-CT and His_6M4-NTCT were casted as as described in Example 6. Fiber was made for His_6-M4-NTCT from 1.14 mg/ml of soluble fusion protein (Fig. 5a). Foam was made for His_6-M4-CT from 30 µl of 1.39 mg/ml of soluble fusion protein (Fig. 5b). The fact that films, fiber and foam of His_6-M4-NTCT and His_6-M4-CT could be obtained although NTCT or CT has been fused to another protein, i.e. the 159 amino acids long M4 domain, demonstrates that NTCT and CT retain their structural forming properties despite being fused to the M4 domain.

Analysis

Spotted films of His_6-M4-NTCT and His_6-M4-CT contain 0.34 nmoles and 0.50 nmoles of target protein molecules, respectively. In order to evaluate the biotin binding ability of the two silk fused M4 constructs, two films for each construct were selected and analyzed by incubating with an equal amount of Atto-565-biotin as compared to the amount of target protein molecules present in the films. Then labeled biotin was removed and the films were washed three times with 100 µl of 1×PBS. Finally, 100 µl of 1×PBS was added to the films before fluorescence microscope analysis using an inverted Nikon Eclipse Ti instrument (excitation at 563 nm, emission at 592 nm). Films of NTCT (0.55 nmoles) and CT (1.28 nmoles) were used as control material, and were treated in the same way.

M4-NTCT and M4-CT films contain monomeric streptavidin (M4), which has an inherent affinity of binding to biotin. As labeled biotin was used in the analysis, binding of biotin to films could be detected by fluorescence microscopy at the wavelengths indicated above. Fig. 6 shows fluorescence microscopy pictures at 2× magnification for detection of Atto-565-biotin bound to M4-NTCT and M4-CT films. A: M4-NTCT; B: NTCT; C: M4-CT and D: CT. Fluorescence can be seen only from M4-NTCT and M4-CT films in Fig. 6 (A, C) but not from the control films (B, D). This confirms that the biotin binding ability of M4 is retained despite being fused to NTCT and CT.

Example 8 - Cloning, expression and formation of solid structures of scFv1-NTCT and scFv1-CT fusion proteins

NTCT and CT were produced in fusion with an engineered antibody fragment named single chain fragment variable (scFv1). scFv1 is a 27-kDa monovalent, engineered antibody fragment that recognizes the antigens specific for an autoimmune disease, Systemic Lupus Erythematosus (SLE). Our aim was to investigate whether it is possible to produce structures, such as fibers, foams and films, from the fusion proteins consisting of the scFv1 protein domain fused to NTCT (denoted His_6-scFv1-NTCT, SEQ ID NO: 76) and to CT (denoted His_6-scFv1-CT, SEQ ID NO: 78), respectively, and still retain the antigen detection ability of scFv1 domain as well as the structure forming properties of NTCT and CT. In order to do so, two fusion proteins consisting of the scFv1 domain fused N-terminally to NTCT and to CT were cloned.

Cloning

A gene (SEQ ID NO: 77) encoding the His_6-scFv1-CT fusion protein (SEQ ID NO: 78) was constructed as set out in Example 6, but the primers were designed in order to generate PCR fragments of domain scFv1 from a vector containing such a scFv1 sequence. The target vector was denoted pAff8His_6TrxHis_6CT, where the His_6TrxHis_6 part was cleaved off upon treatment with *NdeI* and *EcoRI*. The correctly ligated vector is denoted pT7His_6scFv1-CT.

Cloning of a gene (SEQ ID NO: 75) encoding the His_6-scFv1-NTCT fusion protein (SEQ ID NO: 76) was constructed in the same way as described for His_6-scFv1-CT, but the primers used for the amplification of NTCT contained sites for the restriction endonucleases EcoRI and HindIII and the target vector here was denoted by T7His_6scFv1-RepCT, where the RepCT part was cleaved off upon treatment with EcoRI and HindIII. The correctly ligated vector is denoted as pT7His_6scFv1-NTCT.

Production

Production of His_6-scFv1-CT and His_6-scFv1-NTCT was performed in the same way as described in Example 6, except for that production of pT7His_6-scFv1-NTCT was performed in a culture media total volume of 6 liters.

Purification

Purification of His_6-scFv1-CT and pHis_6-scFv1-NTCT was performed in the same way as described in Example 6.

The eluate contained 0.93 mg of His_6-scFv1-CT protein. After protein concentration to 0.87 mg/ml, a final amount of 0.348 mg His_6-scFv1-CT fusion protein was obtained.

The eluate content of His_6-scFv1-NTCT protein was 4.86 mg. After protein concentration to 2.14 mg/ml, a final amount of 2.57 mg His_6-scFv1-NTCT fusion protein was obtained.

Film, foam and fiber formation

Films of His_6-scFv1-CT were spotted onto microarray slides (plastic MaxiSorp, Nunc) from 1 µl of 5 µM soluble fusion protein per film. The films were then allowed to solidify over night in a climate controlled room. The same procedure was followed for casting films of His_6-scFv1-NTCT from 1 µl of 5 µM protein solution.

Fiber was made for His_6-scFv1-NTCT from 0.49 mg/ml (data not shown) and foams were made for both His_6-scFv1-NTCT and His_6-scFv1-CT from 30 µl of 0.22 and 0.38 mg/ml of soluble fusion protein, respectively (Fig-

7a and 7b). The fact that macroscopic fiber and foam for His_6-scFv1-NTCT and His_6-scFv1-CT respectively, could be obtained although NTCT or CT has been fused to another protein, i.e. the 263 amino acids long scFv1 domain, demonstrates that NTCT and CT still retains there structure forming properties despite fused to the scFv1 domain.

Analysis

Pure antibody (scFv1, control) and silk fused antibody (scFv1-NTCT) were spotted in the microarray format manually by adding 1 µL of 5 µM protein solution onto clear and black polymer MaxiSorp microarray slides (NUNC, 25x76 mm) resulting in 135 pmoles of pure antibody (scFv1) and 274 pmoles of silk fused antibody (scFv1-NTCT) in the spotted films, respectively. After spotting the proteins in film format, the films were dried overnight in a climate controlled room. The arrays were then blocked by applying 200 µl of sample buffer (1% (w/v) fat-free milk powder and 1% (v/v) Tween-20 in PBS) for 90 min and then washed three times by applying 200-300 µl of wash buffer (0.05% (v/v) Tween-20 in PBS). All incubations were performed at room temperature on gentle agitation. Next, 100-200 µl of biotinylated antigen sample (10 nM) diluted in sample buffer was applied and incubated for 1h. The arrays were then washed three times by applying 200-300 µl of wash buffer and to detect the bound antigens, 100-200 µl of Alexa-647-labeled streptavidin (1 µg/ml) diluted in sample buffer, was applied onto the arrays and incubated for 1 h. Finally, the arrays were washed three times with 200-300 µl of wash buffer and dried under a stream of nitrogen gas. The arrays were then scanned using a confocal microarray fluorescence scanner (ScanArray Express, Perkin-Elmer Life & Analytical Sciences). The ScanArray Express software V2.0 (Perkin-Elmer Life & Analytical Sciences) was used to quantify the intensity of each spot. The same analysis procedure was carried out for analyzing His_6-scFv1-CT fusion protein.

In order to detect the low abundant serum proteins which can be of potential biomarkers, scFv1 was fused to N-terminal of NTCT or CT giving rise to His_6-scFv1-NTCT and His_6-scFv1-CT, respectively. Pure antibody (control) and silk fused antibody fragments were spotted onto the microarray

slide and their antigen binding capacity was analyzed using biotinylated antigen sample. Alexa-647-labeled streptavidin was then used to detect the bound antigens. Fig. 8 shows an antigen binding analysis of pure (control) and silk fused antibody fragments. Intensity of the spots was measured at 5090 detection intensity. The analysis showed that the antigen recognition of silk fused antibody (His_6-scFv1-NTCT) fragment was increased by 25 times compared to the scFv1 control alone, and no sign of cross reactivity with other antigens was observed for His_6-scFv1-NTCT.

Example 9 - Cloning, expression and formation of solid structures of xylanase-NTCT and xylanase-CT fusion proteins

To prove the concept of fusing a protein with enzymatic activity to NTCT and CT, the enzyme xylanase A from *Bacillus subtilis* was produced in fusion with NTCT and CT, respectively. Xylanase A (endo-1,4-beta-xylanase A) is 185 amino acids long (without signal peptide) and belongs to the glycosyl hydrolase 11 (cellulose G) family. The enzymatic function of xylanase A is to cleave beta-1,4-glycosidic linkages of xylan, the main constituent of hemicellulose in plant cell walls. Our aim was to investigate whether it is possible to produce structures, such as fibers, foams and films, from a fusion protein consisting of xylanase A fused to NTCT (denoted Xyl-NTCT, SEQ ID NO: 80) and to CT (denoted Xyl-CT, SEQ ID NO: 82), respectively, and still retain the enzymatic ability of xylanase as well as the structure forming properties of CT. In order to do so two fusion proteins were cloned consisting of xylanase N-terminally to 1) NTCT and 2) CT.

Cloning

A gene (SEQ ID NO: 81) encoding the His_6Xyl-CT fusion protein (SEQ ID NO: 82) was constructed as set out in Example 6, but the primers were designed in order to generate PCR fragments of the xylanase domain from a vector containing such a xylanase sequence. The target vector was denoted pAff8His_6TrxHis_6CT, where the His_6TrxHis_6 part was cleaved off upon treatment with *NdeI* and *EcoRI*. The correctly ligated vector is denoted pT7His_6Xyl-CT.

Cloning of a gene (SEQ ID NO: 79) encoding the His_6Xyl-NTCT fusion protein (SEQ ID NO: 80) was constructed in the same way as described for His_6Xyl-CT, but the target vector here was denoted by pT7His_6scFv1-NTCT, where the pT7His_6scFv1 part was cleaved off upon treatment with *NdeI* and *EcoRI*. The correctly ligated vector is denoted as pT7His_6Xyl-NTCT.

Production

Production of His_6Xyl-CT and His_6Xyl-NTCT was performed in the same way as described in Example 6.

Purification

Purification of His_6Xyl-CT and His_6Xyl-NTCT was performed in the same way as described in Example 6.

The eluate contained 3.6 mg of His_6Xyl-CT protein. After protein concentration to 2.1 mg/ml, a final amount of 2 mg His_6Xyl-CT fusion protein was obtained.

The eluate content of His_6Xyl-NTCT protein was 4.3 mg. After protein concentration to 0.65 mg/ml, a final amount of 0.3 mg His_6Xyl-NTCT fusion protein was obtained.

Film, foam and fiber formation

Films of His_6Xyl-CT were casted in 96-well plates (Tissue culture plate, Suspension cells, 83.1835.500, Sarstedt). Each film was made from 15 µl of 1.0 mg/ml soluble His_6Xyl-CT at both pH 8 and pH 6. The films were then allowed to solidify over night (20°C, 35% relative humidity). The same procedure was followed for casting of His_6Xyl-NTCT films, each film casted from 15 µl of 0.41 mg/ml soluble His_6Xyl-NTCT (both at pH 8 and pH 6).

Foam was made from soluble His_6Xyl-CT (1-2 mg/ml) by introducing air (by pipetting) into 40 µl of the protein solution, followed by overnight drying at room temperature. The appearance of the formed foam of His_6Xyl-CT (Fig. 9), demonstrates that CT retains its structural forming properties despite being fused to the Xyl domain.

Fig. 9 shows a macroscopic foam of the fusion protein His_6Xyl-CT (SEQ ID NO: 82), made from soluble fusion protein. The appearance of the foam demonstrates that the spider silk CT domain has retained its structural forming properties even though produced in fusion with the enzyme xylanase (Xyl).

Analysis of enzymatic activity of xylanase fused to NTCT or CT

Xylanase is an enzyme that cleaves the beta-1,4-glycosidic linkage between two xylose residues. To test the enzymatic ability in films of xylanase (Xyl) fused to NTCT or CT, each film of His_6Xyl-CT and His_6Xyl-NTCT is incubated with 90 µl of McIlvaine buffer (pH 6.0). After a 10 min preincubation at 50°C of all films, 10 µl of 40 mM PNX (p-nitrophenyl-xylopyranoside) substrate is added, followed by an additional incubation at 50°C for at least 10 min. Then, 100 µl of stop solution (0.5 M Na_2CO_3) is added to each film, followed by absorbance measurements at 410 nm to identify the product from the enzymatic reaction. Films of NTCT and CT, casted according to the same procedure previously stated, are included as controls.

Example 10 - Cloning, expression and formation of solid structures of xylanase-NTCT and xylanase-CT fusion proteins

To prove the concept of fusing a peptide with cell stimulating effect to NTCT and CT, the human epidermal growth factor (EGF) was produced in fusion with NTCT and CT, respectively. EGF is a 53 amino acid residues long growth factor with high affinity for the Epidermal Growth Factor Receptor (EGFR) found on the cell surface of many cell types, e.g. keratinoytes. Upon binding to EGFR, protein-tyrosin kinase activity is stimulated, resulting in a variety of biochemical changes in the cell, triggering cell growth and proliferation. Our aim was to investigate whether it is possible to produce structures, such as fibers and films, from a fusion protein consisting of EGF fused to NTCT (denoted EGF-NTCT, SEQ ID NO: 84) and to CT (denoted EGF-CT, SEQ ID NO: 86), respectively, and still retain the cell stimulating effect of EGF as well as the structure forming properties of CT. In order to do

so two fusion proteins were cloned consisting of EGF N-terminally to 1) NTCT and 2) CT.

Cloning

A gene (SEQ ID NO: 85) encoding the His_6EGF-CT (SEQ ID NO: 86) fusion protein was constructed as set out in Example 6, but the primers were designed in order to generate PCR fragments of EGF from a vector containing such an EGF sequence. The target vector was denoted pAff8His_6TrxHis_6CT, where the His_6TrxHis_6 part was cleaved off upon treatment with *NdeI* and *EcoRI*. The correctly ligated vector is denoted pT7His_6EGF-CT.

Cloning of a gene (SEQ ID NO: 83) encoding the His_6EGF-NTCT fusion protein (SEQ ID NO: 84) was constructed in the same way as described for His_6EGF-CT, but the target vector here was denoted by pT7His_6scFv1-NTCT, where the pT7His_6scFv1 part was cleaved off upon treatment with *NdeI* and *EcoRI*. The correctly ligated vector is denoted as pT7His_6EGF-NTCT.

Production

Production of His_6EGF-CT and His_6EGF-NTCT is performed in the same way as described in Example 6.

Purification

Purification of His_6EGF-CT and His_6EGF-NTCT is performed in the same way as described in Example 6.

Film and fiber formation

Films and fibers are made from both His_6EGF-CT and His_6EGF-NTCT as described in Example 6.

Analysis of the cell stimulating ability of EGF fused to NTCT or CT

The cell stimulating ability towards keratinocytes of EGF when fused to CT (denoted His_6EGF-CT) or NTCT (denoted His_6EGF-NTCT) is investigated.

For this purpose, films of the fusion proteins His_6EGF-CT and His_6EGF-NTCT, respectively are used. Normal human epidermal keratinocytes (primary cells) are seeded onto matrices at a density of 3500 or 7000 cells/cm^2 in cell culture medium (KGM-GOLD, Lonza) with or without recombinant human EGF. The medium is exchanged every second day. Viable and dead cells are stained with Live/Dead assay (Molecular probes) after 24, 48, 72 and 96 h. Micrographs are taken in an Inverted Fluorescent microscope (Nikon Eclipse Ti) at 10× magnification.

2013255815 07 Mar 2018

CLAIMS:

1. A recombinant fusion protein comprising the moieties **B** and **CT**, wherein:

B is a non-spidroin moiety which provides the capacity of selective interaction with an organic target, wherein the **B** moiety is a protein or polypeptide fragment comprising more than 30 amino acid residues; and wherein the **B** moiety has less than 30% identity to any of SEQ ID NOS: 6-10; and

CT is a moiety of from 70 to 120 amino acid residues and has at least 80% identity to any one of SEQ ID NO: 7 and SEQ ID NOS: 14-44; wherein the **CT** moiety provides the capacity of forming a polymer;

with the proviso that the fusion protein is not comprising any moiety derived from the repetitive fragment of a spider silk protein.

2. A recombinant fusion protein according to claim 1, wherein the organic target is selected from the group consisting of immunoglobulins and molecules comprising immunoglobulin or derivatives thereof.

3. A recombinant fusion protein according to claim 2, wherein the immunoglobulins are selected from the immunoglobulin subclasses IgG1, IgG2, IgG4, IgA and IGM from human.

4. A recombinant fusion protein according to claim 2 or 3, wherein the **B** moiety is selected from the group consisting of the Z domain derived from staphylococcal protein A and the E, D, A, B and C domains thereof, streptococcal protein G and the C1, C2 and C3 domains thereof; and protein fragments having at least 70% identity to any of these amino acid sequences.

5. A recombinant fusion protein according to claim 4, wherein the **B** moiety is selected from the group consisting of the Z domain derived from staphylococcal protein A, the B domain of staphylococcal protein A, and the C2 domain of streptococcal protein G; and protein fragments having at least 70% identity to any of these amino acid sequences.

6. A recombinant fusion protein according to claim 1, wherein the organic target is selected from the group consisting of albumin and molecules comprising albumin or derivatives thereof.

7. A recombinant fusion protein according to claim 6, wherein the **B** moiety is selected from streptococcal protein G, the albumin-binding domain of streptococcal protein G, GA

modules from *Finegoldia magna*; and protein fragments having at least 70% identity to any of these amino acid sequences.

8. A recombinant fusion protein according to claim 1, wherein the organic target is selected from the group consisting of biotin and molecules comprising biotin or derivatives or analogues thereof.

9. A recombinant fusion protein according to claim 8, wherein the **B** moiety is selected from the group consisting of streptavidin, monomeric streptavidin (M4); and protein fragments having at least 70% identity to any of these amino acid sequences.

10. A recombinant fusion protein according to any preceding claim, selected from the group of proteins defined by the formula $\mathbf{B_x\text{-}CT\text{-}B_z}$,

wherein x and z are integers from 0 to 5;

and $x+z \geq 1$.

11. A recombinant fusion protein according to any previous claim, selected from the group consisting of SEQ ID NOS: 61-66; and proteins having at least 80% identity to any of these sequences.

12. An isolated nucleic acid encoding a fusion protein according to any one of claims 1 to 11.

13. A protein structure capable of selective interaction with an organic target, wherein said protein structure is a polymer comprising as a repeating structural unit a recombinant fusion protein according to any one of claims 1 to 11, wherein the **B** moiety provides the capacity of selective interaction with the organic target.

14. A protein structure according to claim 13, wherein said protein structure is in a physical form selected from the group consisting of fiber, film, foam, net, mesh, sphere and capsule.

15. Use of a protein structure according to any one of claims 13 or 14 in cultivation of cells.

16. A method of producing a fusion protein, comprising the following steps:

a) expressing in a suitable host a fusion protein according to any one of claims 1 to11; and

2013255815 07 Mar 2018

b) obtaining a mixture containing the fusion protein, and optionally isolating the fusion protein.

17. A method for providing a protein structure according to any one of claims 12 to 14, displaying a binding activity towards an organic target, comprising the steps of:

(a) providing a recombinant fusion protein according to any one of claims 1 to11;

(b) subjecting the fusion protein to conditions to achieve formation of a polymer comprising the recombinant fusion protein.

Spiber Technologies AB

Patent Attorneys for the Applicant/Nominated Person

SPRUSON & FERGUSON

```
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 SGG-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 QVSASNPGSS
AY953075_Lh2    SALSSPTTHA RISSHASTLL SSG-PTNAAA LSNVISNAVS QVSASNPGSS
AY654293_Nc2    SRLASPDSGA RVASAVSNLV SSG-PTSSAA LSSVISNAVS QIGASNPGLS
AF350278_Nm2    SRLASPDSGA RVASAVSNLV SSG-PTSSAA LSSVISNAVS QIGASNPGLS
AF350280_Ns2    SRLASPDSGA 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 1

```
CThyb_Esp       GCDVLVQALL EVVSALIHIL GSSSIGQVNY GSAGQATQLV GQSVYQALGE F
CTnat_Eau       GCEVIVQALL EVITALVQIV SSSSVGYINP SAVNQITNVV ANAMAQVMG- -
AF350266_At1    GCDVLVQALL EIVSALVHIL GSANIGQVNS SGVGRSASIV GQSINQAFS- -
AY666062_Cm1    GCDVLVQALL EVVSALIHIL GSSSIGQVNY GSAGQATQIV ---------- -
AF350273_Lg1    SCDVLVQALL ELVTALLTII GSSNVGNVNY DSSGQYAQVV SQSVQNAFV- -
AY953074_Lh1    ACDVLVQALL ELVTALLTII 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 SSANVGYINP EASGSLN-AV GSALAAAMG- -
U47853_ADF1     SSEALIQALL EVISALIHVL GSASIGNVSS VGVNSALNAV QNAVGAYAG- -
U47854_ADF2     GCDILVQALL EIISALVHIL GSANIGPVNS SSAGQSASIV GQSVYRALS- -
U47855_ADF3     GCDVLVQALL EVVSALVSIL GSSSIGQINY GASAQYTQMV GQSVAQALA- -
U47856_ADF4     GCDALVQALL ELVSALVAIL SSASIGQVNV SSVSQSTQMI SQALS----- -

Consensus       GCDVLVQALL EVVSALVHIL GSSSIGQVNY GSAGQATQIV GQSVAQALGE F
```

Fig 1 (continued)

```
Ea MaSp1   SHTTPWTNPGLAENFMNSFMQGLSSMPGFTASQLDDMSTIAQSMVQSIQSLAAQGRTSPNKLQALNMAFA
Lg MaSp1   QANTPWSSKANADAFINSFISSAQNTGSFSQDQMDDMSLIGNTLMTAMDNMG--GRITPSKLQALDMAFA
Lh MaSp1   QANTPWSSKANADAFINSFISAASNTGSFSQDQMEDMSLIGNTLMAAMDNMG--GRITPSKLQALDMAFA
Nc MaSp1   -QNTPWSSTELADAFINAFMNEAGRTGAFTADQLDDMSTIGDTIKTAMDKMARSNKSSKGKLQALNMAFA
At MaSp2   QGATPWENSQLAESFISRFLRFIGQSGAFSPNQLDDMSSIGDTLKTAIEKMAQSRKSSKSKLQALNMAFA
Lg MaSp2   ---LRWSSKDNADRFINAFLQAASNSGAFSSDQVDDMSVIGNTLMTAMDNMG--GRITPSKLQALDMAFA
Lh MaSp2   QANTPWSSKENADAFIGAFMNAASQSGAFSSDQIDDMSVISNTLMAAMDNMG--GRITQSKLQALDMAFA
Nim MaSp2  QANTPWSDTATADAFIQNFLGAVSGSGAFTPDQLDDMSTVGDTIMSAMDKMARSNKSSKSKLQALNMAFA
Nc MaSp2   QARSPWSDTATADAFIQNFLAAVSGSGAFTSDQLDDMSTIGDTIMSAMDKMARSNKSSQHKLQALNMAFA
Ab CySp1   AVPSVFSSPNLASGFLQCLTFGIGNSPAFPTQEQQDLDAIAQVILNAVSSNTGATASAR--AQALSTALA
Ncl CySp1  PVPSVFSSPSLASGFLGCLTTGIGLSPAFPFQEQQDLDDLAKVILSAVTSNTDTSKSAR--AQALSTALA
Lh TuSp1   ASVNIFNSPNAATSFLNCLRSNIESSPAFPFQEQADLDSIAEVILSDVSS-VNTASSAT--SLALSTALA
Nc flag    IANSPFSNPNTAEAFARSFVSNIVSSGEFGAQGAEDFDDIIQSLIQAQ-SMGKGRHDTKAKAKAMQVALA
Nlm flag   IVNSPFSNPNTAEAFARSFVSNVVSSGEFGAQGAEDFDDIIQSLIQAQ-SMGKGRHDTKAKAKAMQVALA

Ea MaSp1   SSMAEIAASEEGGGSLSTKTSSIASAMSNAFLQTTGVVNQPFINEITQLVSMFAQAGMNDV
Lg MaSp1   SSVAEIAASEG--GDLGVTTNAIADALTSAFYQTTGVVNNRFISEIRSLISMFAQASANDV
Lh MaSp1   SSVAEIAASEG--GDLGVTTNAIADALTSAFYQTTGVVNSRFISEIRSLIGMFAQASANDV
Nc MaSp1   SSMAEIAAVEQGGLSVDAKTNAIADSLNSAFYQTTGAANPQFVNEIRSLINMFAQSSANEV
At MaSp2   SSMAEIAVAEQGGLSLEAKTNAIASALSAAFLETTGYVNQQFVNEIKTLIFMIAQASSNEI
Lg MaSp2   SSVAEIAVADG--QNVGGATNAISNALRSAFYQTTGVVNNQFISEISNLINMFAQVSANEV
Lh MaSp2   SSVAEIAVADG--QNVGAATNAISDALRSAFYQTTGVVNNQFITGISSLIGMFAQVSGNEV
Nim MaSp2  SSMAEIAAVEQGGQSMDVKTNAIANALDSAFYMTTGSTNQQFVNEMRSLINMLSAAAVNEV
Nc MaSp2   SSMAEIAAVEQGGMSMAVKTNAIVDGLNSAFYMTTGAANPQFVNEMRSLISMISAASANEV
Ab CySp1   SSLTDLLIAESAESNYSNQLSELTGILSDCFIQTTGSDNPAFVSRIQSLISVLSQNADTNI
Ncl CySp1  SSLADLLISESSGSSYQTQISALTNILSDCFVTTTGSNNPAFVSRVQTLIGVLSQSSSNAI
Lh TuSp1   SSLAELLVTESAEEDIDNQVVALSTILSQCFVETTGSPNPAFVASVKSLLGVLSQSASNYE
Nc flag    SSIAELVIAESSGGDVQRKTNVISNALRNALMSTTGSPNEEFVHEVQDLIQMLSQEQINEV
Nlm flag   SSIAELVIAESSGGDVQRKTNVISNALRNALMSTTGSPNEEFVHEVQDLIQMLSQEQINEV
```

Fig 2

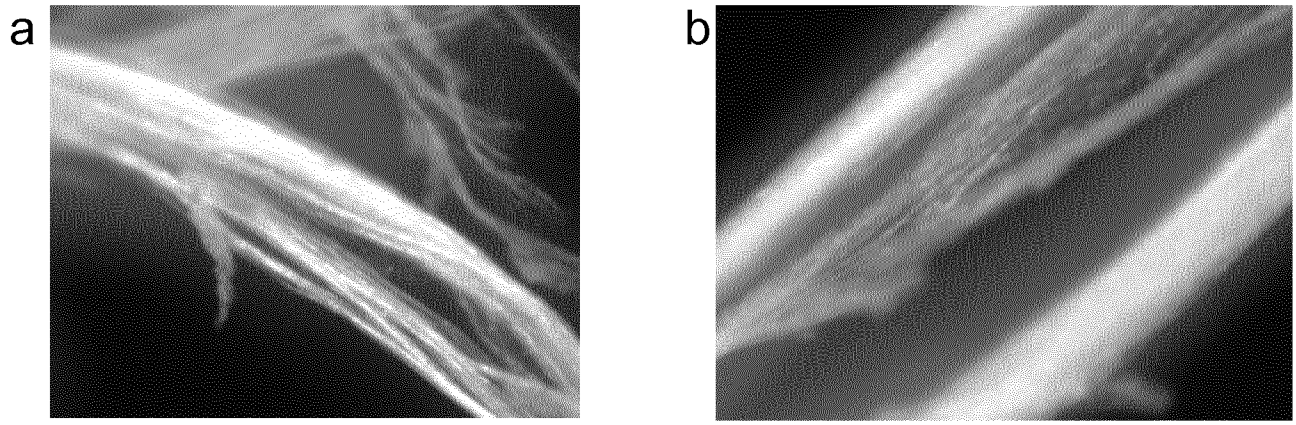
a
b

Fig 3

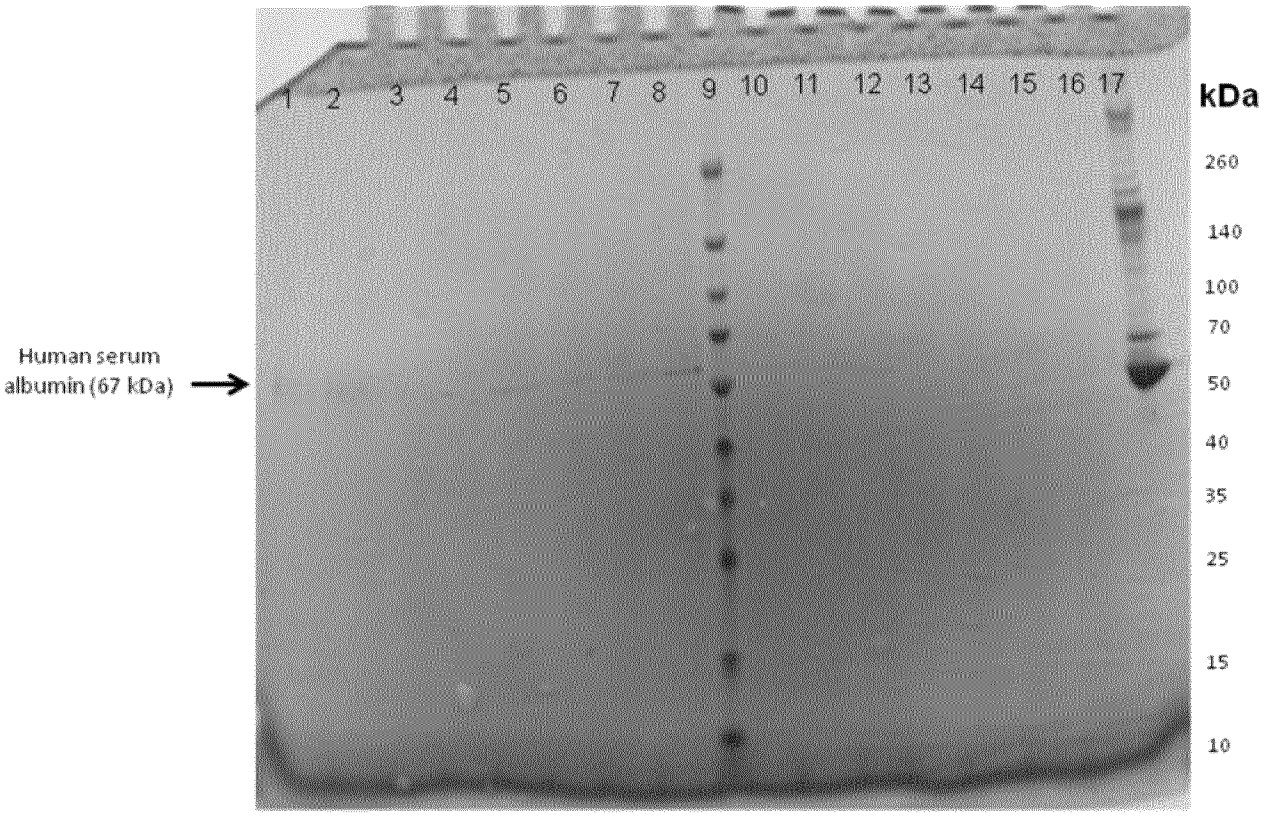
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
kDa
260
140
100
70
50
40
35
25
15
10
Human serum albumin (67 kDa)

Fig 4

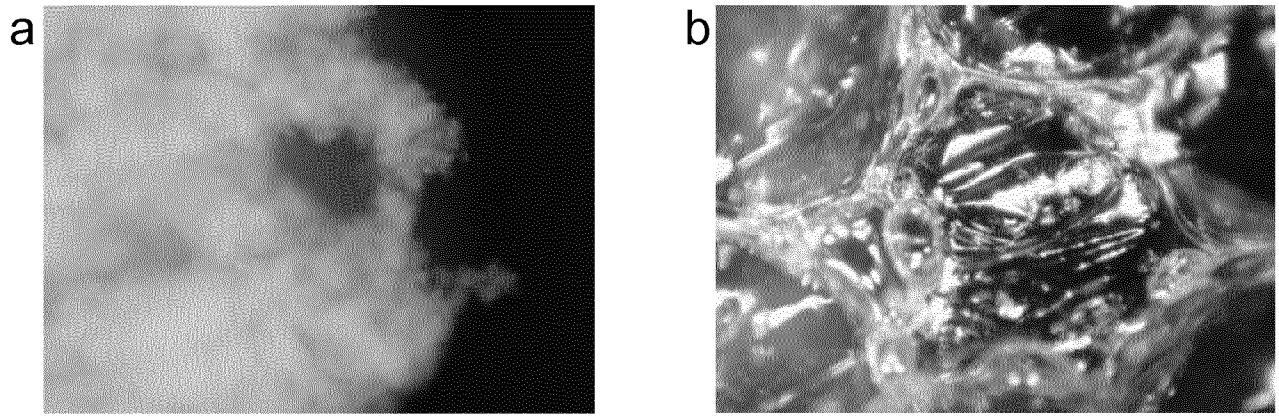

Fig 5

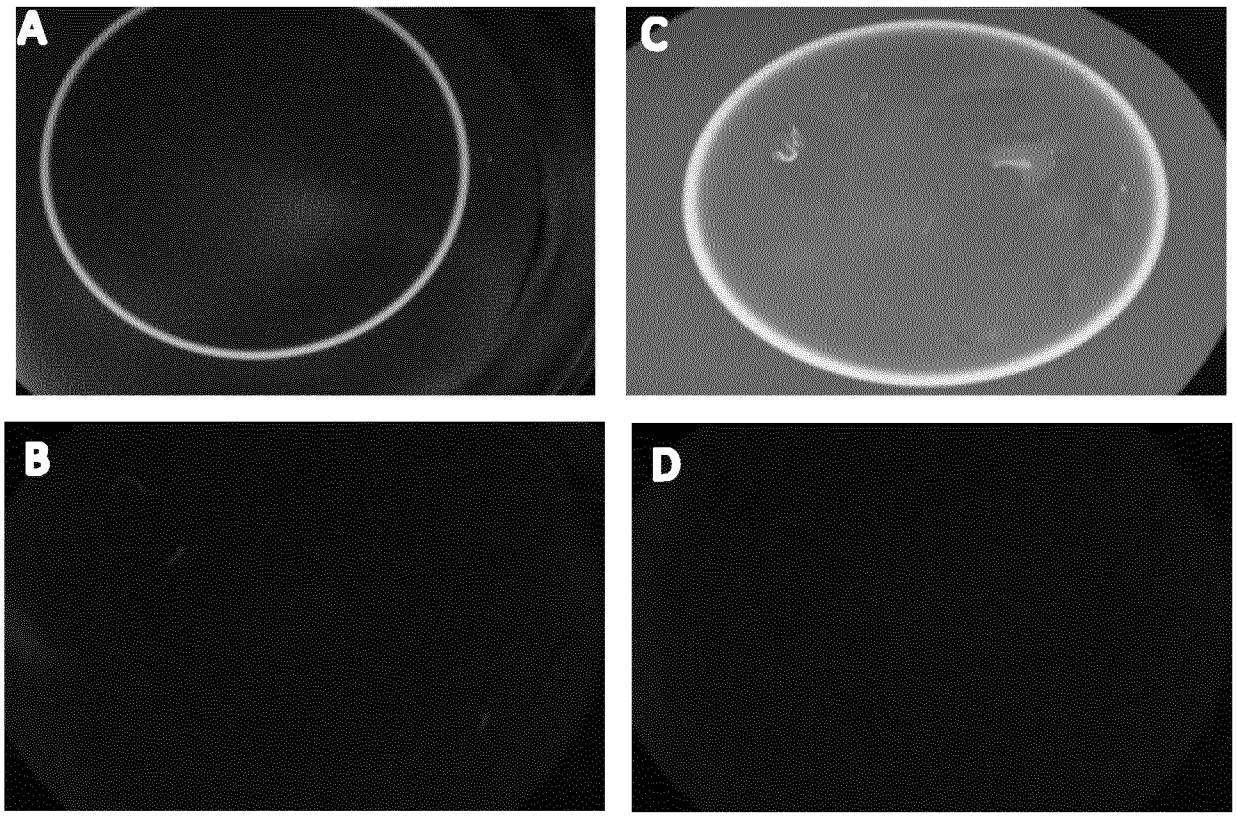

Fig 6

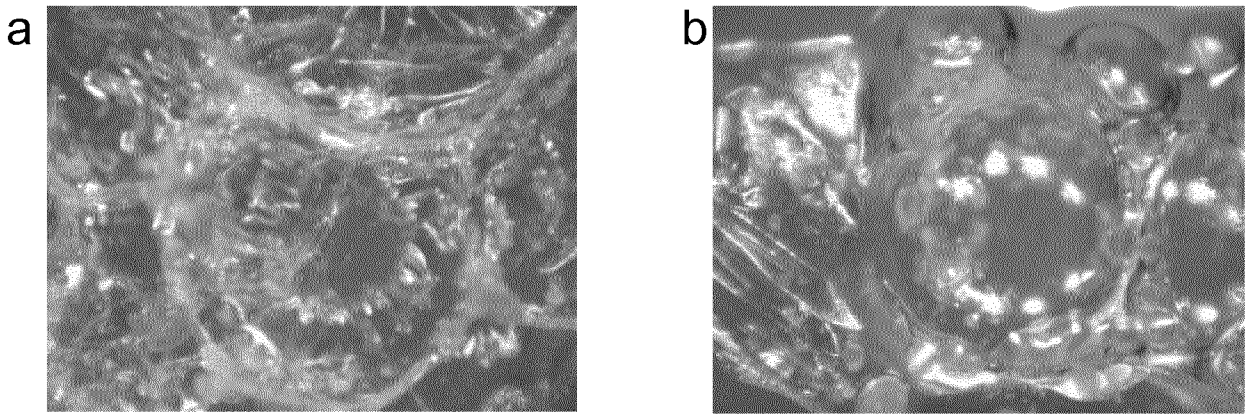
a
b

Fig 7

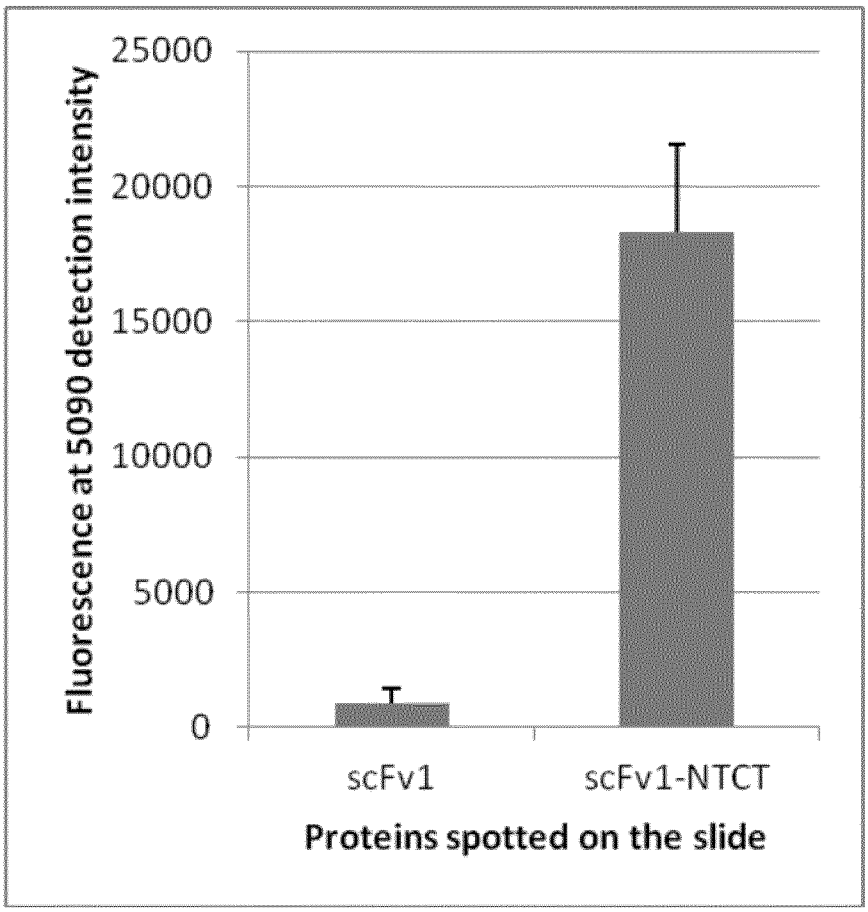
25000
20000
15000
10000
5000
0
Fluorescence at 5090 detection intensity
scFv1
scFv1-NTCT
Proteins spotted on the slide

Fig 8

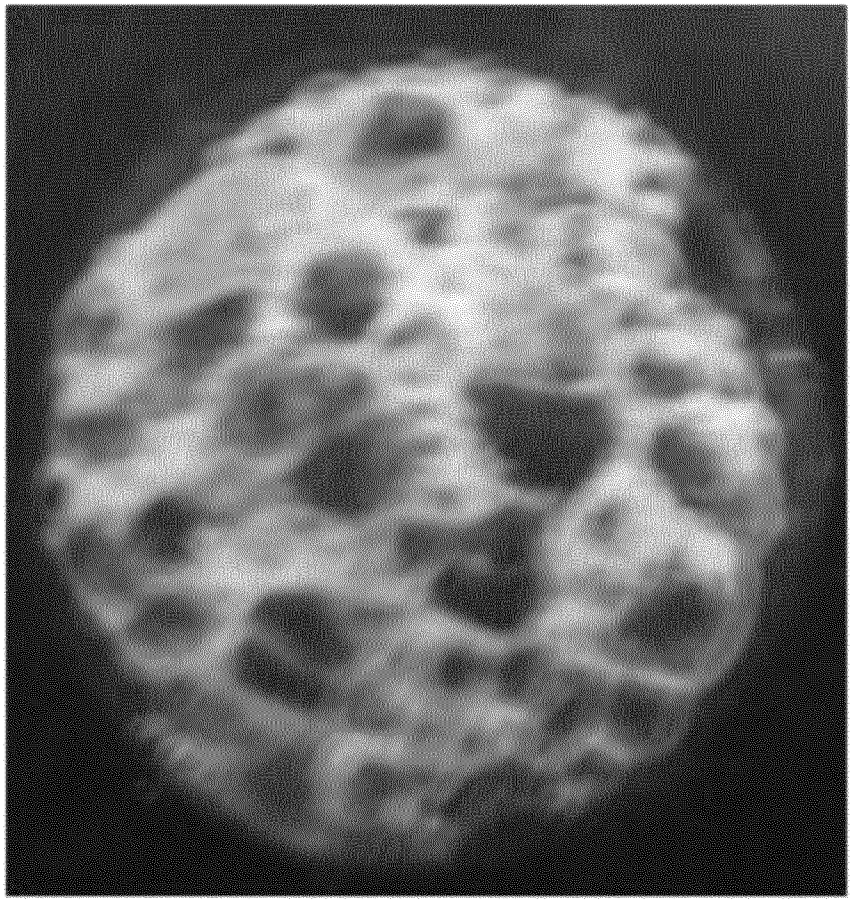

Fig 9

SEQUENCE LISTING

<110> Spiber Technologies AB

<120> SPIDER SILK FUSION PROTEIN STRUCTURES WITHOUT REPETITIVE FRAGMENT FOR BINDING TO AN ORGANIC TARGET

<130> PC-21061941

<150> EP 12166393
<151> 2012-05-02

<160> 86

<170> PatentIn version 3.5

<210> 1
<211> 149
<212> PRT
<213> Euprosthenops 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> Euprosthenops 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 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>  Euprosthenops 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 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 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> Euprosthenops 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>  Euprosthenops 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

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 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>  Euprosthenops 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>  Euprosthenops 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> Euprosthenops 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 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 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 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
        435                 440                 445

Gly Tyr Gly Gln Gly Ala Gly Ser Ser Ala 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 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 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 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 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 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 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
785                 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
865                 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
            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
    930                 935                 940

Ala Gly Gln Gly Gly Gln Gly Gly Tyr Gly Gly Leu Gly Gln Gly Gly
945                 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
    1010                1015                1020

Ser Gly  Gly Asn Ala Ala Ala  Ala Ala Ala Ala Ala  Ala Ala Ala
    1025                1030                1035

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 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 Euprosthenops 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 Euprosthenops 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 Euprosthenops 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> 100
<212> PRT
<213> Euprosthenops sp

<400> 14

```
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> 15
<211> 98
<212> PRT
<213> Euprosthenops australis

<400> 15

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> 16
<211> 99
<212> PRT
<213> Argiope trifasciata

<400> 16

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> 17
<211> 89
<212> PRT
<213> Cyrtophora moluccensis

<400> 17

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> 18
<211> 98
<212> PRT
<213> Latrodectus geometricus

<400> 18

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> 19
<211> 98
<212> PRT
<213> Latrodectus hesperus

<400> 19

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> 20
<211> 93
<212> PRT
<213> Macrothele holsti

<400> 20

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> 21
<211> 98
<212> PRT
<213> Nephila clavipes

<400> 21

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> 22
<211> 89
<212> PRT
<213> Nephila pilipes

<400> 22

```
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> 23
<211> 87
<212> PRT
<213> Nephila madagascariensis

<400> 23

```
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> 24
<211> 87
<212> PRT
<213> Nephila senegalensis

<400> 24

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> 25
<211> 89
<212> PRT
<213> Octonoba varians

<400> 25

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>  26
<211>  89
<212>  PRT
<213>  Psechrus sinensis

<400>  26

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>  27
<211>  88
<212>  PRT
<213>  Tetragnatha kauaiensis

<400>  27

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> 28
<211> 88
<212> PRT
<213> Tetragnatha versicolor

<400> 28

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> 29
<211> 89
<212> PRT
<213> Araneus bicentenarius

<400> 29

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> 30
<211> 97
<212> PRT
<213> Argiope amoena

<400> 30

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> 31
<211> 82
<212> PRT
<213> Argiope aurantia

<400> 31

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> 32
<211> 98
<212> PRT
<213> Argiope trifasciata

<400> 32

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> 33
<211> 89
<212> PRT
<213> Gasteracantha mammosa

<400> 33

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> 34
<211> 90

<212> PRT
<213> Latrodectus geometricus

<400> 34

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> 35
<211> 98
<212> PRT
<213> Latrodectus hesperus

<400> 35

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

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> 36
<211> 97
<212> PRT
<213> Nephila clavipes

<400> 36

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> 37
<211> 82
<212> PRT
<213> Nephila madagascariensis

<400> 37

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

<210> 38
<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> 38

```
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 Ser Ile Gly Gln Val Asn Tyr Gly
65                  70                  75                  80


Ala Ala
```

<210> 39
<211> 71
<212> PRT
<213> Dolomedes tenebrosus

<400> 39

```
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> 40
<211> 97
<212> PRT
<213> Dolomedes tenebrosus

<400> 40

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> 41
<211> 93
<212> PRT
<213> Araneus diadematus

<400> 41

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

<210> 42
<211> 98
<212> PRT
<213> Araneus diadematus

<400> 42

```
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> 43
<211> 98
<212> PRT
<213> Araneus diadematus

<400> 43

```
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> 44
<211> 94
<212> PRT
<213> Araneus diadematus

<400> 44

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> 45
<211> 131
<212> PRT
<213> Euprosthenops australis

<400> 45

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> 46
<211> 127
<212> PRT
<213> Latrodectus geometricus

<400> 46

```
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 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> 47
<211> 127
<212> PRT
<213> Latrodectus hesperus

<400> 47

```
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> 48
<211> 130
<212> PRT
<213> Nephila clavipes

<400> 48

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> 49
<211> 131
<212> PRT
<213> Argiope trifasciata

<400> 49

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>  50
<211>  124
<212>  PRT
<213>  Latrodectus geometricus

<400>  50

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> 51
<211> 127
<212> PRT
<213> Latrodectus hesperus

<400> 51

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> 52
<211> 131
<212> PRT
<213> Nephila inaurata madagascariensis

<400> 52

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>  53
<211>  131
<212>  PRT
<213>  Nephila clavipes

<400>  53

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>  54
<211>  129
<212>  PRT
<213>  Argiope bruennichi

<400>  54

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> 55
<211> 129
<212> PRT
<213> Nephila clavata

<400> 55

```
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> 56
<211> 128
<212> PRT
<213> Latrodectus hesperus

<400> 56

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> 57
<211> 130
<212> PRT
<213> Nephila clavipes

<400> 57

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>  58
<211>  130
<212>  PRT
<213>  Nephila inaurata madagascariensis

<400>  58

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>  59
<211>  283
<212>  PRT
<213>  Euprosthenops australis
```

<400> 59

```
His His His His His His Met Ser His Thr Thr Pro Trp Thr Asn Pro
1               5                  10                  15

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

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

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

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

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

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

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

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

Gly Ala Ser Ala Ala Ala Ser Ala Gly Ala Gly Ser Gly Asn Ser Gly
145                 150                 155                 160

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

Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro Ser
            180                 185                 190

Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn Gly
        195                 200                 205

Gln Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser
    210                 215                 220

Ser Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile Val
225                 230                 235                 240

Gln Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser Ser
                245                 250                 255

Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile Thr Asn
```

260 265 270

Val Val Ala Asn Ala Met Ala Gln Val Met Gly
275 280

<210> 60
<211> 425
<212> PRT
<213> Euprosthenops australis

<400> 60

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 Gly Asn Ser Gly Ile Gln
    290                 295                 300

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

Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro Ser Ala Val
                325                 330                 335

Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn Gly Gln Val
            340                 345                 350

Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser Ser Val
        355                 360                 365

Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile Val Gln Ala
    370                 375                 380

Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser Ser Ser Ser
385                 390                 395                 400

Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile Thr Asn Val Val
                405                 410                 415

Ala Asn Ala Met Ala Gln Val Met Gly
            420                 425


<210>  61
<211>  206
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Fusion protein

<400>  61

Met Gly Ser Ser Gly His His His His His His Met Val Asp Asn Lys
1               5                   10                  15
```

```
Phe Asn Lys Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu His Leu Pro
            20                  25                  30

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

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

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

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

Ala Ala Ser Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser
            100                 105                 110

Ser Pro Ser Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val
        115                 120                 125

Ser Asn Gly Gln Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn
    130                 135                 140

Ile Ser Ser Ser Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu
145                 150                 155                 160

Val Ile Val Gln Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile
                165                 170                 175

Val Ser Ser Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln
            180                 185                 190

Ile Thr Asn Val Val Ala Asn Ala Met Ala Gln Val Met Gly
        195                 200                 205


<210>  62
<211>  206
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Fusion protein

<400>  62

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

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

Ser Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro
        35                  40                  45
```

```
Ser Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn
    50                  55                  60

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

Ser Ser Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile
                85                  90                  95

Val Gln Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser
            100                 105                 110

Ser Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile Thr
        115                 120                 125

Asn Val Val Ala Asn Ala Met Ala Gln Val Met Gly Leu Glu Ala Leu
    130                 135                 140

Phe Gln Gly Pro Val Asp Asn Lys Phe Asn Lys Glu Gln Gln Asn Ala
145                 150                 155                 160

Phe Tyr Glu Ile Leu His Leu Pro Asn Leu Asn Glu Glu Gln Arg Asn
                165                 170                 175

Ala Phe Ile Gln Ser Leu Lys Asp Asp Pro Ser Gln Ser Ala Asn Leu
            180                 185                 190

Leu Ala Glu Ala Lys Lys Leu Asn Asp Ala Gln Ala Pro Lys
        195                 200                 205


<210>  63
<211>  354
<212>  PRT
<213>  Artificial Sequence

<220>
<223>  Fusion protein

<400>  63

Met Gly Ser Ser Gly His His His His His His Met Val Asp Asn Lys
1               5                   10                  15

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

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

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

```
Asp Ala Gln Ala Pro Lys Leu Glu Ala Leu Phe Gln Gly Pro Ser His
65                  70                  75                  80

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

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

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

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

Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly
145                 150                 155                 160

Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser
                165                 170                 175

Asn Ala Phe Leu Gln Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn
            180                 185                 190

Glu Ile Thr Gln Leu Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp
        195                 200                 205

Val Ser Ala Ser Ala Ser Ala Gly Ala Ser Ala Ala Ala Ser Ala Gly
    210                 215                 220

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

Ser Ala Ser Ala Ala Ala Ser Ala Ala Ser Thr Val Ala Asn Ser Val
                245                 250                 255

Ser Arg Leu Ser Ser Pro Ser Ala Val Ser Arg Val Ser Ser Ala Val
            260                 265                 270

Ser Ser Leu Val Ser Asn Gly Gln Val Asn Met Ala Ala Leu Pro Asn
        275                 280                 285

Ile Ile Ser Asn Ile Ser Ser Ser Val Ser Ala Ser Ala Pro Gly Ala
    290                 295                 300

Ser Gly Cys Glu Val Ile Val Gln Ala Leu Leu Glu Val Ile Thr Ala
305                 310                 315                 320

Leu Val Gln Ile Val Ser Ser Ser Ser Val Gly Tyr Ile Asn Pro Ser
                325                 330                 335
```

Ala Val Asn Gln Ile Thr Asn Val Val Ala Asn Ala Met Ala Gln Val
340 345 350

Met Gly

<210> 64
<211> 348
<212> PRT
<213> Artificial Sequence

<220>
<223> Fusion protein

<400> 64

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

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

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

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

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

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

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

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

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

Gly Ala Ser Ala Ala Ala Ser Ala Gly Ala Ser Gly Asn Ser Gly Ile
145 150 155 160

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

Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro Ser Ala
180 185 190

Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn Gly Gln
195 200 205

```
Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser Ser
    210                 215                 220

Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile Val Gln
225                 230                 235                 240

Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser Ser Ser
                245                 250                 255

Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile Thr Asn Val
            260                 265                 270

Val Ala Asn Ala Met Ala Gln Val Met Gly Leu Glu Ala Leu Phe Gln
        275                 280                 285

Gly Pro Val Asp Asn Lys Phe Asn Lys Glu Gln Gln Asn Ala Phe Tyr
    290                 295                 300

Glu Ile Leu His Leu Pro Asn Leu Asn Glu Glu Gln Arg Asn Ala Phe
305                 310                 315                 320

Ile Gln Ser Leu Lys Asp Asp Pro Ser Gln Ser Ala Asn Leu Leu Ala
                325                 330                 335

Glu Ala Lys Lys Leu Asn Asp Ala Gln Ala Pro Lys
            340                 345
```

<210> 65
<211> 469
<212> PRT
<213> Artificial Sequence

<220>
<223> Fusion protein

<400> 65

```
Met Gly Ser Ser Gly His His His His His His Met Val Asp Asn Lys
1               5                   10                  15

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

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

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

Asp Ala Gln Ala Pro Lys Leu Glu Ala Leu Phe Gln Gly Pro Ser His
65                  70                  75                  80
```

```
Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser
                85                  90                  95

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

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

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

Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly
145                 150                 155                 160

Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser
                165                 170                 175

Asn Ala Phe Leu Gln Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn
            180                 185                 190

Glu Ile Thr Gln Leu Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp
        195                 200                 205

Gly Gly Gly Thr Pro Trp Thr Asn Pro Gly Leu Ala Glu Asn Phe Met
    210                 215                 220

Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly Phe Thr Ala Ser
225                 230                 235                 240

Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met Val Gln Ser Ile
                245                 250                 255

Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn Lys Leu Gln Ala
            260                 265                 270

Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala Ala Ser Glu
        275                 280                 285

Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser Ile Ala Ser Ala
    290                 295                 300

Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val Asn Gln Pro Phe
305                 310                 315                 320

Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala Gln Ala Gly Met
                325                 330                 335

Asn Asp Val Ser Gly Ser Gly Asn Ser Gly Ile Gln Gly Tyr Gly Gln
            340                 345                 350
```

Ser Ser Ala Ser Ala Ser Ala Ala Ala Ser Ala Ala Ser Thr Val Ala
355 360 365

Asn Ser Val Ser Arg Leu Ser Ser Pro Ser Ala Val Ser Arg Val Ser
370 375 380

Ser Ala Val Ser Ser Leu Val Ser Asn Gly Gln Val Asn Met Ala Ala
385 390 395 400

Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser Ser Val Ser Ala Ser Ala
405 410 415

Pro Gly Ala Ser Gly Cys Glu Val Ile Val Gln Ala Leu Leu Glu Val
420 425 430

Ile Thr Ala Leu Val Gln Ile Val Ser Ser Ser Ser Val Gly Tyr Ile
435 440 445

Asn Pro Ser Ala Val Asn Gln Ile Thr Asn Val Val Ala Asn Ala Met
450 455 460

Ala Gln Val Met Gly
465

<210> 66
<211> 483
<212> PRT
<213> Artificial Sequence

<220>
<223> Fusion protein

<400> 66

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 Gly Asn Ser Gly Ile Gln
    290                 295                 300

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

Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro Ser Ala Val
                325                 330                 335

Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn Gly Gln Val
            340                 345                 350

Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser Ser Val
        355                 360                 365

Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile Val Gln Ala
    370                 375                 380
```

Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser Ser Ser Ser
385 390 395 400

Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile Thr Asn Val Val
405 410 415

Ala Asn Ala Met Ala Gln Val Met Gly Val Asp Asn Lys Phe Asn Lys
420 425 430

Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu His Leu Pro Asn Leu Asn
435 440 445

Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser Leu Lys Asp Asp Pro Ser
450 455 460

Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys Lys Leu Asn Asp Ala Gln
465 470 475 480

Ala Pro Lys

<210> 67
<211> 963
<212> DNA
<213> Artificial Sequence

<220>
<223> Fusion protein

<220>
<221> misc_feature
<222> (37)..(174)
<223> ABD domain

<400> 67
atgggcagca gcggccatca tcatcatcat catatgttag ctgaagctaa agtcttagct 60

aacagagaac ttgacaaata tggagtaagt gactattaca agaacctaat caacaatgcc 120

aaaactgttg aaggtgtaaa agcactgata gatgaaattt tagctgcatt acctgggaat 180

tcaagccata ccaccccgtg gaccaacccg ggcctggcgg aaaactttat gaacagcttt 240

atgcagggcc tgagcagcat gccgggcttt accgcgagcc agctggatga tatgagcacc 300

attgcgcaga gcatggtgca gagcattcag agcctggcgg cgcagggccg caccagcccg 360

aacaaactgc aggcgctgaa catggcgttt gcgagcagca tggcggaaat tgcggcgagc 420

gaagaaggcg gcggcagcct gagcaccaaa accagcagca ttgcgagcgc gatgagcaac 480

gcgtttctgc agaccaccgg cgtggtgaac cagccgttta ttaacgaaat tacccagctg 540

gtgagcatgt ttgcgcaggc gggcatgaac gatgtgagcg cgggcaattc agggatccaa 600

ggttatggtc agagtagtgc ttctgcttca gctgctgcgt cagctgctag tactgtagct 660

aattcggtga gtcgcctctc atcgccttcc gcagtatctc gagtttcttc agcagtttct 720

agcttggttt caaatggtca agtgaatatg gcagcgttac ctaatatcat ttccaacatt 780

tcttcttctg tcagtgcatc tgctcctggt gcttctggat gtgaggtcat agtgcaagct 840

ctactcgaag tcatcactgc tcttgttcaa atcgttagtt cttctagtgt tggatatatt 900

aatccatctg ctgtgaacca aattactaat gttgttgcta atgccatggc tcaagtaatg 960

ggc 963

<210> 68
<211> 321
<212> PRT
<213> Artificial Sequence

<220>
<223> Fusion protein

<220>
<221> MISC_FEATURE
<222> (13)..(58)
<223> ABD domain

<400> 68

Met Gly Ser Ser Gly His His His His His His Met Leu Ala Glu Ala
1 5 10 15

Lys Val Leu Ala Asn Arg Glu Leu Asp Lys Tyr Gly Val Ser Asp Tyr
20 25 30

Tyr Lys Asn Leu Ile Asn Asn Ala Lys Thr Val Glu Gly Val Lys Ala
35 40 45

Leu Ile Asp Glu Ile Leu Ala Ala Leu Pro Gly Asn Ser Ser His Thr
50 55 60

Thr Pro Trp Thr Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe
65 70 75 80

Met Gln Gly Leu Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp
85 90 95

Asp Met Ser Thr Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu
100 105 110

Ala Ala Gln Gly Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met
115 120 125

Ala Phe Ala Ser Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly
130 135 140

Gly Ser Leu Ser Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn
145 150 155 160

```
Ala Phe Leu Gln Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu
                165                 170                 175

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

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

Ala Ser Ala Ala Ala Ser Ala Ala Ser Thr Val Ala Asn Ser Val Ser
    210                 215                 220

Arg Leu Ser Ser Pro Ser Ala Val Ser Arg Val Ser Ser Ala Val Ser
225                 230                 235                 240

Ser Leu Val Ser Asn Gly Gln Val Asn Met Ala Ala Leu Pro Asn Ile
                245                 250                 255

Ile Ser Asn Ile Ser Ser Ser Val Ser Ala Ser Ala Pro Gly Ala Ser
            260                 265                 270

Gly Cys Glu Val Ile Val Gln Ala Leu Leu Glu Val Ile Thr Ala Leu
        275                 280                 285

Val Gln Ile Val Ser Ser Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala
    290                 295                 300

Val Asn Gln Ile Thr Asn Val Val Ala Asn Ala Met Ala Gln Val Met
305                 310                 315                 320

Gly
```

<210> 69
<211> 546
<212> DNA
<213> Artificial Sequence

<220>
<223> Fusion protein

<220>
<221> misc_feature
<222> (28)..(165)
<223> ABD domain

<400> 69

```
atgggccatc atcatcatca tcatatgtta gctgaagcta aagtcttagc taacagagaa       60

cttgacaaat atggagtaag tgactattac aagaacctaa tcaacaatgc caaaactgtt      120

gaaggtgtaa aagcactgat agatgaaatt ttagctgcat tacctgggaa ttcagggatc      180

caaggttatg gtcagagtag tgcttctgct tcagctgctg cgtcagctgc tagtactgta      240

gctaattcgg tgagtcgcct ctcatcgcct tccgcagtat ctcgagtttc ttcagcagtt      300
```

tctagcttgg tttcaaatgg tcaagtgaat atggcagcgt tacctaatat catttccaac 360

atttcttctt ctgtcagtgc atctgctcct ggtgcttctg gatgtgaggt catagtgcaa 420

gctctactcg aagtcatcac tgctcttgtt caaatcgtta gttcttctag tgttggatat 480

attaatccat ctgctgtgaa ccaaattact aatgttgttg ctaatgccat ggctcaagta 540

atgggc 546

<210> 70
<211> 182
<212> PRT
<213> Artificial Sequence

<220>
<223> Fusion protein

<220>
<221> MISC_FEATURE
<222> (10)..(55)
<223> ABD domain

<400> 70

Met Gly His His His His His His Met Leu Ala Glu Ala Lys Val Leu
1 5 10 15

Ala Asn Arg Glu Leu Asp Lys Tyr Gly Val Ser Asp Tyr Tyr Lys Asn
20 25 30

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

Glu Ile Leu Ala Ala Leu Pro Gly Asn Ser Gly Ile Gln Gly Tyr Gly
50 55 60

Gln Ser Ser Ala Ser Ala Ser Ala Ala Ala Ser Ala Ala Ser Thr Val
65 70 75 80

Ala Asn Ser Val Ser Arg Leu Ser Ser Pro Ser Ala Val Ser Arg Val
85 90 95

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

Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser Ser Val Ser Ala Ser
115 120 125

Ala Pro Gly Ala Ser Gly Cys Glu Val Ile Val Gln Ala Leu Leu Glu
130 135 140

Val Ile Thr Ala Leu Val Gln Ile Val Ser Ser Ser Ser Val Gly Tyr
145 150 155 160

Ile Asn Pro Ser Ala Val Asn Gln Ile Thr Asn Val Val Ala Asn Ala
165 170 175

Met Ala Gln Val Met Gly
180

<210> 71
<211> 1302
<212> DNA
<213> Artificial Sequence

<220>
<223> Fusion protein

<220>
<221> misc_feature
<222> (37)..(513)
<223> M4 domain

<400> 71
atgggcagca gcggccatca tcatcatcat catatggatc cgagcaaaga tagcaaagcg 60

caggtgagcg cggcggaagc gggcattacc ggcacctggt ataaccagct gggcagcacc 120

tttattgtga ccgcgggcgc ggatggcgcg ctgaccggca cctatgaaag cgcggtgggc 180

aacgcggaaa gccgctatac cctgaccggc cgctatgata gcgcgccggc gaccgatggc 240

agcggcaccg cgctgggctg gcgcgtggcg tggaaaaaca actatcgcaa cgcgcatagc 300

gcgaccacct ggagcggcca gtatgtgggc ggcgcggaag cgcgcattaa cacccagtgg 360

accctgacca gcggcaccac cgaagcgaac gcgtggaaaa gcaccctgcg cggccatgat 420

acctttacca aagtgaaacc gagcgcggcg agcattgatg cggcgaaaaa agcgggcgtg 480

aacaacggca acccgctgga tgcggtgcag caggggaatt caagccatac caccccgtgg 540

accaacccgg gcctggcgga aaactttatg aacagcttta tgcagggcct gagcagcatg 600

ccgggcttta ccgcgagcca gctggatgat atgagcacca ttgcgcagag catggtgcag 660

agcattcaga gcctggcggc gcagggccgc accagcccga acaaactgca ggcgctgaac 720

atggcgtttg cgagcagcat ggcggaaatt gcggcgagcg aagaaggcgg cggcagcctg 780

agcaccaaaa ccagcagcat tgcgagcgcg atgagcaacg cgtttctgca gaccaccggc 840

gtggtgaacc agccgtttat taacgaaatt acccagctgg tgagcatgtt tgcgcaggcg 900

ggcatgaacg atgtgagcgc gggcaattca gggatccaag gttatggtca gagtagtgct 960

tctgcttcag ctgctgcgtc agctgctagt actgtagcta attcggtgag tcgcctctca 1020

tcgccttccg cagtatctcg agtttcttca gcagtttcta gcttggtttc aaatggtcaa 1080

gtgaatatgg cagcgttacc taatatcatt tccaacattt cttcttctgt cagtgcatct 1140

gctcctggtg cttctggatg tgaggtcata gtgcaagctc tactcgaagt catcactgct 1200

cttgttcaaa tcgttagttc ttctagtgtt ggatatatta atccatctgc tgtgaaccaa 1260

attactaatg ttgttgctaa tgccatggct caagtaatgg gc 1302

<210> 72
<211> 434
<212> PRT
<213> Artificial Sequence

<220>
<223> Fusion protein

<220>
<221> MISC_FEATURE
<222> (13)..(171)
<223> M4 domain

<400> 72

```
Met Gly Ser Ser Gly His His His His His His Met Asp Pro Ser Lys
1               5                   10                  15

Asp Ser Lys Ala Gln Val Ser Ala Ala Glu Ala Gly Ile Thr Gly Thr
            20                  25                  30

Trp Tyr Asn Gln Leu Gly Ser Thr Phe Ile Val Thr Ala Gly Ala Asp
        35                  40                  45

Gly Ala Leu Thr Gly Thr Tyr Glu Ser Ala Val Gly Asn Ala Glu Ser
    50                  55                  60

Arg Tyr Thr Leu Thr Gly Arg Tyr Asp Ser Ala Pro Ala Thr Asp Gly
65                  70                  75                  80

Ser Gly Thr Ala Leu Gly Trp Arg Val Ala Trp Lys Asn Asn Tyr Arg
                85                  90                  95

Asn Ala His Ser Ala Thr Thr Trp Ser Gly Gln Tyr Val Gly Gly Ala
            100                 105                 110

Glu Ala Arg Ile Asn Thr Gln Trp Thr Leu Thr Ser Gly Thr Thr Glu
        115                 120                 125

Ala Asn Ala Trp Lys Ser Thr Leu Arg Gly His Asp Thr Phe Thr Lys
    130                 135                 140

Val Lys Pro Ser Ala Ala Ser Ile Asp Ala Ala Lys Lys Ala Gly Val
145                 150                 155                 160

Asn Asn Gly Asn Pro Leu Asp Ala Val Gln Gln Gly Asn Ser Ser His
                165                 170                 175

Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala Glu Asn Phe Met Asn Ser
            180                 185                 190

Phe Met Gln Gly Leu Ser Ser Met Pro Gly Phe Thr Ala Ser Gln Leu
        195                 200                 205
```

```
Asp Asp Met Ser Thr Ile Ala Gln Ser Met Val Gln Ser Ile Gln Ser
    210                 215                 220

Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn
225                 230                 235                 240

Met Ala Phe Ala Ser Ser Met Ala Glu Ile Ala Ala Ser Glu Glu Gly
                245                 250                 255

Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser Ile Ala Ser Ala Met Ser
            260                 265                 270

Asn Ala Phe Leu Gln Thr Thr Gly Val Val Asn Gln Pro Phe Ile Asn
        275                 280                 285

Glu Ile Thr Gln Leu Val Ser Met Phe Ala Gln Ala Gly Met Asn Asp
    290                 295                 300

Val Ser Ala Gly Asn Ser Gly Ile Gln Gly Tyr Gly Gln Ser Ser Ala
305                 310                 315                 320

Ser Ala Ser Ala Ala Ala Ser Ala Ala Ser Thr Val Ala Asn Ser Val
                325                 330                 335

Ser Arg Leu Ser Ser Pro Ser Ala Val Ser Arg Val Ser Ser Ala Val
            340                 345                 350

Ser Ser Leu Val Ser Asn Gly Gln Val Asn Met Ala Ala Leu Pro Asn
        355                 360                 365

Ile Ile Ser Asn Ile Ser Ser Ser Val Ser Ala Ser Ala Pro Gly Ala
    370                 375                 380

Ser Gly Cys Glu Val Ile Val Gln Ala Leu Leu Glu Val Ile Thr Ala
385                 390                 395                 400

Leu Val Gln Ile Val Ser Ser Ser Ser Val Gly Tyr Ile Asn Pro Ser
                405                 410                 415

Ala Val Asn Gln Ile Thr Asn Val Val Ala Asn Ala Met Ala Gln Val
            420                 425                 430

Met Gly
```

<210> 73
<211> 885
<212> DNA
<213> Artificial Sequence

<220>
<223> Fusion protein

<220>
<221> misc_feature
<222> (28)..(504)
<223> M4 domain

<400> 73
atgggccatc atcatcatca tcatatggat ccgagcaaag atagcaaagc gcaggtgagc 60

gcggcggaag cgggcattac cggcacctgg tataaccagc tgggcagcac ctttattgtg 120

accgcgggcg cggatggcgc gctgaccggc acctatgaaa gcgcggtggg caacgcggaa 180

agccgctata ccctgaccgg ccgctatgat agcgcgccgg cgaccgatgg cagcggcacc 240

gcgctgggct ggcgcgtggc gtggaaaaac aactatcgca acgcgcatag cgcgaccacc 300

tggagcggcc agtatgtggg cggcgcggaa gcgcgcatta acacccagtg gaccctgacc 360

agcggcacca ccgaagcgaa cgcgtggaaa agcaccctgc gcggccatga tacctttacc 420

aaagtgaaac cgagcgcggc gagcattgat gcggcgaaaa aagcgggcgt gaacaacggc 480

aacccgctgg atgcggtgca gcaggggaat tcagggatcc aaggttatgg tcagagtagt 540

gcttctgctt cagctgctgc gtcagctgct agtactgtag ctaattcggt gagtcgcctc 600

tcatcgcctt ccgcagtatc tcgagtttct tcagcagttt ctagcttggt ttcaaatggt 660

caagtgaata tggcagcgtt acctaatatc atttccaaca tttcttcttc tgtcagtgca 720

tctgctcctg gtgcttctgg atgtgaggtc atagtgcaag ctctactcga agtcatcact 780

gctcttgttc aaatcgttag ttcttctagt gttggatata ttaatccatc tgctgtgaac 840

caaattacta atgttgttgc taatgccatg gctcaagtaa tgggc 885

<210> 74
<211> 295
<212> PRT
<213> Artificial Sequence

<220>
<223> Fusion protein

<220>
<221> MISC_FEATURE
<222> (10)..(168)
<223> M4 domain

<400> 74

Met Gly His His His His His His Met Asp Pro Ser Lys Asp Ser Lys
1 5 10 15

Ala Gln Val Ser Ala Ala Glu Ala Gly Ile Thr Gly Thr Trp Tyr Asn
20 25 30

Gln Leu Gly Ser Thr Phe Ile Val Thr Ala Gly Ala Asp Gly Ala Leu
35 40 45

Thr Gly Thr Tyr Glu Ser Ala Val Gly Asn Ala Glu Ser Arg Tyr Thr
50 55 60

```
Leu Thr Gly Arg Tyr Asp Ser Ala Pro Ala Thr Asp Gly Ser Gly Thr
65                  70                  75                  80

Ala Leu Gly Trp Arg Val Ala Trp Lys Asn Asn Tyr Arg Asn Ala His
                85                  90                  95

Ser Ala Thr Thr Trp Ser Gly Gln Tyr Val Gly Gly Ala Glu Ala Arg
            100                 105                 110

Ile Asn Thr Gln Trp Thr Leu Thr Ser Gly Thr Thr Glu Ala Asn Ala
        115                 120                 125

Trp Lys Ser Thr Leu Arg Gly His Asp Thr Phe Thr Lys Val Lys Pro
    130                 135                 140

Ser Ala Ala Ser Ile Asp Ala Ala Lys Lys Ala Gly Val Asn Asn Gly
145                 150                 155                 160

Asn Pro Leu Asp Ala Val Gln Gln Gly Asn Ser Gly Ile Gln Gly Tyr
                165                 170                 175

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

Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro Ser Ala Val Ser Arg
        195                 200                 205

Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn Gly Gln Val Asn Met
    210                 215                 220

Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser Ser Val Ser Ala
225                 230                 235                 240

Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile Val Gln Ala Leu Leu
                245                 250                 255

Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser Ser Ser Ser Val Gly
            260                 265                 270

Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile Thr Asn Val Val Ala Asn
        275                 280                 285

Ala Met Ala Gln Val Met Gly
    290                 295
```

<210> 75
<211> 1614
<212> DNA
<213> Artificial Sequence

<220>

<223> Fusion protein

<220>
<221> misc_feature
<222> (37)..(825)
<223> scFv1 domain

<400> 75
atgggcagca gcggccatca tcatcatcat catatggagg tgcagctgtt ggagtctggg 60

ggaggcttgg tacagcctgg ggggtccctg agactctcct gtgcagcctc tggattcacc 120

ttcagtagtt atgaaatgaa ctgggtccgc caggctccag ggaaggggct ggagtgggtc 180

tcaggcatta gtggtagtgg tggtttcaca tactacgcag actccgtgaa gggccgattc 240

accatctcca gagacaattc caagaacacg ctgtatctgc aaatgaacag cctgagagcc 300

gaggacactg ccatgtatta ctgtgcgaga gaggggtacc aggatgcttt tgatatctgg 360

ggccagggta cactggtcac cgtgagcagc ggtggaggcg gttcaggcgg aggtggatcc 420

ggcggtggcg gatcgcagtc tgtgctgact cagccaccct cagcgtctgg gacccccggg 480

cagagggtca ccatctcctg cactgggagc agctccaaca tcggggcagg ttatgatgta 540

cactggtatc agcagctccc aggaacggcc cccaaactcc tcatctatag taataatcag 600

cggccctcag gggtccctga ccgattctct ggctccaagt ctggcacctc agcctccctg 660

gccatcagtg ggctccggtc cgaggatgag gctgattatt actgtgcagc atgggatgac 720

agcctgagtg gtccgccttg ggtgttcggc ggaggaacca agctgacggt cctaggtgaa 780

caaaaactca tctcagaaga ggatctgtct ggatcagcgg ctgcagggaa ttcaagccat 840

accaccccgt ggaccaaccc gggcctggcg gaaaacttta tgaacagctt tatgcagggc 900

ctgagcagca tgccgggctt taccgcgagc cagctggatg atatgagcac cattgcgcag 960

agcatggtgc agagcattca gagcctggcg gcgcagggcc gcaccagccc gaacaaactg 1020

caggcgctga acatggcgtt tgcgagcagc atggcggaaa ttgcggcgag cgaagaaggc 1080

ggcggcagcc tgagcaccaa aaccagcagc attgcgagcg cgatgagcaa cgcgtttctg 1140

cagaccaccg gcgtggtgaa ccagccgttt attaacgaaa ttacccagct ggtgagcatg 1200

tttgcgcagg cgggcatgaa cgatgtgagc gcgggcaatt cagggatcca aggttatggt 1260

cagagtagtg cttctgcttc agctgctgcg tcagctgcta gtactgtagc taattcggtg 1320

agtcgcctct catcgccttc cgcagtatct cgagtttctt cagcagtttc tagcttggtt 1380

tcaaatggtc aagtgaatat ggcagcgtta cctaatatca tttccaacat ttcttcttct 1440

gtcagtgcat ctgctcctgg tgcttctgga tgtgaggtca tagtgcaagc tctactcgaa 1500

gtcatcactg ctcttgttca aatcgttagt tcttctagtg ttggatatat taatccatct 1560

gctgtgaacc aaattactaa tgttgttgct aatgccatgg ctcaagtaat gggc 1614

<210> 76
<211> 538
<212> PRT
<213> Artificial Sequence

<220>
<223> Fusion protein

<220>
<221> MISC_FEATURE
<222> (13)..(275)
<223> scFv1 domain

<400> 76

```
Met Gly Ser Ser Gly His His His His His His Met Glu Val Gln Leu
1               5                  10                  15

Leu Glu Ser Gly Gly Gly Leu Val Gln Pro Gly Gly Ser Leu Arg Leu
            20                  25                  30

Ser Cys Ala Ala Ser Gly Phe Thr Phe Ser Ser Tyr Glu Met Asn Trp
        35                  40                  45

Val Arg Gln Ala Pro Gly Lys Gly Leu Glu Trp Val Ser Gly Ile Ser
    50                  55                  60

Gly Ser Gly Gly Phe Thr Tyr Tyr Ala Asp Ser Val Lys Gly Arg Phe
65                  70                  75                  80

Thr Ile Ser Arg Asp Asn Ser Lys Asn Thr Leu Tyr Leu Gln Met Asn
                85                  90                  95

Ser Leu Arg Ala Glu Asp Thr Ala Met Tyr Tyr Cys Ala Arg Glu Gly
            100                 105                 110

Tyr Gln Asp Ala Phe Asp Ile Trp Gly Gln Gly Thr Leu Val Thr Val
        115                 120                 125

Ser Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly
    130                 135                 140

Ser Gln Ser Val Leu Thr Gln Pro Pro Ser Ala Ser Gly Thr Pro Gly
145                 150                 155                 160

Gln Arg Val Thr Ile Ser Cys Thr Gly Ser Ser Ser Asn Ile Gly Ala
                165                 170                 175

Gly Tyr Asp Val His Trp Tyr Gln Gln Leu Pro Gly Thr Ala Pro Lys
            180                 185                 190

Leu Leu Ile Tyr Ser Asn Asn Gln Arg Pro Ser Gly Val Pro Asp Arg
        195                 200                 205

Phe Ser Gly Ser Lys Ser Gly Thr Ser Ala Ser Leu Ala Ile Ser Gly
    210                 215                 220
```

```
Leu Arg Ser Glu Asp Glu Ala Asp Tyr Tyr Cys Ala Ala Trp Asp Asp
225                 230                 235                 240

Ser Leu Ser Gly Pro Pro Trp Val Phe Gly Gly Gly Thr Lys Leu Thr
                245                 250                 255

Val Leu Gly Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu Ser Gly Ser
            260                 265                 270

Ala Ala Ala Gly Asn Ser Ser His Thr Thr Pro Trp Thr Asn Pro Gly
        275                 280                 285

Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met
    290                 295                 300

Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln
305                 310                 315                 320

Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser
                325                 330                 335

Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala
            340                 345                 350

Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr
        355                 360                 365

Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly
    370                 375                 380

Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met
385                 390                 395                 400

Phe Ala Gln Ala Gly Met Asn Asp Val Ser Ala Gly Asn Ser Gly Ile
                405                 410                 415

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

Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro Ser Ala
        435                 440                 445

Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn Gly Gln
    450                 455                 460

Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser Ser
465                 470                 475                 480

Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile Val Gln
                485                 490                 495
```

Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser Ser Ser
500 505 510

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

Val Ala Asn Ala Met Ala Gln Val Met Gly
530 535

<210> 77
<211> 1197
<212> DNA
<213> Artificial Sequence

<220>
<223> Fusion protein

<220>
<221> misc_feature
<222> (28)..(816)
<223> scFv1 domain

<400> 77
atgggccatc atcatcatca tcatatggag gtgcagctgt tggagtctgg gggaggcttg 60

gtacagcctg gggggtccct gagactctcc tgtgcagcct ctggattcac cttcagtagt 120

tatgaaatga actgggtccg ccaggctcca gggaaggggc tggagtgggt ctcaggcatt 180

agtggtagtg gtggtttcac atactacgca gactccgtga agggccgatt caccatctcc 240

agagacaatt ccaagaacac gctgtatctg caaatgaaca gcctgagagc cgaggacact 300

gccatgtatt actgtgcgag agaggggtac caggatgctt ttgatatctg gggccagggt 360

acactggtca ccgtgagcag cggtggaggc ggttcaggcg gaggtggatc cggcggtggc 420

ggatcgcagt ctgtgctgac tcagccaccc tcagcgtctg ggacccccgg gcagagggtc 480

accatctcct gcactgggag cagctccaac atcggggcag gttatgatgt acactggtat 540

cagcagctcc caggaacggc ccccaaactc ctcatctata gtaataatca gcggccctca 600

ggggtccctg accgattctc tggctccaag tctggcacct cagcctccct ggccatcagt 660

gggctccggt ccgaggatga ggctgattat tactgtgcag catgggatga cagcctgagt 720

ggtccgcctt gggtgttcgg cggaggaacc aagctgacgg tcctaggtga acaaaaactc 780

atctcagaag aggatctgtc tggatcagcg gctgcaggga attcagggat ccaaggttat 840

ggtcagagta gtgcttctgc ttcagctgct gcgtcagctg ctagtactgt agctaattcg 900

gtgagtcgcc tctcatcgcc ttccgcagta tctcgagttt cttcagcagt ttctagcttg 960

gtttcaaatg gtcaagtgaa tatggcagcg ttacctaata tcatttccaa catttcttct 1020

tctgtcagtg catctgctcc tggtgcttct ggatgtgagg tcatagtgca agctctactc 1080

gaagtcatca ctgctcttgt tcaaatcgtt agttcttcta gtgttggata tattaatcca 1140

tctgctgtga accaaattac taatgttgtt gctaatgcca tggctcaagt aatgggc 1197

<210> 78
<211> 399
<212> PRT
<213> Artificial Sequence

<220>
<223> Fusion protein

<220>
<221> MISC_FEATURE
<222> (10)..(272)
<223> scFv1 domain

<400> 78

```
Met Gly His His His His His His Met Glu Val Gln Leu Leu Glu Ser
1               5                   10                  15

Gly Gly Gly Leu Val Gln Pro Gly Gly Ser Leu Arg Leu Ser Cys Ala
            20                  25                  30

Ala Ser Gly Phe Thr Phe Ser Ser Tyr Glu Met Asn Trp Val Arg Gln
        35                  40                  45

Ala Pro Gly Lys Gly Leu Glu Trp Val Ser Gly Ile Ser Gly Ser Gly
    50                  55                  60

Gly Phe Thr Tyr Tyr Ala Asp Ser Val Lys Gly Arg Phe Thr Ile Ser
65                  70                  75                  80

Arg Asp Asn Ser Lys Asn Thr Leu Tyr Leu Gln Met Asn Ser Leu Arg
                85                  90                  95

Ala Glu Asp Thr Ala Met Tyr Tyr Cys Ala Arg Glu Gly Tyr Gln Asp
            100                 105                 110

Ala Phe Asp Ile Trp Gly Gln Gly Thr Leu Val Thr Val Ser Ser Gly
        115                 120                 125

Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gln Ser
    130                 135                 140

Val Leu Thr Gln Pro Pro Ser Ala Ser Gly Thr Pro Gly Gln Arg Val
145                 150                 155                 160

Thr Ile Ser Cys Thr Gly Ser Ser Ser Asn Ile Gly Ala Gly Tyr Asp
                165                 170                 175

Val His Trp Tyr Gln Gln Leu Pro Gly Thr Ala Pro Lys Leu Leu Ile
            180                 185                 190

Tyr Ser Asn Asn Gln Arg Pro Ser Gly Val Pro Asp Arg Phe Ser Gly
        195                 200                 205
```

```
Ser Lys Ser Gly Thr Ser Ala Ser Leu Ala Ile Ser Gly Leu Arg Ser
    210                 215                 220

Glu Asp Glu Ala Asp Tyr Tyr Cys Ala Ala Trp Asp Asp Ser Leu Ser
225                 230                 235                 240

Gly Pro Pro Trp Val Phe Gly Gly Gly Thr Lys Leu Thr Val Leu Gly
                245                 250                 255

Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu Ser Gly Ser Ala Ala Ala
            260                 265                 270

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

Ala Ala Ala Ser Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu
    290                 295                 300

Ser Ser Pro Ser Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu
305                 310                 315                 320

Val Ser Asn Gly Gln Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser
                325                 330                 335

Asn Ile Ser Ser Ser Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys
            340                 345                 350

Glu Val Ile Val Gln Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln
        355                 360                 365

Ile Val Ser Ser Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn
    370                 375                 380

Gln Ile Thr Asn Val Val Ala Asn Ala Met Ala Gln Val Met Gly
385                 390                 395


<210>  79
<211>  1380
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  Fusion protein


<220>
<221>  misc_feature
<222>  (37)..(591)
<223>  xylanase domain

<400>  79
atgggcagca gcggccatca tcatcatcat catatggcta gcacagacta ctggcaaaat       60

tggactgatg ggggcggtat agtaaacgct gtcaatgggt ctggcgggaa ttacagtgtt      120

aattggtcta ataccggaaa ttttgttgtt ggtaaaggtt ggactacagg ttcgccattt      180
```

aggacgataa actataatgc cggagtttgg gcgccgaatg gcaatggata tttaacttta 240

tatggttgga cgagatcacc tctcatagaa tattatgtag tggattcatg gggtacttat 300

agacctactg gaacgtataa aggtactgta aaaagtgatg ggggtacgta tgacatatat 360

acaactacac gttataacgc accttccatt gatggcgatc gcactacttt tacgcagtac 420

tggagtgttc gccagtcgaa gagaccaacc ggaagcaacg ctacaatcac tttcagcaat 480

catgtgaacg catggaagag ccatggaatg aatctgggca gtaattgggc ttaccaagtc 540

atggcgacag aaggatatca aagtagtgga agttctaacg taacagtgtg gccgaattca 600

agccatacca ccccgtggac caacccgggc ctggcggaaa actttatgaa cagctttatg 660

cagggcctga gcagcatgcc gggctttacc gcgagccagc tggatgatat gagcaccatt 720

gcgcagagca tggtgcagag cattcagagc ctggcggcgc agggccgcac cagcccgaac 780

aaactgcagg cgctgaacat ggcgtttgcg agcagcatgg cggaaattgc ggcgagcgaa 840

gaaggcggcg gcagcctgag caccaaaacc agcagcattg cgagcgcgat gagcaacgcg 900

tttctgcaga ccaccggcgt ggtgaaccag ccgtttatta acgaaattac ccagctggtg 960

agcatgtttg cgcaggcggg catgaacgat gtgagcgcgg gcaattcagg gatccaaggt 1020

tatggtcaga gtagtgcttc tgcttcagct gctgcgtcag ctgctagtac tgtagctaat 1080

tcggtgagtc gcctctcatc gccttccgca gtatctcgag tttcttcagc agtttctagc 1140

ttggtttcaa atggtcaagt gaatatggca gcgttaccta atatcatttc caacatttct 1200

tcttctgtca gtgcatctgc tcctggtgct tctggatgtg aggtcatagt gcaagctcta 1260

ctcgaagtca tcactgctct tgttcaaatc gttagttctt ctagtgttgg atatattaat 1320

ccatctgctg tgaaccaaat tactaatgtt gttgctaatg ccatggctca agtaatgggc 1380

<210> 80
<211> 460
<212> PRT
<213> Artificial Sequence

<220>
<223> Fusion protein

<220>
<221> MISC_FEATURE
<222> (13)..(197)
<223> xylanase domain

<400> 80

Met Gly Ser Ser Gly His His His His His His Met Ala Ser Thr Asp
1 5 10 15

Tyr Trp Gln Asn Trp Thr Asp Gly Gly Gly Ile Val Asn Ala Val Asn
20 25 30

Gly Ser Gly Gly Asn Tyr Ser Val Asn Trp Ser Asn Thr Gly Asn Phe
35 40 45

```
Val Val Gly Lys Gly Trp Thr Thr Gly Ser Pro Phe Arg Thr Ile Asn
    50                  55                  60

Tyr Asn Ala Gly Val Trp Ala Pro Asn Gly Asn Gly Tyr Leu Thr Leu
65                  70                  75                  80

Tyr Gly Trp Thr Arg Ser Pro Leu Ile Glu Tyr Tyr Val Val Asp Ser
                85                  90                  95

Trp Gly Thr Tyr Arg Pro Thr Gly Thr Tyr Lys Gly Thr Val Lys Ser
            100                 105                 110

Asp Gly Gly Thr Tyr Asp Ile Tyr Thr Thr Thr Arg Tyr Asn Ala Pro
        115                 120                 125

Ser Ile Asp Gly Asp Arg Thr Thr Phe Thr Gln Tyr Trp Ser Val Arg
    130                 135                 140

Gln Ser Lys Arg Pro Thr Gly Ser Asn Ala Thr Ile Thr Phe Ser Asn
145                 150                 155                 160

His Val Asn Ala Trp Lys Ser His Gly Met Asn Leu Gly Ser Asn Trp
                165                 170                 175

Ala Tyr Gln Val Met Ala Thr Glu Gly Tyr Gln Ser Ser Gly Ser Ser
            180                 185                 190

Asn Val Thr Val Trp Pro Asn Ser Ser His Thr Thr Pro Trp Thr Asn
        195                 200                 205

Pro Gly Leu Ala Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser
    210                 215                 220

Ser Met Pro Gly Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile
225                 230                 235                 240

Ala Gln Ser Met Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg
                245                 250                 255

Thr Ser Pro Asn Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser
            260                 265                 270

Met Ala Glu Ile Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr
        275                 280                 285

Lys Thr Ser Ser Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr
    290                 295                 300

Thr Gly Val Val Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val
305                 310                 315                 320
```

```
Ser Met Phe Ala Gln Ala Gly Met Asn Asp Val Ser Ala Gly Asn Ser
                325                 330                 335

Gly Ile Gln Gly Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala Ala
            340                 345                 350

Ser Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro
        355                 360                 365

Ser Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn
    370                 375                 380

Gly Gln Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser
385                 390                 395                 400

Ser Ser Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile
                405                 410                 415

Val Gln Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser
            420                 425                 430

Ser Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile Thr
        435                 440                 445

Asn Val Val Ala Asn Ala Met Ala Gln Val Met Gly
    450                 455                 460


<210>  81
<211>  963
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  Fusion protein


<220>
<221>  misc_feature
<222>  (28)..(582)
<223>  xylanase domain

<400>  81
atgggccatc atcatcatca tcatatggct agcacagact actggcaaaa ttggactgat       60

gggggcggta tagtaaacgc tgtcaatggg tctggcggga attacagtgt taattggtct      120

aataccggaa attttgttgt tggtaaaggt tggactacag gttcgccatt taggacgata      180

aactataatg ccggagtttg ggcgccgaat ggcaatggat atttaacttt atatggttgg      240

acgagatcac ctctcataga atattatgta gtggattcat ggggtactta tagacctact      300

ggaacgtata aaggtactgt aaaaagtgat gggggtacgt atgacatata tacaactaca      360

cgttataacg caccttccat tgatggcgat cgcactactt ttacgcagta ctggagtgtt      420

cgccagtcga agagaccaac cggaagcaac gctacaatca ctttcagcaa tcatgtgaac      480
```

gcatggaaga gccatggaat gaatctgggc agtaattggg cttaccaagt catggcgaca 540

gaaggatatc aaagtagtgg aagttctaac gtaacagtgt ggccgaattc agggatccaa 600

ggttatggtc agagtagtgc ttctgcttca gctgctgcgt cagctgctag tactgtagct 660

aattcggtga gtcgcctctc atcgccttcc gcagtatctc gagtttcttc agcagtttct 720

agcttggttt caaatggtca agtgaatatg gcagcgttac ctaatatcat ttccaacatt 780

tcttcttctg tcagtgcatc tgctcctggt gcttctggat gtgaggtcat agtgcaagct 840

ctactcgaag tcatcactgc tcttgttcaa atcgttagtt cttctagtgt tggatatatt 900

aatccatctg ctgtgaacca aattactaat gttgttgcta atgccatggc tcaagtaatg 960

ggc 963

<210> 82
<211> 321
<212> PRT
<213> Artificial Sequence

<220>
<223> Fusion protein

<220>
<221> MISC_FEATURE
<222> (10)..(194)
<223> xylanase domain

<400> 82

Met Gly His His His His His His Met Ala Ser Thr Asp Tyr Trp Gln
1 5 10 15

Asn Trp Thr Asp Gly Gly Gly Ile Val Asn Ala Val Asn Gly Ser Gly
20 25 30

Gly Asn Tyr Ser Val Asn Trp Ser Asn Thr Gly Asn Phe Val Val Gly
35 40 45

Lys Gly Trp Thr Thr Gly Ser Pro Phe Arg Thr Ile Asn Tyr Asn Ala
50 55 60

Gly Val Trp Ala Pro Asn Gly Asn Gly Tyr Leu Thr Leu Tyr Gly Trp
65 70 75 80

Thr Arg Ser Pro Leu Ile Glu Tyr Tyr Val Val Asp Ser Trp Gly Thr
85 90 95

Tyr Arg Pro Thr Gly Thr Tyr Lys Gly Thr Val Lys Ser Asp Gly Gly
100 105 110

Thr Tyr Asp Ile Tyr Thr Thr Thr Arg Tyr Asn Ala Pro Ser Ile Asp
115 120 125

```
Gly Asp Arg Thr Thr Phe Thr Gln Tyr Trp Ser Val Arg Gln Ser Lys
    130                 135                 140

Arg Pro Thr Gly Ser Asn Ala Thr Ile Thr Phe Ser Asn His Val Asn
145                 150                 155                 160

Ala Trp Lys Ser His Gly Met Asn Leu Gly Ser Asn Trp Ala Tyr Gln
                165                 170                 175

Val Met Ala Thr Glu Gly Tyr Gln Ser Ser Gly Ser Ser Asn Val Thr
            180                 185                 190

Val Trp Pro Asn Ser Gly Ile Gln Gly Tyr Gly Gln Ser Ser Ala Ser
        195                 200                 205

Ala Ser Ala Ala Ala Ser Ala Ala Ser Thr Val Ala Asn Ser Val Ser
    210                 215                 220

Arg Leu Ser Ser Pro Ser Ala Val Ser Arg Val Ser Ser Ala Val Ser
225                 230                 235                 240

Ser Leu Val Ser Asn Gly Gln Val Asn Met Ala Ala Leu Pro Asn Ile
                245                 250                 255

Ile Ser Asn Ile Ser Ser Ser Val Ser Ala Ser Ala Pro Gly Ala Ser
            260                 265                 270

Gly Cys Glu Val Ile Val Gln Ala Leu Leu Glu Val Ile Thr Ala Leu
        275                 280                 285

Val Gln Ile Val Ser Ser Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala
    290                 295                 300

Val Asn Gln Ile Thr Asn Val Val Ala Asn Ala Met Ala Gln Val Met
305                 310                 315                 320

Gly
```

<210> 83
<211> 984
<212> DNA
<213> Artificial Sequence

<220>
<223> Fusion protein

<220>
<221> misc_feature
<222> (37)..(195)
<223> hEGF sequence

<400> 83
atgggcagca gcggccatca tcatcatcat catatgaact ccgactccga atgtccattg 60

tcccacgacg gttactgttt gcacgacggt gtttgtatgt acatcgaagc tttggacaag 120

tacgcttgta actgtgttgt tggttacatc ggtgaaagat gtcaatacag agacttcaag 180

tggtgggaat tgagaccgaa ttcaagccat accaccccgt ggaccaaccc gggcctggcg 240

gaaaacttta tgaacagctt tatgcagggc ctgagcagca tgccgggctt taccgcgagc 300

cagctggatg atatgagcac cattgcgcag agcatggtgc agagcattca gagcctggcg 360

gcgcagggcc gcaccagccc gaacaaactg caggcgctga acatggcgtt tgcgagcagc 420

atggcggaaa ttgcggcgag cgaagaaggc ggcggcagcc tgagcaccaa aaccagcagc 480

attgcgagcg cgatgagcaa cgcgtttctg cagaccaccg gcgtggtgaa ccagccgttt 540

attaacgaaa ttacccagct ggtgagcatg tttgcgcagg cgggcatgaa cgatgtgagc 600

gcgggcaatt cagggatcca aggttatggt cagagtagtg cttctgcttc agctgctgcg 660

tcagctgcta gtactgtagc taattcggtg agtcgcctct catcgccttc cgcagtatct 720

cgagtttctt cagcagtttc tagcttggtt tcaaatggtc aagtgaatat ggcagcgtta 780

cctaatatca tttccaacat ttcttcttct gtcagtgcat ctgctcctgg tgcttctgga 840

tgtgaggtca tagtgcaagc tctactcgaa gtcatcactg ctcttgttca aatcgttagt 900

tcttctagtg ttggatatat taatccatct gctgtgaacc aaattactaa tgttgttgct 960

aatgccatgg ctcaagtaat gggc 984

<210> 84
<211> 328
<212> PRT
<213> Artificial Sequence

<220>
<223> Fusion protein

<220>
<221> MISC_FEATURE
<222> (13)..(65)
<223> hEGF

<400> 84

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

Glu Cys Pro Leu Ser His Asp Gly Tyr Cys Leu His Asp Gly Val Cys
20 25 30

Met Tyr Ile Glu Ala Leu Asp Lys Tyr Ala Cys Asn Cys Val Val Gly
35 40 45

Tyr Ile Gly Glu Arg Cys Gln Tyr Arg Asp Phe Lys Trp Trp Glu Leu
50 55 60

Arg Pro Asn Ser Ser His Thr Thr Pro Trp Thr Asn Pro Gly Leu Ala
65 70 75 80

```
Glu Asn Phe Met Asn Ser Phe Met Gln Gly Leu Ser Ser Met Pro Gly
                85                  90                  95

Phe Thr Ala Ser Gln Leu Asp Asp Met Ser Thr Ile Ala Gln Ser Met
            100                 105                 110

Val Gln Ser Ile Gln Ser Leu Ala Ala Gln Gly Arg Thr Ser Pro Asn
        115                 120                 125

Lys Leu Gln Ala Leu Asn Met Ala Phe Ala Ser Ser Met Ala Glu Ile
    130                 135                 140

Ala Ala Ser Glu Glu Gly Gly Gly Ser Leu Ser Thr Lys Thr Ser Ser
145                 150                 155                 160

Ile Ala Ser Ala Met Ser Asn Ala Phe Leu Gln Thr Thr Gly Val Val
                165                 170                 175

Asn Gln Pro Phe Ile Asn Glu Ile Thr Gln Leu Val Ser Met Phe Ala
            180                 185                 190

Gln Ala Gly Met Asn Asp Val Ser Ala Gly Asn Ser Gly Ile Gln Gly
        195                 200                 205

Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala Ala Ser Ala Ala Ser
    210                 215                 220

Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser Pro Ser Ala Val Ser
225                 230                 235                 240

Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser Asn Gly Gln Val Asn
                245                 250                 255

Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile Ser Ser Ser Val Ser
            260                 265                 270

Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val Ile Val Gln Ala Leu
        275                 280                 285

Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val Ser Ser Ser Ser Val
    290                 295                 300

Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile Thr Asn Val Val Ala
305                 310                 315                 320

Asn Ala Met Ala Gln Val Met Gly
                325
```

<210> 85
<211> 567

<212> DNA
<213> Artificial Sequence

<220>
<223> Fusion protein

<220>
<221> misc_feature
<222> (28)..(186)
<223> hEGF

<400> 85
atgggccatc atcatcatca tcatatgaac tccgactccg aatgtccatt gtcccacgac 60

ggttactgtt tgcacgacgg tgtttgtatg tacatcgaag ctttggacaa gtacgcttgt 120

aactgtgttg ttggttacat cggtgaaaga tgtcaataca gagacttcaa gtggtgggaa 180

ttgagaccga attcagggat ccaaggttat ggtcagagta gtgcttctgc ttcagctgct 240

gcgtcagctg ctagtactgt agctaattcg gtgagtcgcc tctcatcgcc ttccgcagta 300

tctcgagttt cttcagcagt ttctagcttg gtttcaaatg gtcaagtgaa tatggcagcg 360

ttacctaata tcatttccaa catttcttct tctgtcagtg catctgctcc tggtgcttct 420

ggatgtgagg tcatagtgca agctctactc gaagtcatca ctgctcttgt tcaaatcgtt 480

agttcttcta gtgttggata tattaatcca tctgctgtga accaaattac taatgttgtt 540

gctaatgcca tggctcaagt aatgggc 567

<210> 86
<211> 189
<212> PRT
<213> Artificial Sequence

<220>
<223> Fusion protein

<220>
<221> MISC_FEATURE
<222> (10)..(62)
<223> hEGF

<400> 86

Met Gly His His His His His His Met Asn Ser Asp Ser Glu Cys Pro
1 5 10 15

Leu Ser His Asp Gly Tyr Cys Leu His Asp Gly Val Cys Met Tyr Ile
20 25 30

Glu Ala Leu Asp Lys Tyr Ala Cys Asn Cys Val Val Gly Tyr Ile Gly
35 40 45

Glu Arg Cys Gln Tyr Arg Asp Phe Lys Trp Trp Glu Leu Arg Pro Asn
50 55 60

Ser Gly Ile Gln Gly Tyr Gly Gln Ser Ser Ala Ser Ala Ser Ala Ala
65 70 75 80

```
Ala Ser Ala Ala Ser Thr Val Ala Asn Ser Val Ser Arg Leu Ser Ser
                85                  90                  95

Pro Ser Ala Val Ser Arg Val Ser Ser Ala Val Ser Ser Leu Val Ser
            100                 105                 110

Asn Gly Gln Val Asn Met Ala Ala Leu Pro Asn Ile Ile Ser Asn Ile
        115                 120                 125

Ser Ser Ser Val Ser Ala Ser Ala Pro Gly Ala Ser Gly Cys Glu Val
    130                 135                 140

Ile Val Gln Ala Leu Leu Glu Val Ile Thr Ala Leu Val Gln Ile Val
145                 150                 155                 160

Ser Ser Ser Ser Val Gly Tyr Ile Asn Pro Ser Ala Val Asn Gln Ile
                165                 170                 175

Thr Asn Val Val Ala Asn Ala Met Ala Gln Val Met Gly
            180                 185
```