

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

(19) World Intellectual Property Organization

International Bureau



(10) International Publication Number

WO 2013/071356 A1

(43) International Publication Date

23 May 2013 (23.05.2013)

(51) International Patent Classification:

C12N 15/63 (2006.01) A61K 47/42 (2006.01)  
C07K 14/435 (2006.01)

(74) Agent: FB RICE; Level 23, 200 Queen Street, Melbourne, Victoria 3000 (AU).

(21) International Application Number:

PCT/AU2012/001412

(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.

(22) International Filing Date:

15 November 2012 (15.11.2012)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/560,649 16 November 2011 (16.11.2011) US

(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 2013/071356 A1

(54) Title: COLLAGEN-LIKE SILK GENES

(57) Abstract: The present invention relates to silk proteins which can be used to produce silk with a collagen-like structure, as well as nucleic acids encoding such proteins. The present invention also relates to recombinant cells and/or organisms which synthesize silk proteins. Silk proteins of the invention can be used for a variety of purposes such as in the production of personal care products, plastics, textiles, and biomedical products.

## **COLLAGEN-LIKE SILK GENES**

### **FIELD OF THE INVENTION**

The present invention relates to silk proteins which can be used to produce silk  
5 with a collagen-like structure, as well as nucleic acids encoding such proteins. The present invention also relates to recombinant cells and/or organisms which synthesize silk proteins. Silk proteins of the invention can be used for a variety of purposes such as in the production of personal care products, plastics, textiles, and biomedical products.

10

### **BACKGROUND OF THE INVENTION**

'Silk' has become a single, all encompassing description for an extremely wide range of biological materials (Sutherland et al., 2010) that are an ancient product in evolution. Silks are produced by a wide range insects, for example by the larvae of  
15 insects where a cocoon is formed for protection during metamorphosis, by adult insects such as webspinners that spin silk with structures on their front legs to make a web-like pouch or gallery in which they live, in the Hymenoptera (which includes bees, wasps and ants) where silk is part of nest construction, and by large numbers of other arthropods, most notably the various arachnids such as spiders where orb-webs  
20 used for catching prey are common (Sutherland et al., 2010).

Silks are fibrous protein secretions that exhibit exceptional strength and toughness and as such have been the target of extensive study. Silks are produced by over 30,000 species of spiders and by many insects. Furthermore, in comparison with other arthropods, spiders produce more than one silk type, typically between 5 and 7,  
25 each with different properties for different purposes, with most involved with web construction. Of these the major ampullate silk, also known as the 'dragline silk' which is used as a lifeline and for the web's outer rim and spokes has attracted much attention since it can be as strong per unit weight as steel, but much tougher.

Despite the diversity of structures and distributions, silks have features in  
30 common, notably, being semicrystalline materials, that is materials with regions of ordered molecular structure (crystallites) within an amorphous matrix and also, all show typically similar protein compositions, often rich in alanine, serine, and/or glycine (Sutherland et al., 2010).

Overall, very few of these silks have been characterised, with most research  
35 concentrating on the cocoon silk of the domesticated silkworm, *Bombyx mori* and on the dragline silks of the orb-weaving spider *Nephila clavipes*, the European garden spider, *Araneus diadematus*, and the nursery web spider, *Euprosthenops australis*.

In the Lepidoptera and spider, the fibroin silk genes code for proteins that are generally large with prominent hydrophilic terminal domains at either end spanning an extensive region of alternating hydrophobic and hydrophilic blocks (Bini et al., 2004). Generally these proteins comprise different combinations of crystalline arrays 5 of  $\beta$ -pleated sheets loosely associated with  $\beta$ -sheets,  $\beta$ -spirals,  $\alpha$ -helices and amorphous regions (see Craig and Riek, 2002 for review).

The ready commercial availability of domesticated silk worm, *B. mori*, silk, has meant that there is no commercial driver to make a recombinant silkworm silk, especially considering the difficulties in making a recombinant product for a protein 10 that is so large and is built around a highly repetitive structure. This silk comprises 2 chains; a heavy chain of ~390kDa and a light chain of ~26kDa in a 1:1 ratio, with these 2 chains linked by a critical disulphide bond.

Very limited complete sequence data is available for spider silks. This is because of the difficulties in studying highly repetitive structures (Arcidiacono et al., 15 1998). The main examples include dragline silk from black widow spider (Ayoub et al., 2007) and flagelliform silk from *Nephila clavipes* (Hayashi and Lewis, 2000). Whereas cultivation has proved highly successful for cocoon silks, farming is not an option for spider silks. The main problem is that most spiders are very territorial, aggressive and are cannibals. This has meant, however, that these have been 20 considerable efforts to use recombinant technologies to produce spider silks. Native spider dragline silk is remarkably strong, although dragline silks from different species show different properties, examples exist of silks that are five times stronger by weight than steel, and/or three times tougher than Kevlar (Dupont) (Gosline et al., 1999, Volrath and Knight, 2001). All dragline silks have a high MW, 250-320 kDa 25 (Ayoub et al., 2007), which on its own provides difficulties for recombinant expression. Dragline silks are typically composed of two main proteins, the major ampulate spidroins.

Spidroins have highly repetitive structures; they are modular, and contain hundreds of tandem repeats of distinct consensus motifs. MaSp1 spidroins generally 30 comprise two motifs, polyalanine, and GlyGlyXaa, where Xaa is frequently Leu, Tyr, Gln or Ala. MaSp2 spidroins also contain polyalanine, as well as GlyProGlyXaaXaa repeats, where Xaa is frequently Gly, Gln or Tyr. The polyalanine or poly(glycylalanine) sequences form into tightly packed  $\beta$ -sheet crystallites.

More recently, a novel structural silk from the honey bee, *Apis mellifera*, has 35 attracted attention. Early X-ray evidence (Rudall, 1962) and different amino acid composition suggested that distinct class of silk molecule with an alpha helical structure was present in honeybee silk. Further analysis suggested that a four-stranded coiled-coil was present (Atkins, 1967). Recent molecular studies have

confirmed this (Sutherland et al., 2006). Four silk genes have been identified (AmelFibroin 1-4), that each comprising a single exon with the genes separated by short regions of 1659-1796 bases are clustered sequentially in the *A. mellifera* genome. They do not contain the highly repetitive sequences of other silks. Hence the chains 5 are of a size and structure that can be more readily produced by recombinant methods. Honeybee silk may be useful as a new biomedical material. For example, it can be electrospun into sheets with uniform fibres of around 200 nm.

As silk fibres represent some of the strongest natural fibres known, they have been a subject to extensive research in attempts to reproduce their synthesis. 10 However, a recurrent problem with expression of Lepidopteran and spider fibroin genes has been low expression rates in various recombinant expression systems due to the combination of repeating nucleotide motifs that lead to deleterious recombination events, large gene size and the small number of codons for each amino acid which leads to depletion of tRNA pools. Recombinant expression leads to difficulties during 15 translation such as translational pauses as a result of codon preferences and codon demands and extensive recombination rates leading to truncation of the genes. Shorter, less repetitive sequences would avoid many of the problems associated with silk gene expression to date.

Silks have long found applications as biomedical materials, as they are 20 typically biocompatible, biodegradable and have low immunogenicity. For biomedical applications, recombinant silk can be fabricated into various formats. Included in these is fabrication of a natural fibre, as well as hydrogels formed using connectivity through either physical or chemical crosslinking. Variations in properties can also be produced to match specific clinical needs, so that silks that can 25 be produced for application that need high stress prior to failure, or where extensibility is required, such as in blood vessels, and where an appropriate modulus is required to modify or control cell response, for example for tissue engineering (Vepari and Kaplan, 2007). For example, this silk has been used to form non-woven mats (Dal Pra et al., 2004), and has been electrospun into fibres and fibre mats with 30 fibre sizes from nanometres to microns (Jin et al., 2002). Silk fibroin films can be cast from aqueous and non-aqueous solvents (Minoura et al., 1990). Porous sponges can be made from silk solutions, for example, by using salt or sugar as porogens with fibroin in HFIP (Nazarov et al., 2004) or in fully aqueous system (Kim et al., 2005).

In the early 1960s, the silk of some sawflies (Hymenopteran) was suggested to 35 have a collagen structure by X-ray diffraction patterns obtained cocoons or from silk fibres drawn from the salivary gland of *Nematus ribesii* (Rudall 1967 and 1968; Lucas and Rudall, 1968). The X-ray diffraction patterns suggest twisted cables of collagen molecules with dimensions of 30A diameter – suggesting a two or three stranded

cable (Rudall, 1967 and 1968). Proteins with X-ray diffraction patterns characteristic of collagen were also found in the silk gland of other species within the tribe Nematini: in *Hemicroa*, *Pristiphora*, *Pachynemalus*, *Pikonema* and *Nematus* species (subfamily Nematinae); in the anterior half of the silk gland of *Tomostethus* and 5 *Tethida* species (subfamily Blennocampinae) (Rudall and Kenchington, 1971).

After the contents of the silk gland of *Nematus ribesii* were dissolved in 0.2M borate buffer and then precipitated in ethanol, the following amino acid analysis (from 4 different preps) was obtained (Gly: 336 (std dev=16); Ala: 122 (2); Ser: 33 (12); Pro: 100 (3); Hydroxy-Lys: 37 (3); Lys: 14 (4): Rudall and Kenchington, 1971). This 10 analysis indicated that the silk had a high Gly content, characteristic of collagens, along with a high Ala content. However, these amino acids are also found in high abundance in cocoon and spider silks.

Considering the unique properties of silks produced by insects, and that they are available naturally in only minute amounts, there is a need for the identification of 15 further novel nucleic acids encoding silk proteins.

### **SUMMARY OF THE INVENTION**

The present inventors have identified numerous polynucleotides encoding silk proteins which are distinct from other silk proteins that have been characterized at the 20 primary amino acid sequence level.

Thus, in a first aspect the present invention provides an isolated and/or exogenous polynucleotide which encodes a collagen-like silk polypeptide.

In an embodiment, the polynucleotide comprises one or more of:

- i) a sequence of nucleotides as provided in any one of SEQ ID NO's 7 to 12, 22 25 or 23;
- ii) a sequence of nucleotides encoding a polypeptide comprising an amino acid sequence as provided in any one of SEQ ID NO's 1 to 6;
- iii) a sequence of nucleotides encoding a polypeptide comprising an amino acid sequence which is at least 30% identical to any one or more of SEQ ID NO's 1 to 30 6;
- iv) a sequence of nucleotides encoding a biologically active fragment of ii) or iii),
- v) a sequence of nucleotides which is at least 30% identical to any one or more of SEQ ID NO's 7 to 12, 22 or 23, or
- vi) a sequence which hybridizes to any one or more of i) to v) under stringent 35 conditions.

More preferably, the polynucleotide comprises a sequence of nucleotides encoding a polypeptide comprising an amino acid sequence which is at least 95% identical to any one or more of SEQ ID NO's 1 to 6.

5 In a particularly preferred embodiment, the polynucleotide encodes a polypeptide of the invention.

In another aspect, the present invention provides a vector comprising at least one polynucleotide of the invention.

10 In a preferred embodiment, the vector is an expression vector. More preferably, the polynucleotide is operably linked to a promoter in the expression vector.

In a further aspect, the present invention provides a host cell comprising at least one polynucleotide of the invention, and/or at least one vector of the invention.

15 The host cell can be any cell type. Examples include, but are not limited to, a bacterial, yeast, animal or plant cell. In a preferred embodiment, the cell is a bacterial cell.

In a further aspect, the present invention provides a substantially purified and/or recombinant collagen-like silk polypeptide.

In an embodiment, the polypeptide comprises one or more of:

- i) an amino acid sequence as provided in any one of SEQ ID NO's 1 to 6;
- 20 ii) an amino acid sequence which is at least 30% identical to any one or more of SEQ ID NO's 1 to 6; or
- iii) a biologically active fragment of i) or ii).

In an embodiment, the collagen-like silk polypeptide has, or is capable of forming under suitable conditions, a triple helical structure.

25 Preferably, the polypeptide can be purified from a species of Hymenoptera.

Preferably, the Genera of Hymenoptera is a species of *Hemichroa*, *Pristiphora*, *Pachynemalus*, *Pikonema*, *Tomostethus*, *Tethida* or *Nematus* (such as *Nematus ribesii* or *Nematus oligospilus*).

30 In a further embodiment, the polypeptide is fused to at least one other polypeptide. In a preferred embodiment, the at least one other polypeptide is selected from the group consisting of: a polypeptide that enhances the stability of a polypeptide of the present invention, a polypeptide that assists in the purification of the fusion protein, a polypeptide which promotes the formation of a triple helix, and a polypeptide which assists in the polypeptide of the invention being secreted from a 35 cell (for example secreted from a bacterial cell). Examples of such fusion proteins are provided as SEQ ID NO's 16 to 18.

In yet another aspect, the present invention provides a non-human transgenic organism comprising an exogenous polynucleotide of the invention, the polynucleotide encoding at least one polypeptide according of the invention.

5 In an embodiment, the transgenic organism is a plant or transgenic non-human animal.

Also provided is a process for preparing a polypeptide of the invention, the process comprising cultivating one or more of a host cell of the invention, a vector of the invention, a transgenic organism of the invention, under conditions which allow expression of the polynucleotide encoding the polypeptide, and recovering the 10 expressed polypeptide. As the skilled person would appreciate, the cultivation of a vector in an expression system is also known as cell-free expression.

Mammalian collagen to mimic the natural protein needs to be co-expressed with a prolyl-4-hydroxylase. Co-expression of prolyl-4-hydroxylase has been very problematic in bacteria, with the typical expression being achieved in yeast, especially 15 *Pichia*. However, *Pichia* needs methanol for induction, which in the presence of added oxygen to get proper hydroxylation means that fermentation requires a flame-proof system. Thus, in a preferred embodiment, a process for preparing a polypeptide of the invention does not comprise the expression/presence of a prolyl-4-hydroxylase.

20 In a further embodiment, the process further comprises producing a product from the polypeptides such as a personal care product, textiles, plastics, or a biomedical product.

In another aspect, the present invention provides an isolated and/or recombinant antibody which specifically binds a polypeptide of the invention.

25 In another aspect, the present invention provides a silk fibre, sponge, film, hydrogel or particle comprising at least one polypeptide of the invention.

Preferably, the polypeptide is a recombinant polypeptide.

In a further aspect, the present invention provides a copolymer comprising at least two polypeptides of the invention.

30 Preferably, the polypeptides are recombinant polypeptides.

In another aspect, the present invention provides a product comprising one or more of at least one polypeptide of the invention, at least one silk fibre, sponge, film, hydrogel or particle of the invention, or at least one copolymer of the invention.

Examples of products of the invention include, but are not limited to, a personal care product, textiles, plastics, and biomedical products.

35 In an embodiment, at least some of the polypeptides in the silk fibre, sponge, film, hydrogel, particle, copolymer, or product of the invention are crosslinked.

In a preferred embodiment, a product of the invention is not a product, such as a cocoon, produced by an insect such as a sawfly.

In a further aspect, the present invention provides a composition comprising one or more of at least one polypeptide of the invention, at least one silk fibre, sponge, film, hydrogel or particle of the invention, or at least one copolymer of the invention, and one or more acceptable carriers.

5 In an embodiment, the composition further comprises a drug.

In another embodiment, the composition is for use as a medicine, a medical device or a cosmetic.

In yet another aspect, the present invention provides a composition comprising at least one polynucleotide of the invention, and one or more acceptable carriers.

10 In a further aspect, the present invention provides a process for producing a product comprising collagen-like silk polypeptides, the process comprising;

i) obtaining collagen-like silk polypeptides, and

ii) processing the polypeptides to produce the product.

15 As the skilled person will appreciate, the nature of the processing step will depend upon the form of the final product and the processing step can take a large number of different forms well within the capability of the skilled reader. For example, a product comprising silk fibres may be produced through spinning the polypeptides to produce the fibres and then weaving the fibres to produce the product.

20 Typically, the collagen-like silk polypeptides will be obtained by performing a process for preparing a polypeptide of the invention, or from a third party who has performed this process.

25 In a further aspect, the present invention provides a method of treating or preventing a disease, the method comprising administering a composition comprising one or more of at least one drug for treating or preventing the disease and a pharmaceutically acceptable carrier, wherein the pharmaceutically acceptable carrier is selected from at least one polypeptide of the invention, at least one silk fibre, sponge, film, hydrogel or particle of the invention, at least one copolymer of the invention, at least one product of the invention, or at least one composition of the invention.

30 Also provided is the use of one or more of at least one polypeptide of the invention, at least one silk fibre, sponge, film, hydrogel or particle of the invention, at least one copolymer of the invention, at least one product of the invention, or at least one composition of the invention, and at least one drug, for the manufacture of a medicament for treating or preventing a disease.

35 Furthermore, provided is the use of one or more of at least one polypeptide of the invention, at least one silk fibre, sponge, film, hydrogel or particle of the invention, at least one copolymer of the invention, at least one product of the

invention, or at least one composition of the invention, and at least one drug, as a medicament for treating or preventing a disease.

In yet another aspect, the present invention provides a kit comprising one or more of at least one polypeptide of the invention, at least one polynucleotide of the invention, at least one vector of the invention, at least one silk fibre, sponge, film, hydrogel or particle of the invention, at least one copolymer of the invention, at least one product of the invention, or at least one composition of the invention.

Any embodiment herein shall be taken to apply *mutatis mutandis* to any other embodiment unless specifically stated otherwise.

10 The present invention is not to be limited in scope by the specific embodiments described herein, which are intended for the purpose of exemplification only. Functionally-equivalent products, compositions and methods are clearly within the scope of the invention, as described herein.

15 Throughout this specification, unless specifically stated otherwise or the context requires otherwise, reference to a single step, composition of matter, group of steps or group of compositions of matter shall be taken to encompass one and a plurality (i.e. one or more) of those steps, compositions of matter, groups of steps or group of compositions of matter.

20 The invention is hereinafter described by way of the following non-limiting Examples and with reference to the accompanying figures.

#### **BRIEF DESCRIPTION OF THE ACCOMPANYING DRAWINGS**

**Figure 1.** Alignment of the collagen-like domains using the ExPASy (ETH) program (SEQ ID NOs: 13 to 15).

25 **Figure 2.** Coomassie Blue stained SDS PAGE gels of sawfly silk protein expression. MW: molecular weight markers; A: expression of cDNA SF21; B: expression of SF9 cDNA; C: expression of V-SF30 cDNA.

30 **Figure 3.** Binding of SF9, SF21, SF30, bovine collagen and fibronectin to fibroblasts.

35 **Figure 4.** Biophysical evidence that recombinant sawfly silk proteins are collagen. Top (A) Full length and pepsin resistant (+P) fragments of each silk protein after SDS PAGE separation. Incomplete digestion was observed for the SfC B chain. Molecular weight standards (kDa) are indicated on the left. (B) circular dichroism spectra of purified recombinant sawfly cocoon collagens after pepsin treatment showing characteristic collagen maxima at about 220 nm, SfC A (---), SfC B (-·

• — • • ) and SfC C (— — —). The solid line ( ————— ) shows a representative collagen spectra.

#### **KEY TO THE SEQUENCE LISTING**

- 5 SEQ ID NO:1 - Sawfly collagen-like silk polypeptide type A sequence.  
SEQ ID NO:2 - Sawfly collagen-like silk polypeptide type A sequence without signal sequence.  
SEQ ID NO:3 - Sawfly collagen-like silk polypeptide type B sequence.  
SEQ ID NO:4 - Sawfly collagen-like silk polypeptide type B sequence without signal
- 10 sequence.  
SEQ ID NO:5 - Sawfly collagen-like silk polypeptide type C sequence.  
SEQ ID NO:6 - Sawfly collagen-like silk polypeptide type C sequence without signal sequence.  
SEQ ID NO:7 - Open reading frame encoding sawfly collagen-like silk polypeptide
- 15 type A sequence.  
SEQ ID NO:8 - Open reading frame encoding sawfly collagen-like silk polypeptide type A sequence without signal sequence.  
SEQ ID NO:9 - Open reading frame encoding sawfly collagen-like silk polypeptide type B sequence.
- 20 SEQ ID NO:10 - Open reading frame encoding sawfly collagen-like silk polypeptide type B sequence without signal sequence.  
SEQ ID NO:11 - Open reading frame encoding sawfly collagen-like silk polypeptide type C sequence.  
SEQ ID NO:12 - Open reading frame encoding sawfly collagen-like silk polypeptide
- 25 type C sequence without signal sequence.  
SEQ ID NO:13 - Collagen-like domain of type A silk protein (Figure 1).  
SEQ ID NO:14 - Collagen-like domain of type B silk protein (Figure 1).  
SEQ ID NO:15 - Collagen-like domain of type C silk protein (Figure 1).  
SEQ ID NO:16 - Type A collagen-like silk fusion protein (Example 2).
- 30 SEQ ID NO:17 - Type B collagen-like silk fusion protein (Example 2).  
SEQ ID NO:18 - Type C collagen-like silk fusion protein (Example 2).  
SEQ ID NO:19 - Open reading frame encoding SEQ ID NO:16.  
SEQ ID NO:20 - Open reading frame encoding SEQ ID NO:17.  
SEQ ID NO:21 - Open reading frame encoding SEQ ID NO:18.
- 35 SEQ ID NO:22 - Codon optimized open reading frame type A for bacterial expression (without signal).  
SEQ ID NO:23 - Codon optimized open reading frame type B for bacterial expression (without signal).

**DETAILED DESCRIPTION OF THE INVENTION****General Techniques and Definitions**

Unless specifically defined otherwise, all technical and scientific terms used 5 herein shall be taken to have the same meaning as commonly understood by one of ordinary skill in the art (e.g., in cell culture, molecular genetics, recombinant biology, silk technology, immunology, protein chemistry, and biochemistry).

Unless otherwise indicated, the recombinant protein, cell culture, and 10 immunological techniques utilized in the present invention are standard procedures, well known to those skilled in the art. Such techniques are described and explained throughout the literature in sources such as, J. Perbal, *A Practical Guide to Molecular Cloning*, John Wiley and Sons (1984), J. Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbour Laboratory Press (1989), T.A. Brown (editor), *Essential Molecular Biology: A Practical Approach*, Volumes 1 and 2, IRL 15 Press (1991), D.M. Glover and B.D. Hames (editors), *DNA Cloning: A Practical Approach*, Volumes 1-4, IRL Press (1995 and 1996), and F.M. Ausubel et al. (editors), *Current Protocols in Molecular Biology*, Greene Pub. Associates and Wiley-Interscience (1988, including all updates until present), Ed Harlow and David Lane (editors) *Antibodies: A Laboratory Manual*, Cold Spring Harbour Laboratory, (1988), 20 and J.E. Coligan et al. (editors) *Current Protocols in Immunology*, John Wiley & Sons (including all updates until present).

The term "and/or", e.g., "X and/or Y" shall be understood to mean either "X 25 and Y" or "X or Y" and shall be taken to provide explicit support for both meanings or for either meaning. Furthermore, a list or features including the phrase "and/or" between the second last and last feature means that any one or more the listed features may be present in any combination.

As used herein, the term about, unless stated to the contrary, refers to +/- 20%, more preferably +/- 10%, more preferably +/- 5%, more preferably +/- 1%, of the designated value.

30 As used herein, the terms "silk protein" and "silk polypeptide" refer to a fibrous protein/polypeptide that can be used to produce a silk fibre, and/or a fibrous protein complex.

Collagen is a group of naturally occurring proteins found in animals, especially 35 in the flesh and connective tissues of mammals. It is the main component of connective tissue, and is the most abundant protein in mammals, making up about 25% to 35% of the whole-body protein content. In mammals, including human, there are 29 defined types, each with a different structure (composition) and function. Collagen, in the form of elongated fibrils, is mostly found in fibrous tissues such as

tendon, ligament and skin. Collagen is a composed of a triple helix, which often exists as two identical chains ( $\alpha 1$ ) and an additional chain that differs slightly in its chemical composition ( $\alpha 2$ ). The amino acid composition of collagen is atypical for proteins, particularly with respect to its high hydroxyproline content. The most 5 common motifs in the amino acid sequence of collagen are Glycine-X-Hydroxyproline and Glycine-Proline-Y, where X and Y can be any amino acid, although preferences exist for certain combinations (Ramshaw et al., 1998). As used herein, the term "collagen-like" refers to a polypeptide comprising Gly-X-Y triplets, where X and Y can be any amino acid. A silk protein of the invention could also be 10 referred to as a "collagen silk" protein. In a particularly preferred embodiment, unlike collagen, a collagen-like silk protein of the invention does not have any hydroxyproline. Preferably, collagen-like silk proteins of the invention comprise at least about 40, more preferably at least about 50, Gly-X-Y triplets. Furthermore, preferably the Gly-X-Y triplets constitute at least about 40%, more preferably at least 15 about 50%, of the primary amino acid sequence of the proteins. In an embodiment, a collagen-like silk polypeptide of the invention has, or is capable of forming under suitable conditions, a triple helical structure. In an embodiment, a collagen-like silk protein of the invention is resistant to trypsin digestion.

As used herein, a "silk fibre" refers to filaments comprising proteins of the 20 invention which can be woven into various items such as textiles. This term excludes naturally occurring silk fibres such as cocoons of insects.

As used herein, a "copolymer" is a composition comprising two or more different silk proteins (for example type A and type B collagen-like silk proteins defined herein) of the invention. This term excludes naturally occurring copolymers 25 such as cocoons of insects.

As used herein, the term "at least one" when referring to, for example, a polypeptide in a silk fibre of the invention clearly does not mean the silk fibre comprises only a single polypeptide molecule, but means that a homogeneous population of the polypeptides are present (for example type A collagen-like silk 30 proteins) and other related molecules of the invention (such as type B and/or type C collagen-like silk proteins) are absent.

The term "plant" includes whole plants, vegetative structures (for example, leaves, stems, roots), floral organs/structures, seed (including embryo, endosperm, and seed coat), plant tissue (for example, vascular tissue, ground tissue, and the like), cells 35 and progeny of the same.

A "transgenic plant" refers to a plant that contains a gene construct ("transgene") not found in a wild-type plant of the same species, variety or cultivar. A "transgene" as referred to herein has the normal meaning in the art of biotechnology

and includes a genetic sequence which has been produced or altered by recombinant DNA or RNA technology and which has been introduced into the plant cell. The transgene may include genetic sequences derived from a plant cell. Typically, the transgene has been introduced into the plant by human manipulation such as, for 5 example, by transformation but any method can be used as one of skill in the art recognizes.

"Polynucleotide" refers to an oligonucleotide, nucleic acid molecule or any fragment thereof. It may be DNA or RNA of genomic or synthetic origin, double-stranded or single-stranded, and combined with carbohydrate, lipids, protein, or other 10 materials to perform a particular activity defined herein.

"Operably linked" as used herein refers to a functional relationship between two or more nucleic acid (e.g., DNA) segments. Typically, it refers to the functional relationship of transcriptional regulatory element to a transcribed sequence. For example, a promoter is operably linked to a coding sequence, such as a polynucleotide 15 defined herein, if it stimulates or modulates the transcription of the coding sequence in an appropriate host cell. Generally, promoter transcriptional regulatory elements that are operably linked to a transcribed sequence are physically contiguous to the transcribed sequence, i.e., they are *cis*-acting. However, some transcriptional regulatory elements, such as enhancers, need not be physically contiguous or located 20 in close proximity to the coding sequences whose transcription they enhance.

The term "signal peptide" refers to an amino terminal polypeptide preceding a secreted mature protein. The signal peptide is cleaved from and is therefore not present in the mature protein. Signal peptides have the function of directing and *trans*-locating secreted proteins across cell membranes. The signal peptide is also 25 referred to as signal sequence.

As used herein, "transformation" is the acquisition of new genes in a cell by the incorporation of a polynucleotide.

As used herein, the term "drug" refers to any compound that can be used to treat or prevent a particular disease, examples of drugs which can be formulated with 30 a silk protein of the invention include, but are not limited to, proteins, nucleic acids, anti-tumor agents, analgesics, antibiotics, anti-inflammatory compounds (both steroid and non-steroid), hormones, vaccines, labeled substances, and the like.

### Polypeptides

35 By "substantially purified polypeptide" or "purified polypeptide" we mean a polypeptide that has generally been separated from the lipids, nucleic acids, other polypeptides, and other contaminating molecules such as wax with which it is associated in its native state. With the exception of other proteins of the invention, it

is preferred that the substantially purified polypeptide is at least 60% free, more preferably at least 75% free, and more preferably at least 90% free from other components with which it is naturally associated. In one aspect, the present invention relates to a homogenous population of, or composition comprising, a single type of 5 collagen-like silk protein of the invention which does not exist in nature.

The term "recombinant" in the context of a polypeptide refers to the polypeptide when produced by a cell, or in a cell-free expression system, in an altered amount or at an altered rate compared to its native state. In one embodiment the cell is a cell that does not naturally produce the polypeptide. However, the cell may be a 10 cell which comprises a non-endogenous gene that causes an altered, preferably increased, amount of the polypeptide to be produced. A recombinant polypeptide of the invention includes polypeptides which have not been separated from other components of the transgenic (recombinant) cell, or cell-free expression system, in which it is produced, and polypeptides produced in such cells or cell-free systems 15 which are subsequently purified away from at least some other components.

The terms "polypeptide" and "protein" are generally used interchangeably and refer to a single polypeptide chain which may or may not be modified by addition of non-amino acid groups. In an embodiment, the terms "proteins" and "polypeptides" as used herein also include variants, mutants, modifications, analogous and/or 20 derivatives of the polypeptides of the invention as described herein.

The % identity of a polypeptide is determined by GAP (Needleman and Wunsch, 1970) analysis (GCG program) with a gap creation penalty=5, and a gap extension penalty=0.3. The query sequence is at least 50 amino acids in length, and the GAP analysis aligns the two sequences over a region of at least 50 amino acids. 25 More preferably, the query sequence is at least 100 amino acids in length and the GAP analysis aligns the two sequences over a region of at least 100 amino acids. Even more preferably, the query sequence is at least 250 amino acids in length and the GAP analysis aligns the two sequences over a region of at least 250 amino acids. Even more preferably, the GAP analysis aligns the two sequences over their entire 30 length.

As used herein a "biologically active" fragment is a portion of a polypeptide of the invention which maintains a defined activity of the full-length polypeptide, namely the ability to be used to produce silk. Biologically active fragments can be any size as long as they maintain the defined activity. Preferably, biologically active 35 fragments are, where relevant, at least 100, more preferably at least 200, and even more preferably at least 225 amino acids in length. In an embodiment, a "biologically active" fragment of the invention comprises a collagen-like domain of a polypeptide of the invention such as those provided in Figure 1.

With regard to a defined polypeptide, it will be appreciated that % identity figures higher than those provided above will encompass preferred embodiments. Thus, where applicable, in light of the minimum % identity figures, it is preferred that the polypeptide comprises an amino acid sequence which is at least 40%, more 5 preferably at least 45%, more preferably at least 50%, more preferably at least 55%, more preferably at least 60%, more preferably at least 65%, more preferably at least 70%, more preferably at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 91%, more preferably at least 92%, more preferably at least 93%, more preferably at least 94%, 10 more preferably at least 95%, more preferably at least 96%, more preferably at least 97%, more preferably at least 98%, more preferably at least 99%, more preferably at least 99.1%, more preferably at least 99.2%, more preferably at least 99.3%, more preferably at least 99.4%, more preferably at least 99.5%, more preferably at least 99.6%, more preferably at least 99.7%, more preferably at least 99.8%, and even more 15 preferably at least 99.9% identical to the relevant nominated SEQ ID NO.

Amino acid sequence mutants of the polypeptides of the present invention can be prepared by introducing appropriate nucleotide changes into a nucleic acid of the present invention, or by *in vitro* synthesis of the desired polypeptide. Such mutants include, for example, deletions, insertions or substitutions of residues within the 20 amino acid sequence. A combination of deletion, insertion and substitution can be made to arrive at the final construct, provided that the final polypeptide product possesses the desired characteristics.

Mutant (altered) polypeptides can be prepared using any technique known in the art. For example, a polynucleotide of the invention can be subjected to *in vitro* 25 mutagenesis. Such *in vitro* mutagenesis techniques include sub-cloning the polynucleotide into a suitable vector, transforming the vector into a "mutator" strain such as the *E. coli* XL-1 red (Stratagene) and propagating the transformed bacteria for a suitable number of generations. In another example, the polynucleotides of the invention are subjected to DNA shuffling techniques as broadly described by 30 Harayama (1998). These DNA shuffling techniques may include genes of the invention possibly in addition to genes related to those of the present invention, such as silk genes from Hymenopteran species other than the specific species characterized herein. Products derived from mutated/ altered DNA can readily be screened using techniques described herein to determine if they can be used as silk proteins.

In designing amino acid sequence mutants, the location of the mutation site 35 and the nature of the mutation will depend on characteristic(s) to be modified. The sites for mutation can be modified individually or in series, e.g., by (1) substituting first with conservative amino acid choices and then with more radical selections

depending upon the results achieved, (2) deleting the target residue, or (3) inserting other residues adjacent to the located site.

Amino acid sequence deletions generally range from about 1 to about 150 residues.

5 Substitution mutants have at least one amino acid residue in the polypeptide molecule removed and a different residue inserted in its place. The sites of greatest interest for substitutional mutagenesis include sites identified as important for function. Other sites of interest are those in which particular residues obtained from various strains or species are identical. These positions may be important for 10 biological activity. These sites, especially those falling within a sequence of at least three other identically conserved sites, are preferably substituted in a relatively conservative manner. Such conservative substitutions are shown in Table 1 under the heading of "exemplary substitutions".

15 **Table 1.** Exemplary substitutions

Original Residue	Exemplary Substitutions
Ala (A)	val; leu; ile; ser; thr
Arg (R)	lys
Asn (N)	gln; his
Asp (D)	glu
Cys (C)	ser; thr; ala; gly; val
Gln (Q)	asn; his
Glu (E)	asp
Gly (G)	pro; ala; ser; val; thr
His (H)	asn; gln
Ile (I)	leu; val; ala; met
Leu (L)	ile; val; met; ala; phe
Lys (K)	arg
Met (M)	leu; phe
Phe (F)	leu; val; ala
Pro (P)	ala; ser; thr
Ser (S)	thr; ala; val; gln
Thr (T)	ser; gln; ala
Trp (W)	tyr
Tyr (Y)	trp; phe
Val (V)	ile; leu; met; phe; ala; ser; thr

In a particularly preferred embodiment, unlike collagen, a collagen-like silk protein of the invention does not have any hydroxyproline.

Preferably, collagen-like silk proteins of the invention comprise at least about 40, more preferably at least about 50, Gly-X-Y triplets, where X and Y can be any 5 amino acid. In a further embodiment, collagen-like silk proteins of the invention comprise between about 40 and about 150, more preferably between about 50 and about 100, Gly-X-Y triplets.

In a further preferred embodiment, the Gly-X-Y triplets constitute at least about 40%, more preferably at least about 50%, of the primary amino acid sequence 10 of the proteins.

Further guidance regarding amino acid substitutions, particularly in the collagen-like domain, can be found in Persikov et al. (2005).

In an embodiment, it is preferred that a substitution in the Y position is not with an amino acid with a branched  $\beta$ -carbon.

15 In an embodiment, the polypeptides provided as SEQ ID NO's 2, 4 and 6 comprise an N-terminal methionine.

20 In an embodiment, a collagen-like silk polypeptide of the invention has, or is capable of forming under suitable conditions, a triple helical structure. Folding of the silk proteins into a triple helical structure may be enhanced by addition (for example production as a fusion protein) of a triple helix promoting sequence, for example the Scl2 V-domain from *S. pyogenes*. Other examples include, but are not limited to, foldon or trimeric coiled-coil sequences.

25 In another embodiment, a polypeptide (and/or fusion protein) of the invention comprises multimers of the collagen-like domain, or a fragment thereof capable of forming a triple helical structure (examples of such domains are provided in Figure 1). The multimer may comprise collagen-like domains, or fragments thereof, from numerous different polypeptides of the invention, or comprise identical repeats of the same domain or a fragment thereof. The number of repeats may be, for example, 4 to 15, or 5 to 12.

30 Furthermore, if desired, unnatural amino acids or chemical amino acid analogues can be introduced as a substitution or addition into the polypeptides of the present invention. Such amino acids include, but are not limited to, the D-isomers of the common amino acids, 2,4-diaminobutyric acid,  $\alpha$ -amino isobutyric acid, 4-aminobutyric acid, 2-aminobutyric acid, 6-aminohexanoic acid, 2-aminoisobutyric acid, 3-aminopropionic acid, ornithine, norleucine, norvaline, hydroxyproline, 35 sarcosine, citrulline, homocitrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine,  $\beta$ -alanine, fluoro-amino acids, designer amino

acids such as  $\beta$ -methylamino acids,  $\text{C}\alpha$ -methyl amino acids,  $\text{N}\alpha$ -methyl amino acids, and amino acid analogues in general.

Also included within the scope of the invention are polypeptides of the present invention which are differentially modified during or after synthesis, e.g., by 5 biotinylation, benzylation, glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, etc. These modifications may serve to increase the stability and/or bioactivity of the polypeptide of the invention.

Polypeptides of the present invention can be produced in a variety of ways, 10 including production and recovery of natural polypeptides, production and recovery of recombinant polypeptides, and chemical synthesis of the polypeptides. In one embodiment, an isolated polypeptide of the present invention is produced by culturing a cell capable of expressing the polypeptide under conditions effective to produce the polypeptide, and recovering the polypeptide. A preferred cell to culture is a 15 recombinant cell of the present invention. Effective culture conditions include, but are not limited to, effective media, bioreactor, temperature, pH and oxygen conditions that permit polypeptide production. An effective medium refers to any medium in which a cell is cultured to produce a polypeptide of the present invention. Such medium typically comprises an aqueous medium having assimilable carbon, nitrogen 20 and phosphate sources, and appropriate salts, minerals, metals and other nutrients, such as vitamins. Cells of the present invention can be cultured in conventional fermentation bioreactors, shake flasks, test tubes, microtiter dishes, and Petri plates. Culturing can be carried out at a temperature, pH and oxygen content appropriate for a recombinant cell. Such culturing conditions are within the expertise of one of 25 ordinary skill in the art.

#### Polynucleotides

By an "isolated polynucleotide", including DNA, RNA, or a combination of these, single or double stranded, in the sense or antisense orientation or a combination 30 of both, dsRNA or otherwise, we mean a polynucleotide which is at least partially separated from the polynucleotide sequences with which it is associated or linked in its native state. Preferably, the isolated polynucleotide is at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated. Furthermore, the term "polynucleotide" is used 35 interchangeably herein with the term "nucleic acid".

The term "exogenous" in the context of a polynucleotide refers to the polynucleotide when present in a cell, or in a cell-free expression system, in an altered amount compared to its native state. In one embodiment, the cell is a cell that does

not naturally comprise the polynucleotide. However, the cell may be a cell which comprises a non-endogenous polynucleotide resulting in an altered, preferably increased, amount of production of the encoded polypeptide. An exogenous polynucleotide of the invention includes polynucleotides which have not been 5 separated from other components of the transgenic (recombinant) cell, or cell-free expression system, in which it is present, and polynucleotides produced in such cells or cell-free systems which are subsequently purified away from at least some other components.

The % identity of a polynucleotide is determined by GAP (Needleman and 10 Wunsch, 1970) analysis (GCG program) with a gap creation penalty=5, and a gap extension penalty=0.3. Unless stated otherwise, the query sequence is at least 45 nucleotides in length, and the GAP analysis aligns the two sequences over a region of at least 45 nucleotides. Preferably, the query sequence is at least 150 nucleotides in length, and the GAP analysis aligns the two sequences over a region of at least 150 15 nucleotides. More preferably, the query sequence is at least 300 nucleotides in length and the GAP analysis aligns the two sequences over a region of at least 300 nucleotides. Even more preferably, the GAP analysis aligns the two sequences over their entire length.

With regard to the defined polynucleotides, it will be appreciated that % 20 identity figures higher than those provided above will encompass preferred embodiments. Thus, where applicable, in light of the minimum % identity figures, it is preferred that a polynucleotide of the invention comprises a sequence which is at least 40%, more preferably at least 45%, more preferably at least 50%, more preferably at least 55%, more preferably at least 60%, more preferably at least 65%, 25 more preferably at least 70%, more preferably at least 75%, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 91%, more preferably at least 92%, more preferably at least 93%, more preferably at least 94%, more preferably at least 95%, more preferably at least 96%, more preferably at least 97%, more preferably at least 98%, more preferably at least 30 99%, more preferably at least 99.1%, more preferably at least 99.2%, more preferably at least 99.3%, more preferably at least 99.4%, more preferably at least 99.5%, more preferably at least 99.6%, more preferably at least 99.7%, more preferably at least 99.8%, and even more preferably at least 99.9% identical to the relevant nominated SEQ ID NO.

35 Polynucleotides of the present invention may possess, when compared to naturally occurring molecules, one or more mutations which are deletions, insertions, or substitutions of nucleotide residues. Mutants can be either naturally occurring (that

is to say, isolated from a natural source) or synthetic (for example, by performing site-directed mutagenesis on the nucleic acid).

Oligonucleotides and/or polynucleotides of the invention hybridize to a silk gene of the present invention, or a region flanking said gene, under stringent conditions. The term "stringent hybridization conditions" and the like as used herein refers to parameters with which the art is familiar, including the variation of the hybridization temperature with length of an oligonucleotide. Nucleic acid hybridization parameters may be found in references which compile such methods, Sambrook, et al. (*supra*), and Ausubel, et al. (*supra*). For example, stringent hybridization conditions, as used herein, can refer to hybridization at 65°C in hybridization buffer (3.5×SSC, 0.02% Ficoll, 0.02% polyvinyl pyrrolidone, 0.02% Bovine Serum Albumin (BSA), 2.5 mM NaH<sub>2</sub>PO<sub>4</sub> (pH7), 0.5% SDS, 2 mM EDTA), followed by one or more washes in 0.2×SSC, 0.01% BSA at 50°C. Alternatively, the nucleic acid and/or oligonucleotides (which may also be referred to as "primers" or "probes") hybridize to the region of the an insect genome of interest, such as the genome of a *Nematus* sp., under conditions used in nucleic acid amplification techniques such as PCR.

Oligonucleotides of the present invention can be RNA, DNA, or derivatives of either. Although the terms polynucleotide and oligonucleotide have overlapping meaning, oligonucleotides are typically relatively short single stranded molecules. The minimum size of such oligonucleotides is the size required for the formation of a stable hybrid between an oligonucleotide and a complementary sequence on a target nucleic acid molecule. Preferably, the oligonucleotides are at least 15 nucleotides, more preferably at least 18 nucleotides, more preferably at least 19 nucleotides, more preferably at least 20 nucleotides, even more preferably at least 25 nucleotides in length.

Usually, monomers of a polynucleotide or oligonucleotide are linked by phosphodiester bonds or analogs thereof to form oligonucleotides ranging in size from relatively short monomeric units, e.g., 12-18, to several hundreds of monomeric units. Analogs of phosphodiester linkages include: phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phosphoranilidate, phosphoramidate.

The present invention includes oligonucleotides that can be used as, for example, probes to identify nucleic acid molecules, or primers to produce nucleic acid molecules. Oligonucleotides of the present invention used as a probe are typically conjugated with a detectable label such as a radioisotope, an enzyme, biotin, a fluorescent molecule or a chemiluminescent molecule.

Recombinant Vectors

One embodiment of the present invention includes a recombinant vector, which comprises at least one isolated polynucleotide molecule of the present invention, inserted into any vector capable of delivering the polynucleotide molecule into a host cell. Such a vector contains heterologous polynucleotide sequences, which are polynucleotide sequences that are not naturally found adjacent to polynucleotide molecules of the present invention and that preferably are derived from a species other than the species from which the polynucleotide molecule(s) are derived. The vector can be either RNA or DNA, either prokaryotic or eukaryotic, and typically is a transposon (such as described in US 5,792,294), a virus or a plasmid.

One type of recombinant vector comprises a polynucleotide molecule of the present invention operatively linked to an expression vector. The phrase operatively linked refers to insertion of a polynucleotide molecule into an expression vector in a manner such that the molecule is able to be expressed when transformed into a host cell. As used herein, an expression vector is a DNA or RNA vector that is capable of transforming a host cell and of effecting expression of a specified polynucleotide molecule. Preferably, the expression vector is also capable of replicating within the host cell. Expression vectors can be either prokaryotic or eukaryotic, and are typically viruses or plasmids. Expression vectors of the present invention include any vectors that function (i.e., direct gene expression) in recombinant cells of the present invention, including in bacterial, fungal, endoparasite, arthropod, animal, and plant cells. Particularly preferred expression vectors of the present invention can direct gene expression in plants cells. Vectors of the invention can also be used to produce the polypeptide in a cell-free expression system, such systems are well known in the art.

In particular, expression vectors of the present invention contain regulatory sequences such as transcription control sequences, translation control sequences, origins of replication, and other regulatory sequences that are compatible with the recombinant cell and that control the expression of polynucleotide molecules of the present invention. In particular, recombinant molecules of the present invention include transcription control sequences. Transcription control sequences are sequences which control the initiation, elongation, and termination of transcription. Particularly important transcription control sequences are those which control transcription initiation, such as promoter, enhancer, operator and repressor sequences. Suitable transcription control sequences include any transcription control sequence that can function in at least one of the recombinant cells of the present invention. A variety of such transcription control sequences are known to those skilled in the art. Preferred transcription control sequences include those which function in bacterial,

yeast, arthropod, plant or mammalian cells, such as, but not limited to, tac, lac, trp, trc, oxy-pro, omp/lpp, rrnB, bacteriophage lambda, bacteriophage T7, T7lac, bacteriophage T3, bacteriophage SP6, bacteriophage SP01, metallothionein, alpha-mating factor, *Pichia* alcohol oxidase, alphavirus subgenomic promoters (such as 5 Sindbis virus subgenomic promoters), antibiotic resistance gene, baculovirus, *Heliothis zea* insect virus, vaccinia virus, herpesvirus, raccoon poxvirus, other poxvirus, adenovirus, cytomegalovirus (such as intermediate early promoters), simian virus 40, retrovirus, actin, retroviral long terminal repeat, Rous sarcoma virus, heat shock, phosphate and nitrate transcription control sequences as well as other 10 sequences capable of controlling gene expression in prokaryotic or eukaryotic cells.

Examples of useful expression vectors include, but are not limited to, pColdI, pColdII, pColdIII and pColdIV (pCold Vector Information: [http://catalog.takara-bio.co.jp/en/product/basic\\_info.asp?unitid=U100005856](http://catalog.takara-bio.co.jp/en/product/basic_info.asp?unitid=U100005856)).

Particularly preferred transcription control sequences are promoters active in 15 directing transcription in plants, either constitutively or stage and/or tissue specific, depending on the use of the plant or parts thereof. These plant promoters include, but are not limited to, promoters showing constitutive expression, such as the 35S promoter of Cauliflower Mosaic Virus (CaMV), those for leaf-specific expression, such as the promoter of the ribulose bisphosphate carboxylase small subunit gene, 20 those for root-specific expression, such as the promoter from the glutamine synthase gene, those for seed-specific expression, such as the cruciferin A promoter from *Brassica napus*, those for tuber-specific expression, such as the class-I patatin promoter from potato or those for fruit-specific expression, such as the polygalacturonase (PG) promoter from tomato.

25 Recombinant molecules of the present invention may also (a) contain secretory signals (i.e., signal segment nucleic acid sequences) to enable an expressed polypeptide of the present invention to be secreted from the cell that produces the polypeptide and/or (b) contain fusion sequences which lead to the expression of nucleic acid molecules of the present invention as fusion proteins. Examples of 30 suitable signal segments include any signal segment capable of directing the secretion of a polypeptide of the present invention. Preferred signal segments include, but are not limited to, tissue plasminogen activator (t-PA), interferon, interleukin, growth hormone, viral envelope glycoprotein signal segments, *Nicotiana nectarin* signal peptide (US 5,939,288), tobacco extensin signal, the soy oleosin oil body binding 35 protein signal, *Arabidopsis thaliana* vacuolar basic chitinase signal peptide, as well as native signal sequences of a polypeptide of the invention. In addition, a nucleic acid molecule of the present invention can be joined to a fusion segment that directs the encoded polypeptide to the proteosome, such as an ubiquitin fusion segment.

Recombinant molecules may also include intervening and/or untranslated sequences surrounding and/or within the nucleic acid sequences of the present invention.

#### Host Cells

5 Another embodiment of the present invention includes a recombinant cell comprising a host cell transformed with one or more recombinant molecules of the present invention, or progeny cells thereof. Transformation of a polynucleotide molecule into a cell can be accomplished by any method by which a polynucleotide molecule can be inserted into the cell. Transformation techniques include, but are not  
10 limited to, transfection, electroporation, microinjection, lipofection, adsorption, and protoplast fusion. A recombinant cell may remain unicellular or may grow into a tissue, organ or a multicellular organism. Transformed polynucleotide molecules of the present invention can remain extrachromosomal or can integrate into one or more sites within a chromosome of the transformed (i.e., recombinant) cell in such a  
15 manner that their ability to be expressed is retained.

Suitable host cells to transform include any cell that can be transformed with a polynucleotide of the present invention. Host cells of the present invention either can be endogenously (i.e., naturally) capable of producing polypeptides of the present invention or can be capable of producing such polypeptides after being transformed  
20 with at least one polynucleotide molecule of the present invention. Host cells of the present invention can be any cell capable of producing at least one protein of the present invention, and include bacterial, fungal (including yeast), parasite, arthropod, animal and plant cells. Examples of host cells include *Salmonella*, *Escherichia*, *Bacillus*, *Listeria*, *Saccharomyces*, *Spodoptera*, *Mycobacteria*, *Trichoplusia*, BHK  
25 (baby hamster kidney) cells, MDCK cells, CRFK cells, CV-1 cells, COS (e.g., COS-7) cells, and Vero cells. Further examples of host cells are *E. coli*, including *E. coli* K-12 derivatives; *Salmonella typhi*; *Salmonella typhimurium*, including attenuated strains; *Spodoptera frugiperda*; *Trichoplusia ni*; and non-tumorigenic mouse myoblast G8 cells (e.g., ATCC CRL 1246). Additional appropriate mammalian cell hosts  
30 include other kidney cell lines, other fibroblast cell lines (e.g., human, murine or chicken embryo fibroblast cell lines), myeloma cell lines, Chinese hamster ovary cells, mouse NIH/3T3 cells, LMTK cells and/or HeLa cells. Particularly preferred host cells are plant cells such as those available from Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (German Collection of Microorganisms  
35 and Cell Cultures).

Recombinant DNA technologies can be used to improve expression of a transformed polynucleotide molecule by manipulating, for example, the number of copies of the polynucleotide molecule within a host cell, the efficiency with which

those polynucleotide molecules are transcribed, the efficiency with which the resultant transcripts are translated, and the efficiency of post-translational modifications. Recombinant techniques useful for increasing the expression of polynucleotide molecules of the present invention include, but are not limited to, operatively linking 5 polynucleotide molecules to high-copy number plasmids, integration of the polynucleotide molecule into one or more host cell chromosomes, addition of vector stability sequences to plasmids, substitutions or modifications of transcription control signals (e.g., promoters, operators, enhancers), substitutions or modifications of translational control signals (e.g., ribosome binding sites, Shine-Dalgarno sequences), 10 modification of polynucleotide molecules of the present invention to correspond to the codon usage of the host cell, and the deletion of sequences that destabilize transcripts.

#### Transgenic Plants

The term "plant" refers to whole plants, plant organs (e.g. leaves, stems, roots 15 etc), seeds, plant cells and the like. Plants contemplated for use in the practice of the present invention include both monocotyledons and dicotyledons. Target plants include, but are not limited to, the following: cereals (wheat, barley, rye, oats, rice, sorghum and related crops); beet (sugar beet and fodder beet); pomes, stone fruit and soft fruit (apples, pears, plums, peaches, almonds, cherries, strawberries, raspberries 20 and blackberries); leguminous plants (beans, lentils, peas, soybeans); oil plants (rape, mustard, poppy, olives, sunflowers, coconut, castor oil plants, cocoa beans, groundnuts); cucumber plants (marrows, cucumbers, melons); fibre plants (cotton, flax, hemp, jute); citrus fruit (oranges, lemons, grapefruit, mandarins); vegetables (spinach, lettuce, asparagus, cabbages, carrots, onions, tomatoes, potatoes, paprika); 25 lauraceae (avocados, cinnamon, camphor); or plants such as maize, tobacco, nuts, coffee, sugar cane, tea, vines, hops, turf, bananas and natural rubber plants, as well as ornamentals (flowers, shrubs, broad-leaved trees and evergreens, such as conifers).

Transgenic plants, as defined in the context of the present invention, include 30 plants (as well as parts and cells of said plants) and their progeny which have been genetically modified using recombinant techniques to cause production of at least one polypeptide of the present invention in the desired plant or plant organ. Transgenic plants can be produced using techniques known in the art, such as those generally described in A. Slater et al., *Plant Biotechnology - The Genetic Manipulation of Plants*, Oxford University Press (2003), and P. Christou and H. Klee, *Handbook of 35 Plant Biotechnology*, John Wiley and Sons (2004).

A polynucleotide of the present invention may be expressed constitutively in the transgenic plants during all stages of development. Depending on the use of the

plant or plant organs, the polypeptides may be expressed in a stage-specific manner. Furthermore, the polynucleotides may be expressed tissue-specifically.

Regulatory sequences which are known or are found to cause expression of a gene encoding a polypeptide of interest in plants may be used in the present invention.

5 The choice of the regulatory sequences used depends on the target plant and/or target organ of interest. Such regulatory sequences may be obtained from plants or plant viruses, or may be chemically synthesized. Such regulatory sequences are well known to those skilled in the art.

Constitutive plant promoters are well known. Further to previously mentioned 10 promoters, some other suitable promoters include but are not limited to the nopaline synthase promoter, the octopine synthase promoter, CaMV 35S promoter, the ribulose-1,5-bisphosphate carboxylase promoter, Adh1-based pEmu, Act1, the SAM synthase promoter and Ubi promoters and the promoter of the chlorophyll a/b binding protein. Alternatively it may be desired to have the transgene(s) expressed in a 15 regulated fashion. Regulated expression of the polypeptides is possible by placing the coding sequence of the silk protein under the control of promoters that are tissue-specific, developmental-specific, or inducible. Several tissue-specific regulated genes and/or promoters have been reported in plants. These include genes encoding the seed storage proteins (such as napin, cruciferin,  $\beta$ -conglycinin, glycinin and phaseolin), 20 zein or oil body proteins (such as oleosin), or genes involved in fatty acid biosynthesis (including acyl carrier protein, stearoyl-ACP desaturase, and fatty acid desaturases (FAD 2-1)), and other genes expressed during embryo development (such as Bce4). Particularly useful for seed-specific expression is the pea vicilin promoter. Other useful promoters for expression in mature leaves are those that are switched on at the 25 onset of senescence, such as the SAG promoter from *Arabidopsis*. A class of fruit-specific promoters expressed at or during anthesis through fruit development, at least until the beginning of ripening, is discussed in US 4,943,674. Other examples of tissue-specific promoters include those that direct expression in tubers (for example, patatin gene promoter), and in fibre cells (an example of a developmentally-regulated 30 fibre cell protein is E6 fibre).

Other regulatory sequences such as terminator sequences and polyadenylation signals include any such sequence functioning as such in plants, the choice of which would be obvious to the skilled addressee. The termination region used in the expression cassette will be chosen primarily for convenience, since the termination 35 regions appear to be relatively interchangeable. The termination region which is used may be native with the transcriptional initiation region, may be native with the polynucleotide sequence of interest, or may be derived from another source. The termination region may be naturally occurring, or wholly or partially synthetic.

Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions or from the genes for  $\beta$ -phaseolin, the chemically inducible plant gene, pIN.

Several techniques are available for the introduction of an expression construct 5 containing a nucleic acid sequence encoding a polypeptide of interest into the target plants. Such techniques include but are not limited to transformation of protoplasts using the calcium/polyethylene glycol method, electroporation and microinjection or (coated) particle bombardment. In addition to these so-called direct DNA transformation methods, transformation systems involving vectors are widely 10 available, such as viral and bacterial vectors (e.g. from the genus *Agrobacterium*). After selection and/or screening, the protoplasts, cells or plant parts that have been transformed can be regenerated into whole plants, using methods known in the art. The choice of the transformation and/or regeneration techniques is not critical for this invention.

15 To confirm the presence of the transgenes in transgenic cells and plants, a polymerase chain reaction (PCR) amplification or Southern blot analysis can be performed using methods known to those skilled in the art. Expression products of the transgenes can be detected in any of a variety of ways, depending upon the nature of the product, and include Western blot and enzyme assay. One particularly useful 20 way to quantitate protein expression and to detect replication in different plant tissues is to use a reporter gene, such as GUS. Once transgenic plants have been obtained, they may be grown to produce plant tissues or parts having the desired phenotype. The plant tissue or plant parts may be harvested, and/or the seed collected. The seed 25 may serve as a source for growing additional plants with tissues or parts having the desired characteristics.

#### Transgenic Non-Human Animals

Techniques for producing transgenic animals are well known in the art. A 30 useful general textbook on this subject is Houdebine, *Transgenic animals – Generation and Use* (Harwood Academic, 1997).

Heterologous DNA can be introduced, for example, into fertilized mammalian ova. For instance, totipotent or pluripotent stem cells can be transformed by microinjection, calcium phosphate mediated precipitation, liposome fusion, retroviral 35 infection or other means, the transformed cells are then introduced into the embryo, and the embryo then develops into a transgenic animal. In a highly preferred method, developing embryos are infected with a retrovirus containing the desired DNA, and transgenic animals produced from the infected embryo. In a most preferred method, however, the appropriate DNAs are coinjected into the pronucleus or cytoplasm of

embryos, preferably at the single cell stage, and the embryos are allowed to develop into mature transgenic animals.

Another method used to produce a transgenic animal involves microinjecting a nucleic acid into pro-nuclear stage eggs by standard methods. Injected eggs are then 5 cultured before transfer into the oviducts of pseudopregnant recipients.

Transgenic animals may also be produced by nuclear transfer technology. Using this method, fibroblasts from donor animals are stably transfected with a plasmid incorporating the coding sequences for a binding domain or binding partner of interest under the control of regulatory sequences. Stable transfecants are then 10 fused to enucleated oocytes, cultured and transferred into female recipients.

In one embodiment, the transgenic non-human animal is *Bombyx* silkworms (Tomita et al., 2011).

#### Recovery Methods and Production of Silk

15 The silk proteins of the present invention may be extracted and purified from recombinant cells, such as plant, animal, bacteria or yeast cells, producing said protein by a variety of methods. In one embodiment, the method involves removal of proteins from homogenized cells/tissues/plants etc by lowering pH and heating (to no more than 4°C, preferably no more than 10°C, below the melting temperature of the 20 triple helix), followed by ammonium sulfate fractionation. Briefly, total soluble proteins are extracted by homogenizing cells/tissues/plants. Proteins are removed by precipitation at pH 4.7 and then at 60°C. The resulting supernatant is then fractionated with ammonium sulfate at 40% saturation. The resulting protein will be of the order of 95% pure. Additional purification may be achieved with conventional 25 gel or affinity chromatography.

In another example, cell lysates are treated with high concentrations of acid e.g. HCl or propionic acid to reduce pH to ~1-2 for 1 hour or more which will solubilise the silk proteins but precipitate other proteins.

30 Fibrillar aggregates will form from solutions by spontaneous self-assembly of silk proteins of the invention when the protein concentration exceeds a critical value with selected pH and ionic strength conditions. The aggregates may be gathered and mechanically spun into macroscopic fibres according to the method of O'Brien et al. ("Design, Synthesis and Fabrication of Novel Self-Assembling Fibrillar Proteins", in Silk Polymers: Materials Science and Biotechnology, pp. 104-117, Kaplan, Adams, 35 Farmer and Viney, eds., (1994) by American Chemical Society, Washington, D.C.).

Products of the invention have a low processing requirement. The silk proteins of the invention require minimal processing, e.g. spinning, to form a strong fibre. This contrasts with *B. mori* and spider recombinant silk polypeptides which require

sophisticated spinning techniques in order to obtain the secondary structure ( $\beta$ -sheet) and strength of the fibre.

However, fibres may be spun from solutions having properties characteristic of a liquid crystal phase. The fibre concentration at which phase transition can occur is dependent on the composition of a protein or combination of proteins present in the solution. Phase transition, however, can be detected by monitoring the clarity and birefringence of the solution. Onset of a liquid crystal phase can be detected when the solution acquires a translucent appearance and registers birefringence when viewed through crossed polarizing filters.

10 In one fibre-forming technique, fibres can first be extruded from the protein solution through an orifice into methanol, until a length sufficient to be picked up by a mechanical means is produced. Then a fibre can be pulled by such mechanical means through a methanol solution, collected, and dried. Methods for drawing fibres are considered well-known in the art.

15 In one embodiment, extrusion into precipitants such as ammonium sulphate of polyethylene glycol is used.

Further examples of methods which may be used for producing products of the present are described in US 2004/0170827 and US 2005/0054830.

20 In a preferred embodiment, silk proteins of the invention, such as when in the form of a tube, rod or sponge, are crosslinked. In one embodiment, the proteins are crosslinked to a surface/article/product etc of interest using techniques known in the art. In another embodiment (or in combination with the previous embodiment), at least some silk proteins are crosslinked to each other. Such crosslinking can be performed using chemical and/or enzymatic techniques known in the art. For 25 example, enzymatic cross-links can be catalysed by lysyl oxidase, whereas nonenzymatic cross-links can be generated from glycated lysine residues (Reiser et al., 1992) or using transglutaminase. In another embodiment, the silk proteins of the invention are cross-linked by glutaraldehydes. With the preponderance of Tyr residues particularly in the B and C chains, photochemical cross-linking which 30 involves just this residue could be useful (Elvin et al., 2010). See Ramshaw et al. (2000) Stabilisation of collagen in clinical applications. In: *Handbook of Biomaterials Engineering*, Wise, D.L. (Ed.), Marcel Dekker Inc., New York, pp. 717-738, for further details of cross-linking.

35 **Antibodies**

The term "antibody" as used in this invention includes polyclonal antibodies, monoclonal antibodies, bispecific antibodies, diabodies, triabodies, heteroconjugate antibodies, chimeric antibodies including intact molecules as well as fragments

thereof, such as Fab, F(ab')2, and Fv which are capable of binding the epitopic determinant, and other antibody-like molecules.

Antibody fragments retain some ability to selectively bind with its antigen or receptor and are defined as follows:

- 5 (1) Fab, the fragment which contains a monovalent antigen-binding fragment of an antibody molecule can be produced by digestion of whole antibody with the enzyme papain to yield an intact light chain and a portion of one heavy chain;
- 10 (2) Fab', the fragment of an antibody molecule can be obtained by treating whole antibody with pepsin, followed by reduction, to yield an intact light chain and a portion of the heavy chain; two Fab' fragments are obtained per antibody molecule;
- 15 (3) (Fab')2, the fragment of the antibody that can be obtained by treating whole antibody with the enzyme pepsin without subsequent reduction; F(ab)2 is a dimer of two Fab' fragments held together by two disulfide bonds;
- 20 (4) Fv, defined as a genetically engineered fragment containing the variable region of the light chain and the variable region of the heavy chain expressed as two chains; and
- (5) Single chain antibody ("SCA"), defined as a genetically engineered molecule containing the variable region of the light chain, the variable region of the heavy chain, linked by a suitable polypeptide linker as a genetically fused single chain molecule.

Methods of making these fragments are known in the art (see for example, Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, New York (1988)).

- 25 (6) Single domain antibody, typically a variable heavy domain devoid of a light chain.

The phrase "specifically binds" means that under particular conditions, the compound binds a polypeptide of the invention and does not bind to a significant amount to other, for example, proteins or carbohydrates. Specific binding may require an antibody that is selected for its specificity. In another embodiment, an antibody is considered to "specifically binds" if there is a greater than 10 fold difference, and preferably a 25, 50 or 100 fold greater difference between the binding of the antibody to a polypeptide of the invention when compared to another protein, especially a silk protein or a collagen protein (for example from mammals).

As used herein, the term "epitope" refers to a region of a polypeptide of the invention which is bound by the antibody. An epitope can be administered to an animal to generate antibodies against the epitope, however, antibodies of the present invention preferably specifically bind the epitope region in the context of the entire polypeptide.

If polyclonal antibodies are desired, a selected mammal (e.g., mouse, rabbit, goat, horse, etc) is immunised with an immunogenic polypeptide of the invention. Serum from the immunised animal is collected and treated according to known procedures. If serum containing polyclonal antibodies contains antibodies to other 5 antigens, the polyclonal antibodies can be purified by immunoaffinity chromatography. Techniques for producing and processing polyclonal antisera are known in the art. In order that such antibodies may be made, the invention also provides polypeptides of the invention or fragments thereof haptensed to another polypeptide for use as immunogens in animals.

10 Monoclonal antibodies directed against polypeptides of the invention can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies by hybridomas is well known. Immortal antibody-producing cell lines can be created by cell fusion, and also by other techniques such as direct transformation of B lymphocytes with oncogenic DNA, or transfection with Epstein- 15 Barr virus. Panels of monoclonal antibodies produced can be screened for various properties; i.e., for isotype and epitope affinity.

An alternative technique involves screening phage display libraries where, for example the phage express scFv fragments on the surface of their coat with a large variety of complementarity determining regions (CDRs). This technique is well 20 known in the art.

Other techniques for producing antibodies of the invention are known in the art.

Antibodies of the invention may be bound to a solid support and/or packaged 25 into kits in a suitable container along with suitable reagents, controls, instructions and the like.

In an embodiment, antibodies of the present invention are detectably labeled. Exemplary detectable labels that allow for direct measurement of antibody binding include radiolabels, fluorophores, dyes, magnetic beads, chemiluminescers, colloidal particles, and the like. Examples of labels which permit indirect measurement of 30 binding include enzymes where the substrate may provide for a coloured or fluorescent product. Additional exemplary detectable labels include covalently bound enzymes capable of providing a detectable product signal after addition of suitable substrate. Examples of suitable enzymes for use in conjugates include horseradish peroxidase, alkaline phosphatase, malate dehydrogenase and the like. Where not 35 commercially available, such antibody-enzyme conjugates are readily produced by techniques known to those skilled in the art. Further, exemplary detectable labels include biotin, which binds with high affinity to avidin or streptavidin; fluorochromes (e.g., phycobiliproteins, phycoerythrin and allophycocyanins; fluorescein and Texas

red), which can be used with a fluorescence activated cell sorter; haptens; and the like. Preferably, the detectable label allows for direct measurement in a plate luminometer, for example, biotin. Such labeled antibodies can be used in techniques known in the art to detect polypeptides of the invention.

5

### Compositions

Compositions of the present invention may include an "acceptable carrier". Examples of such acceptable carriers include water, saline, Ringer's solution, dextrose solution, Hank's solution, and other aqueous physiologically balanced salt solutions.

10 Nonaqueous vehicles, such as fixed oils, sesame oil, ethyl oleate, or triglycerides may also be used.

In one embodiment, the "acceptable carrier" is a "pharmaceutically acceptable carrier". The term pharmaceutically acceptable carrier refers to molecular entities and compositions that do not produce an allergic, toxic or otherwise adverse reaction 15 when administered to an animal, particularly a mammal, and more particularly a human. Useful examples of pharmaceutically acceptable carriers or diluents include, but are not limited to, solvents, dispersion media, coatings, stabilizers, protective colloids, adhesives, thickeners, thixotropic agents, penetration agents, sequestering agents and isotonic and absorption delaying agents that do not affect the activity of 20 the polypeptides of the invention. The proper fluidity can be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. More generally, the polypeptides of the invention can be combined with any non-toxic solid or liquid additive corresponding to the usual formulating techniques.

25 As outlined herein, in some embodiments a product comprising silk proteins of the invention is used as a pharmaceutically acceptable carrier.

Other suitable compositions are described below with specific reference to specific uses of the polypeptides of the invention.

30 Uses

Silk proteins are useful for the creation of new biomaterials because of their exceptional toughness and strength. However, the fibrous proteins of spiders and insects are generally large proteins (over 100kDa) and consist of highly repetitive amino acid sequences. These proteins are encoded by large genes containing highly 35 biased codons making them particularly difficult to produce in recombinant systems. By comparison, the silk proteins of the invention are short and less-repetitive. These properties make the genes encoding these proteins particularly attractive for recombinant production of new biomaterials.

The silk proteins of the invention can be used for a broad and diverse array of medical, military, industrial and commercial applications. Silk proteins, for example in the form of silk fibres, can be used in the manufacture of medical devices such as sutures, skin grafts, cellular growth matrices, replacement ligaments, and surgical mesh, and in a wide range of industrial and commercial products, such as, for example, cable, rope, netting, fishing line, clothing fabric, bullet-proof vest lining, container fabric, backpacks, knapsacks, bag or purse straps, adhesive binding material, non-adhesive binding material, strapping material, tent fabric, tarpaulins, pool covers, vehicle covers, fencing material, sealant, construction material, weatherproofing material, flexible partition material, sports equipment; and, in fact, in nearly any use of fibre or fabric for which high tensile strength and elasticity are desired characteristics. The silk proteins, for example in the form of silk fibres, and/or copolymers of the present invention also have applications in compositions for personal care products such as cosmetics, skin care, hair care and hair colouring; and in coating of particles, such as pigments.

The silk proteins may be used in their native form or they may be modified to form derivatives, which provide a more beneficial effect. For example, the silk protein may be modified by conjugation to a polymer to reduce allergenicity as described in US 5,981,718 and US 5,856,451. Suitable modifying polymers include, but are not limited to, polyalkylene oxides, polyvinyl alcohol, polycarboxylates, polyvinylpyrrolidone, and dextrans. In another example, the silk proteins may be modified by selective digestion and splicing of other protein modifiers. For example, the silk proteins may be cleaved into smaller peptide units by treatment with acid at an elevated temperature of about 60°C. The useful acids include, but are not limited to, dilute hydrochloric, sulfuric or phosphoric acids. Alternatively, digestion of the silk proteins may be done by treatment with a base, such as sodium hydroxide, or enzymatic digestion using a suitable protease may be used.

The proteins may be further modified to provide performance characteristics that are beneficial in specific applications for personal care products. The modification of proteins for use in personal care products is well known in the art. For example, commonly used methods are described in US 6,303,752, US 6,284,246, and US 6,358,501. Examples of modifications include, but are not limited to, ethoxylation to promote water-oil emulsion enhancement, siloxylation to provide lipophilic compatibility, and esterification to aid in compatibility with soap and detergent compositions. Additionally, the silk proteins may be derivatized with functional groups including, but not limited to, amines, oxiranes, cyanates, carboxylic acid esters, silicone copolyols, siloxane esters, quaternized amine aliphatics, urethanes, polyacrylamides, dicarboxylic acid esters, and halogenated esters. The silk

proteins may also be derivatized by reaction with diimines and by the formation of metal salts.

Consistent with the above definitions of "polypeptide" (and "protein"), such derivatized and/or modified molecules are also referred to herein broadly as 5 "polypeptides" and "proteins".

Silk proteins of the invention can be spun together and/or bundled or braided with other fibre types. Examples include, but are not limited to, polymeric fibres (e.g., polypropylene, nylon, polyester), fibres and silks of other plant and animal 10 sources (e.g., cotton, wool, *Bombyx mori*, spider silk or honey bee (for example see, 15 WO 2007/038837), and glass fibres. A preferred embodiment is silk fibre braided with 10% polypropylene fibre. The present invention contemplates that the production of such combinations of fibres can be readily practiced to enhance any desired characteristics, e.g., appearance, softness, weight, durability, water-repellant properties, improved cost-of-manufacture, that may be generally sought in the 20 manufacture and production of silk protein comprising products, for example fibres, for medical, industrial, or commercial applications.

Silk proteins of the invention can be used to stabilize compounds, for example therapeutic drugs such as vaccines and antibiotics, during shipping and/or storage (see, for example, Zhang et al., 2012).

20

#### *Personal Care Products*

Cosmetic and skin care compositions may be anhydrous compositions comprising an effective amount of silk protein in a cosmetically acceptable medium.

The uses of these compositions include, but are not limited to, skin care, skin 25 cleansing, make-up, and anti-wrinkle products. An effective amount of a silk protein for cosmetic and skin care compositions is herein defined as a proportion of from

about  $10^{-4}$  to about 30% by weight, but preferably from about  $10^{-3}$  to 15% by weight, relative to the total weight of the composition. This proportion may vary as a function 30 of the type of cosmetic or skin care composition. Suitable compositions for a

cosmetically acceptable medium are described in US 6,280,747. For example, the 35 cosmetically acceptable medium may contain a fatty substance in a proportion generally of from about 10 to about 90% by weight relative to the total weight of the composition, where the fatty phase containing at least one liquid, solid or semi-solid fatty substance. The fatty substance includes, but is not limited to, oils, waxes, gums,

and so-called pasty fatty substances. Alternatively, the compositions may be in the 40 form of a stable dispersion such as a water-in-oil or oil-in-water emulsion. Additionally, the compositions may contain one or more conventional cosmetic or dermatological additives or adjuvants, including but not limited to, antioxidants,

preserving agents, fillers, surfactants, UVA and/or UVB sunscreens, fragrances, thickeners, wetting agents and anionic, nonionic or amphoteric polymers, and dyes or pigments.

Emulsified cosmetics and quasi drugs which are producible with the use of emulsified materials comprising at least one silk protein of the present invention include, for example, cleansing cosmetics (beauty soap, facial wash, shampoo, rinse, and the like), hair care products (hair dye, hair cosmetics, and the like), basic cosmetics (general cream, emulsion, shaving cream, conditioner, cologne, shaving lotion, cosmetic oil, facial mask, and the like), make-up cosmetics (foundation, 5 eyebrow pencil, eye cream, eye shadow, mascara, and the like), aromatic cosmetics (perfume and the like), tanning and sunscreen cosmetics (tanning and sunscreen cream, tanning and sunscreen lotion, tanning and sunscreen oil, and the like), nail cosmetics (nail cream and the like), eyeliner cosmetics (eyeliner and the like), lip cosmetics (lipstick, lip cream, and the like), oral care products (tooth paste and the 10 like) bath cosmetics (bath products and the like), and the like.

The cosmetic composition may also be in the form of products for nail care, such as a nail varnish. Nail varnishes are herein defined as compositions for the treatment and colouring of nails, comprising an effective amount of silk protein in a cosmetically acceptable medium. An effective amount of a silk protein for use in a 20 nail varnish composition is herein defined as a proportion of from about  $10^{-4}$  to about 30% by weight relative to the total weight of the varnish. Components of a cosmetically acceptable medium for nail varnishes are described in US 6,280,747. The nail varnish typically contains a solvent and a film forming substance, such as cellulose derivatives, polyvinyl derivatives, acrylic polymers or copolymers, vinyl 25 copolymers and polyester polymers. The composition may also contain an organic or inorganic pigment.

Hair care compositions are herein defined as compositions for the treatment of hair, including but not limited to shampoos, conditioners, lotions, aerosols, gels, and mousses, comprising an effective amount of silk protein in a cosmetically acceptable 30 medium. An effective amount of a silk protein for use in a hair care composition is herein defined as a proportion of from about  $10^{-2}$  to about 90% by weight relative to the total weight of the composition. Components of a cosmetically acceptable medium for hair care compositions are described in US 2004/0170590, US 6,280,747, US 6,139,851, and US 6,013,250. For example, these hair care compositions can be 35 aqueous, alcoholic or aqueous-alcoholic solutions, the alcohol preferably being ethanol or isopropanol, in a proportion of from about 1 to about 75% by weight relative to the total weight, for the aqueous-alcoholic solutions. Additionally, the hair

care compositions may contain one or more conventional cosmetic or dermatological additives or adjuvants, as given above.

Hair colouring compositions are herein defined as compositions for the colouring, dyeing, or bleaching of hair, comprising an effective amount of silk protein 5 in a cosmetically acceptable medium. An effective amount of a silk protein for use in a hair colouring composition is herein defined as a proportion of from about  $10^{-4}$  to about 60% by weight relative to the total weight of the composition. Components of a cosmetically acceptable medium for hair colouring compositions are described in US 2004/0170590, US 6,398,821 and US 6,129,770. For example, hair colouring 10 compositions generally contain a mixture of inorganic peroxygen-based dye oxidizing agent and an oxidizable coloring agent. The peroxygen-based dye oxidizing agent is most commonly hydrogen peroxide. The oxidative hair coloring agents are formed by oxidative coupling of primary intermediates (for example p-phenylenediamines, p-aminophenols, p-diaminopyridines, hydroxyindoles, aminoindoles, aminothymidines, 15 or cyanophenols) with secondary intermediates (for example phenols, resorcinols, m-aminophenols, m-phenylenediamines, naphthols, pyrazolones, hydroxyindoles, catechols or pyrazoles). Additionally, hair colouring compositions may contain oxidizing acids, sequestrants, stabilizers, thickeners, buffers carriers, surfactants, solvents, antioxidants, polymers, non-oxidative dyes and conditioners.

20 The silk proteins can also be used to coat pigments and cosmetic particles in order to improve dispersibility of the particles for use in cosmetics and coating compositions. Cosmetic particles are herein defined as particulate materials such as pigments or inert particles that are used in cosmetic compositions. Suitable pigments and cosmetic particles include, but are not limited to, inorganic color pigments, 25 organic pigments, and inert particles. The inorganic color pigments include, but are not limited to, titanium dioxide, zinc oxide, and oxides of iron, magnesium, cobalt, and aluminium. Organic pigments include, but are not limited to, D&C Red No. 36, D&C Orange No. 17, the calcium lakes of D&C Red Nos. 7, 11, 31 and 34, the barium lake of D&C Red No. 12, the strontium lake D&C Red No. 13, the aluminium 30 lake of FD&C Yellow No. 5 and carbon black particles. Inert particles include, but are not limited to, calcium carbonate, aluminium silicate, calcium silicate, magnesium silicate, mica, talc, barium sulfate, calcium sulfate, powdered Nylon<sup>TM</sup>, perfluorinated alkanes, and other inert plastics.

35 The silk proteins may also be used in dental floss (see, for example, US 2005/0161058). The floss may be monofilament yarn or multifilament yarn, and the fibres may or may not be twisted. The dental floss may be packaged as individual pieces or in a roll with a cutter for cutting pieces to any desired length. The dental floss may be provided in a variety of shapes other than filaments, such as but not

limited to, strips and sheets and the like. The floss may be coated with different materials, such as but not limited to, wax, polytetrafluoroethylene monofilament yarn for floss.

The silk proteins may also be used in soap (see, for example, US 5 2005/0130857).

#### *Pigment and Cosmetic Particle Coating*

The effective amount of a silk protein for use in pigment and cosmetic particle coating is herein defined as a proportion of from about  $10^{-4}$  to about 50%, but 10 preferably from about 0.25 to about 15% by weight relative to the dry weight of particle. The optimum amount of the silk protein to be used depends on the type of pigment or cosmetic particle being coated. For example, the amount of silk protein used with inorganic color pigments is preferably between about 0.01% and 20% by weight. In the case of organic pigments, the preferred amount of silk protein is 15 between about 1% to about 15% by weight, while for inert particles, the preferred amount is between about 0.25% to about 3% by weight. Methods for the preparation of coated pigments and particles are described in US 5,643,672. These methods include: adding an aqueous solution of the silk protein to the particles while tumbling or mixing, forming a slurry of the silk protein and the particles and drying, spray 20 drying a solution of the silk protein onto the particles or lyophilizing a slurry of the silk protein and the particles. These coated pigments and cosmetic particles may be used in cosmetic formulations, paints, inks and the like.

#### *Biomedical*

25 The silk proteins may be used as a coating on a bandage to promote wound healing. For this application, the bandage material is coated with an effective amount of the silk protein. For the purpose of a wound-healing bandage, an effective amount of silk protein is herein defined as a proportion of from about  $10^{-4}$  to about 30% by weight relative to the weight of the bandage material. The material to be coated may 30 be any soft, biologically inert, porous cloth or fibre. Examples include, but are not limited to, cotton, silk, rayon, acetate, acrylic, polyethylene, polyester, and combinations thereof. The coating of the cloth or fibre may be accomplished by a number of methods known in the art. For example, the material to be coated may be dipped into an aqueous solution containing the silk protein. Alternatively, the 35 solution containing the silk protein may be sprayed onto the surface of the material to be coated using a spray gun. Additionally, the solution containing the silk protein may be coated onto the surface using a roller coat printing process. The wound bandage may include other additives including, but not limited to, disinfectants such

as iodine, potassium iodide, povidon iodine, acrinol, hydrogen peroxide, benzalkonium chloride, and chlorohexidine; cure accelerating agents such as allantoin, dibucaine hydrochloride, and chlorophenylamine malate; vasoconstrictor agents such as naphazoline hydrochloride; astringent agents such as zinc oxide; and 5 crust generating agents such as boric acid.

The silk proteins of the present invention may also be used in the form of a film as a wound dressing material. The use of silk proteins, in the form of an amorphous film, as a wound dressing material is described in US 6,175,053. The amorphous film comprises a dense and nonporous film of a crystallinity below 10% 10 which contains an effective amount of silk protein. For a film for wound care, an effective amount of silk protein is herein defined as between about 1 to 99% by weight. The film may also contain other components including but not limited to other proteins such as sericin, and disinfectants, cure accelerating agents, vasoconstrictor agents, astringent agents, and crust generating agents, as described 15 above. Other proteins such as sericin may comprise 1 to 99% by weight of the composition. The amount of the other ingredients listed is preferably below a total of about 30% by weight, more preferably between about 0.5 to 20% by weight of the composition. The wound dressing film may be prepared by dissolving the above mentioned materials in an aqueous solution, removing insolubles by filtration or 20 centrifugation, and casting the solution on a smooth solid surface such as an acrylic plate, followed by drying.

The silk proteins of the present invention may also be used in sutures (see, for example, US 2005/0055051). Such sutures can feature a braided jacket made of ultrahigh molecular weight fibres and silk fibres. The polyethylene provides strength. 25 Polyester fibres may be woven with the high molecular weight polyethylene to provide improved tie down properties. The silk may be provided in a contrasting color to provide a trace for improved suture recognition and identification. Silk also is more tissue compliant than other fibres, allowing the ends to be cut close to the knot without concern for deleterious interaction between the ends of the suture and 30 surrounding tissue. Handling properties of the high strength suture also can be enhanced using various materials to coat the suture. The suture advantageously has the strength of Ethibond No. 5 suture, yet has the diameter, feel and tie-ability of No. 2 suture. As a result, the suture is ideal for most orthopedic procedures such as rotator 35 cuff repair, Achilles tendon repair, patellar tendon repair, ACL/PCL reconstruction, hip and shoulder reconstruction procedures, and replacement for suture used in or with suture anchors. The suture can be uncoated, or coated with wax (beeswax, petroleum wax, polyethylene wax, or others), silicone (Dow Corning silicone fluid 202A or others), silicone rubbers, PBA (polybutylate acid), ethyl cellulose (Filodel) or

other coatings, to improve lubricity of the braid, knot security, or abrasion resistance, for example.

The silk proteins of the present invention may also be used in stents (see, for example, US 2004/0199241). For example, a stent graft is provided that includes an 5 endoluminal stent and a graft, wherein the stent graft includes silk. The silk induces a response in a host who receives the stent graft, where the response can lead to enhanced adhesion between the silk stent graft and the host's tissue that is adjacent to the silk of the silk stent graft. The silk may be attached to the graft by any of various means, e.g., by interweaving the silk into the graft or by adhering the silk to the graft 10 (e.g., by means of an adhesive or by means of suture). The silk may be in the form of a thread, a braid, a sheet, powder, etc. As for the location of the silk on the stent graft, the silk may be attached only the exterior of the stent, and/or the silk may be attached to distal regions of the stent graft, in order to assist in securing those distal regions to neighbouring tissue in the host. A wide variety of stent grafts may be utilized within 15 the context of the present invention, depending on the site and nature of treatment desired. Stent grafts may be, for example, bifurcated or tube grafts, cylindrical or tapered, self-expandable or balloon-expandable, unibody or, modular, etc.

In addition to silk, the stent graft may contain a coating on some or all of the 20 silk, where the coating degrades upon insertion of the stent graft into a host, the coating thereby delaying contact between the silk and the host. Suitable coatings include, without limitation, gelatin, degradable polyesters (e.g., PLGA, PLA, MePEG-PLGA, PLGA-PEG-PLGA, and copolymers and blends thereof), cellulose and cellulose derivatives (e.g., hydroxypropyl cellulose), polysaccharides (e.g., hyaluronic acid, dextran, dextran sulfate, chitosan), lipids, fatty acids, sugar esters, nucleic acid 25 esters, polyanhydrides, polyorthoesters and polyvinylalcohol (PVA). The silk-containing stent grafts may contain a biologically active agent (drug), where the agent is released from the stent graft and then induces an enhanced cellular response (e.g., cellular or extracellular matrix deposition) and/or fibrotic response in a host into which the stent graft has been inserted.

The silk proteins of the present invention may also be used in a matrix for 30 producing ligaments and tendons *ex vivo* (see, for example, US 2005/0089552). A silk-fibre-based matrix can be seeded with pluripotent cells, such as bone marrow stromal cells (BMSCs). The bioengineered ligament or tendon is advantageously characterized by a cellular orientation and/or matrix crimp pattern in the direction of 35 applied mechanical forces, and also by the production of ligament and tendon specific markers including collagen type I, collagen type III, and fibronectin proteins along the axis of mechanical load produced by the mechanical forces or stimulation, if such forces are applied. In a preferred embodiment, the ligament or tendon is characterized

by the presence of fibre bundles which are arranged into a helical organization. Some examples of ligaments or tendons that can be produced include anterior cruciate ligament, posterior cruciate ligament, rotator cuff tendons, medial collateral ligament of the elbow and knee, flexor tendons of the hand, lateral ligaments of the ankle and 5 tendons and ligaments of the jaw or temporomandibular joint. Other tissues that may be produced by methods of the present invention include cartilage (both articular and meniscal), bone, muscle, skin and blood vessels.

The silk proteins of the present invention may also be used in hydrogels (see, for example, US 2005/0266992). Silk fibroin hydrogels can be characterized by an 10 open pore structure which allows their use as tissue engineering scaffolds, substrate for cell culture, wound and burn dressing, soft tissue substitutes, bone filler, and as well as support for pharmaceutical or biologically active compounds.

The silk proteins may also be used in dermatological compositions (see, for example, US 2005/0019297). Furthermore, the silk proteins of the invention and 15 derivatives thereof may also be used in sustained release compositions (see, for example, US 2004/0005363 and US 2012/0195967).

Products comprising silk proteins of the invention may also more broadly be used as scaffolds for the culture (such as cell and/or tissue culture) and/or implantation of cells, such as stem cells (see, for example, US 2012/0189669, US 20 2012/0172985, US 2012/0171257, US 2011/0293685, US 20090214614, US 20110238179 and US 20070041952).

### *Textiles*

The silk proteins of the present invention may also be applied to the surface of 25 fibres for subsequent use in textiles. This provides a monolayer of the protein film on the fibre, resulting in a smooth finish. US 6,416,558 and US 5,232,611 describe the addition of a finishing coat to fibres. The methods described in these disclosures provide examples of the versatility of finishing the fibre to provide a good feel and a smooth surface. For this application, the fibre is coated with an effective amount of 30 the silk protein. For the purpose of fibre coating for use in textiles, an effective amount of silk protein is herein defined as a proportion of from about 1 to about 99% by weight relative to the weight of the fibre material. The fibre materials include, but are not limited to textile fibres of cotton, polyesters such as rayon and Lycra<sup>TM</sup>, nylon, 35 wool, and other natural fibres including native silk. Compositions suitable for applying the silk protein onto the fibre may include co-solvents such as ethanol, isopropanol, hexafluorans, isothiocyanouranates, and other polar solvents that can be mixed with water to form solutions or microemulsions. The silk protein-containing solution may be sprayed onto the fibre or the fibre may be dipped into the solution.

While not necessary, flash drying of the coated material is preferred. An alternative protocol is to apply the silk protein composition onto woven fibres. An ideal embodiment of this application is the use of silk proteins to coat stretchable weaves such as used for stockings.

5

#### *Composite Materials*

Silk proteins, for example in the form of fibres, can be added to polyurethane, other resins or thermoplastic fillers to prepare panel boards and other construction material or as moulded furniture and benchtops that replace wood and particle board.

10 The composites can be also be used in building and automotive construction, especially rooftops and door panels. The silk re-enforces the resin making the material much stronger and allowing lighterweight construction which is of equal or superior strength to other particle boards and composite materials. The silk proteins, for example in the form of fibres, may be isolated and added to a synthetic composite-forming resin or be used in combination with plant-derived proteins, starch and oils to produce a biologically-based composite materials. Processes for the production of such materials are described in JP 2004284246, US 2005175825, US 4,515,737, JP 47020312 and WO 2005/017004.

20 *Paper Additives*

The fibre properties of the silk of the invention can add strength and quality texture to paper making. Silk papers are made by mottling silk threads in cotton pulp to prepare extra smooth handmade papers is used for gift wrapping, notebook covers, carry bags. Processes for production of paper products which can include silk proteins of the invention are generally described in JP 2000139755.

#### *Advanced Materials*

Silks of the invention have considerable toughness and stands out among other silks in maintaining these properties when wet (Hepburn et al., 1979).

30 Areas of substantial growth in the clothing textile industry are the technical and intelligent textiles. There is a rising demand for healthy, high value functional, environmentally friendly and personalized textile products. Fibres, such as those of the invention, that do not change properties when wet and in particular maintain their strength and extensibility are useful for functional clothing for sports and leisure wear as well as work wear and protective clothing.

35 Developments in the weapons and surveillance technologies are prompting innovations in individual protection equipments and battle-field related systems and structures. Besides conventional requirements such as material durability to

prolonged exposure, heavy wear and protection from external environment, silk textiles of the invention can be processed to resist ballistic projectiles, fire and chemicals. Processes for the production of such materials are described in WO 2005/045122 and US 2005268443.

5

## **EXAMPLES**

### **Example 1 - Preparation and analysis of late last instar salivary gland cDNAs**

The proteins that are found in sawfly (*Nematus oligospilus*, the willow sawfly) silks were identified from cDNAs isolated from the salivary gland of late final instar 10 larvae. Willow sawfly larvae (*N. oligospilus* from the genus *Nematus*) were collected from *Salix* (willow) species around lake Burley Griffin in Canberra (Australia). Larvae were maintained on a diet of fresh willow leaves in the laboratory until required.

A *N. oligospilus* silk gland cDNA library was constructed after mRNA was 15 isolated using the Micro-FastTrack™ 2.0 mRNA Isolation kit (Invitrogen) from 28.1 µg total RNA that had been isolated from the labial glands of 10 larval sawflies using the RNAqueous-4PCR kit (Ambion). The cDNA library was constructed from the mRNA using the CloneMiner™ cDNA kit (Invitrogen). The cDNA library comprised approximately  $2.9 \times 10^7$  colony forming units (cfu) with less than 1% of the original 20 vector. The average insert size within the libraries was 1.1 kbp.

Fifty randomly chosen clones were sent for sequence analysis (Table 2). Introduced cDNA sequences were found in 28 of the clones with the remainder of the 25 inserts the original cDNA cloning vector. Twenty of the identified cDNA sequences contained sequences where the majority of the translated sequence contained the 3 residue repeating sequence Gly-Xaa-Yaa which is characteristic of collagen and other triple-helix containing structures. The Gly is required every third residue as only this amino acid is small enough to fit within the centre of a triple-helix. The sequences containing the Gly-Xaa-Yaa repeat were grouped into 3 distinct classes and were termed Collagens A, B and C. The potential identity of other sequences was predicted 30 from database comparisons.

An alignment of the collagen-like domains using the Expasy(ETH) program is provide as Figure 1. For the 3 collagen segments, 126 residues in A and B and 120 in C chain. Of the non-Gly positions, 120-126 amino acids, 20 are common to all three chains. Between the A- and B-chains there are a further 21 positions in common 35 (=41). Between the A- and C-chains there are a further 24 positions in common (=44). Between the B- and C-chains there are a further 13 positions in common (=33). There is also a tyrosine rich region at the C-terminus of the proteins, with each protein ending with a tyrosine.

**Table 2.** Identity of proteins contained within cDNA from *N. oligospilus* silk gland.

Sequence type	Number of cDNA
Collagen A	4
Collagen B	8
Collagen C	8
Highly repetitive sequence	3
Peptidase	1
RNA helicase	1
Esterase	1
Actin	1
Translation elongation factor	1

5 **Example 2 - Recombinant expression of sawfly collagen-like silk proteins**

Three cDNA sequences, one from each of the three amino acid sequence types A (SF21), B (SF9) and C (SF30), were selected for expression studies. The cDNA were in pDONR 222. Restriction enzyme digestion sites were introduced into the original constructs by PCR mutagenesis using a Stratagene Site-directed Mutagenesis 10 Kit (#200519) according to manufactures directions. The introduction of restriction enzyme digestion sites allowed isolation of the DNA for each sawfly silk gene and its insertion into an expression vector.

Sawfly collagen-like silk type A gene was inserted into pColdI vector via NdeI and EcoRI. Sawfly collagen-like silk type B gene was inserted pColdIV vector via 15 NdeI and EcoRI. Sawfly collagen-like silk type C gene was inserted into pCold III via BamHI and SalI, and including a triple helix promoting sequence (V-domain) derived from the bacterial collagen Scl2 from *S. pyogenes*.

For expression of the sawfly silk protein, one colony of transformed *E. coli* BL21 cells was added to 100ml starter culture medium, 2x YT-Amp and incubated at 20 37°C with 200 rpm shaking overnight. This culture then had 100ml fresh 2x YT-2% Glucose-Amp added, and was induced with 1mM IPTG at 25°C for 10 hour, then 20°C for another 16 hours. The cell paste was harvested by centrifugation (3000×g for 30 min). The protein was associated with the cell pellet. For extraction, 1 gram of cell paste was resuspended in 10ml of 40mM Na/Phosphate buffer pH 8.0, and the 25 cells burst by sonication. The cell lysate mixture was clarified by centrifugation (20k × g for 40 min) and the clear supernatant retained.

Expressed silk proteins were purified used affinity purification using Pall 50ml IMAC HyperCel column. Individual stock solutions of 200mM Na/Phosphate, 4.4 M NaCl and 2 M imidazole were added to the clear lysate to final concentration of 20mM Na/Phosphate buffer pH 7.6, 500mM NaCl and 30mM imidazole. This 5 solution was loaded onto the affinity column, followed by washing the column with 5 column volumes of buffer A (20mM Na/Phosphate buffer pH7.6, 300mM NaCl and 40mM Imidazole), sufficient for there to be no residual elution of material absorbing at 280 nm. The column was then eluted with a series of 1 column volume of buffer A, containing sequential stepwise increases of various amount of imidazole (70, 100, 10 150, 250 and 500mM imidazole. Eluted fractions were monitored by absorption at 214 nm or 280 nm and were then analysed further by SDS-PAGE analysis. Samples containing the silk protein were then pooled and concentrated, and the buffer exchanged to 20mM Phosphate pH 8.0, by membrane filtration (10KDa cut off). This material was then further purified by ion-exchange chromatography on an anion-1ml 15 Q column and Cation-1m SP column then further purify by Gel filtration on a Superdex 200 column. Many methods could be used to purify the protein.

Examples of sawfly silk protein expression from a pCold vector are shown in Figure 2. In this example, SF30 was expressed as a fusion protein with a V-domain.

20 **Example 3 - Collagen-like silks bind fibroblasts**

For cell adhesion studies, tissue culture 96 well microtiter plates were coated with 2 $\mu$ g of proteins in PBS overnight at 4°C. After blocking with PBS containing 2% BSA for 2h, mouse fibroblast L929 cells were seeded onto the coated surface at 250,000 cells cm<sup>-2</sup>. Prior to seeding, cells were harvested by brief exposure to 25 Tryple<sup>TM</sup>Express (Gibco #12605) and resuspended in serum-free medium following by washing the cells with serum-free medium three times. Following 2h exposure at 37°C/5% CO<sub>2</sub>, the cells in the well were rinsed extensively with warm PBS twice, and apply 100 $\mu$ l of MTS solution to the plates for further 1-3 hour incubation. Cells were also stained with phalloidin (for Actin) and DAPI (for DNA).

30 The collagen-like silk proteins of the invention support cell attachment of the L929 cells, presumably via the integrin receptor (Figure 3). Silks of the invention were as good as mammalian collagen and fibronectin. These results show that the collagen-like silk proteins of the invention will be useful in the production of biomedical implants.

35

**Example 4 - Recombinant sawfly silk proteins form a collagen-like structure**

To demonstrate that sawfly collagen-like silk type A to C proteins are able to fold into triple-helical structures, the sequence of each type was expressed

individually in *E. coli* (Figure 4). The full length, expressed silk proteins were purified using affinity chromatography and gel permeation chromatography as follows: Cells were lysed by sonication in 40mM sodium phosphate buffer, pH8.0, and the cell lysate clarified by centrifugation (20,000 xg for 40 min) and the clear supernatant retained. The expressed silk proteins were purified by absorbing the clarified lysates on an IMAC HyperCel™ column (Pall), with elution by stepwise increments up to 500 mM imidazole, adjusted to pH 8 with HCl. Cross-flow filtration was used to lower the salt content and to concentrate the protein solution. Final purification was by gel permeation chromatography on a HiPrep Sephacryl™ S-200 column (GE Healthcare). The individual triple-helical collagen segments were prepared by digestion of the proteins with 0.1 mg/ml pepsin in 50 mM acetic acid. Purity of all products was assessed by SDS-PAGE.

Generally collagen is resistant to pepsin digestion and the treatment is commonly used to purify collagen molecules. Consistent with a collagen structure, a large proportion of the recombinantly produced sawfly silk proteins were resistant to pepsin digestion at 20°C (Figure 4A). In comparison, after denaturation of the proteins structure at high temperature, the entire molecule was protease sensitive. The circular dichroism spectra of the pepsin-resistant fragments confirmed the proteins were folded in the collagen triple-helix, with each spectrum showing positive ellipticity with maxima around 220 nm (Figure B).

It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.

The present application claims priority from US 61/560,649 filed 16 November 2011, the entire contents of which are incorporated herein by reference.

All publications discussed and/or referenced herein are incorporated herein in their entirety.

Any discussion of documents, acts, materials, devices, articles or the like which has been included in the present specification is solely for the purpose of providing a context for the present invention. It is not to be taken as an admission that any or all of these matters form part of the prior art base or were common general knowledge in the field relevant to the present invention as it existed before the priority date of each claim of this application.

**REFERENCES**

Antonicelli (2007) *Curr. Top. Dev. Biol.* 79:99-155.

Atkins (1967) *J. Mol. Biol.* 24:139-41.

5 Ayoub et al. (2007) *PLoS One* 13:e514.

Bini et al. (2004) *J. Mol. Biol.* 335:27-40.

Craig (2002) *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 133:493-507.

Craig and Riek (2002) *Comp Biochem. and Phys. Part B* 133:247-255.

Dal Pra et al. (2005) *Biomaterials* 26:1987-99.

10 Elvin et al. (2010) *Biomaterials* 31:8323-8331.

Gosline et al. (1999) *J. Exp. Biol.* 202:3295-303.

Hayashi and Lewis (2000) *Science* 287:1477-9.

Hepburn and Kurstjens (1988) *Apidologie* 19:25-36.

Hepurn et al. (1979) *Insect Biochem.* 9:66.

15 Jin et al. (2002) *Biomacromolecules* 3:1233-9.

Kim et al. (2005) *Biomaterials* 26:2775-85.

Lucas and Rudall (1968) Extracellular fibrous proteins: the silks. In *Comprehensive Biochemistry* eds. M Florkin, EH Stotz, Chapter 7. 26B:475-558. Amsterdam, London, New York: Elsevier.

20 Mathews (1975) Connective tissue: macromolecular structure and evolution. Springer-Verlag.

Nazarov et al. (2004) *Biomacromolecules* 5:718-26.

Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443-453.

Persikov et al. (2005) *J. Biol. Chem.* 280:19343-19349.

25 Reiser et al. (1992) *Nucleic Acids Research* 32 (Web Server issue): W321-W326.

Ramshaw et al. (2000) *J. Struct. Biol.* 122:86-91.

Rudall (1962) In *Comparative Biochemistry* (ed: Florkin and Mason) 4:297-435 Academic Press, New York.

Rudall (1968) Comparative biology and biochemistry of collagen. In: *Treatise on* 30 *Collagen* ed BS Gould Chapter 2 2A:83-137. London and New York: Academic.

Rudall and Kenchington (1971) *Annual Review of Entomology* 16:73-96.

Sutherland et al. (2006) *Genome Res.* 16:1414-1421.

Sutherland et al. (2010) *Annu. Rev. Entomol.* 55:171-88.

Tomita et al. (2011) *Biotechnol. Lett.* 33:645-654.

35 Vepari and Kaplan (2007) *Prog. Polym. Sci.* 32:991-1007.

Vollrath and Knight (2001) *Nature* 410: 541-8.

Zhang et al. (2012) *PNAS* 109:11981-11986.

**CLAIMS**

1. An isolated and/or exogenous polynucleotide which encodes a collagen-like silk polypeptide.

5

2. The polynucleotide of claim 1 which comprises one or more of:

i) a sequence of nucleotides as provided in any one of SEQ ID NO's 7 to 12, 22 or 23;

10 ii) a sequence of nucleotides encoding a polypeptide comprising an amino acid sequence as provided in any one of SEQ ID NO's 1 to 6;

iii) a sequence of nucleotides encoding a polypeptide comprising an amino acid sequence which is at least 30% identical to any one or more of SEQ ID NO's 1 to 6;

15 iv) a sequence of nucleotides encoding a biologically active fragment of ii) or iii),

v) a sequence of nucleotides which is at least 30% identical to any one or more of SEQ ID NO's 7 to 12, 22 or 23, or

20 vi) a sequence which hybridizes to any one or more of i) to v) under stringent conditions.

20

3. The polynucleotide of claim 2, wherein the polynucleotide comprises a sequence of nucleotides encoding a polypeptide comprising an amino acid sequence which is at least 95% identical to any one or more of SEQ ID NO's 1 to 6.

25

4. A vector comprising at least one polynucleotide according to any one of claims 1 to 3.

5. The vector of claim 4 which is an expression vector.

30

6. A host cell comprising at least one polynucleotide according to any one of claims 1 to 3, and/or at least one vector of claim 4 or claim 5.

7. The host cell of claim 6 which is a bacterial, yeast, animal or plant cell.

35

8. A substantially purified and/or recombinant collagen-like silk polypeptide.

9. The polypeptide of claim 8 which comprises one or more of:

i) an amino acid sequence as provided in any one of SEQ ID NO's 1 to 6;

- ii) an amino acid sequence which is at least 30% identical to any one or more of SEQ ID NO's 1 to 6; or
  - iii) a biologically active fragment of i) or ii).

5 10. The polypeptide of claim 8 or claim 9 which can be purified from a species of Hymenoptera.

11. The polypeptide of claim 10, wherein the Genera of Hymenoptera is *Nematus*.

10 12. The polypeptide according to any one of claims 8 to 11 which is fused to at least one other polypeptide.

13. A transgenic non-human organism comprising an exogenous polynucleotide according to any one of claims 1 to 3, the polynucleotide encoding at least one 15 polypeptide according to any one of claims 8 to 12.

20 14. A process for preparing a polypeptide according to any one of claims 8 to 12, the process comprising cultivating one or more of a host cell of claim 6 or claim 7, a vector of claim 5, a transgenic organism of claim 13, under conditions which allow expression of the polynucleotide encoding the polypeptide, and recovering the expressed polypeptide.

15. An isolated and/or recombinant antibody which specifically binds a polypeptide according to any one of claims 8 to 12.

25 16. A silk fibre, sponge, film, hydrogel or particle comprising at least one polypeptide according to any one of claims 8 to 12.

17. The fibre, sponge, film, hydrogel or particle of claim 16, wherein at least some 30 of the polypeptides are crosslinked.

18. A copolymer comprising at least two polypeptides according to any one of claims 8 to 12.

35 19. The copolymer of claim 18, wherein at least some of the polypeptides are crosslinked.

20. A product comprising one or more of at least one polypeptide according to any one of claims 8 to 12, at least one silk fibre, sponge, film, hydrogel or particle of claim 16 or claim 17, or at least one copolymer of claim 18 or claim 19.

5 21. The product of claim 20, wherein the product is selected from the group consisting of: a personal care product, textiles, plastics, and a biomedical product.

22. The product of claim 20 or claim 21, wherein at least some of the polypeptides in the product are crosslinked.

10 23. A composition comprising one or more of at least one polypeptide according to any one of claims 8 to 12, at least one silk fibre, sponge, film, hydrogel or particle of claim 16 or claim 17, at least one copolymer of claim 18 or claim 19, or at least one product according to any one of claims 20 to 22, and one or more acceptable carriers.

15 24. The composition of claim 23 which further comprises a drug.

25. The composition of claim 24 for use as a medicine, a medical device or a cosmetic.

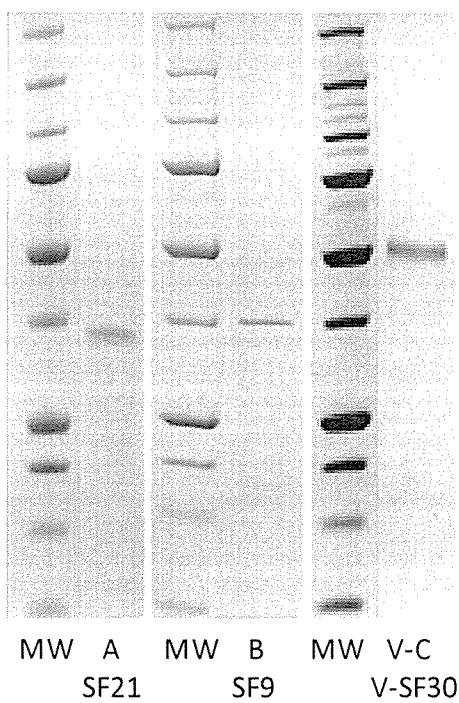
20 26. A composition comprising at least one polynucleotide according to any one of claims 1 to 3, and one or more acceptable carriers.

25 27. A process for producing a product comprising collagen-like silk polypeptides, the process comprising;  
i) obtaining collagen-like silk polypeptides, and  
ii) processing the polypeptides to produce the product.

30 28. A method of treating or preventing a disease, the method comprising administering a composition comprising at least one drug for treating or preventing the disease and a pharmaceutically acceptable carrier, wherein the pharmaceutically acceptable carrier is selected from one or more of at least one polypeptide according to any one of claims 8 to 12, at least one silk fibre, sponge, film, hydrogel or particle of claim 16 or claim 17, at least one copolymer of claim 18 or claim 19, at least one product according to any one of claims 20 to 22 or at least one composition according to any one of claims 23 to 26.

29. Use of one or more of at least one polypeptide according to any one of claims 8 to 13, at least one silk fibre, sponge, film, hydrogel or particle of claim 16 or claim 17, at least one copolymer of claim 18 or claim 19, at least one product according to any one of claims 20 to 22 or at least one composition according to any one of claims 5 23 to 26, and at least one drug, for the manufacture of a medicament for treating or preventing a disease.
30. Use of one or more of at least one polypeptide according to any one of claims 8 to 12, at least one silk fibre, sponge, film, hydrogel or particle of claim 16 or claim 17, at 10 least one copolymer of claim 18 or claim 19, at least one product according to any one of claims 20 to 22 or at least one composition according to any one of claims 23 to 26, and at least one drug, as a medicament for treating or preventing a disease.
31. A kit comprising one or more of at least one polypeptide according to any one 15 of claims 8 to 12, at least one polynucleotide according to any one of claims 1 to 3, at least one vector of claim 4 or claim 5, at least one silk fibre, sponge, film, hydrogel or particle of claim 16 or claim 17, at least one copolymer of claim 18 or claim 19, at least one product according to any one of claims 20 to 22, or at least one composition according to any one of claims 23 to 26.

Figure 1



**Figure 2**

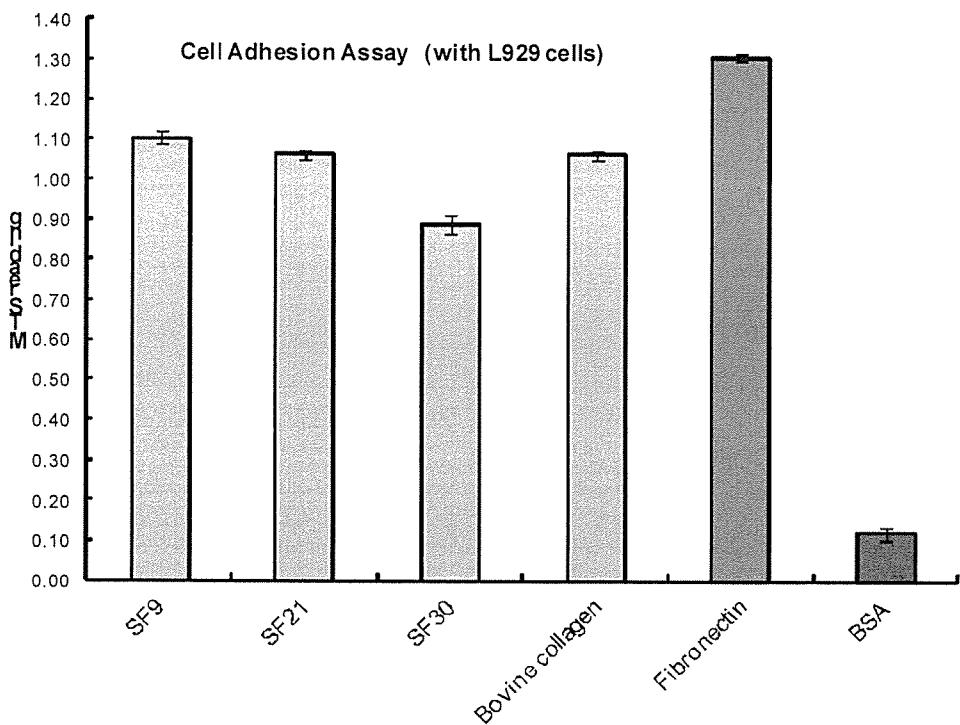


Figure 3

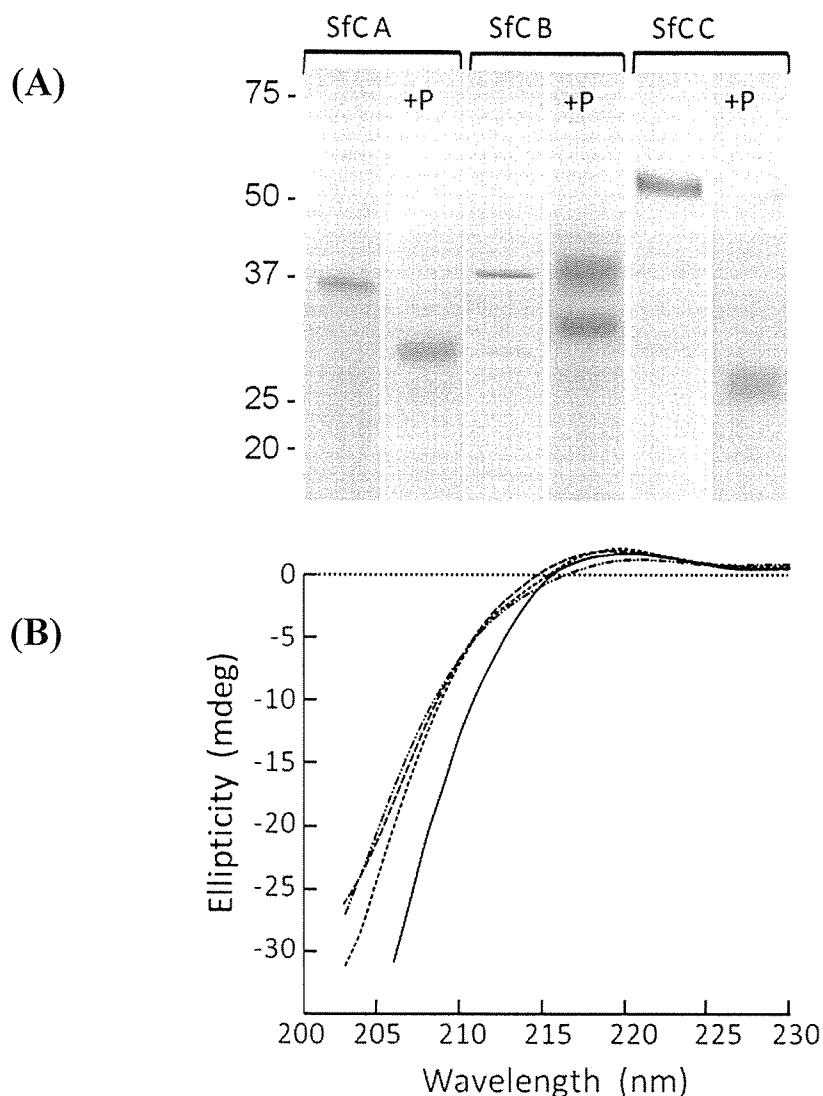


Figure 4

# INTERNATIONAL SEARCH REPORT

International application No.

**PCT/AU2012/001412**

## A. CLASSIFICATION OF SUBJECT MATTER

**C12N 15/63 (2006.01) C07K 14/435 (2006.01) A61K 47/42 (2006.01)**

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

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

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

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

WPI, Epodoc, Medline, hcaplus, embase, Agricola, biosis & Keywords: silk, fibroin, sericin, sericulture, collagen, spidroin, sawfly, nematus, hymenoptera, hemichroa, pristiphora, pachynemalus, pikonema, tomostethus, tethida, vector, construct, plasmid, transgenic and like words.

GenomeQuest: SEQ ID NOs: 1-12, 22 and 23

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
	Documents are listed in the continuation of Box C	

Further documents are listed in the continuation of Box C

See patent family annex

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

Date of the actual completion of the international search  
10 January 2013

Date of mailing of the international search report  
10 January 2013

## Name and mailing address of the ISA/AU

AUSTRALIAN PATENT OFFICE  
PO BOX 200, WODEN ACT 2606, AUSTRALIA  
Email address: pct@ipaustralia.gov.au  
Facsimile No.: +61 2 6283 7999

## Authorised officer

Esther Ng  
AUSTRALIAN PATENT OFFICE  
(ISO 9001 Quality Certified Service)  
Telephone No. +612 6283 3129

INTERNATIONAL SEARCH REPORT		International application No. <b>PCT/AU2012/001412</b>
C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2007/057207 A1 (TECHNISCHE UNIVERSITAET MUENCHEN) 24 May 2007 See Abstract; Pages 6, 7; Page 9, paragraph 2; Page 10, paragraph 2; pages 12-14; Page 17, paragraph 1; examples and SEQ ID NO: 8 comprises a sequence with >48% ID with present SEQ ID NO: 2 and >44% ID with present SEQ ID NO: 1; SEQ ID NO: 4 comprises a sequence with >40% ID with present SEQ ID NO: 4 and >38% with present SEQ ID NO: 3; SEQ ID NO: 10 comprises a sequence with >36% with present SEQ ID NO: 6 and >34% with present SEQ ID NO: 5	1, 2, 4-9 and 12-31
X	US 2009/0280071 A1 (KUMAR et al) 12 November 2009 See Paragraphs [0011]; [0015]-[0020]; [0033]; [0040] and SEQ ID NO: 29 shares >33% ID with present SEQ ID NO: 1-5	1, 2, 4-9 and 12-31
X	SPIRO, R. G. et al., 'Glycosylation of hydroxylysine in collagens', Nature New Biology, 1971, vol. 231, no. 19, pages 54-55 See page 54, col. 2	1-31
A	SEHNAL, F. and SUTHERLAND, T., 'Silks produced by insect labial glands', Prion, 2008, vol. 2, issue 4, pages 145-153 See whole document	
A	KENCHINGTON, W., 'Silk secretion in sawflies', Journal of Morphology, 1969, vol. 127, no. 3, pages 355-362 See whole document	

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International application No.

**PCT/AU2012/001412**

This Annex lists known patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

<b>Patent Document/s Cited in Search Report</b>		<b>Patent Family Member/s</b>	
<b>Publication Number</b>	<b>Publication Date</b>	<b>Publication Number</b>	<b>Publication Date</b>
WO 2007/057207 A1	24 May 2007	AU 2006314708 A1	24 May 2007
		AU 2006314708 B2	13 Sep 2012
		CA 2629821 A1	24 May 2007
		CN 101316862 A	03 Dec 2008
		EP 1787995 A1	23 May 2007
		EP 1948684 A1	30 Jul 2008
		JP 2009515540 A	16 Apr 2009
		KR 20080074134 A	12 Aug 2008
		RU 2008123706 A	27 Dec 2009
		US 2009162896 A1	25 Jun 2009
		WO 2007057207 A1	24 May 2007
US 2009/0280071 A1	12 Nov 2009	BR PI0408244 A	01 Mar 2006
		CA 2517849 A1	23 Sep 2004
		CN 1774260 A	17 May 2006
		EP 1610807 A2	04 Jan 2006
		KR 20050118282 A	16 Dec 2005
		MX PA05009666 A	20 Oct 2005
		US 2004180027 A1	16 Sep 2004
		US 7297678 B2	20 Nov 2007
		US 2009280071 A1	12 Nov 2009
		US 8048859 B2	01 Nov 2011
		US 2008226706 A1	18 Sep 2008
		US 8273704 B2	25 Sep 2012
		US 2005142094 A1	30 Jun 2005
		US 2012276021 A1	01 Nov 2012
		WO 2004080426 A2	23 Sep 2004
		WO 2005094868 A1	13 Oct 2005

**End of Annex**

## 摘要

本發明涉及可用於生產具有膠原蛋白樣結構的絲的絲蛋白、以及編碼這種蛋白的核酸。本發明還涉及合成絲蛋白的重組細胞和/或生物體。本發明的絲蛋白可用於各種目的，如在生產個人護理產品、塑料、紡織品和生物醫學產品中。