



(86) **Date de dépôt PCT/PCT Filing Date:** 2006/10/04
(87) **Date publication PCT/PCT Publication Date:** 2007/04/12
(45) **Date de délivrance/Issue Date:** 2018/01/09
(85) **Entrée phase nationale/National Entry:** 2008/04/03
(86) **N° demande PCT/PCT Application No.:** AU 2006/001453
(87) **N° publication PCT/PCT Publication No.:** 2007/038837
(30) **Priorité/Priority:** 2005/10/05 (US60/723,766)

(51) **Cl.Int./Int.Cl. C07K 14/435** (2006.01),
C07H 21/04 (2006.01), **C07K 19/00** (2006.01),
C12N 15/63 (2006.01)

(72) **Inventeurs/Inventors:**
SUTHERLAND, TARA D., AU;
HARITOS, VICTORIA S., AU;
TRUEMAN, HOLLY, AU;
SRISKANTHA, ALAGACONE, AU;
WEISMAN, SARAH, AU;
CAMPBELL, PETER M., AU

(73) **Propriétaire/Owner:**
COMMONWEALTH SCIENTIFIC AND INDUSTRIAL
RESEARCH ORGANISATION, AU

(74) **Agent:** SIM & MCBURNEY

(54) **Titre : PROTEINES DE SOIE CONTENANT UNE REGION BISPIRALEE**

(54) **Title: SILK PROTEINS CONTAINING COILED COIL REGION**

(57) **Abrégé/Abstract:**

A substantially purified and/or recombinant silk polypeptide is provided, wherein at least a portion of the polypeptide has a coiled coil structure and wherein the polypeptide comprises a sequence selected from: i) an amino acid sequence as provided in any one of SEQ ID NOs:1-8, 22-29, 40-47, 56-63, 72, or 73; ii) an amino acid sequence which is at least 40% identical to any one or more of SEQ ID NOs:1-8, 22-29, 40-47, 56-63, 72, or 73; and iii) a biologically active fragment of i) or ii).



ABSTRACT

A substantially purified and/or recombinant silk polypeptide is provided, wherein at least a portion of the polypeptide has a coiled coil structure and wherein the polypeptide comprises a sequence selected from:

i) an amino acid sequence as provided in any one of SEQ ID NOs:1-8, 22-29, 40-47, 56-63, 72, or 73;

ii) an amino acid sequence which is at least 40% identical to any one or more of SEQ ID NOs:1-8, 22-29, 40-47, 56-63, 72, or 73; and

iii) a biologically active fragment of i) or ii).

DEMANDES OU BREVETS VOLUMINEUX

**LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS
COMPREND PLUS D'UN TOME.**

CECI EST LE TOME __1__ DE __2__

NOTE: Pour les tomes additionels, veuillez contacter le Bureau Canadien des Brevets.

JUMBO APPLICATIONS / PATENTS

**THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE
THAN ONE VOLUME.**

THIS IS VOLUME __1__ OF __2__

NOTE: For additional volumes please contact the Canadian Patent Office.

SILK PROTEINS CONTAINING COILED COIL REGION

FIELD OF THE INVENTION

The present invention relates to silk proteins, as well as nucleic acids encoding
5 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

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. Very few of these silks have been
characterised, with most research concentrating on the cocoon silk of the domesticated
15 silkworm, *Bombyx mori* and on the dragline silk of the orb-weaving spider *Nephila*
clavipes.

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).
20 Generally these proteins comprise different combinations of crystalline arrays of β -
pleated sheets loosely associated with β -sheets, β -spirals, α -helices and amorphous
regions (see Craig and Riekel, 2002 for review).

As silk fibres represent some of the strongest natural fibres known, they have
been subject to extensive research in attempts to reproduce their synthesis. However, a
25 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 the repeating nucleotide motifs in the silk gene that lead to deleterious recombination
events, the large gene size and the small number of codons used for each amino acid in
the gene which leads to depletion of tRNA pools in the host cells. Recombinant
30 expression leads to difficulties during 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.

In contrast to the extensive knowledge that has accumulated about the
35 Lepidopteran (in particular the cocoon silk of *Bombyx mori*) and spider (in particular

the dragline silk of *Nephila clavipes*) little is known about the chemical composition and molecular organisation of other insect silks.

In the early 1960s, the silk of the aculeate Hymenopteran was shown to have an alpha-helical structure by X-ray diffraction patterns obtained from silk fibres drawn
5 from the salivary gland of honeybee larvae (Rudall, 1962). As well as demonstrating that this silk was helical, the patterns obtained were indicative of a coiled-coil system of alpha-helical chains (Atkins, 1967). Similar X-ray diffraction patterns have been obtained for cocoon silks from other Aculeata species including the wasp *Pseudopompilus humboldti* (Rudall, 1962) and the bumblebee, *Bombus lucorum* (Lucas
10 and Rudall, 1967).

In contrast to the alpha-helical structure described in the Aculeata silks, the silks characterised from a related clade to the aculeata, the Ichneumonoidea, have parallel- β structures. X-ray diagrams for four examples of this structure have been described in the Braconidae (*Cotesia*(=*Apanteles*) *glomerata*; *Cotesia*(=*Apanteles*) *gonopterygis*;
15 *Apanteles bignelli*) and three in Ichneumonidae (*Dusona* sp.; *Phytodietris* sp.; *Branchus femoralis*) (Lucas and Rudall, 1967). In addition the sequence of a single Braconidae (*Cotesia glomerata*) silk has been described (Genbank database accession number AB188680; Yamada et al., 2004). This partial protein sequence consists of a highly conserved 28 X-asparagine repeat (where X is alanine or serine) and is not
20 predicted to contain coiled coil forming heptad repeats. Extensive analysis of the amino acid composition of the cocoon silks of the Braconidae has shown that the silks from the subfamily Microgastrinae are unique in their high asparagine and serine content (Lucas et al., 1960; Quicke et al., 2004). Related subfamilies produce silks with significantly different amino acid compositions suggesting that the Microgastrinae silks
25 have evolved specifically in this subfamily (Yamada et al., 2004). The partial cDNA of *Cotesia glomerata* was isolated using PCR primers designed from sequence obtained from internal peptides derived from isolated cocoon silk proteins. The predicted amino acid composition of this partial sequence closely resembles the amino acid composition of the extensively washed silk from this species.

30 The structure of many of the silks within other non aculeate Apocrita and within the rest of the Hymenoptera (Symphata) are most commonly parallel- β sheets, with both collagen-like and polyglycine silks produced by the Tenthredinidae (Lucas and Rudall, 1967).

Honeybee silk proteins are synthesised in the middle of the final instar and can
35 be imaged as a mix of depolymerised silk proteins (Silva-Zacarin et al., 2003). As the instar progresses, water is removed from the gland and dehydration results in the

polymerisation of the silk protein to form well-organised and insoluble silk filaments labelled tactoids (Silva-Zacarin et al., 2003). Progressive dehydration leads to further reorganisation of the tactoids (Silva-Zacarin et al., 2003) and possibly new inter-filamentary bonding between filaments (Rudall, 1962). Electron microscope images of
5 fibrils isolated from the honeybee silk gland show structures of approximately 20-25 angstroms diameter (Flower and Kenchington, 1967). This value is consistent with three-, four-, or five-stranded coiled coils.

The amino acid composition of the silks of various aculeate Hymenopteran species was determined by Lucas and Rudall (1967) and found to contain high contents
10 of alanine, serine, the acid residues, aspartic acid and glutamic acids, and reduced amounts of glycine in comparison to classical fibroins. It was considered that the helical content of the aculeate Hymenoptera silk was a consequence of a reduced glycine content and increased content of acidic residues (Rudall and Kenchington, 1971).

15 Little is known about the larval silk of the lacewings (Order: Neuroptera). The cocoon is comprised of two layers, an inner solid layer and an outer fibrous layer. Previously the cocoon was described as being comprised of a cuticulin silk (Rudall and Kenchington, 1971), a description that only related to the inner solid layer. LaMunyon (1988) described a substance excreted from the malpighian tubules that made up the
20 outer fibres. After deposition of this layer, the solid inner wall was constructed from secretions from the epithelial cells in the highly villous lumen (LaMunyon, 1988).

It is also known that lacewing larva produce a proteinaceous adhesive substance from the malpighian tubules throughout all instars to stick the larvae to substrates, to glue items of camouflage on to the larvae's back or to entrap prey (Speilger, 1962). In
25 the genus *Lomamyia* (Bethothidae), the larvae produce the silk and adhesive substance at the same time and it has been postulated that these two substances may well be the same product (Speilger, 1962). The adhesive secretion is highly soluble and is also thought to be associated with defense against predators (LaMunyon & Adams, 1987).

Considering the unique properties of silks produced by insects such as
30 Hymenopterans and Neuropterans, there is a need for the identification of novel nucleic acids encoding silk proteins from these organisms.

SUMMARY OF THE INVENTION

The present inventors have identified numerous silk proteins from insects.
35 These silk proteins are surprisingly different to other known silk proteins in their primary sequence, secondary structure and/or amino acid content.

Thus, in a first aspect the present invention provides a substantially purified and/or recombinant silk polypeptide, wherein at least a portion of the polypeptide has a coiled coil structure.

As known in the art, coiled coil structures of polypeptides are characterized by heptad repeats represented by the consensus sequence $(abcdefg)_n$, with generally hydrophobic residues in position a and d , and generally polar residues at the remaining positions. Surprisingly, the heptads of the polypeptides of the present invention have a novel composition when viewed collectively - with an unusually high abundance of alanine in the 'hydrophobic' heptad positions a and d . Additionally, there are high levels of small polar residues in these positions. Furthermore, the e position also has high levels of alanine and small hydrophobic residues.

Accordingly, in a particularly preferred embodiment, the portion of the polypeptide that has a coiled coil structure comprises at least 10 copies of the heptad sequence $abcdefg$, and at least 25% of the amino acids at positions a and d are alanine residues.

In a further preferred embodiment, the portion of the polypeptide that has a coiled coil structure comprises at least 10 copies of the heptad sequence $abcdefg$, and at least 25% of the amino acids at positions a , d and e are alanine residues.

In a further preferred embodiment, the portion of the polypeptide that has a coiled coil structure comprises at least 10 copies of the heptad sequence $abcdefg$, and at least 25% of the amino acids at position a are alanine residues.

In a further preferred embodiment, the portion of the polypeptide that has a coiled coil structure comprises at least 10 copies of the heptad sequence $abcdefg$, and at least 25% of the amino acids at position d are alanine residues.

In a further preferred embodiment, the portion of the polypeptide that has a coiled coil structure comprises at least 10 copies of the heptad sequence $abcdefg$, and at least 25% of the amino acids at position e are alanine residues.

In a particularly preferred embodiment, the at least 10 copies of the heptad sequence are contiguous.

In a further preferred embodiment, the portion of the polypeptide that has a coiled coil structure comprises at least 5 copies of the heptad sequence $abcdefg$, and at least 15% of the amino acids at positions a and d are alanine residues.

In a further preferred embodiment, the portion of the polypeptide that has a coiled coil structure comprises at least 5 copies of the heptad sequence $abcdefg$, and at least 15% of the amino acids at positions a , d and e are alanine residues.

In a further preferred embodiment, the portion of the polypeptide that has a coiled coil structure comprises at least 5 copies of the heptad sequence *abcdefg*, and at least 15% of the amino acids at position *a* are alanine residues.

In a further preferred embodiment, the portion of the polypeptide that has a coiled coil structure comprises at least 5 copies of the heptad sequence *abcdefg*, and at least 15% of the amino acids at position *d* are alanine residues.

In a further preferred embodiment, the portion of the polypeptide that has a coiled coil structure comprises at least 5 copies of the heptad sequence *abcdefg*, and at least 15% of the amino acids at position *e* are alanine residues.

10 In a particularly preferred embodiment, the at least 5 copies of the heptad sequence are contiguous.

In one embodiment, the polypeptide comprises a sequence selected from:

i) an amino acid sequence as provided in any one of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:56, and SEQ ID NO:57;

ii) an amino acid sequence which is at least 30% identical to any one or more of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:56, and SEQ ID NO:57; and

iii) a biologically active fragment of i) or ii).

20 In another embodiment, the polypeptide comprises a sequence selected from:

i) an amino acid sequence as provided in any one of SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:58, and SEQ ID NO:59;

ii) an amino acid sequence which is at least 30% identical to any one or more of SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:58, and SEQ ID NO:59; and

iii) a biologically active fragment of i) or ii).

In another embodiment, the polypeptide comprises a sequence selected from:

i) an amino acid sequence as provided in any one of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:60, and SEQ ID NO:61;

ii) an amino acid sequence which is at least 30% identical to any one or more of SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:60, and SEQ ID NO:61; and

35 iii) a biologically active fragment of i) or ii).

In another embodiment, the polypeptide comprises a sequence selected from:

i) an amino acid sequence as provided in any one of SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:62, and SEQ ID NO:63;

ii) an amino acid sequence which is at least 30% identical to any one or more of
5 SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:62, and SEQ ID NO:63; and

iii) a biologically active fragment of i) or ii).

In a further embodiment, the polypeptide comprises a sequence selected from:

i) an amino acid sequence as provided in SEQ ID NO:72 or SEQ ID NO:73;

10 ii) an amino acid sequence which is at least 30% identical to SEQ ID NO:72 and/or SEQ ID NO:73; and

iii) a biologically active fragment of i) or ii).

Further silk proteins which co-associate with proteins of the first aspect have been identified. One of these proteins (SEQ ID NO:10) is predicted to have 41%
15 alpha-helical, 8% beta-sheet and 50% loop secondary structure by PROFsec, and therefore is classified as a mixed structure protein. MARCOIL analysis of this protein predicted only a short region of heptad repeats characteristic of proteins with a coiled coil structure.

Accordingly, in a second aspect, the present invention provides a substantially
20 purified and/or recombinant silk polypeptide which comprises a sequence selected from:

i) an amino acid sequence as provided in any one of SEQ ID NO:9, SEQ ID NO:10 and SEQ ID NO:30;

ii) an amino acid sequence which is at least 30% identical to any one or more of
25 SEQ ID NO:9, SEQ ID NO:10 and SEQ ID NO:30; and

iii) a biologically active fragment of i) or ii).

Without wishing to be limited by theory, it appears that four proteins of the first aspect become intertwined to form a bundle with helical axes almost parallel to each other, and this bundle extends axially into a fibril. Furthermore, it is predicted that in at
30 least some species such as the honeybee and bumblebee the proteins of the second aspect act as a "glue" assisting in binding various bundles of coiled coil proteins of the first aspect together to form a fibrous protein complex. However, silk fibers and copolymers can still be formed without a polypeptide of second aspect.

In a preferred embodiment, a polypeptide of the invention can be purified from,
35 or is a mutant of a polypeptide purified from, a species of Hymenoptera or Neuroptera. Preferably, the species of Hymenoptera is *Apis mellifera*, *Oecophylla smaragdina*,

Myrmecia foricata or *Bombus terrestris*. Preferably, the species of Neuroptera is *Mallada signata*.

In another aspect, the present invention provides a polypeptide of the invention fused to at least one other polypeptide.

5 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, and a polypeptide which assists in the polypeptide of the invention being secreted from a cell (for example secreted from a plant cell).

10 In another aspect, the present invention provides an isolated and/or exogenous polynucleotide which encodes a silk polypeptide, wherein at least a portion of the polypeptide has a coiled coil structure.

In one embodiment, the polynucleotide comprises a sequence selected from:

i) a sequence of nucleotides as provided in any one of SEQ ID NO:11, SEQ ID
15 NO:12, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:64, and SEQ ID NO:65;

ii) a sequence of nucleotides encoding a polypeptide of the invention,

iii) a sequence of nucleotides which is at least 30% identical to any one or more
of SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:48,
20 SEQ ID NO:49, SEQ ID NO:64, and SEQ ID NO:65, and

iv) a sequence which hybridizes to any one of i) to iii) under stringent conditions.

In another embodiment, the polynucleotide comprises a sequence selected from:

i) a sequence of nucleotides as provided in any one of SEQ ID NO:13, SEQ ID
25 NO:14, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:66, and SEQ ID NO:67;

ii) a sequence of nucleotides encoding a polypeptide of the invention,

iii) a sequence of nucleotides which is at least 30% identical to any one or more
of SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:50,
30 SEQ ID NO:51, SEQ ID NO:66, and SEQ ID NO:67, and

iv) a sequence which hybridizes to any one of i) to iii) under stringent conditions.

In another embodiment, the polynucleotide comprises a sequence selected from:

i) a sequence of nucleotides as provided in any one of SEQ ID NO:15, SEQ ID
35 NO:16, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:68, and SEQ ID NO:69;

ii) a sequence of nucleotides encoding a polypeptide of the invention,
iii) a sequence of nucleotides which is at least 30% identical to any one or more of SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:68, and SEQ ID NO:69, and

5 iv) a sequence which hybridizes to any one of i) to iii) under stringent conditions.

In a further embodiment, the polynucleotide comprises a sequence selected from:

i) a sequence of nucleotides as provided in any one of SEQ ID NO:17, SEQ ID
10 NO:18, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:70, SEQ ID NO:71 and SEQ ID NO:76;

ii) a sequence of nucleotides encoding a polypeptide of the invention,

iii) a sequence of nucleotides which is at least 30% identical to any one or more of SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:54,
15 SEQ ID NO:55, SEQ ID NO:70, SEQ ID NO:71 and SEQ ID NO:76, and

iv) a sequence which hybridizes to any one of i) to iii) under stringent conditions.

In another embodiment, the polynucleotide comprises a sequence selected from:

i) a sequence of nucleotides as provided in SEQ ID NO:74 or SEQ ID NO:75;

20 ii) a sequence of nucleotides encoding a polypeptide of the invention,

iii) a sequence of nucleotides which is at least 30% identical to SEQ ID NO:74 and/or SEQ ID NO:75, and

iv) a sequence which hybridizes to any one of i) to iii) under stringent conditions.

25 In a further aspect, the present invention provides an isolated and/or exogenous polynucleotide, the polynucleotide comprising a sequence selected from:

i) a sequence of nucleotides as provided in any one of SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, and SEQ ID NO:39;

ii) a sequence of nucleotides encoding a polypeptide of the invention,

30 iii) a sequence of nucleotides which is at least 30% identical to any one or more of SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, and SEQ ID NO:39, and

iv) a sequence which hybridizes to any one of i) to iii) under stringent conditions.

In a preferred embodiment, a polynucleotide can be isolated from, or is a mutant
35 of a polynucleotide isolated from, a species of Hymenoptera or Neuroptera. Preferably, the species of Hymenoptera is *Apis mellifera*, *Oecophylla smaragdina*, *Myrmecia*

foricata or *Bombus terrestris*. Preferably, the species of Neuroptera is *Mallada signata*.

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

5 Preferably, the vector is an expression vector.

In another 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.

The host cell can be any type of cell. Examples include, but are not limited to, a bacterial, yeast or plant cell.

10 Also provided is a process for preparing a polypeptide according to the invention, the process comprising cultivating a host cell of the invention, or a vector of the invention, under conditions which allow expression of the polynucleotide encoding the polypeptide, and recovering the expressed polypeptide.

It is envisaged that transgenic plants will be particularly useful for the
15 production of polypeptides of the invention. Thus, in yet another aspect, the present provides a transgenic plant comprising an exogenous polynucleotide, the polynucleotide encoding at least one polypeptide of the invention.

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

In yet another aspect, the present invention provides an antibody which specifically binds a polypeptide of the invention.

In a further aspect, the present invention provides a silk fiber comprising at least one polypeptide of the invention.

25 Preferably, the polypeptide is a recombinant polypeptide.

In an embodiment, at least some of the polypeptides are crosslinked. In an embodiment, at least some of the lysine residues of the polypeptides are crosslinked.

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

30 Preferably, the polypeptides are recombinant polypeptides.

In an embodiment, the copolymer comprises at least four different polypeptide of the first aspect. In another embodiment, the copolymer further comprises a polypeptide of the second aspect.

In an embodiment, at least some of the polypeptides are crosslinked. In an
35 embodiment, at least some of the lysine residues of the polypeptides are crosslinked.

As the skilled addressee will appreciate, the polypeptides of the invention have a wide variety of uses as is known in the art for other types of silk proteins. Thus, in a further aspect, the present invention provides a product comprising at least one polypeptide of the invention, a silk fiber of the invention and/or a copolymer of the
5 invention.

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

In yet a further aspect, the present invention provides a composition comprising at least one polypeptide of the invention, a silk fiber of the invention and/or a
10 copolymer of the invention, and one or more acceptable carriers.

In one embodiment, the composition further comprises a drug.

In another embodiment, the composition is used as a medicine, in a medical device or a cosmetic.

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

In a preferred embodiment, a composition, silk fiber, copolymer and/or product of the invention does not comprise a royal jelly protein produced by an insect.

In a further aspect, the present invention provides a method of treating or preventing a disease, the method comprising administering a composition comprising a
20 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, a silk fiber of the invention and/or a copolymer of the invention.

In yet another aspect, the present invention provides for the use of at least one
25 polypeptide of the invention, a silk fiber of the invention and/or a copolymer of the invention, and a drug, for the manufacture of a medicament for treating or preventing a disease.

In a further aspect, the present invention provides a kit comprising at least one polypeptide of the invention, at least one polynucleotide of the invention, at least one
30 vector of the invention, at least one silk fiber of the invention and/or a copolymer of the invention.

Preferably, the kit further comprises information and/or instructions for use of the kit.

According to another aspect, there is provided a substantially purified and/or recombinant silk polypeptide, wherein at least a portion of the polypeptide has a coiled coil structure and wherein the polypeptide comprises a sequence selected from:

i) an amino acid sequence as provided in any one of SEQ ID NO:1, SEQ ID
 5 NO:2, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:40, SEQ ID NO:41, SEQ ID
 NO:56, SEQ ID NO:57; SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:24, SEQ ID
 NO:25, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:58, SEQ ID NO:59, SEQ ID
 NO:5, SEQ ID NO:6, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:44, SEQ ID
 NO:45, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:7, SEQ ID NO:8, SEQ ID
 10 NO:28, SEQ ID NO:29, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:62, SEQ ID
 NO:63, SEQ ID NO:72 or SEQ ID NO:73; or

ii) an amino acid sequence which is at least 40% identical to any one or more
 of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:40,
 SEQ ID NO:41, SEQ ID NO:56, SEQ ID NO:57; SEQ ID NO:3, SEQ ID NO:4, SEQ
 15 ID NO:24, SEQ ID NO:25, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:58, SEQ ID
 NO:59, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:26, SEQ ID NO:27, SEQ ID
 NO:44, SEQ ID NO:45, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:7, SEQ ID
 NO:8, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:46, SEQ ID NO:47, SEQ ID
 NO:62, SEQ ID NO:63, SEQ ID NO:72 or SEQ ID NO:73.

20 According to another aspect, there is provided an isolated and/or exogenous
 polynucleotide which encodes a silk polypeptide, wherein at least a portion of the
 polypeptide has a coiled coil structure, wherein the polypeptide comprises a sequence
 selected from:

i) a sequence of nucleotides as provided in any one of SEQ ID NO:11, SEQ ID
 25 NO:12, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:48, SEQ ID NO:49, SEQ ID
 NO:64, SEQ ID NO:65, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:33, SEQ ID
 NO:34, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:66, SEQ ID NO:67, SEQ ID
 NO:15, SEQ ID NO:16, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:52, SEQ ID
 NO:53, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:17, SEQ ID NO:18, SEQ ID
 30 NO:37, SEQ ID NO:38, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:70, SEQ ID
 NO:71, SEQ ID NO:76, SEQ ID NO:74 or SEQ ID NO:75;

ii) a sequence of nucleotides encoding the polypeptide of the invention; or

iii) a sequence of nucleotides which is at least 40% identical to any one or more of SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:76, SEQ ID NO:74 or SEQ ID NO:75.

10 According to another aspect, there is provided a non-human animal cell comprising an exogenous polynucleotide, the polynucleotide encoding at least one polypeptide of the invention.

According to another aspect, there is provided an antibody which specifically binds a purified and/or recombinant silk polypeptide, wherein at least a portion of the polypeptide has a coiled coil structure and wherein the polypeptide consists of a sequence selected from an amino acid sequence as provided in any one of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:56, SEQ ID NO:57; SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:72 or SEQ ID NO:73.

25 According to another aspect, there is provided a composition comprising at least one isolated and/or exogenous polynucleotide of the invention, and one or more acceptable carriers.

According to another aspect, there is provided a use of at least one polypeptide of the invention, the silk fiber of the invention and/or the copolymer of the invention, as a pharmaceutically acceptable carrier.

30 As will be apparent, preferred features and characteristics of one aspect of the invention are applicable to many other aspects of the invention.

Throughout this specification the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps.

5 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. Fourier transform infrared spectra of the amide I and II regions of the silks:
10 1) honeybee silk, 2) bumblebee silk, 3) bulldog ant silk, 4) weaver ant silk 5) lacewing larval silk. All the silks have spectra expected of helical proteins. The Hymenopteran silks (ants and bees) have spectral maxima at $1645\text{-}1646\text{ cm}^{-1}$ (labelled), shifted approximately 10 cm^{-1} lower than a classical alpha-helical signal and broadened, as is typical of coiled-coil proteins (Heimburg et al., 1999).

15

Figure 2. Comparison of amino acid composition of SDS washed honeybee brood comb silk with amino acid composition of Xenospira proteins (namely, Xenospira1, Xenospira2, Xenospira3 and Xenospira4) (equimolar amounts totalling 65%) and Xenosin (35%).

20

Figure 3. Comparison of amino acid composition of silk with amino acid composition predicted from proteins encoded by silk genes.

Figure 4. Prediction of coiled coil regions in honeybee silk proteins. COILS is a
25 program that compares a sequence to a database of known parallel two-stranded coiled-coils and derives a similarity score. By comparing this score to the distribution of scores in globular and coiled-coil proteins, the program then calculates the probability that the sequence will adopt a coiled-coil conformation as described in Lupas et al. (1991). Using a window size of 28 this program predicts the following numbers of
30 residues exist in each protein in coiled coil domains: Xenospira3: 77; Xenospira4: 35; Xenospira1: 28; Xenospira2: 80.

Figure 5. Alignment of honey bee silk proteins showing MARCOIL prediction of major heptads that form a coiled-coil structure. Heptad sequences are shown above the
35 amino acids, and alanine residues in positions *a* and *d* are highlighted.

Figure 6. Alignment of Marciol predicted coiled coil regions of hymenopteran (bees and ants) silk proteins showing the heptad position assignment. Amel, honeybee; BB, bumblebee; BA, bulldog ant; WA, weaver ant; F1-4, silk fibroins 1-4. Heptad sequences are shown above the amino acids, and alanine residues in positions *a*, *d* and *e* are highlighted.

Figure 7. The amino acid character of heptad positions in the predicted coiled coil regions of the *Mallada signata* larval silk protein and the orthologous clusters of the Hymenopteran silk proteins.

10

Figure 8. SDS polyacrylamide gel electrophoresis of late last instar salivary gland proteins. Proteins were identified after tryptic digest and analysis of mass spectral data set using Agilent's Spectrum Mill software to match the data with predictions of protein sequences from proteins identified from cDNA sequences. The software generated scores for the quality of each match between experimentally observed sets of masses of fragments of peptides and the predictions of fragments that might be generated according to the sequences of proteins in a provided database. All the sequence matches shown here received scores greater than 20 by the Spectrum Mill software, where a score of 20 would be sufficient for automatic, confident acceptance of a valid match.

Figure 9. Parsimony analysis of the coiled coil region of silk proteins. The relatedness of the four coiled-coil proteins suggests that the genes evolved from a common ancestor predating the divergence of the Euaculeata. The area bound by the dashed line indicates variation that occurred before the ants and wasps (Vespoidea) diverged from the bees (Apoidea) in the Late Jurassic (155 myrs; Grimaldi and Engel, 2005). Numbers indicating bootstrap values from 1000 iterations are shown.

Figure 10. A) *Apis mellifera* silk proteins identified by mass spectral analysis of peptides generated from bee silk after digestion with trypsin. Shading indicates peptides identified by the mass spectral analysis. All the sequence matches shown here received scores greater than 20 by the Spectrum Mill software, where a score of 20 would be sufficient for automatic, confident acceptance of a valid match.

B) Full length amino sequences of bumblebee, bulldog ant, weaver and lacewing silk proteins.

35

Figure 11. Open reading frames encoding honeybee, bumblebee, bulldog ant, weaver ant and lacewing silk proteins.

Figure 12. Sequence of gene encoding Xenosin. Entire coding sequence provided
5 which is interrupted by a single intron (highlighted).

Figure 13. Expression of silk protein in tobacco. Detection of histidine tagged proteins after western blot analysis of proteins from: 1. *E. coli* transformed with empty expression vector, 2. *E. coli* transformed with expression vector containing AmelF4
10 (Xenospira4) coding region, 3. tobacco transformed with empty expression vector, 4. tobacco transformed with expression vector containing AmelF4 coding region.

Figure 14. Fibres made from recombinant honeybee silk proteins showing birefringent threads. Biorefringence indicates structure is present in the threads. Different
15 recombinant honeybee threads are shown in each panel A-D, and recombinant lacewing thread is shown in panel E.

KEY TO THE SEQUENCE LISTING

- SEQ ID NO:1 - Honeybee silk protein termed herein Xenospira1 (also termed herein
20 AmelF1) (minus signal peptide).
SEQ ID NO:2 - Honeybee silk protein termed herein Xenospira1.
SEQ ID NO:3 - Honeybee silk protein termed herein Xenospira2 (also termed herein AmelF2) (minus signal peptide).
SEQ ID NO:4 - Honeybee silk protein termed herein Xenospira2.
25 SEQ ID NO:5 - Honeybee silk protein termed herein Xenospira3 (also termed herein AmelF3) (minus signal peptide).
SEQ ID NO:6 - Honeybee silk protein termed herein Xenospira3.
SEQ ID NO:7 - Honeybee silk protein termed herein Xenospira4 (also termed herein AmelF4) (minus signal peptide).
30 SEQ ID NO:8 - Honeybee silk protein termed herein Xenospira4.
SEQ ID NO:9 - Honeybee silk protein termed herein Xenosin (also termed herein AmelSA1) (minus signal peptide).
SEQ ID NO:10 - Honeybee silk protein termed herein Xenosin.
SEQ ID NO:11 - Nucleotide sequence encoding honeybee silk protein Xenospira1
35 (minus region encoding signal peptide).
SEQ ID NO:12 - Nucleotide sequence encoding honeybee silk protein Xenospira1.

- SEQ ID NO:13 - Nucleotide sequence encoding honeybee silk protein Xenospira2 (minus region encoding signal peptide).
- SEQ ID NO:14 - Nucleotide sequence encoding honeybee silk protein Xenospira2.
- SEQ ID NO:15 - Nucleotide sequence encoding honeybee silk protein Xenospira3
5 (minus region encoding signal peptide).
- SEQ ID NO:16 - Nucleotide sequence encoding honeybee silk protein Xenospira3.
- SEQ ID NO:17 - Nucleotide sequence encoding honeybee silk protein Xenospira4 (minus region encoding signal peptide).
- SEQ ID NO:18 - Nucleotide sequence encoding honeybee silk protein Xenospira4.
- 10 SEQ ID NO:19 - Nucleotide sequence encoding honeybee silk protein Xenosin (minus region encoding signal peptide).
- SEQ ID NO:20 - Nucleotide sequence encoding honeybee silk protein Xenosin.
- SEQ ID NO:21 - Gene sequence encoding honeybee silk protein Xenosin.
- SEQ ID NO:22 - Bumblebee silk protein termed herein BBF1 (minus signal peptide).
- 15 SEQ ID NO:23 - Bumblebee silk protein termed herein BBF1.
- SEQ ID NO:24 - Bumblebee silk protein termed herein BBF2 (minus signal peptide).
- SEQ ID NO:25 - Bumblebee silk protein termed herein BBF2.
- SEQ ID NO:26 - Bumblebee silk protein termed herein BBF3 (minus signal peptide).
- SEQ ID NO:27 - Bumblebee silk protein termed herein BBF3.
- 20 SEQ ID NO:28 - Bumblebee silk protein termed herein BBF4 (minus signal peptide).
- SEQ ID NO:29 - Bumblebee silk protein termed herein BBF4.
- SEQ ID NO:30 - Partial amino acid sequence of bumblebee silk protein termed herein BBSA1.
- SEQ ID NO:31 - Nucleotide sequence encoding bumblebee silk protein BBF1 (minus
25 region encoding signal peptide).
- SEQ ID NO:32 - Nucleotide sequence encoding bumblebee silk protein BBF1.
- SEQ ID NO:33 - Nucleotide sequence encoding bumblebee silk protein BBF2 (minus region encoding signal peptide).
- SEQ ID NO:34 - Nucleotide sequence encoding bumblebee silk protein BBF2.
- 30 SEQ ID NO:35 - Nucleotide sequence encoding bumblebee silk protein BBF3 (minus region encoding signal peptide).
- SEQ ID NO:36 - Nucleotide sequence encoding bumblebee silk protein BBF3.
- SEQ ID NO:37 - Nucleotide sequence encoding bumblebee silk protein BBF4 (minus region encoding signal peptide).
- 35 SEQ ID NO:38 - Nucleotide sequence encoding bumblebee silk protein BBF4.
- SEQ ID NO:39 - Partial nucleotide sequence encoding bumblebee silk protein BBSA1.

- SEQ ID NO:40 - Bulldog ant silk protein termed herein BAF1 (minus signal peptide).
SEQ ID NO:41 - Bulldog ant silk protein termed herein BAF1.
SEQ ID NO:42 - Bulldog ant silk protein termed herein BAF2 (minus signal peptide).
SEQ ID NO:43 - Bulldog ant silk protein termed herein BAF2.
5 SEQ ID NO:44 - Bulldog ant silk protein termed herein BAF3 (minus signal peptide).
SEQ ID NO:45 - Bulldog ant silk protein termed herein BAF3.
SEQ ID NO:46 - Bulldog ant silk protein termed herein BAF4 (minus signal peptide).
SEQ ID NO:47 - Bulldog ant silk protein termed herein BAF4.
SEQ ID NO:48 - Nucleotide sequence encoding bulldog ant silk protein BAF1 (minus
10 region encoding signal peptide).
SEQ ID NO:49 - Nucleotide sequence encoding bulldog ant silk protein BAF1.
SEQ ID NO:50 - Nucleotide sequence encoding bulldog ant silk protein BAF2 (minus
region encoding signal peptide).
SEQ ID NO:51 - Nucleotide sequence encoding bulldog ant silk protein BAF2.
15 SEQ ID NO:52 - Nucleotide sequence encoding bulldog ant silk protein BAF3 (minus
region encoding signal peptide).
SEQ ID NO:53 - Nucleotide sequence encoding bulldog ant silk protein BAF3.
SEQ ID NO:54 - Nucleotide sequence encoding bulldog ant silk protein BAF4 (minus
region encoding signal peptide).
20 SEQ ID NO:55 - Nucleotide sequence encoding bulldog ant silk protein BAF4.
SEQ ID NO:56 - Weaver ant silk protein termed herein GAF1 (minus signal peptide).
SEQ ID NO:57 - Weaver ant silk protein termed herein GAF1.
SEQ ID NO:58 - Weaver ant silk protein termed herein GAF2 (minus signal peptide).
SEQ ID NO:59 - Weaver ant silk protein termed herein GAF2.
25 SEQ ID NO:60 - Weaver ant silk protein termed herein GAF3 (minus signal peptide).
SEQ ID NO:61 - Weaver ant silk protein termed herein GAF3.
SEQ ID NO:62 - Weaver ant silk protein termed herein GAF4 (minus signal peptide).
SEQ ID NO:63 - Weaver ant silk protein termed herein GAF4.
SEQ ID NO:64 - Nucleotide sequence encoding weaver ant silk protein GAF1 (minus
30 region encoding signal peptide).
SEQ ID NO:65 - Nucleotide sequence encoding weaver ant silk protein GAF1.
SEQ ID NO:66 - Nucleotide sequence encoding weaver ant silk protein GAF2 (minus
region encoding signal peptide).
SEQ ID NO:67 - Nucleotide sequence encoding weaver ant silk protein GAF2.
35 SEQ ID NO:68 - Nucleotide sequence encoding weaver ant silk protein GAF3 (minus
region encoding signal peptide).

SEQ ID NO:69 - Nucleotide sequence encoding weaver ant silk protein GAF3.

SEQ ID NO:70 - Nucleotide sequence encoding weaver ant silk protein GAF4 (minus region encoding signal peptide).

SEQ ID NO:71 - Nucleotide sequence encoding weaver ant silk protein GAF4.

5 SEQ ID NO:72 - Lacewing silk protein termed herein MalF1 (minus signal peptide).

SEQ ID NO:73 - Lacewing silk protein termed herein MalF1.

SEQ ID NO:74 - Nucleotide sequence encoding lacewing silk protein MalF1 (minus region encoding signal peptide).

SEQ ID NO:75 - Nucleotide sequence encoding lacewing silk protein MalF1.

10 SEQ ID NO:76 - Nucleotide sequence encoding honeybee silk protein termed herein Xenospira4 codon-optimized for plant expression (before subcloning into pET14b and pVEC8).

SEQ ID NO:77 - Honeybee silk protein (Xenospira4) open reading frame optimized for plant expression (without translational fusion).

15

DETAILED DESCRIPTION OF THE INVENTION

General Techniques and Definitions

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

Unless otherwise indicated, the recombinant protein, cell culture, and immunological techniques utilized in the present invention are standard procedures, well known to those skilled in the art. Such techniques are described and explained
25 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 Press (1991), D.M. Glover and B.D. Hames (editors), DNA Cloning: A Practical
30 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), and J.E. Coligan et al. (editors) Current Protocols in Immunology, John Wiley & Sons
35 (including all updates until present).

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. Naturally occurring silk proteins of the invention form part of the brood comb silk of insects such as honeybees, however, as described herein variants of these proteins could readily be produced which would perform the same function if expressed within an appropriate insect.

As used herein, a "silk fibre" refers to filaments comprising proteins of the invention which can be woven into various items such as textiles.

As used herein, a "copolymer" is composition comprising two or more silk proteins of the invention. This term excludes naturally occurring copolymers such as the brood comb of insects.

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 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 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 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 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 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
5 in the mature protein. Signal peptides have the function of directing and *trans*-locating secreted proteins across cell membranes. The signal peptide is also referred to as signal sequence.

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

10 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 a silk protein of the invention include, but are not limited to, proteins, nucleic acids, anti-tumor agents, analgesics, antibiotics, anti-inflammatory compounds (both steroidal and non-steroidal), hormones, vaccines, labeled substances, and the like.

15

Polypeptides

By "substantially 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.
20 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.

The term "recombinant" in the context of a polypeptide refers to the polypeptide
25 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 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
30 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 which are subsequently purified away from at least some other components.

The terms "polypeptide" and "protein" are generally used interchangeably and
35 refer to a single polypeptide chain which may or may not be modified by addition of non-amino acid groups. The terms "proteins" and "polypeptides" as used herein also

include variants, mutants, modifications, analogous and/or 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 15 amino acids in length, and the GAP analysis aligns the two sequences over a region of at least 15 amino acids. More preferably, 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. 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 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.

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 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%, 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 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 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* 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 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 or Neuroptean 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 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 15 residues, more preferably about 1 to 10 residues and typically about 1 to 5 contiguous residues.

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 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".

As outlined above, a portion of some of the polypeptides of the invention have a coiled coil structure. Coiled coil structures of polypeptides are characterized by heptad repeats represented by the consensus sequence $(abcdefg)_n$. In a preferred embodiment, the portion of the polypeptide that has a coiled coil structure comprises at least 10

copies of the heptad sequence *abcdefg*, and at least 25% of the amino acids at positions *a* and *d* are alanine residues.

Table 1. Exemplary substitutions

Original Residue	Exemplary Substitutions
Ala (A)	val; leu; ile; gly; cys; 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)	gly
Ser (S)	thr; ala; gly; val; gln
Thr (T)	ser; gln; ala
Trp (W)	tyr
Tyr (Y)	trp; phe
Val (V)	ile; leu; met; phe; ala; ser; thr

5

In a preferred embodiment, the polypeptide that has a coiled coil structure comprises at least 12 consecutive copies, more preferably at least 15 consecutive copies, and even more preferably at least 18 consecutive copies of the heptad. In further embodiments, the polypeptide that has a coiled coil structure can have up to at least 28 copies of the heptad. Typically, the copies of the heptad will be tandemly repeated. However, they do not necessarily have to be perfect tandem repeats, for

10

example, as shown in Figures 5 and 6 a few amino acids may be found between two heptads, or a few truncated heptads may be found (see, for example, Xenospiral in Figure 5).

Guidance regarding amino acid substitutions which can be made to the polypeptides of the invention which have a coiled coil structure is provided in Figures 5 and 6, as well as Tables 6 to 10. Where a predicted useful amino acid substitution based on the experimental data provided herein is in anyway in conflict with the exemplary substitutions provided in Table 1 it is preferred that a substitution based on the experimental data is used.

Coiled coil structures of polypeptides of the invention have a high content of alanine residues, particularly at amino acid positions *a*, *d* and *e* of the heptad. However, positions *b*, *c*, *f* and *g* also have a high frequency of alanine residues. In a preferred embodiment, at least 15% of the amino acids at positions *a*, *d* and/or *e* of the heptads are alanine residues, more preferably at least 25%, more preferably at least 30%, more preferably at least 40%, and even more preferably at least 50%. In a further preferred embodiment, at least 25% of the amino acids at both positions *a* and *d* of the heptads are alanine residues, more preferably at least 30%, more preferably at least 40%, and even more preferably at least 50%. Furthermore, it is preferred that at least 15% of the amino acids at positions *b*, *c*, *f* and *g* of the heptads are alanine residues, more preferably at least 20%, and even more preferably at least 25%.

Typically, the heptads will not comprise any proline or histidine residues. Furthermore, the heptads will comprise few (1 or 2), if any, phenylalanine, methionine, tyrosine, cysteine, glycine or tryptophan residues. Apart from alanine, common (for example greater than 5%, more preferably greater than 10%) amino acids in the heptads include leucine (particularly at positions *b* and *d*), serine (particularly at positions *b*, *e* and *f*), glutamic acid (particularly at positions *c*, *e* and *f*), lysine (particularly at positions *b*, *c*, *d*, *f* and *g*) as well as arginine at position *g*.

Polypeptides (and polynucleotides) of the invention can be purified (isolated) from a wide variety of Hymenopteran and Neuropteran species. Examples of Hymenopterans include, but are not limited to, any species of the Suborder Apocrita (bees, ants and wasps), which include the following Families of insects; Chrysididae (cuckoo wasps), Formicidae (ants), Mutillidae (velvet ants), Pompilidae (spider wasps), Scoliidae, Vespidae (paper wasps, potter wasps, hornets), Agaonidae (fig wasps), Chalcididae (chalcidids), Eucharitidae (eucharitids), Eupelmidae (eupelmids), Pteromalidae (pteromalids), Evaniidae (ensign wasps), Braconidae, Ichneumonidae (ichneumons), Megachilidae, Apidae, Colletidae, Halictidae, and Melittidae (oil

collecting bees). Examples of Neuropterans include species from the following insect Families: Mantispidae, Chrysopidae (lacewings), Myrmeleontidae (antlions), and Ascalaphidae (owlflies). Such further polypeptides (and polynucleotides) can be characterized using the same procedures described herein for silks from *Bombus*
5 *terrestris*, *Myrmecia forficata*, *Oecophylla smaragdina* and *Mallada signata*.

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, α -amino isobutyric acid, 4-
10 aminobutyric acid, 2-aminobutyric acid, 6-amino hexanoic acid, 2-amino isobutyric acid, 3-amino propionic acid, ornithine, norleucine, norvaline, hydroxyproline, sarcosine, citrulline, homocitrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine, β -alanine, fluoro-amino acids, designer amino acids such as β -methyl amino acids, C α -methyl amino acids, N α -methyl amino acids, and
15 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 biotinylation, benzylation, glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an
20 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, including production and recovery of natural polypeptides, production and recovery of recombinant polypeptides, and chemical synthesis of the polypeptides. In one
25 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 recombinant cell of the present invention. Effective culture conditions include, but are not limited to, effective media, bioreactor, temperature, pH and oxygen conditions that
30 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 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
35 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 ordinary skill in the art.

Polynucleotides

5 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 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
10 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 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
15 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
20 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
25 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
30 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 % identity
35 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%, 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 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.

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 (3.5xSSC, 0.02% Ficoll, 0.02% polyvinyl pyrrolidone, 0.02% Bovine Serum Albumin (BSA), 2.5 mM NaH₂PO₄ (pH7), 0.5% SDS, 2 mM EDTA), followed by one or more washes in 0.2xSSC, 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 honeybee, 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 a 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, that is 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
5 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
10 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
15 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-
20 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 Sindbis virus subgenomic promoters), antibiotic resistance gene, baculovirus, *Heliothis zea* insect virus, vaccinia virus, herpesvirus, raccoon poxvirus, other poxvirus, adenovirus, cytomegalovirus
25 (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 sequences capable of controlling gene expression in prokaryotic or eukaryotic cells.

Particularly preferred transcription control sequences are promoters active in
30 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 biphosphate carboxylase small subunit gene, those for root-
35 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.

Recombinant molecules of the present invention may also (a) contain secretory signals (i.e., signal segment nucleic acid sequences) to enable an expressed polypeptide
5 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 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
10 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 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
15 invention can be joined to a fusion segment that directs the encoded polypeptide to the proteasome, such as a 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.

20 Host Cells

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
25 molecule can be inserted into the cell. Transformation techniques include, but are not 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
30 within a chromosome of the transformed (i.e., recombinant) cell in such a 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
35 invention or can be capable of producing such polypeptides after being transformed 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 (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 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 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 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), 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.

30 Transgenic Plants

The term "plant" refers to whole plants, plant organs (e.g. leaves, stems roots, 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

and black-berries); 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,
5 lettuce, asparagus, cabbages, carrots, onions, tomatoes, potatoes, paprika); 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
10 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*,
15 Oxford University Press (2003), and P. Christou and H. Klee, *Handbook of 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.
20 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. 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
25 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 promoters, some other suitable promoters include but are not limited to the nopaline synthase promoter, the octopine synthase promoter, CaMV 35S promoter, the ribulose-
30 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 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,
35 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, β -conglycinin, glycinin and phaseolin), zein or oil body proteins (such as oleosin), or genes involved in fatty acid biosynthesis (including acyl carrier protein, stearyl-ACP desaturase, and fatty acid desaturases (fad 2- 1)), and other genes expressed during embryo development (such as Bce4). Particularly useful
5 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 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
10 include those that direct expression in tubers (for example, patatin gene promoter), and in fiber cells (an example of a developmentally-regulated fiber cell protein is E6 fiber).

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
15 expression cassette will be chosen primarily for convenience, since the termination 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.
20 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 β -phaseolin, the chemically inducible lant gene, pIN.

Several techniques are available for the introduction of an expression construct containing a nucleic acid sequence encoding a polypeptide of interest into the target
25 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 available, such as viral and bacterial vectors (e.g. from the genus *Agrobacterium*). After selection
30 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.

To confirm the presence of the transgenes in transgenic cells and plants, a polymerase chain reaction (PCR) amplification or Southern blot analysis can be
35 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 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
5 plant tissue or plant parts, may be harvested, and/or the seed collected. The seed may serve as a source for growing additional plants with tissues or parts having the desired characteristics.

Transgenic Non-Human Animals

10 Techniques for producing transgenic animals are well known in the art. A 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
15 microinjection, calcium phosphate mediated precipitation, liposome fusion, retroviral 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,
20 however, the appropriate DNAs are coinjected into the pronucleus or cytoplasm of embryos, preferably at the single cell stage, and the embryos 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
25 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 transfectants are then fused to
30 enucleated oocytes, cultured and transferred into female recipients.

Recovery Methods and Production of Silk

The silk proteins of the present invention may be extracted and purified from recombinant cells, such as plant, bacteria or yeast cells, producing said protein by a
35 variety of methods. In one embodiment, the method involves removal of native cell proteins from homogenized cells/tissues/plants etc. by lowering pH and heating,

followed by ammonium sulfate fractionation. Briefly, total soluble proteins are extracted by homogenizing cells/tissues/plants. Native 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 5 95% pure. Additional purification may be achieved with conventional 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.

10 Fibrillar aggregates will form from solutions by spontaneous self-assembly of silk proteins of the invention when the protein concentration exceeds a critical value. The aggregates may be gathered and mechanically spun into macroscopic fibers according to the method of O'Brien et al. (I. O'Brien et al., "Design, Synthesis and Fabrication of Novel Self-Assembling Fibrillar Proteins", in Silk Polymers: Materials 15 Science and Biotechnology, pp. 104-117, Kaplan, Adams, Farmer and Viney, eds., c. 1994 by American Chemical Society, Washington, D.C.).

By nature of the inherent coiled coil secondary structure, proteins such as Xenospira1-4, BBF1-4, BAF1-4 and GAF1-4 will spontaneously form the coiled coil secondary structure upon dehydration. As described below, the strength of the coiled 20 coil can be enhanced through enzymatic or chemical cross-linking of lysine residues in close proximity.

Silk fibres and/or copolymers 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 as they spontaneously forms strong coiled coils which 25 can be reinforced with crosslinks such as lysine crosslinks. This contrasts with *B. mori* and spider recombinant silk polypeptides which require sophisticated spinning techniques in order to obtain the secondary structure (β -sheet) and strength of the fibre.

However, fibers may be spun from solutions having properties characteristic of a liquid crystal phase. The fiber concentration at which phase transition can occur is 30 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.

35 In one fiber-forming technique, fibers 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 fiber can be pulled by such mechanical means through a methanol solution, collected, and dried. Methods for drawing fibers are considered well-known in the art.

Further examples of methods which may be used for producing silk fibres
5 and/or copolymers of the present are described in US 2004/0170827 and US 2005/0054830.

In a preferred embodiment, silk fibres and/or copolymers of the invention are crosslinked. In one embodiment, the silk fibres and/or copolymers are crosslinked to a surface/article/product etc of interest using techniques known in the art. In another
10 embodiment (or in combination with the previous embodiment), at least some silk proteins in the silk fibres and/or copolymers are crosslinked to each other. Preferably, the silk proteins are crosslinked via lysine residues in the proteins. Such crosslinking can be performed using chemical and/or enzymatic techniques known in the art. For example, enzymatic cross links can be catalysed by lysyl oxidase, whereas
15 nonenzymatic cross links can be generated from glycated lysine residues (Reiser et al., 1992).

Antibodies

The invention also provides monoclonal or polyclonal antibodies to
20 polypeptides of the invention or fragments thereof. Thus, the present invention further provides a process for the production of monoclonal or polyclonal antibodies to polypeptides of the invention.

The term "binds specifically" refers to the ability of the antibody to bind to at least one polypeptide of the present invention but not other known silk proteins.

As used herein, the term "epitope" refers to a region of a polypeptide of the
25 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 antigens, the polyclonal antibodies can be purified by immunoaffinity chromatography.
30
35 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 haptenised to another polypeptide for use as immunogens in animals.

Monoclonal antibodies directed against polypeptides of the invention can also be readily produced by one skilled in the art. The general methodology for making
5 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-Barr virus. Panels of monoclonal antibodies produced can be screened for various properties; i.e., for isotype and epitope affinity.

10 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 known in the art.

For the purposes of this invention, the term "antibody", unless specified to the
15 contrary, includes fragments of whole antibodies which retain their binding activity for a target antigen. Such fragments include Fv, F(ab') and F(ab')₂ fragments, as well as single chain antibodies (scFv). Furthermore, the antibodies and fragments thereof may be humanised antibodies, for example as described in EP-A-239400.

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

Preferably, 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
25 particles, and the like. Examples of labels which permit indirect measurement of 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,
30 alkaline phosphatase, malate dehydrogenase and the like. Where not 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),
35 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,

e.g., biotin. Such labeled antibodies can be used in techniques known in the art to detect polypeptides of the invention.

Compositions

- 5 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. Nonaqueous vehicles, such as fixed oils, sesame oil, ethyl oleate, or triglycerides may also be used.
- 10 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 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
- 15 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 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
- 20 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.

As outlined herein, in some embodiments a polypeptide, a silk fiber and/or a copolymer of the invention is used as a pharmaceutically acceptable carrier.

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

Uses

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

The silk proteins, silk fibers and/or copolymers of the invention can be used for a broad and diverse array of medical, military, industrial and commercial applications. The fibers 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 fiber or fabric for which high tensile strength and elasticity are desired characteristics. The silk proteins, silk fibers 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, poly-carboxylates, poly(vinylpyrrolidone), 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 fiber types. Examples include, but are not limited to, polymeric fibers (e.g., polypropylene, nylon, polyester), fibers and silks of other plant and animal sources (e.g., cotton, wool, *Bombyx mori* or spider silk), and glass fibers. A preferred
10 embodiment is silk fiber braided with 10% polypropylene fiber. The present invention contemplates that the production of such combinations of fibers 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 manufacture and production of fibers for medical, industrial, or
15 commercial applications.

Personal Care Products

Cosmetic and skin care compositions may be anhydrous compositions comprising an effective amount of silk protein in a cosmetically acceptable medium.
20 The uses of these compositions include, but are not limited to, skin care, skin 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 of the type of
25 cosmetic or skin care composition. Suitable compositions for a cosmetically acceptable medium are described in US 6,280,747. For example, the 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
30 substance includes, but is not limited to, oils, waxes, gums, and so-called pasty fatty substances. Alternatively, the compositions may be in the 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
35 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
5 (general cream, emulsion, shaving cream, conditioner, cologne, shaving lotion, cosmetic oil, facial mask, and the like), make-up cosmetics (foundation, 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
10 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 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
15 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 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
20 typically contains a solvent and a film forming substance, such as cellulose derivatives, polyvinyl derivatives, acrylic polymers or copolymers, vinyl 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
25 mousses, comprising an effective amount of silk protein in a cosmetically acceptable 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
30 6,139,851, and US 6,013,250. For example, these hair care compositions can be 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
35 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 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 5 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 compositions generally contain a mixture of inorganic peroxygen-based dye oxidizing agent and an oxidizable coloring agent. The peroxygen-based dye oxidizing agent is 10 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, or cyanophenols) with secondary intermediates (for example phenols, resorcinols, m-aminophenols, m-phenylenediamines, naphthols, pyrazolones, hydroxyindoles, 15 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.

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 20 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, 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 25 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 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, 30 mica, talc, barium sulfate, calcium sulfate, powdered NylonTM, perfluorinated alkanes, and other inert plastics.

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 fibers may or may not be twisted. The dental floss may be packaged as individual 35 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 2005/0130857).

5

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 preferably from about 0.25 to about 15% by weight relative to the dry weight of
10 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 between about 1% to about 15% by weight, while for inert particles, the preferred
15 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 drying a solution of the silk protein onto the particles or lyophilizing a slurry of the silk protein and the
20 particles. These coated pigments and cosmetic particles may be used in cosmetic formulations, paints, inks and the like.

Biomedical

The silk proteins may be used as a coating on a bandage to promote wound
25 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 be any soft, biologically inert, porous cloth or fiber. Examples include, but are not limited
30 to, cotton, silk, rayon, acetate, acrylic, polyethylene, polyester, and combinations thereof. The coating of the cloth or fiber 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 solution containing the silk protein may be sprayed onto the surface of the material to be coated using a spray
35 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
5 hydrochloride; astringent agents such as zinc oxide; and 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
10 comprises a dense and nonporous film of a crystallinity below 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,
15 and crust generating agents, as described 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
20 insolubles by filtration or 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 fibers and silk fibers. The polyethylene provides strength.
25 Polyester fibers 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 fibers, allowing the ends to be cut close to the knot without concern for deleterious interaction between the ends of the suture and surrounding
30 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 cuff repair, Achilles tendon repair, patellar tendon repair, ACL/PCL reconstruction, hip and
35 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 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 (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 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 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 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 producing ligaments and tendons *ex vivo* (see, for example, US 2005/0089552). A silk-fiber-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 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 fiber bundles which are arranged into a helical organization. Some
5 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 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
10 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 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
15 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 derivatives thereof may also be used in sustained release compositions (see, for example, US 2004/0005363).

20

Textiles

The silk proteins of the present invention may also be applied to the surface of fibers for subsequent use in textiles. This provides a monolayer of the protein film on the fiber, resulting in a smooth finish. US 6,416,558 and US 5,232,611 describe the
25 addition of a finishing coat to fibers. The methods described in these disclosures provide examples of the versatility of finishing the fiber to provide a good feel and a smooth surface. For this application, the fiber is coated with an effective amount of the silk protein. For the purpose of fiber 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
30 relative to the weight of the fiber material. The fiber materials include, but are not limited to textile fibers of cotton, polyesters such as rayon and LycraTM, nylon, wool, and other natural fibers including native silk. Compositions suitable for applying the silk protein onto the fiber may include co-solvents such as ethanol, isopropanol, hexafluoranol, isothiocyanouranates, and other polar solvents that can be mixed with
35 water to form solutions or microemulsions. The silk protein-containing solution may be sprayed onto the fiber or the fiber 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 fibers. 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 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. The composites can be also be used in building and automotive construction especially rooftops and door panels. The silk
10 fibres re-enforce 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. Silk 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
15 production of such materials are described in JP 2004284246, US 2005175825, US 4,515,737, JP 47020312 and WO 2005/017004.

Paper Additives

The fibre properties of the silk of the invention can add strength and quality
20 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.

25 *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).

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,
30 environmentally friendly and personalized textile products. Fibers, 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.

Developments in the weapons and surveillance technologies are prompting
35 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 euaculeatan and neuropteran (*Apis mellifera*, *Bombus terrestris*, *Myrmecia forficata*, *Oecophylla smaragdina*, *Mallada signata*) silks
10 were identified by matching ion trap consecutive mass spectral (MS/MS) fragmentation patterns of peptides obtained by trypsin digestion of the silk with the predicted mass spectral data of proteins encoded by cDNAs isolated from the salivary gland of late final instar larvae. For confirmation that no proteins were missed by this analysis for the honeybee, the peptide mass spectral data were also compared to virtual tryptic
15 digests of *Apis mellifera* proteins predicted by the bee genome project and translations of the Amel3 honeybee genomic sequences in all six reading frames.

Honeybee

Apis mellifera larvae were obtained from domestic hives. Previously it was
20 shown that silk production in *Apis mellifera* is confined to the salivary gland during the latter half of the final instar (Silva-Zacarin et al., 2003). During this period, RNA is more abundant in the posterior end of the gland (Flower and Kenchington, 1967). The cubical cell regions of 50 salivary glands were dissected from late fifth instar *Apis mellifera* immersed in phosphate buffered saline. The posterior end of the dissected
25 gland was immediately placed into RNA*later*® (Ambion, Austin, TX, USA), to stabilise the mRNA, and subsequently stored at 4°C.

Total RNA (35µg) was isolated from the late final instar salivary glands using the RNAqueous for PCR kit from Ambion (Austin, TX, USA). Message RNA was isolated from the total RNA using the Micro-FastTrack™ 2.0 mRNA Isolation kit from
30 Invitrogen (Calsbad, CA, USA) according to the manufacturer's directions with the isolated mRNA being eluted into 10ul RNase free water.

A cDNA library was constructed from the mRNA isolated from *Apis mellifera* larvae using the CloneMiner™ cDNA library construction kit of Invitrogen (Calsbad, CA, USA) with the following modifications from the standard protocol: For the first
35 strand synthesis, 0.5µl of Biotin-attB2-Oligo(dT) primer at 6pmol.µl⁻¹ and 0.5µl of dNTPs at 2mM each was added to the 10µl mRNA. After incubation at 65°C for 5 min

then 45°C for 2 min, 2µl 5XFirst strand buffer, 1µl of 0.1M DTT, and 0.5µl SuperScript™ II RT at 200U.µl⁻¹ were added. For second strand synthesis, the total volume of all reagents was halved and after ethanol precipitation, the cDNA was resuspended in 5µl of DEPC-treated water. The *aatB1* adapter (1µl) was ligated in a
5 total volume of 10µl to the 5µl cDNA with 2µl 5X Adapter buffer, 1µl 0.1M DTT and 1µl T4 DNA ligase (1U.µl⁻¹) at 16°C for 48 hrs with an additional 0.5µl T4 DNA ligase (1U.µl⁻¹) added after 16 hrs. The cDNA was size fractionated according to the manufactures instructions with samples eluting between 300 -500µl being precipitated with ethanol, resuspended and transformed into the provided *E. coli* DH10B™ T1
10 phage resistant cells as recommended. The cDNA library comprised approximately 1,200,000 colony forming units (cfu) with approximately 1% the original vector. The average insert size was 1.3 ± 1.4 kbp.

Eighty two clones were randomly selected and sequenced using the GenomeLab™ DTCS Quick start kit (BeckmanCoulter, Fullerton CA USA) and run on
15 a CEQ8000 Biorad sequencer. These clustered into fifty four groups (Table 2). Identification of the cDNAs that encoded the silk proteins is described below.

Other Species

Total RNA was isolated from 4 bumblebee (*Bombus terrestris*) (2µg RNA), 4
20 bulldog ant (*Myrmecia forficata*) (3µg RNA), approximately 100 Weaver ants (*Oecophylla smaragdina*) (0.4µg RNA) and approximately 50 green lacewing (*Mallada signata*) late larval labial glands using the RNAqueous for PCR kit from Ambion (Austin, TX, USA). mRNA was isolated from the total RNA using the Micro-FastTrack™ 2.0 mRNA Isolation kit from Invitrogen (Calsbad, CA, USA) into a final
25 volume of 10µl water. cDNA libraries were constructed from the mRNA using the CloneMiner™ cDNA kit of Invitrogen (Calsbad, CA, USA) with the following modifications from the standard protocol: For the first strand synthesis, 3pmol of Biotin-attB2-Oligo(dT) primer and 1nmol each dNTPs were added to the 10µl mRNA. After 5 min at 65°C followed by 2 min at 45°C, 2 µl 5XFirst strand buffer, 50 nmol
30 DTT, and 100U SuperScript™ II RT were added.

Table 2. *A. mellifera* final instar salivary gland cDNAs and MS ion trap fragmentation patterns of peptides from trypsin digestion of SDS treated brood comb silk.

Number of cDNA's in cluster	Abundance in salivary gland library (%)	Protein or gene synonyms	Number of tryptic peptides identified in the silk	Distinct summed MS/MS search score	Coverage of protein sequence (% protein)	Protein identification
Proteins identified in cDNA library and in honeybee silk						
10	13	Xenosin; GB15233-PA	9	143.89	25	AC004701
8	11	Xenospiral; GB12184-PA	10	165.13	37	No matches
6	7	Xenospira4; GB19585-PA	8	142.16	35	No matches
6	7	Xenospira2; GB12348-PA	9	145.91	28	No matches
5	6	Xenospira3; GB17818-PA	9	147.02	31	No matches
Proteins identified in cDNA library only						
4	4	GB14261-PA	0			
2	2	Contig 2504	0			
2	2	GB17108-PA	0			
1	1	Contig 68	0			
1	1	Contig 110	0			
1	1	Contig 487	0			
1	1	GB14199-PA	0			
1	1	GB10847-PA	0			
1	1	Contig 1047	0			
1	1	GB17558-PA	0			
1	1	Contig 1471	0			
1	1	GB16480-PA	0			
1	1	Contig 1818	0			
1	1	GB16911-PA	0			
1	1	Contig 2046	0			
1	1	Contig 2136	0			
1	1	Contig 2196	0			
1	1	GB11234-PA	0			
1	1	GB11199-PA	0			
1	1	GB18183-PA	0			
1	1	Contig 2938	0			
1	1	Contig 2976	0			
1	1	Contig 3263	0			
1	1	Contig 3527	0			
1	1	GB16412-PA	0			
1	1	GB18750-PA	0			
1	1	GB16132-PA	0			
1	1	Contig 4536	0			
1	1	GB19431-PA	0			
1	1	Contig 4704	0			
1	1	Contig 4758	0			
1	1	Contig 4830	0			
1	1	Contig 4968	0			
1	1	Contig 5402	0			
1	1	Contig 5971	0			
1	1	GB11274-PA	0			
1	1	GB14693-PA	0			
1	1	GB19585-PA	0			
1	1	GB15606-PA	0			
1	1	GB16801-PA	0			
1	1	GB12085-PA	0			
1	1	Contig 7704	0			
1	1	Contig 8630	0			
1	1	Contig 9774	0			
1	1	GB16452-PA	0			
1	1	GB10420-PA	0			
1	1	GB14724-PA	0			

For second strand synthesis, the total volume of all reagents was halved from the manufacturer's recommended amounts and after ethanol precipitation, the cDNA was resuspended in 5 μ l of DEPC-treated water. The *aatB1* adapter (1 μ l) was ligated in a total volume of 10 μ l to the 5 μ l cDNA with 2 μ l 5X Adapter buffer, 50 nmol DTT and 1U T4 DNA ligase at 16°C for 12 hrs. The cDNA libraries comprised approximately 2.4x10⁷ (bumblebee), 5.0x10⁷ (bulldog ant) and 6000 (green ant) colony forming units (cfu) with less than 1% the original vector for the bulldog ant and bumblebee libraries and greater than 80% original vector in the green ant library. The average insert size within the libraries was 1.3 Kbp.

Sequence data was obtained from more than 100 random clones from the cDNA libraries from bumblebee and bulldog ant, 82 clones from the honeybee and 60 clones from the lacewing. The technical difficulties of obtaining salivary glands from the minute green ants (approximately 1mm in length) reduced the efficiency of the library from this species and as such only 40 sequences were examined. A summary of the silk proteins identified is provided in Table 3.

Table 3. Identification and properties of the euaculeatan silk proteins.

Species	Protein name	Length of protein (amino acids)	% cDNA library clones	Distinct summed MS/MS identification score	% helical structure**	MARCOIL predicted coiled coil length*** (amino acids)
Honeybee	AmelF1*	333	6	52	76	117
Honeybee	AmelF2*	290	7	51	88	175
Honeybee	AmelF3*	335	11	107	81	154
Honeybee	AmelF4*	342	7	88	76	174
Honeybee	AmelSA1*	578	13	40	41	45
Bumblebee	BBF1	327	4	180	86	147
Bumblebee	BBF2	313	14	100	84	199
Bumblebee	BBF3	332	20	218	86	146
Bumblebee	BBF4	357	32	137	80	188
Bumblebee	BBSA1	>501	3	138	21	0
Bulldog ant	BAF1	422	16	99	69	121
Bulldog ant	BAF2	411	30	90	76	132
Bulldog ant	BAF3	394	26	88	79	131
Bulldog ant	BAF4	441	24	116	76	157
Weaver ant	GAF1	391	35	228	74	177
Weaver ant	GAF2	400	22	191	79	158
Weaver ant	GAF3	395	13	156	72	103
Weaver ant	GAF4	443	17	148	74	166
Lacewing	MalF1	596	23	45	89	151

* also referred to herein as Xenospiral-4 and Xenosin respectively, ** predicted by PROFsec, *** predicted by MARCOIL at 90% threshold

EXAMPLE 2 - Preparation and proteomic analysis of native silk

5 Honeybee brood comb after the removal of larvae, bumblebee cocoons after the removal of larvae, bulldog ant cocoons after the removal of larvae, or weaver ant silk sheets were washed extensively three times in warm water to remove water soluble contaminants and then washed extensively three times in chloroform to remove wax. Chloroform was removed by rinsing in distilled water and a subset of this silk was
10 retained for analysis. A subset of the Hymenopteran (ants and bees) silk samples was further washed by boiling for 30 minutes in 0.05% sodium carbonate solution, a standard procedure for degumming silkworm silk, then rinsed in distilled water. Lacewing silk was rinsed in distilled water only. A subset of the lacewing silk samples was degummed by boiling for 30 minutes in 0.05% sodium carbonate solution.

15 A subset of the honeybee material was soaked overnight in 2% SDS at 95°C, followed by three washes in distilled water. Extraction in hot SDS solution solubilises most proteins, but in this case the silk sheets retained their conformation.

The clean silks were analysed by liquid chromatography followed by tandem mass spectrometry (LCMS) as described below.

20 Pieces of cleaned silk were placed in a well of a Millipore 'zipplate', a 96 well microtitre tray containing a plug of C18 reversed phase chromatography medium through the bottom of each well to which was added 20µl 25mM ammonium bicarbonate containing 160 ng of sequencing grade trypsin (Promega). Then the tray was incubated overnight in a humidified plastic bag at 30°C.

25 The C18 material was wetted by pipetting acetonitrile (10µl) to the sides of each well and incubating the plate at 37°C for 15 min. Formic acid solution (130µl, 1% v/v) was added to each well and after 30 min peptides from the digested bee proteins were captured on the C18 material by slowly drawing the solutions from each well through the base of the plate under a reduced vacuum. The C18 material was washed twice by
30 drawing through 100 µl of formic acid solution. Peptides were eluted with 6 µl of 1% formic acid in 70% methanol pipetted directly onto the C18 material and promptly centrifuged through the C18 plug to an underlying microtitre tray. This tray was placed under vacuum till the volume in each well was reduced about 2-fold by evaporation. Formic acid solution (10 µl) was added to each well and the tray was transferred to the
35 well plate sampler of an Agilent 1100 capillary liquid chromatography system.

Peptides (8 μ l) from the silk extract were bound to an Agilent Zorbax SB-C18 5 μ m 150x0.5mm column with a flow rate of 0.1% formic acid/5% acetonitrile at 20 μ l.min⁻¹ for one min then eluted with gradients of increasing acetonitrile concentration to 0.1% formic acid/20% acetonitrile over one minute at 5 μ l.min⁻¹, then
5 to 0.1% formic acid/50% acetonitrile over 28 minutes, then to 0.1% formic acid/95% acetonitrile over one minute. The column was washed with 0.1% formic acid/95%-100% acetonitrile over 5 mins at 20 μ l.min⁻¹ and reequilibrated with 0.1% formic acid/5% acetonitrile for 7 mins before peptides from the next well were sampled.

Eluate from the column was introduced to an Agilent XCT ion trap mass
10 spectrometer through the instrument's electrospray ion source fitted with a micronebuliser. Briefly, as peptides were eluting from the column, the ion trap collected full spectrum positive ion scans (100-2200m/z) followed by two MS/MS scans of ions observed in the full spectrum avoiding the selection of ions that carried only a single charge. When an ion was selected for MS/MS analysis all others were
15 excluded from the ion trap, the selected ion was fragmented according to the instrument's recommended "SmartFrag" and "Peptide Scan" settings. Once two fragmentation spectra were collected for any particular m/z value it was excluded from selection for analysis for a further 30 seconds to avoid collecting redundant data.

Mass spectral data sets from the entire experiment were analysed using
20 Agilent's Spectrum Mill software to match the data with predictions of protein sequences from the cDNA libraries. The software generated scores for the quality of each match between experimentally observed sets of masses of fragments of peptides and the predictions of fragments that might be generated according to the sequences of proteins in a provided database. All the sequence matches reported here received
25 scores greater than 20, the default setting for automatic, confident acceptance of valid matches.

This analysis identified that five proteins expressed at high levels in the labial gland matched the silk from each of the cognate bee species (shown in Tables 2 and 3) and four proteins expressed at high levels in the labial gland matched the silk from each
30 of the cognate ant species (shown in Table 3). The abundance of message RNA encoding these proteins in the labial gland of the larvae was consistent with the proteins being abundantly produced (abundance of message shown in Table 3).

To ensure that none of the honeybee silk proteins were missed by this identification process, we also compared the honeybee silk trypsin peptide mass
35 spectral data to a set of publicly available predicted protein sequences from the honeybee genome project, generated by a computer algorithm that tries to recognise

transcribed genes in the complete genomic DNA sequences of the bee. Additionally, we generated a database of translations in the six possible reading frames of each contiguous genomic DNA sequence provided by the bee genome project (Amel3 release). These translated DNA sequences were presented to the Spectrum Mill
5 software as if they were the sequences of very large proteins. Matching MS/MS peptide data identified open reading frames within the genomic sequences that had encoded parts of the isolated bee proteins without the need to first predict the organisation of genes. No additional proteins were identified in the silk by this analysis.

10

EXAMPLE 3 – Structural analysis of the native silk

Native silk samples were prepared as described in Example 2. Silk samples were examined using a Bruker Tensor 37 Fourier transform infrared spectrometer with a Pike Miracle diamond attenuated total reflection accessory. Analysis of the amide I
15 and II regions of the spectra of honeybee, bumblebee, green ant, bulldog ant silks and lacewing larval silk (Figure 1) shows that all these silks have a predominantly alpha-helical secondary structure. The silks of the Euaculeatan species have dominant peaks in the FT-IR spectra at 1645-1646 cm^{-1} , shifted approximately 10 cm^{-1} lower than a classical α -helical signal and broadened. This shift in the α -helical signal is typical of
20 coiled-coil proteins (Heimburg et al., 1999). Spectra from samples that were degummed were unchanged.

EXAMPLE 4 - The amino acid composition of native silks closely resembles that of the identified silk proteins

25 The amino acid composition of the native silks was determined after 24 hr gas phase hydrolysis at 110°C using the Waters AccQTag chemistry by Australian Proteome Analysis Facility Ltd (Macquarie University, Sydney).

The measured amino acid composition of the SDS washed silk was similar to that predicted from the identified silks protein sequences (Figures 2 and 3).

30

EXAMPLE 5 - Structural analysis of the silk proteins

Predicted secretory peptides

As expected for silk proteins, the SignalP 3.0 signal prediction program (Bendtsen et al., 2004), which uses two models to identify signal peptides predicted
35 that all the identified silk genes encoded proteins which contain signal peptides that

targeted them for secretion from a cell (data not shown). The predicted cleavage sites of the polypeptides are as follows:

- Xenospira1 (AmelF1) - between pos 19 and 20 (ASA-GL),
- Xenospira2 (AmelF2) - between pos 19 and 20 (AEG-RV),
- 5 Xenospira3 (AmelF3) - between pos 19 and 20 (VHA-GV),
- Xenospira4 (AmelF4) - between pos 19 and 20 (ASG-AR),
- Xenosin (AmelSA1) - between pos 19 and 20 (VCA-GV),
- BBF1 - between pos 19 and 20 (ASA-GQ),
- BBF2 - between pos 20 and 21 (AEG-HV),
- 10 BBF3 - between pos 19 and 20 (VHA-GS),
- BBF4 - between pos 19 and 20 (ASA-GK),
- BAF1 - between pos 19 and 20 (ASA-SG),
- BAF2 - between pos 19 and 20 (ASG-RV),
- BAF3 - between pos 19 and 20 (ASG-NL),
- 15 BAF4 - between pos 19 and 20 (VGA-SE),
- GAF1 - between pos 19 and 20 (ADA-SK),
- GAF2 - between pos 19 and 20 (ASG-GV),
- GAF3 - between pos 19 and 20 (ASG-GV),
- GAF4 - between pos 19 and 20 (VGA-SE),
- 20 MalF1 – between pos 26 and 27 (SST-AV).

All four of the ant and four of the five bee silk proteins are helical and formed coiled coils

Protein modelling and results from pattern recognition algorithms confirmed
25 that the majority of the identified honeybee silk proteins were helical proteins that formed coiled coils.

PROFsec (Rost and Sander, 1993) and NNPredict (McClelland and Rumelhart, 1988; Kneller et al., 1990), algorithms were used to investigate the secondary structure of the identified silk genes. These algorithms identified Xenospira1 [GB12184-PA]
30 (SEQ ID NO:1), Xenospira2 [GB12348-PA] (SEQ ID NO:3), Xenospira3 [GB17818-PA] (SEQ ID NO:5), and Xenospria4 [GB19585-PA] (SEQ ID NO:7), as highly helical proteins, with between 76-85 % helical structure (Table 4). Xenosin [GB15233-PA] (SEQ ID NO:10) had significantly less helical structure.

Table 4. The secondary structure of *Apis mellifera* silk proteins predicted by PROFsec (Rost and Sander, 1993) showing percentages of helices, extended sheets and loops.

Protein	helical		extended		loop	
	PROFsec	NNPredict	PROFsec	NNPredict	PROFsec	NNPredict
Xenospira3	77	70	3	6	20	27
Xenospira4	85	82	2	6	14	16
Xenospira1	80	73	1	4	19	26
Xenospira2	77	69	2	5	21	29
Xenosin	41	41	8	9	51	50

Further protein modelling and results from pattern recognition algorithms confirmed that the majority of the identified silk proteins were helical proteins that formed coiled coils. PredictProtein (Rost et al., 2004) algorithms were used to investigate the secondary structure of the identified silk genes. These algorithms identified Xenospira1 (SEQ ID NO:1), Xenospira2 (SEQ ID NO:3), Xenospira3 (SEQ ID NO:5), Xenospira4 (SEQ ID NO:7), BBF1 (SEQ ID NO:22), BBF2 (SEQ ID NO:24), BBF3 (SEQ ID NO:26), BBF4 (SEQ ID NO:28), BAF1 (SEQ ID NO:40), BAF2 (SEQ ID NO:42), BAF3 (SEQ ID NO:44), BAF4 (SEQ ID NO:46), GAF1 (SEQ ID NO:56), GAF2 (SEQ ID NO:58), GAF3 (SEQ ID NO:60), GAF4 (SEQ ID NO:62), and MalF1 (SEQ ID NO:72) as highly helical proteins, with between 69-88 % helical structure (Table 3). AmelSA1 [GB15233-PA] (Xenosin) (SEQ ID NO:10) and BBSA1 (SEQ ID NO:30) had significantly less helical structure.

Super-coiling of helical proteins (coiled coils) arises from a characteristic heptad repeat sequence normally denoted as $(abcdefg)_n$ with generally hydrophobic residues in position *a* and *d*, and generally charged or polar residues at the remaining positions. The pattern recognition programs (MARCOIL (Delorenzi and Speed, 2002), COILS (Lupas et al., 1991)) identified numerous heptad repeats typical of coiled-coils in Xenospira1 [GB12184-PA] (SEQ ID NO:1), Xenospira2 [GB12348-PA] (SEQ ID NO:3), Xenospira3 [GB17818-PA] (SEQ ID NO:5), and Xenospira4 [GB19585-PA] (SEQ ID NO:7) (MARCOIL: Table 5; COILS: Figure 4), as well as BBF1 (SEQ ID NO:22), BBF2 (SEQ ID NO:24), BBF3 (SEQ ID NO:26), BBF4 (SEQ ID NO:28), BAF1 (SEQ ID NO:40), BAF2 (SEQ ID NO:42), BAF3 (SEQ ID NO:44), BAF4 (SEQ ID NO:46), GAF1 (SEQ ID NO:56), GAF2 (SEQ ID NO:58), GAF3 (SEQ ID NO:60), GAF4 (SEQ ID NO:62), and MalF1 (SEQ ID NO:72) (MARCOIL: Table 3).

Identification of a novel coiled coil sequence in the honeybee silk proteins

The heptad repeats of amino acid residues identified in the sequences of Xenospira1 [GB12184-PA], Xenospira2 [GB12348-PA], Xenospira3 [GB17818-PA],

Xenospria4 [GB19585-PA], were each highly indicative of a coiled coil secondary structure (Figure 5) (see Table 5 for confidence levels). The fact that the heptads are found consecutively and numerous suggests the proteins adopt a very regular structure. Overlapping heptads were identified in two of the honeybee proteins: the major coiled coil region of Xenospira1 contained overlapping heptads with a 3 residue offset followed by a space of 5 residues and then four consecutive heptads; and the entire coiled coil region of Xenospira2 had multiple overlapping heptads with a single offset and 4 residue offset (equivalent to 3 residue offset). The composition of amino acids in the various positions of the major heptad are shown in the first column in Table 6, with the positions of the overlapping heptads indicated in adjacent columns.

Table 5. Percent of residues in the identified silk proteins predicted to exist as coiled coil by the MARCOIL (Delorenzi and Speed, 2002) pattern recognition algorithm.

Protein	Length of mature protein (amino acids)	Percent protein that exists as coiled coil		
		50% threshold	90% threshold	99% threshold
Xenospira3	315	64% (residues 68 to 268)	34% (residues 128-223 and 235-246)	20% (residues 149-211)
Xenospira4	290	73% (residues 83-293)	60% (residues 98-168 and 182-285)	27% (residues 113-154 and 212-247)
Xenospira1	316	69% (residues 67-282)	49% (residues 103-256)	18% (residues 113-169)
Xenospira2	328	65% (residues 89-298)	54% (residues 110-283)	45% (residues 127-270)
Xenosin	350	26% (residues 32-127)	9% (residues 42-75)	2% (residues 59-67)

15

Surprisingly the major heptads have a novel composition when viewed collectively - with an unusually high abundance of alanine in the 'hydrophobic' heptad positions *a* and *d* (see Table 6 and Figure 5). Additionally, a high proportion of heptads have alanine at both *a* and *d* positions within the same heptad (33% in Xenospira1 [GB12184-PA]; 36% in Xenospira2 [GB12348-PA]; 27% in Xenospira3 [GB17818-PA]; and 38% in Xenospira4 [GB19585-PA]; see Tables 6 and 7).

20

Table 6. Summary of the number of each amino acid residues in the various heptad positions in coiled coil regions of honeybee silk proteins.

Xenospira4																		
	A	I	R	L	K	T	E	V	F	S	Q	N	D	G	M	Y	W	Total
a	23	0	1	1	0	1	1	1	0	1	0	0	0	0	0	0	0	29
b	12	0	0	2	2	2	3	1	0	3	1	1	1	1	0	0	0	29
c	12	0	0	1	5	1	3	1	0	3	1	1	0	1	0	0	0	29
d	17	0	0	5	1	0	1	2	0	2	1	0	0	0	0	0	0	29
e	12	0	1	0	0	2	4	2	0	5	2	1	0	0	0	0	0	29
f	13	1	0	1	2	0	7	1	0	1	1	2	0	0	0	0	0	29
g	9	3	4	0	2	1	2	1	0	2	0	1	2	2	0	0	0	29

Xenospira3																		
	A	I	R	L	K	T	E	V	F	S	Q	N	D	G	M	Y	W	Total
a	19	0	0	1	0	4	2	0	0	1	1	1	0	0	0	1	0	30
b	8	0	0	5	1	2	2	0	0	5	4	2	1	0	0	0	0	30
c	13	0	1	0	3	2	2	3	0	1	2	0	1	1	0	0	1	30
d	13	3	0	2	2	0	2	2	0	4	0	1	1	0	0	0	0	30
e	8	0	0	2	2	2	4	0	0	7	4	0	0	1	0	0	0	30
f	7	0	2	3	4	2	4	0	0	4	1	2	1	0	0	0	0	30
g	9	0	5	2	3	0	1	2	0	5	0	2	1	0	0	0	0	30

Xenospira2																		
	A	I	R	L	K	T	E	V	F	S	Q	N	D	G	M	Y	W	Total
a	20	0	0	1	0	3	1	1	1	1	0	0	0	0	0	0	0	28
b	7	2	2	2	2	2	2	4	0	1	1	3	0	0	0	0	0	28
c	9	0	2	0	4	1	2	4	0	1	3	2	0	0	1	1	1	28
d	16	0	0	3	3	1	0	1	0	1	2	0	1	0	0	0	0	28
e	11	0	1	3	0	3	4	1	0	2	2	0	1	0	0	0	0	28
f	10	2	1	0	1	2	6	1	0	3	1	1	0	0	0	0	0	28
g	8	4	1	0	1	1	5	0	0	0	2	4	0	1	1	0	0	28

Xenospira1																		
	A	I	R	L	K	T	E	V	F	S	Q	N	D	G	M	Y	W	Total
a	13	3	0	1	2	0	1	1	0	2	1	1	0	2	0	0	0	27
b	7	1	1	1	6	0	2	1	0	3	1	0	4	0	0	0	0	27
c	8	1	2	1	1	1	7	2	0	1	1	1	0	1	0	0	0	27
d	18	0	0	2	1	2	1		0	2	1	0	0	0	0	0	0	27
e	11	1	2	1	1	2	3	2	0	4	0	0	0	0	0	0	0	27
f	7	0	3	0	2	1	3	3	0	7	0	1	0	0	0	0	0	27
g	13	0	0	3	3	0	2	1	0	3	1	0	0	0	1	0	0	27

Table 7. Summary of alanine residues in heptads of honeybee silk proteins.

Protein	Amount of helical structure (%) ¹	Number of major heptads	Amount of protein in major heptad (%)	Amount of Ala in major heptads (%)	Amount of Ala in position a of major heptads (%)	Amount of Ala in position d of major heptads (%)	Amount of Ala in position a and d of major heptads (%)
Xenospira1	77 (70)	27		41	44	74	33
Xenospira2	85 (82)	28		37	71	57	36
Xenospira3	80 (73)	30		37	63	43	27
Xenospira4	77 (69)	29		48	79	58	38
Xenosin	41 (41)			n/a	n/a	n/a	n/a

¹PROFsec predictions with NNPredict predictions shown in brackets.

5 The composition of amino acids in the various heptad positions in the coiled coil region of the hymenopteran silks are summarised in Figures 6 and 7. As noted above, the positions within the heptads have a novel composition - the 'hydrophobic' heptad positions *a* and *d* of the bee and ant silks contain very high levels of alanine (average 58%) and high levels of small polar residues (average 21%) in comparison to other
10 coiled coils. Additionally, position *e* is unusually small and hydrophobic (Table 8, Figure 7). Topographically this position is located adjacent to the *a* residues within the helices. Its compositional similarity with the *a* and *d* residues suggest that the silks adopt a coiled coil structure with three core residues per α -helix. Three residue cores contribute a larger hydrophobic interface than two residues in the core (Deng et al.,
15 2006) - a feature that would assist coiled coil formation and stability.

 In addition, when viewed collectively the positions *b*, *c*, *e*, *f* and *g* within the heptad are generally more hydrophobic, less polar and less charged than protein coiled coil regions previously characterised (see Figure 7, and Tables 8 and 9). Therefore, although historically it was regarded that the helical content of the aculeate
20 Hymenopteran silk was a consequence of a reduced glycine content and increased content of acidic residues (Rudall and Kenchington, 1971), we have discovered that it is not the glycine/acid residues that are responsible for the novel silk structure but rather the position of the alanine residues within the polypeptide chains.

Table 8. Average size and hydrophobicity at each heptad position of the orthologous hymenopteran silk proteins and of the green lacewing silk protein (MalF1) showing that *a*, *d*, and *e* positions (core) are smaller and more hydrophobic than other positions. In some cases the *b* position (partially submerged) is also small and hydrophobic.

5

	Heptad position						
	a	b	c	d	e	f	g
Amel F1 orthologs							
Average residue side chain hydrophobicity	0.36	0.20	0.20	0.30	0.26	-0.16	0.03
Average residue side chain length	1.7	2.5	2.5	2.1	2.3	3.0	2.6
Amel F2 orthologs							
Average residue side chain hydrophobicity	0.53	0.20	0.03	0.36	0.24	0.05	0.12
Average residue side chain length	1.5	2.6	2.6	2.0	2.2	2.5	3.0
Amel F3 orthologs							
Average residue side chain hydrophobicity	0.44	0.36	0.06	0.41	0.27	-0.10	0.00
Average residue side chain length	1.9	2.3	2.4	2.1	2.3	2.8	2.8
Amel F4 orthologs							
Average residue side chain hydrophobicity	0.46	0.17	-0.13	0.61	0.04	0.06	0.06
Average residue side chain length	1.4	2.2	2.6	2.04	2.3	2.6	2.7
MalF1							
Average residue side chain hydrophobicity	-0.05	0.14	-0.61	0.27	0.59	0.23	-0.22
Average residue side chain length	2.1	1.7	2.5	1.4	1.5	1.7	3.5

EXAMPLE 6 - The bee silk proteins are likely to be extensively cross-linked

The bee silk proteins all contain a high proportion of lysine (6.5% – 16.3%). A comparison between the measured amino acid composition of bee silk and the sequences of the identified silk proteins reveals a substantial mismatch in the number of lysine residues, with much less lysine detected in the silk than expected (Figures 2 and 3). This suggests that lysine residues in the silk have been modified, so are not being identified by standard amino acid analysis. Lysine is known to form a variety of cross-links: either enzymatic cross links catalysed by lysyl oxidase or nonenzymatic cross

15

links generated from glycated lysine residues (Reiser et al., 1992). The under-representation of lysine in the honeybee and bumblebee silk amino acid analysis is consistent with the presence of lysine cross-linking

5 **Table 9.** Number of residues in each class of amino acids at various heptad positions in coiled coil regions of silk proteins.

Xenospira4						Heptad position
Nonpolar	Polar	Charged	Small	Medium	Large	
25	2	2	26	2	1	a
16	7	6	19	10	0	b
15	6	8	18	11	0	c
24	3	2	21	8	0	d
14	10	5	21	7	1	e
16	4	9	15	14	0	f
15	4	10	15	10	4	g

Xenospira3						
Nonpolar	Polar	Charged	Small	Medium	Large	
20	8	2	24	5	1	a
13	13	4	15	15	0	b
17	6	7	20	8	2	c
20	5	5	19	11	0	d
11	13	6	18	12	0	e
10	9	11	13	15	2	f
13	7	10	16	9	5	g

Xenospira2						
Nonpolar	Polar	Charged	Small	Medium	Large	
23	4	1	25	2	1	a
15	7	6	14	12	2	b
13	7	8	15	11	2	c
20	4	4	19	9	0	d
15	7	6	17	10	1	e
13	7	8	16	11	1	f
14	7	7	10	17	1	g

Xenospira1						
Nonpolar	Polar	Charged	Small	Medium	Large	
20	4	3	18	9	0	a
10	4	13	11	15	1	b
13	4	10	13	12	2	c
20	5	2	22	5	0	d
15	6	6	19	6	2	e
10	9	8	18	6	3	f
18	4	5	17	10	0	g

Covalently cross-linked proteins subjected to SDS polyacrylamide gel electrophoresis (PAGE) are expected to migrate according to the molecular weight of the cross-linked complex. We subjected late last instar honeybee labial gland proteins to SDS PAGE and measured the migration of the silk proteins in relation to standard
5 protein markers. Bands were observed corresponding to monomers of each of the identified silk proteins, however higher molecular weight bands containing these proteins were also present, as expected in a cross-linked system (Figure 8).

As described above, the honeybee labial gland contains a mixture of organised and disorganised silk proteins. The cross-linked proteins observed probably correspond
10 to the protein population of the anterior region of the gland, where the silk is prepared for extrusion. It is reasonable to assume that extracellular honeybee silk contains a substantially higher proportion of cross-linked proteins than is observed in a heterogenous mixture of all stages of salivary gland silk proteins. The bonds are unlikely to be cysteine cross-links, as the silk was unaffected by reductive treatment,
15 and the identified silk proteins contain few or no cysteine residues.

EXAMPLE 7 - The euaculeatan silk proteins differ significantly from the other silk proteins

The euaculeatan silk is significantly different from other described silk genes in
20 relation to amino acid composition (Table 10), molecular weight of the proteins involved, secondary structure and physical properties (Tables 11 and 12). The lepidopteran silks are primarily composed of the small amino acid residues alanine, serine and glycine (for example the silk of *Bombyx mori*, Table 10) and are dominated by extended beta sheet secondary structure. The *Cotesia glomerata* silk protein is high
25 in asparagine and serine – the abundance of the latter residue being characteristic of Lepidopteran silk sericins (glues) (Table 10). Modelling of the *Cotesia glomerata* silk protein does not identify helices or coiled coils in the secondary structure. In contrast, the bee, ant and lacewing silks are high in alanine (Table 10) and are comprised of a high level of helical secondary structure that forms coiled coils.

30

35

Table 10. Amino acid composition of silk from various Insects with most abundant residues shown in boldface.

	Honeybee	Euaculeatan silk	Mallada silk	Cotesia glomerata	Bombyx mori
Alanine	22.6	27.5	26.9	12.5	29.3
Glutamic acid + Glutamine	16.1	13.9	7.4	0.6	0.9
Aspartic acid + Asparagine	13.2	8.6	15.0	37.6 (Asn 33.7)	1.2
Serine	10.4	11.5	8.5	37.1	11.3
Leucine	9.0	7.2	5.9	0.4	0.4
Valine	6.6	4.8	4.1	0.3	2.1
Glycine	5.7	6.6	11.2	5.5	46.0
Isoleucine	5.6	4.0	3.9	0.4	0.6
Threonine	5.1	4.9	5.3	0.5	0.8
Lysine	3.7	3.7	3.2	0.1	0.3
Phenylalanine	2.0	1.0	0.5	0.5	0.6
Tyrosine	0	0.9	0.5	3.1	5.3
Proline	0	0	0	0.7	0.4
Histidine	0	0.5	0.5	0.4	0.2
Arginine	0	3.3	5.4	0.2	0.4
Methionine	0	1.0	1.6	0	0.1
Tryptophan	0	Not reported	Not reported	Not reported	0.2
Cysteine	0	0.4	0.3	Not reported	0.1

5 Table 11. Differences between insect silks.

	Ant and bee silk	Mallada silk	Cotesia sp.	Lepidoptera For example Bombyx mori
Most abundant amino acids	Ala	Ala	Ser, Asn	Gly, Ala
Size of fibroin proteins	25-35 kDa	57KDa	Approx 500KDa	>100KDa
Secondary structure	Coiled coil	Coiled coil	Most likely beta sheets. Secondary structure prediction programs PROFsec and MARCOIL do not recognise any helical structure or coiled coil regions.	beta-pleated sheets loosely associated with beta-sheets, beta-spirals, alpha helices and amorphous regions

Table 12. Solubility of insect silks.

Solvent	Ant and bee silk		Mallada silk		Cotesia sp.		Bombyx mori	
	20°C	95°C	20°C	95°C	20°C	95°C	20°C	95°C
LiBr 54%	-	-	-	-	-	part	-	✓
LiSCN saturated	-	-	-	-	-	part	-	✓
8M urea	-	-	-	-	-	-	-	part
6M guanidine HCl	-	-	-	-	-	-	-	part
1M NaOH	-	part	?	?	-	part	part	✓
6M HCl	-	part	part	✓	-	part	-	✓
3M HCl / 50% propanoic acid	-	part	?	?	-	part	part	✓

Cladistic analysis of the coiled coil regions of the silk proteins of the four Hymenopteran species (Figure 9) suggests that the genes evolved in a common ancestor that predates the divergence of the Euaculeata from the parasitic wasps. The sequences of the silk have diverged extensively and we were only able to align the 210 amino acids that comprise the coiled coil region of each protein. The amino acid sequence identity between the coiled coil regions of each of the silk proteins provided herein is shown in Table 13 and DNA identity in the corresponding region is shown in Table 14. Whilst the proteins have similar amino acid contents (especially high levels of alanine) and tertiary structure, the primary amino acid sequence identity is very low. In fact, the gene encoding the *Mallada* silk protein has evolved independently and as such the silk protein sequence cannot be aligned to the Hymenopteran sequences. This indicates that considerable variety in the identity of the amino acids can occur, whilst not affecting the biological function of the proteins.

The cladistic analysis predicts that silk of euaculeatan wasps is comprised of related proteins to the silk of ants and bees and that although these proteins will have similar composition and architecture to the proteins described here, they will have highly diverged primary sequence.

The amino acid sequences of the silk proteins provided herein (Figure 10) were subjected to comparisons with protein databases, however, no prior art proteins were identified with any reasonable level of sequence identity (for example, none greater than 30% identical over the length of the silk protein sequence).

Table 13. Percent identity between protein sequences of the coiled coil region of the fibre proteins in ants and bees.

	Honeybee				Bumblebee				Bulldog ant				Green ant			
	F1	F2	F3	F4	F1	F2	F3	F4	F1	F2	F3	F4	F1	F2	F3	F4
beeF1	100															
beeF2	26.7	100														
beeF3	23.3	31.4	100													
beeF4	34.8	32.4	30.0	100												
BBF1	65.7	28.1	24.8	35.7	100											
BBF2	28.6	71.4	28.6	31.9	31.0	100										
BBF3	25.2	31.0	65.7	27.6	27.1	29.5	100									
BBF4	33.3	31.0	29.5	64.8	34.8	31.4	28.1	100								
BAF1	37.1	20.0	20.0	32.4	39.5	21.4	21.4	29.1	100							
BAF2	25.2	44.3	29.5	33.8	28.1	38.1	28.6	27.6	27.1	100						
BAF3	23.8	26.2	36.7	28.1	24.8	25.2	36.7	28.1	21.0	27.6	100					
BAF4	28.1	33.8	24.8	45.2	28.6	33.8	23.3	43.8	26.1	27.6	25.2	100				
GAF1	33.8	20.0	23.8	32.9	36.2	22.9	23.8	29.1	66.7	28.1	25.2	28.6	100			
GAF2	24.8	41.9	27.6	29.5	28.1	39.5	29.0	26.7	21.9	66.2	23.8	26.7	23.8	100		
GAF3	26.9	28.8	40.1	31.6	25.5	28.3	38.2	30.2	24.0	28.3	62.7	27.4	27.4	26.4	100	
GAF4	24.7	32.4	24.3	37.6	27.1	32.4	24.8	38.1	23.9	29.5	21.0	63.3	24.8	27.6	24.1	100

Table 14. Percent identity between nucleotide sequences encoding coiled coil region of the fibre proteins in ants and bees.

	Honeybee				Bumblebee				Bulldog ant				Green ant			
	F1	F2	F3	F4	F1	F2	F3	F4	F1	F2	F3	F4	F1	F2	F3	F4
beeF1	100															
beeF2	39.4	100														
beeF3	37.0	40.2	100													
beeF4	45.1	44.8	41.0	100												
BBF1	68.9	40.9	37.5	45.2	100											
BBF2	42.5	72.9	42.5	44.9	42.2	100										
BBF3	40.6	40.0	67.6	40.5	38.4	41.0	100									
BBF4	45.4	41.0	41.7	66.0	45.9	43.6	40.0	100								
BAF1	45.7	35.1	35.9	41.1	47.9	36.5	36.0	38.7	100							
BAF2	38.1	49.8	41.4	44.6	38.7	47.3	40.0	41.0	40.6	100						
BAF3	33.3	36.7	45.4	40.3	36.3	36.8	46.2	39.4	36.0	40.5	100					
BAF4	39.5	43.3	41.4	46.8	43.0	47.6	39.8	49.4	42.5	41.7	40.3	100				
GAF1	45.6	35.1	37.3	42.4	47.6	38.5	37.8	41.4	68.9	41.7	36.7	43.0	100			
GAF2	38.5	47.8	38.4	43.2	38.1	46.5	41.4	40.0	37.5	69.7	38.9	40.6	39.4	100		
GAF3	39.0	40.1	46.1	41.8	37.7	39.3	46.1	40.0	37.7	41.7	65.1	41.2	40.0	41.7	100	
GAF4	38.9	42.4	38.1	44.9	38.9	43.8	38.4	44.3	37.3	42.7	36.7	67.8	38.2	40.3	37.7	100

The open reading frames encoding the silk proteins (provided on Figure 11) were subjected to similar database searching as that described above. The only related molecules that were identified have been published as part of the honeybee genome project (www.ncbi.nlm.nih.gov/genome/guide/bee). The open reading frames had been predicted by the bee genome project, however, the function of the encoded proteins had not been suggested. Furthermore, there is no evidence that a polynucleotide

comprising the open reading frame of the mRNA had ever been produced for any of these molecules.

The genes encoding Xenospira1, Xenospira2, Xenospira3 and Xenospira4 comprise an exon covering the entire single open reading frame, whereas the gene
5 encoding Xenosin comprises at least one intron (see Figure 12).

EXAMPLE 8 - Expression of silk proteins in transgenic plants

A plant expression vector encoding a silk protein of the invention may consist of a recombinant nucleic acid molecule coding for said protein (for example a
10 polynucleotide provided in any one of SEQ ID NO's:11 to 21, 31 to 39, 48 to 55, 64 to 71, 74 or 75) placed downstream of the CaMV 35S promoter in a binary vector backbone containing a kanamycin-resistance gene (NptII).

For the polynucleotides comprising any one of SEQ ID NO's 11, 13, 15, 17, 19, 31, 33, 35, 37, 48, 50, 52, 54, 64, 66, 68, 70 or 74 the construct further may comprise a
15 signal peptide encoding region such as *Arabidopsis thaliana* vacuolar basic chitinase signal peptide, which is placed in-frame and upstream of the sequence encoding the silk protein.

The construct carrying a silk protein encoding polypeptide is transformed separately into *Agrobacterium tumefaciens* by electroporation prior to transformation
20 into *Arabidopsis thaliana*. The hypocotyl method of transformation can be used to transform *A. thaliana* which can be selected for survival on selective media comprising kanamycin media. After roots are formed on the regenerates they are transferred to soil to establish primary transgenic plants.

Verification of the transformation process can be achieved via PCR screening.
25 Incorporation and expression of polynucleotide can be measured using PCR, Southern blot analysis and/or LC/MS of trypsin-digested expressed proteins.

Two or more different silk protein encoding constructs can be provided in the same vector, or numerous different vectors can be transformed into the plant each encoding a different protein.

30 As an experimental example of plant expression, a codon-optimised version of AmelF4 (Xenospira4) (SEQ ID NO:76) was cloned into pET14b (Novagen), generating pET14b-6xHis:F4op, forming an in-frame translational fusion with a 6xhistidine at the N-terminal of the protein. The sequence encoding the protein "6xHistidine:F4op" was cloned into pVEC8 (Wang et al., 1992) under the control of the CaMV 35S promoter
35 and ocs polyadenylation regulatory apparatus, generating pVEC8-35S-6xHis:F4op-ocs. pET14b-6xHis:F4op was transformed into chemically-competent *E. coli* and pVEC8-

35S-6xHis:F4op was transformed into tobacco leaf discs by *Agrobacterium* mediated transformation. Proteins from antibiotic resistant *E. coli* (induced expression) and tobacco leaves were isolated and subjected to western blot analysis using the Tetra-Histidine antibody (Qiagen, Karlsruhe, Germany) for detection. The empty vectors
5 pET14b and pVEC8-35S-ocs were used as negative controls in there respective host backgrounds. As shown in Figure 13, these experiments resulted in the plant producing the Xenospira4 (AmelF4) protein.

EXAMPLE 9 - Fermentation and purification of silk proteins

10 Expression constructs were constructed after the silk coding regions of honeybee genes AmelF1-F4 (Xenospira1 to 4 respectively) and lacewing MalF1 genes were amplified by PCR and cloned into pET14b expression vectors (Novagen, Madison, WI). The resultant expression plasmids were then electroporated into *E. coli* BL21 (DE3) Rosetta cells and grown overnight on LB agar containing ampicilin. A
15 single colony was then used to inoculate LB broth containing ampicilin then grown at 37 °C overnight. Cells were harvested by centrifugation and lysed with detergent (Bugbuster, Novagen). Inclusion bodies were washed extensively and re-solubilised in 6M guanidinium.

This procedure yielded proteins mixtures with greater than 95% purity of the
20 honeybee proteins and greater than 50% purity of the lacewing MalF1 protein. Yields of up to 50% of the wet weight of the *E. coli* cell pellet were regularly obtained, indicating that the proteins are easy to express in this manner.

The solubilised honeybee recombinant proteins were applied to a Talon resin column prepared according to manufactures directions. They were then eluted off the
25 column in 100mM Tris.HCL, 150mM imidazole pH 8.

EXAMPLE 10 – Processing of silk proteins into threads

The honeybee and lacewing silk proteins have been readily made into threads using a variety of methods (see Figure 14) using the following procedure.

30 The anterior segment of the salivary gland from late final instar *Apis mellifera* was dissected under phosphate buffered saline and removed to a flat surface in a droplet of buffer. Forceps were used to grasp either end of the segment. One end was raised out of the droplet and away from the other at a steady rate. This enabled the drawing of a fine thread that rapidly solidified in air.

The honeybee and lacewing larval recombinant silk proteins formed threads or sheets after dehydration or concentration. For example, by dropping soluble protein into a butanol solution or by concentrating proteins on the Talon resin column.

Threads were also obtained after honeybee or lacewing recombinant silk proteins were mixed with an organic solvent (such as hexane) to concentrate them at the interface in the correct conformation, and then addition of a reagent to exclude them from the interface (such as butanol). The threads formed by this procedure had similar FT-IR spectra to the native silk indicating that they were comprised of the same coiled coil structure.

Silk proteins from other species described herein can also be processed by this procedure.

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 scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.

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.

25

REFERENCES

- Atkins E.D.T. (1967) *J Mol Biol* 24:139-141.
- Bendtsen J.D., Nielsen H., von Heijne G. and Brunak S. (2004) *J. Mol. Biol.* 340:783-795.
- Bini E., Knight D.P. and Kaplan D.L. (2004) *J. Mol. Biol.* 335:27-40.
- Craig C.L. and Riekel C. (2002) *Comparative Biochemistry and Physiology Part B* 133:493-507.
- Delorenzi M. and Speed T. (2002) *Bioinformatics* 18:617-625.
- 10 Deng Y., Liu J., Zheng Q., Eliezer D., Kallenbach N.R. and Lu M. (2006) *Structure* 14:247-255.
- Flower N.E. and Kenchington W.R. (1967) *Journal of the Royal Microscopical Society* 86:297.
- Grimaldi D. and Engel M.S. (2005) *Evolution of insects*. Cambridge University Press, New York.
- 15 Harayama S. (1998) *Trends Biotech.*, 16;76-82.
- Heimburg T, Schunemann J., Weber K., and Geisler N. (1999) *Biochemistry* 38:12727-12734.
- Hepburn H.R., Chandler H. D. and Davidoff M.R. (1979) *Insect Biochem.* 9:66.
- 20 Kneller D.G., Cohen F.E. and Langridge R. (1990) *J. Mol. Biol.* 214:171-182.
- LaMunyon C.W. (1988) *Psyche* 95:203-209.
- LaMunyon C.W. and Adams P.A. (1987) *Annals of the Entomological Society of America* 80:804-808.
- Lucas F. Shaw J.T.B. and Smith S.G. (1960) *J. Mol. Biol.* 2:339-349.
- 25 Lucas F. and Rudall K.M. (1967) In *Comprehensive Biochemistry* (Ed. Florkin M and Stotz H) Vol 26B pp 475-559 Elsevier Amsterdam.
- Lupas A., Van Dyke M. and Stock J. (1991) *Science* 252:1162-1164.
- McClelland J.L. and Rumelhart D.E. (1988) *Explorations in Parallel Distributed Processing* vol 13. pp 318-362. MITPress, Cambridge MA.
- 30 Needleman, S.B. and Wunsch, C.D. (1970) *J. Mol. Biol.*, 48;443-453.
- Quicke D.L.J., Shaw M.R., Takahashi M. and Yanechin B. (2004) *Journal of Natural History* 38:2167-2181.
- Reiser K., McCormick, Rucker R.B. (1992) *The FASEB Journal* 6:2439-2449.
- Rost B. and Sander C. (1993) *J. Molecular Biology* 232:584-599.
- 35 Rost B., Yachdav G. and Liu J. (2004) *Nucleic Acids Research* 32(Web Server issue):W321-W326.

Rudall K.M. (1962) In Comparative Biochemistry (Ed. By Florkin M and Mason HS) Vol 4, pp. 297-435. Academic Press, New York.

Rudall K.M. and Kenchington W. (1971) Annual Reviews in Entomology 16:73-96.

Silva-Zacarin E.C.M., Silva De Moraes R.L.M. and Taboga S.R. (2003) J. Biosci.
5 6:753-764.

Speilger P.E. (1962) Annals of the Entomological Society of America. 55: 69-77.

Wang M.B., Li Z.Y. et al. (1998). Acta Hort. 461: 401-407.

Yamada H., Shigesada K., Igarashi Y., Takasu Y., Tsubouchi K. and Kato Y. (2004)
Int. J. Wild Silkmoth and Silk 9:61-66.

DEMANDES OU BREVETS VOLUMINEUX

**LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVETS
COMPREND PLUS D'UN TOME.**

CECI EST LE TOME __1__ DE __2__

NOTE: Pour les tomes additionels, veuillez contacter le Bureau Canadien des Brevets.

JUMBO APPLICATIONS / PATENTS

**THIS SECTION OF THE APPLICATION / PATENT CONTAINS MORE
THAN ONE VOLUME.**

THIS IS VOLUME __1__ OF __2__

NOTE: For additional volumes please contact the Canadian Patent Office.

CLAIMS

1. A substantially purified and/or recombinant silk polypeptide, wherein at least a portion of the polypeptide has a coiled coil structure and wherein the polypeptide
5 comprises a sequence selected from:
 - i) an amino acid sequence as provided in any one of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:56, SEQ ID NO:57; SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:58, SEQ ID NO:59, SEQ ID
10 NO:5, SEQ ID NO:6, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:72 or SEQ ID NO:73; or
 - ii) an amino acid sequence which is at least 40% identical to any one or more
15 of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:56, SEQ ID NO:57; SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:7, SEQ ID
20 NO:8, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:72 or SEQ ID NO:73.
2. The polypeptide of claim 1, wherein the portion of the polypeptide that has a coiled coil structure comprises at least 10 copies of a heptad sequence *abcdefg*, and
25 wherein at least 25% of the amino acids at positions *a* and *d* are alanine residues.
3. The polypeptide of claim 1, wherein the portion of the polypeptide that has a coiled coil structure comprises at least 10 copies of a heptad sequence *abcdefg*, and at least 25% of the amino acids at positions *a*, *d* and *e* are alanine residues.
30
4. The polypeptide according to any one of claims 1 to 3 which comprises an amino acid sequence which is at least 80% identical to any one or more of SEQ ID

NO:1, SEQ ID NO:2, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:56, SEQ ID NO:57; SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:72 or SEQ ID NO:73.

5. The polypeptide according to any one of claims 1 to 3 which comprises an amino acid sequence which is at least 90% identical to any one or more of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:56, SEQ ID NO:57; SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:72 or SEQ ID NO:73.

6. The polypeptide according to any one of claims 1 to 5 which is fused to at least one other polypeptide.

7. An isolated and/or exogenous polynucleotide which encodes a silk polypeptide, wherein at least a portion of the polypeptide has a coiled coil structure, wherein the polypeptide comprises a sequence selected from:

i) a sequence of nucleotides as provided in any one of SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:76, SEQ ID NO:74 or SEQ ID NO:75;

ii) a sequence of nucleotides encoding the polypeptide according to any one of claims 1 to 6; or

iii) a sequence of nucleotides which is at least 40% identical to any one or more of SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:31, SEQ ID NO:32, SEQ ID
 5 NO:48, SEQ ID NO:49, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:54, SEQ ID
 10 NO:55, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:76, SEQ ID NO:74 or SEQ ID NO:75.

8. The polynucleotide of claim 7 which comprises a sequence of nucleotides which is at least 80% identical to any one or more of SEQ ID NO:11, SEQ ID NO:12,
 15 SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:37,
 20 SEQ ID NO:38, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:76, SEQ ID NO:74 or SEQ ID NO:75.

9. The polynucleotide of claim 7 or claim 8 which comprises a sequence of nucleotides which is at least 90% identical to any one or more of SEQ ID NO:11,
 25 SEQ ID NO:12, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:17, SEQ ID NO:18,
 30 SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:76, SEQ ID NO:74 or SEQ ID NO:75.

10. A vector comprising at least one polynucleotide according to any one of claims 7 to 9.

11. The vector of claim 10 which is an expression vector.

5

12. A host cell comprising at least one polynucleotide according to any one of claims 7 to 9, and/or at least one vector of claim 10 or claim 11.

13. The host cell of claim 12 which is a bacterial, yeast or plant cell.

10

14. A process for preparing the polypeptide according to any one of claims 1 to 6, the process comprising cultivating the host cell according to claim 12 or claim 13 under conditions which allow expression of the polynucleotide encoding the polypeptide, and recovering the expressed polypeptide.

15

15. A process for preparing a polypeptide according to any one of claims 1 to 6, the process comprising expressing in a cell-free expression system the vector of claim 11 comprising a polynucleotide encoding the polypeptide, under conditions which allow expression of the polynucleotide encoding the polypeptide, and
20 recovering the expressed polypeptide.

16. A plant cell comprising an exogenous polynucleotide, the polynucleotide encoding at least one polypeptide according to any one of claims 1 to 9.

25 17. A non-human animal cell comprising an exogenous polynucleotide, the polynucleotide encoding at least one polypeptide according to any one of claims 1 to 6.

18. An antibody which specifically binds a purified and/or recombinant silk
30 polypeptide, wherein at least a portion of the polypeptide has a coiled coil structure and wherein the polypeptide consists of a sequence selected from an amino acid sequence as provided in any one of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:22,

SEQ ID NO:23, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:56, SEQ ID NO:57;
 SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:42, SEQ
 ID NO:43, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:5, SEQ ID NO:6, SEQ ID
 NO:26, SEQ ID NO:27, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:60, SEQ ID
 5 NO:61, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:28, SEQ ID NO:29, SEQ ID
 NO:46, SEQ ID NO:47, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:72 or SEQ ID
 NO:73.

19. A silk fiber comprising polypeptides according to any one of claims 1 to 6
 10 wherein the silk fiber does not comprise a royal jelly protein produced by an insect.

20. The fibre of claim 19, wherein at least two of the polypeptides are crosslinked.

21. The fibre of claim 20, wherein at least two of the lysine residues of the
 15 polypeptides are crosslinked.

22. A copolymer comprising at least two different polypeptides according to any
 one of claims 1 to 6, wherein the copolymer does not comprise a royal jelly protein
 produced by an insect.

20

23. The copolymer of claim 22, wherein at least two of the polypeptides are
 crosslinked.

24. The copolymer of claim 23, wherein at least two of the lysine residues of the
 25 polypeptides are crosslinked.

25. A product comprising at least one polypeptide according to any one of claims
 1 to 6, the silk fiber according to any one of claims 19 to 21 and/or the copolymer
 according to any one of claims 22 to 24.

30

26. The product of claim 25, wherein the product is selected from the group
 consisting of: a personal care product, textiles, plastics, and biomedical products.

27. A composition comprising at least one polypeptide according to any one of claims 1 to 6, the silk fiber according to any one of claims 19 to 21 and/or the copolymer according to any one of claims 22 to 24, and one or more acceptable carriers.

28. The composition of claim 27 which further comprises a drug.

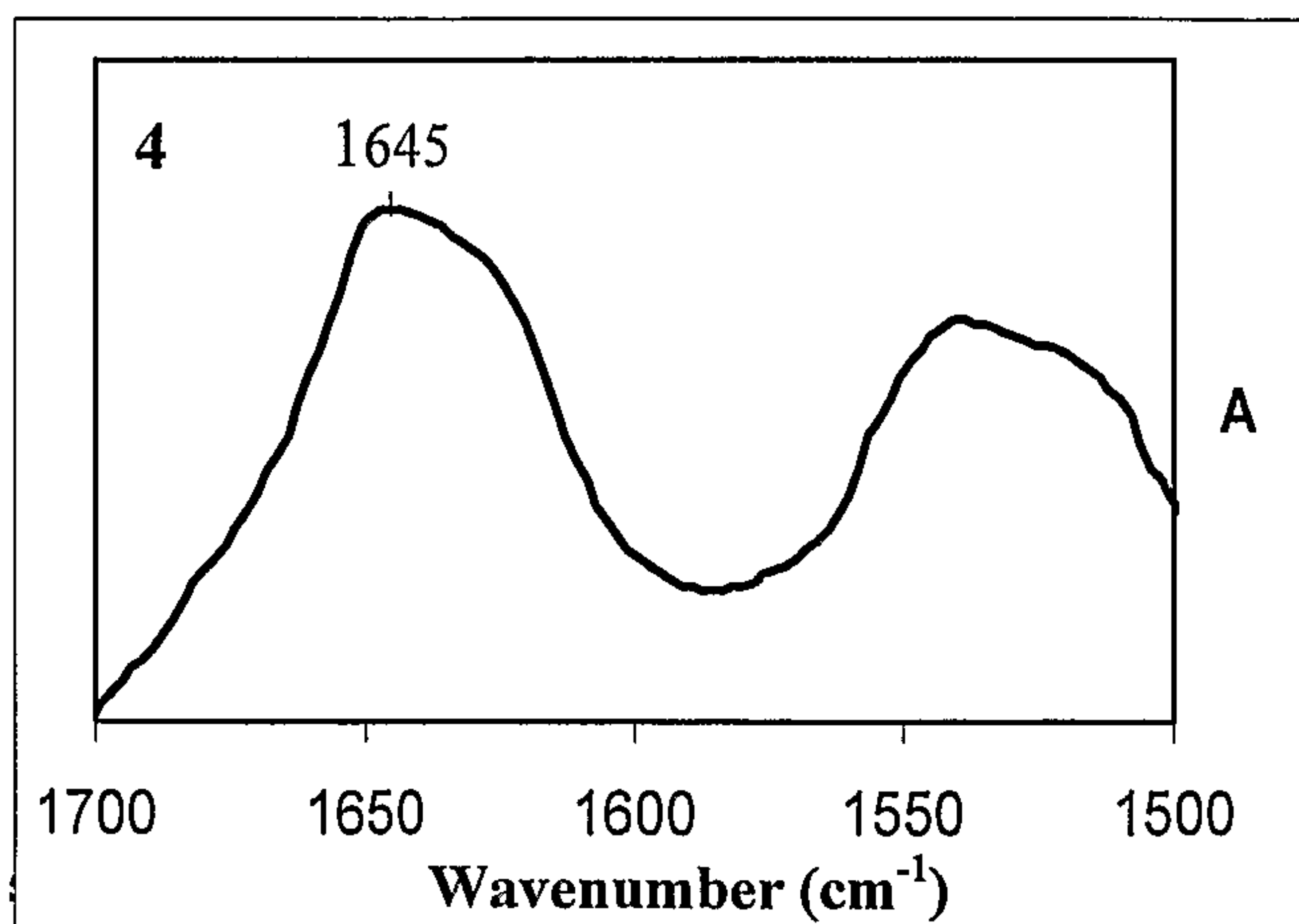
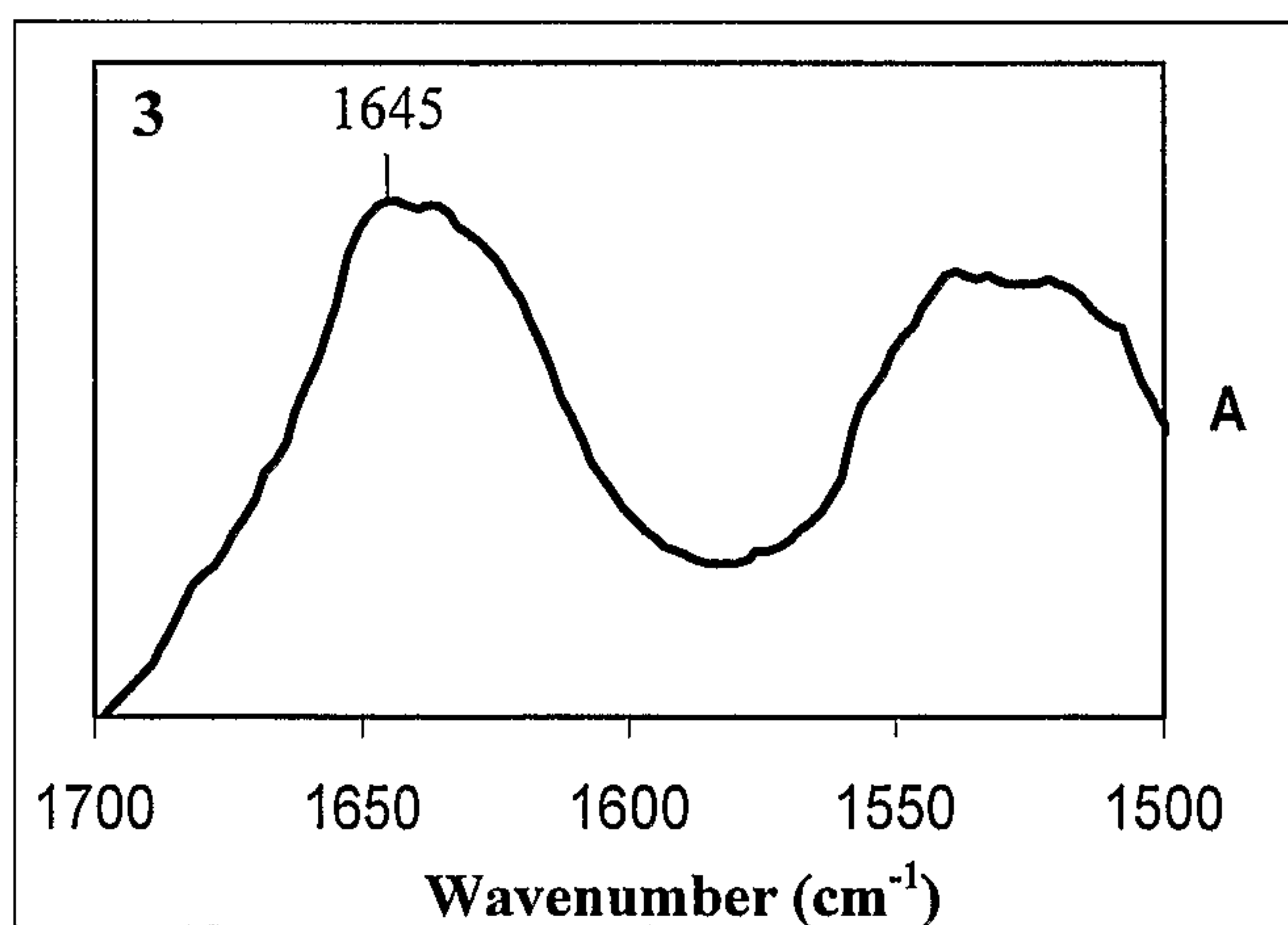
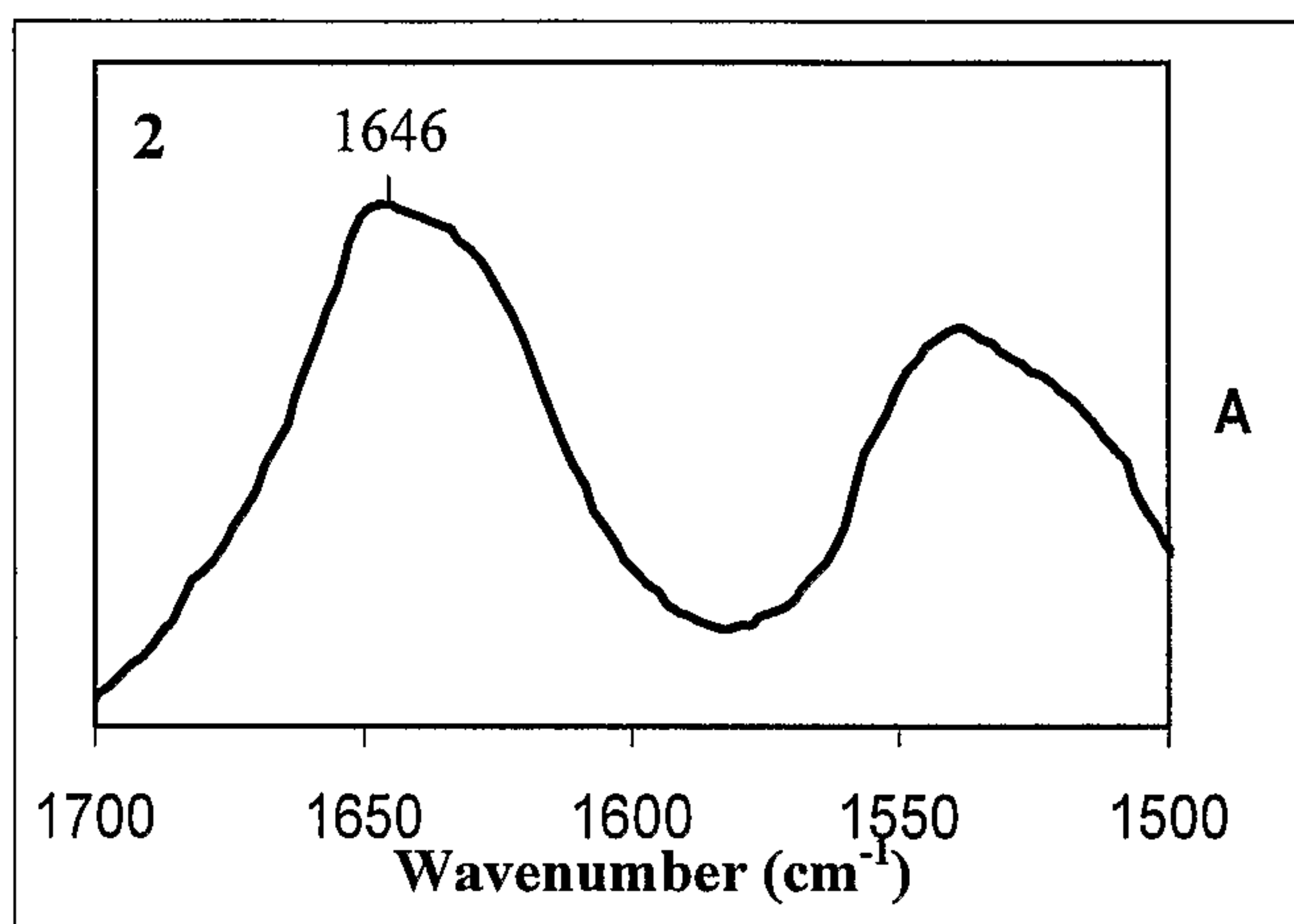
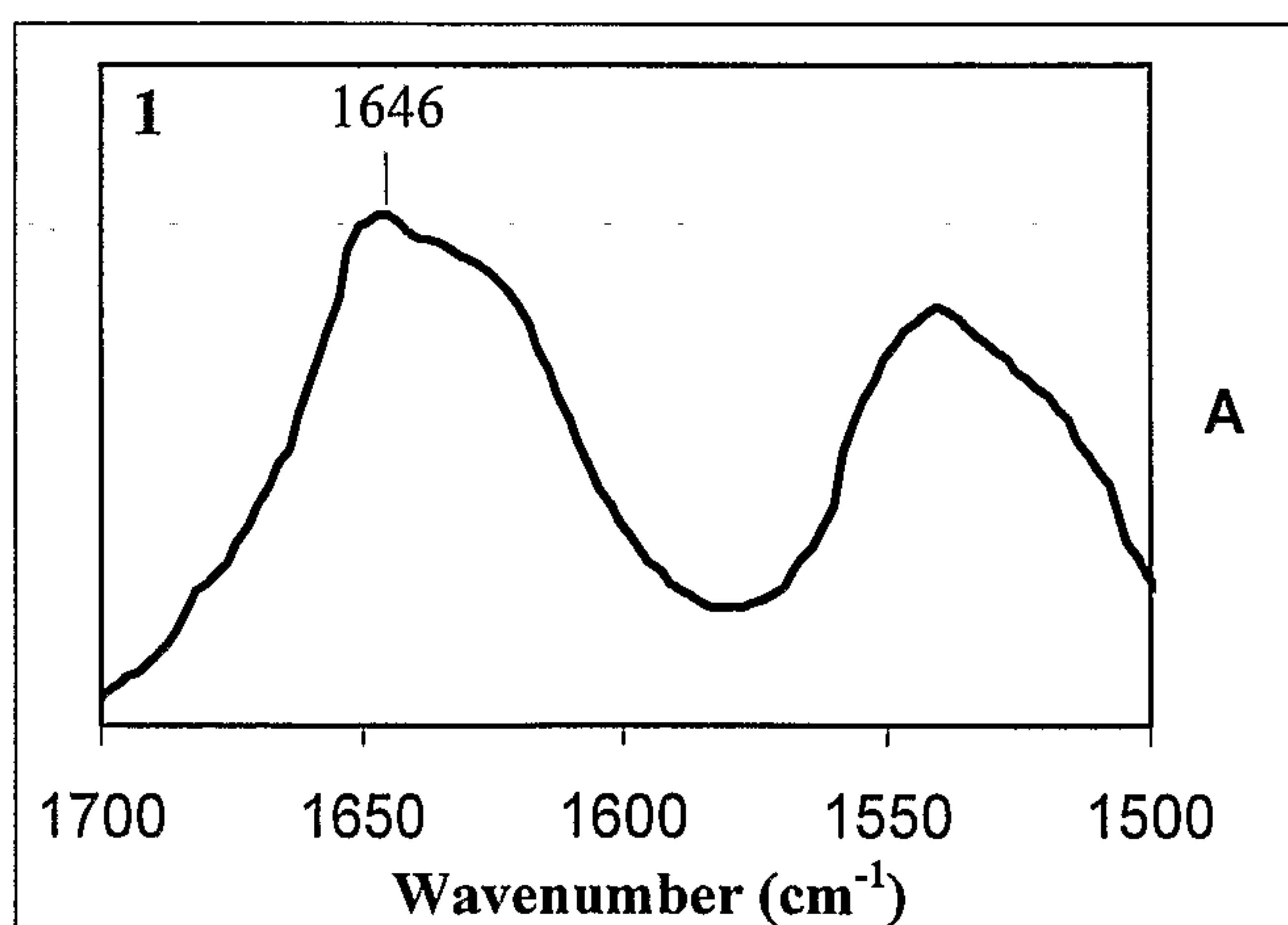
29. The composition of claim 27 for use as a medical device or a cosmetic.

30. A composition comprising at least one isolated and/or exogenous polynucleotide according to any one of claims 7 to 9, and one or more acceptable carriers.

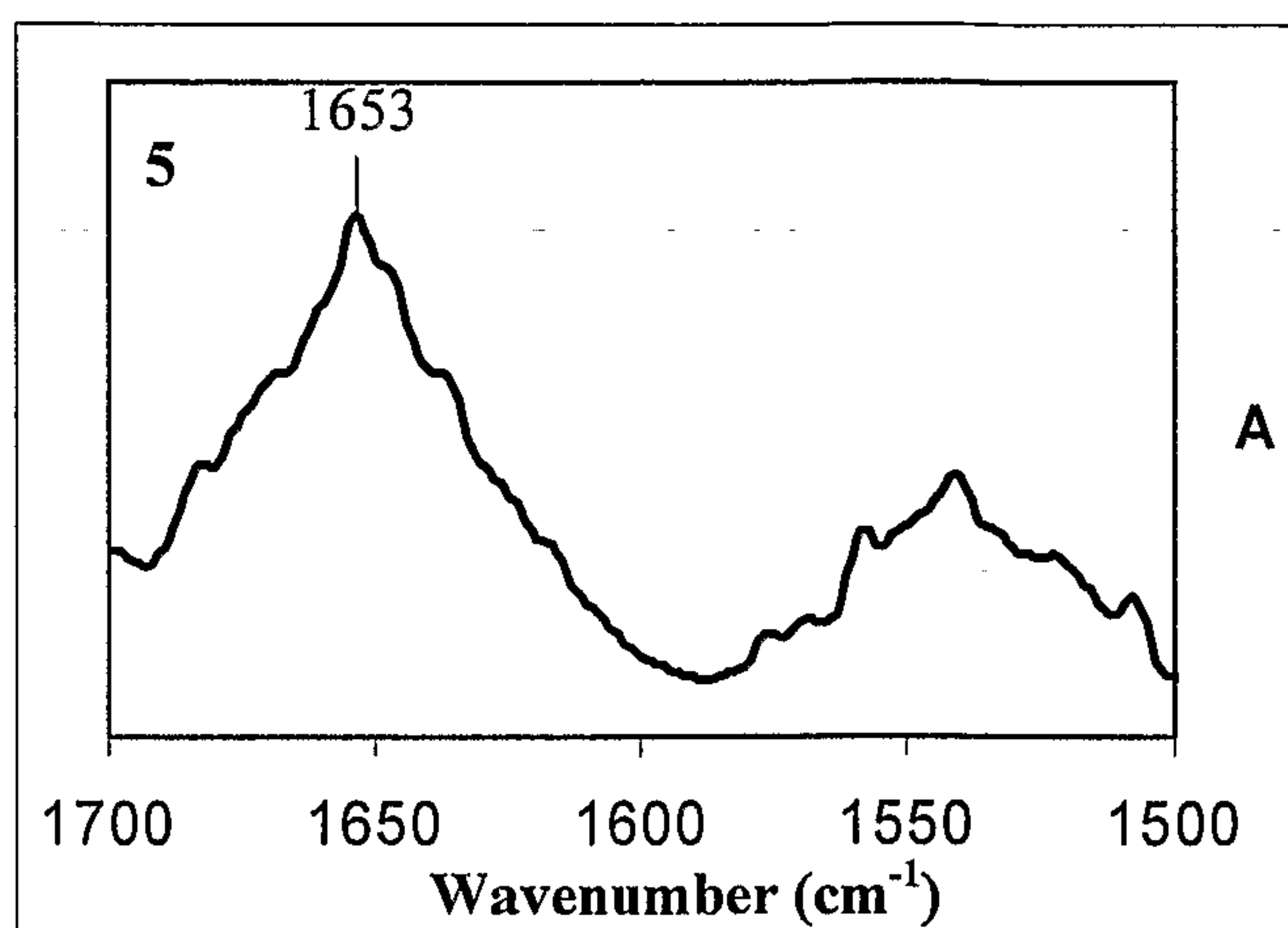
31. The composition of claim 30 which further comprises a drug.

32. Use of at least one polypeptide according to any one of claims 1 to 6, the silk fiber according to any one of claims 19 to 21 and/or the copolymer according to any one of claims 22 to 24, as a pharmaceutically acceptable carrier.

33. A kit comprising two or more of the following: at least one polypeptide according to any one of claims 1 to 6; at least one polynucleotide according to any one of claims 7 to 9; at least one vector of claim 10 or claim 11; at least one silk fiber according to any one of claims 19 to 21; and the copolymer according to any one of claims 22 to 24.



2 / 22

**Figure 1**

3 / 22

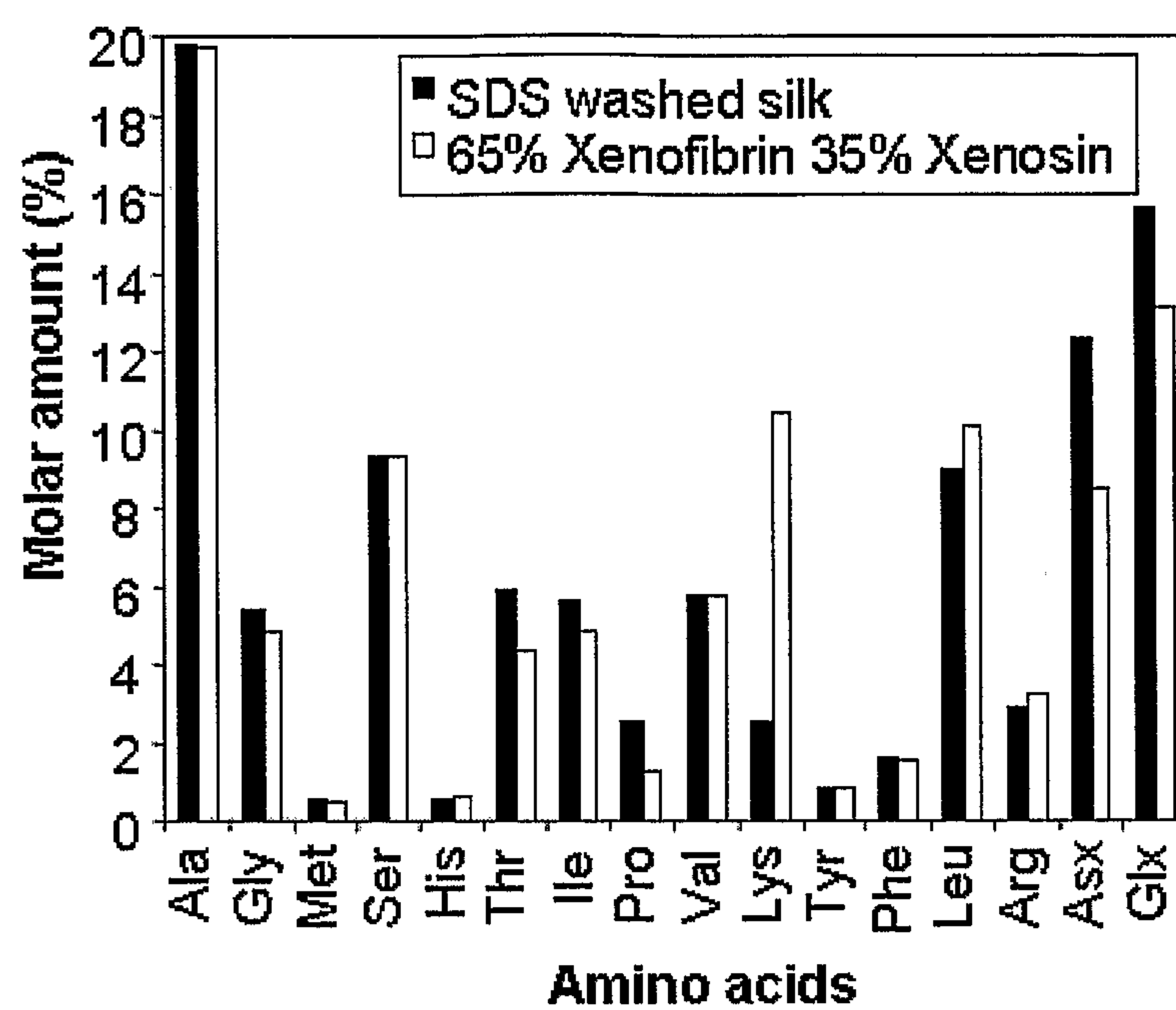


Figure 2

4 / 22

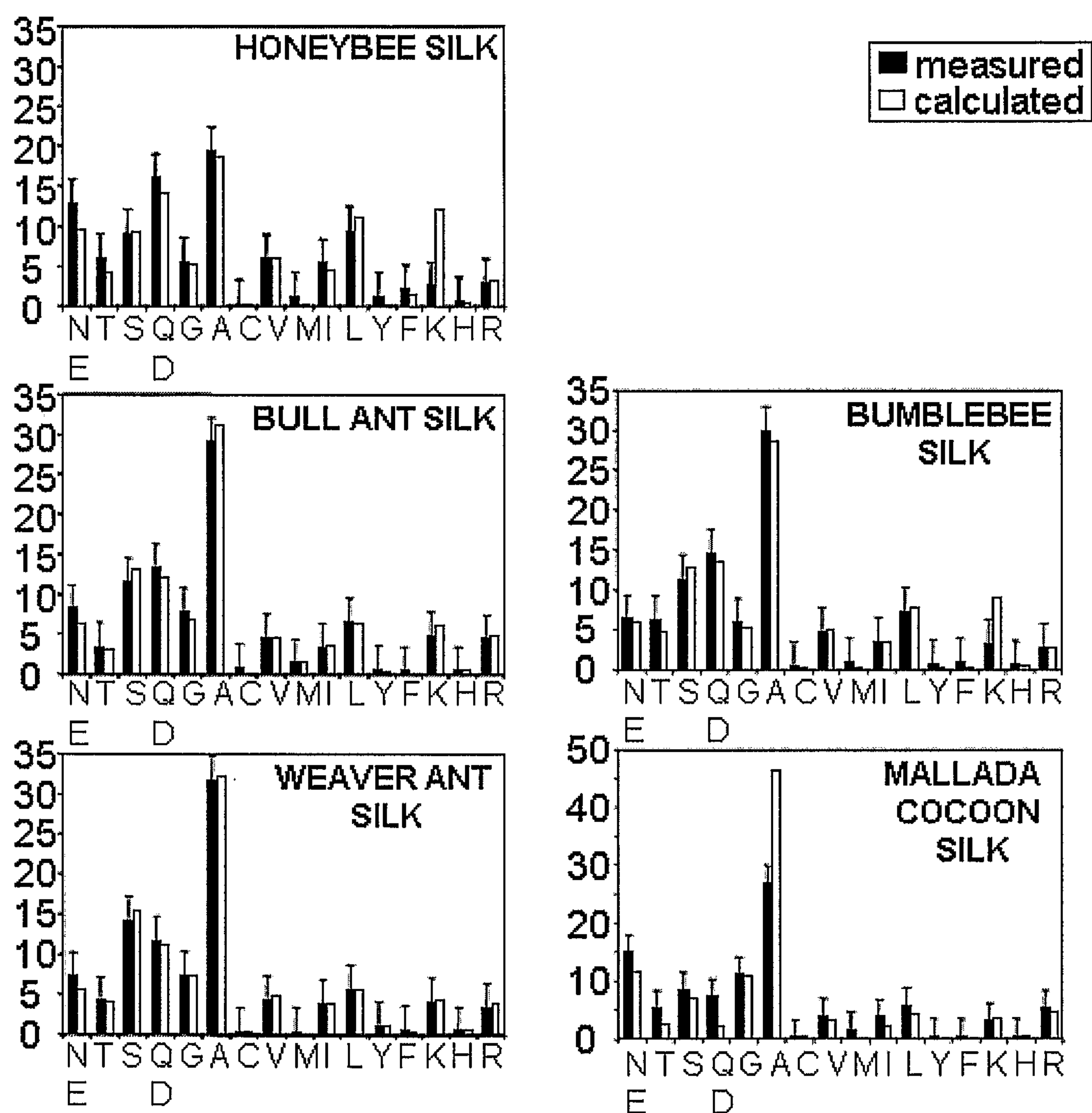


Figure 3

5 / 22

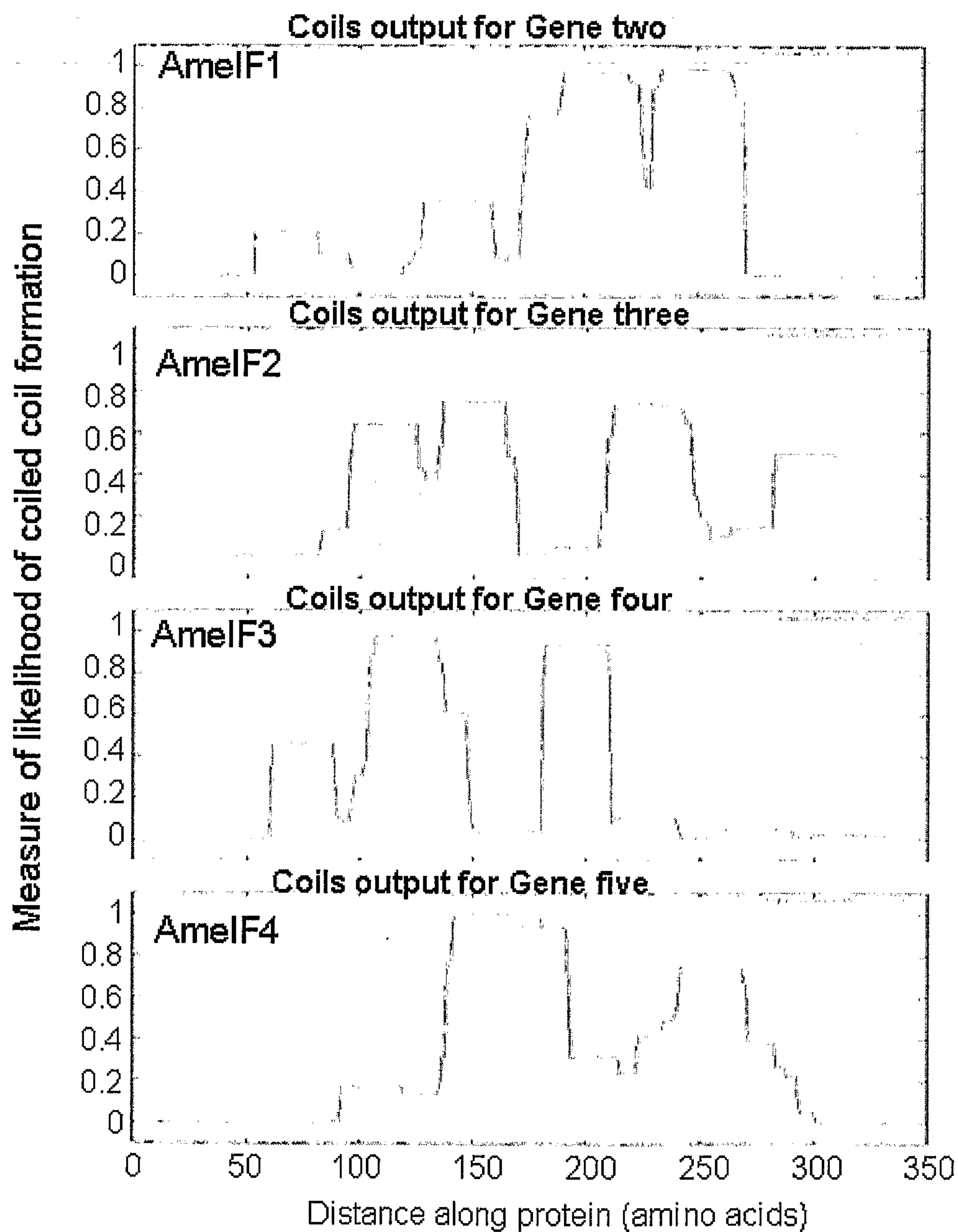
**Figure 4**

Figure 5

1 11 21 31 41 51 61 71

BBF1 AVASQKDATLQADASAAAAAARASADQSASLAQQSASLQSKAAARAKSAEESAAATAKAELQAESIAASASS

AmelF1 ALASQKDAVLQAQAAASAASEARAAADLTAKLSQESASVQSQAAAKGKETEEAAVGOARAGLESVSMASATS

BAF1 TAEEQREALEMLTSLADKNAEARILADDTAVLVQGSAAEQSVAAAKTVAVEEESASLDAAAVEAEVAAATSKS

GAF1 TVEEQREALQLLTASAEKNAEARSLADDAAVLVQGAEEAQSVAAAKTVAVEQGSNSLDAAAAEAEAAAAASRV

BBF2 ARAAAAAATKQAKDTEAAEAGAAAAIAIAIAKREEAIKASEIANQLLTNAAKAAEATVSATKRAAQLTAAASA

AmelF2 ARAASVAAAAQAKNTEAAEAGANAALAAIAIAKREEAIKASEIANQLLTNAAKAAEATVSATKRAAQLTAAAKE

BAF2 ANEAEILAEQAATSAQAGAVANSAAERAIAMEMADRTEYIAALVTTKAAKAAEATMAATARATAAASASKI

GAF2 GNSAESLARAQASSSASAGAKANALIKQSTAAIEITEKAEYLASIVATKAAKAAEATAAATARATAVAEAAKV

BBF3 SAKVKASALALAEAFLRASAAFAAASAKAAAVKEATQAQLLAQEKALIALKTQSEQQAASARADAAAAAASVS

AmelF3 AAVVKASALALAEAYLRASALSAAASAKAAALKNAQQAQLNAQEKSLAALKAQSEEEAASARANAATAATQS

BAF3 LAMIKKIAMARSSAYVQAAWASAQASADALASARVAQASQEAEEAKGRAASEALSRATEASSRADAAAAATLD

GAF3 LAIIKKIVALLISAYVRAEEAARASAEALATVRAAEQAKIAEAKGRAAAEALSELVEASQKADAAAAGTTD

BBF4 AAEAQAAAAAQAKSAAAAANAGESSNSAAALVAAAAAQKAAAAAAATKASLEAADAAEEAESEAVALARA

AmelF4 AAESQALAASQSKTAANAAIGASELTNKVAALVAGATGAQARATAASSSALKASLATEEEAAEEAEAAVADAKA

BAF4 AAEAQAAAAGQAEVAAKSCELADKTTAKAVAMVEAAAEAEIEVANQEVAAVKLSTWAAKAARIVEEDSAAVRA

GAF4 AAEAQAAAAGQAAIAAKSCALAAKSTAQAVALVEKVARAEVDLAESARKATRLSAEAAKAAAEVEKDLVLGR

81 91 101 111 121 131 141

BBF1 NAREAAASAKASASAMSSAAVQAKLAEKTAKNQALASEEAKLKAA~~AAASAAAAASAAEAALKAERIAEEA

AmelF1 AAKEASTAAKAAASALSTAVVQAKIAERAAKAEAVASDEAKAKAI~~AAANLAAEASVAAEAALKAEKVAAEA

BAF1 SAGQALQSAQTAASALRTSARSALTALKKLARLQGAASSNAARMME~~KALAATQDANAAAQQAAMAEASAAEA

GAF1 SAQQALQAAQTSAAAIQTAAGSALTALKKLARKQEAESNNAAEQAN~~KALALSRAASAATQRAVAAQNAAAAS

BBF2 AAKASASASEASAEAOVRANAENIAKKASAAEAKAAEAQVKAE~~LAKKAAAGFLAKARLAASAESEATKL

AmelF2 ATRASAAAAEAATEAQVKANADSIITKRAAIAEAQAAEAQVKAA~~IARKSAANFLAKAQIAAAAESEATKL

BAF2 SSQESAASAANAANAEEAKANAASI IANKANAVLAEAAVLAATAA~~KAKESAMKSLSAAQAAKAQARNAEA

GAF2 SSEQFAAEARAAADAEAKANAASI IANKANAVLAEATGLSASAG~~KAQQSATRALQAARAAKAQAEELTQK

BBF3 ALERAQASSRAATTAQDISDLEKRVATSAAAEAGATLRAEQSAA~~QSKWSAALAAQTAAAAAAIEAKATAS

AmelF3 ALERAQASSRLATVAQNVASDLQKRTSTKAAAEAAATLRQLQDAE~~RTKWSANAALVSAAAAAAETKTTAS

BAF3 AMDRTMENARAANAQAQOASQGAENANRSAAAILAALLRIAESA~~LNNEAAVNAAAAAAAASALQAKANAA

GAF3 AIERTYQDARAATSAQTKASGEAENANRNAAATLAAVLISIKAASGGGTRAAVDAAAAAAAAAALHAKANAV

BBF4 ASAKAEALASTAAANTRAAALQAEKSNELAQAEAAAAEAQAKAA~~AAAKATQLALKVAETAVKTEADAAAA

AmelF4 AAEKAESLAKNLSASARAALSSERANELAQAESAAAAEAQAKTA~~AAAKAEIALKVAEIAVKAEADAAAA

BAF4 AAGKLLLAAARAAAAAERRANEESAAANELAQASSAAAAEAEEKAN~~AGREAAAAALIAEAAVAIEQEAVIL

GAF4 AAGKLNLAARAGSKAQERANEDSIEANELAQATAAAGAEAEKAN~~AAQEAGASALAIQAALNIEQETVKL

151 161 171 181 191 201 210

BBF1 IAKAAAAKAAARAAAAALNSAKEAATSSARSAAEAEAKSEVAILISELDKKSREVAASASAKARAA

AmelF1 IARAASAKAAARAAAAALASSKEAATASARNAAESEARNEVAVLIAEIDKKSREIDAASSLNARAA

BAF1 AAIAAAKQSEARDAGAEAKAAMAALITAQRNLVQANARAEMASEEAELDSKSRASDAKVNARAA

GAF1 AASAGAAQAEARNAYAKAKAAIAALTAAQRNYAAAKASASAGSVVAEQDAQSRAADAENVAVAQAA

BBF2 AAEAEVALAKARVAVDQSQSAQATATAQAATAVQLQSQAANAEASAVAQAETLLVTAEAVSAAEAE

AmelF2 AAEAVVALTNAEVAVNQARNAQANASTQASMAVRVDSQAANAEAAVAQAETLLVTAEAVAAAEAE

BAF2 SAEAQIKLSQARAAVARAADQAVCSSQAQAASQIQSRASASESAASAQSE TNTAAAEAVATADAE

GAF2 AAQILVLIAEAKAAVSRASADQSVCTSQAQAASQIQSRASAAESAASAQSEANTIAAEAVARADAE

BBF3 SESTAAATSKAAVLTA DTSSAEAAAAEAQSASRIAGTAATEGSANWASENSRTAQLEASASAKAT

AmelF3 SEANAAAKKAAAIASDADGAERSASTEQAQSAAKIESVAAAEGSANSASEDSRAAQLEASTAARAN

BAF3 SQATARAAGQASTAAEEAQSAQEADKNAELTTVMLEKASADQQAASARADYYTASTEAEAAQAS

GAF3 SQATSKAAAEARVAAEEAASAQASASASAQLTAQLEEKVSADQQAASASTDTSAAIAEAEAAALAS

BBF4 AVAAAKARAVADAAASRATAVNIAEAEERDSAQAENTAGVAQAALAAAEAQDSCIGAAATPRHSS

AmelF4 AVAAAKARAVADAAARAAAVNAIAKAEEEASAQAENAAGVLQAAASAAAEASRAAAAAAATSEAA

BAF4 ARKAQDARLNAEAAAAAAMNARVIASAESEASEDLENRASVARASAAGAAEAKAIATDAGATAEIA

GAF4 TRQAQNTRLRSENILAAASNARAIASAEAEASSDLNNRANAARSNARAAAE TRAVATEAASTAEIA

Figure 6

8 / 22

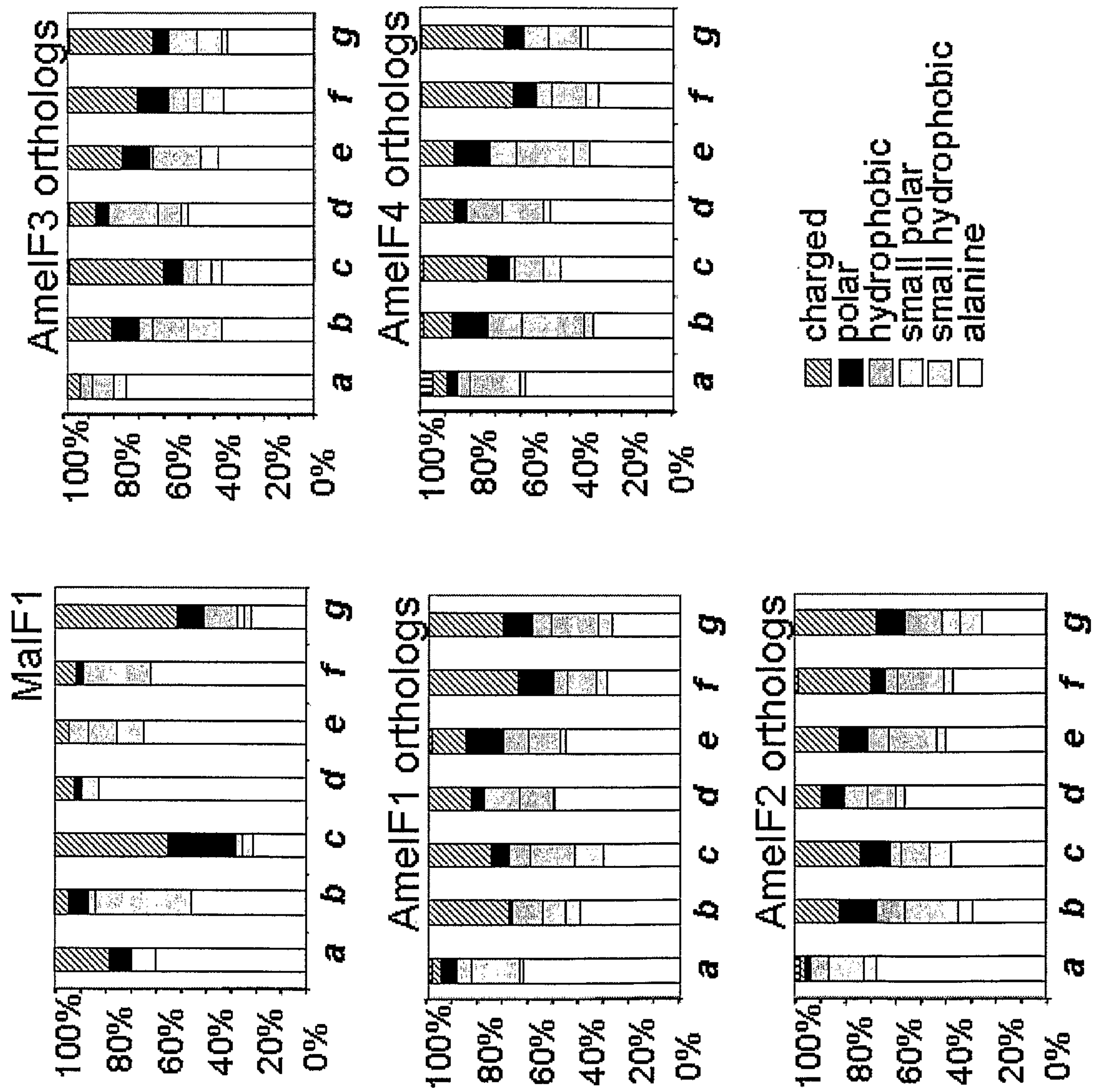


Figure 7

9 / 22

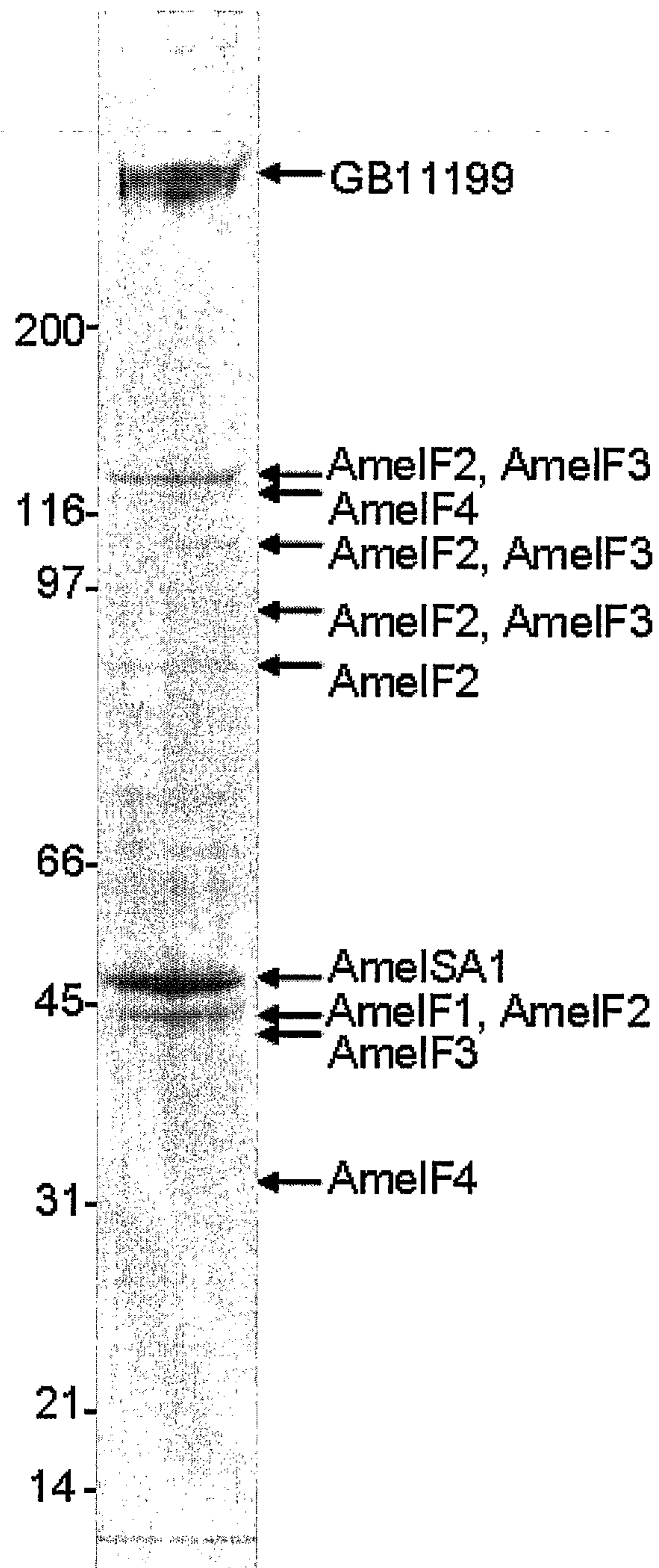


Figure 8

10 / 22

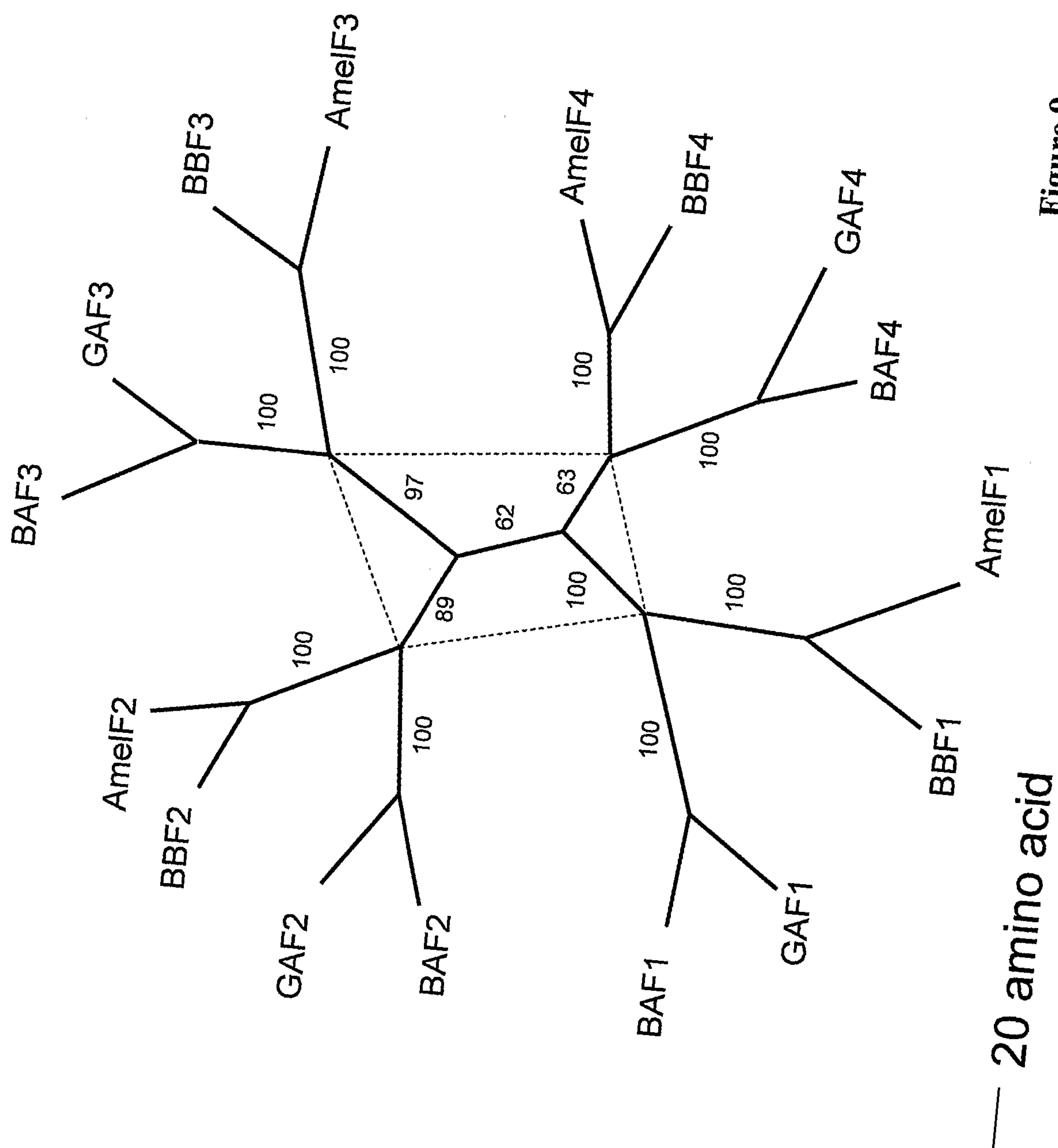


Figure 9

11 / 22

A)

Xenospiral (SEQ ID NO:2)

MKIPVLLATCLYL CGFASAGLEGPGNSLP ELVKGSASATASTAVTARSGLRAGQVALASQKDAVLQAQAA
 ASAASEARAAADLTAKLSQESASVQSQA AAKGKETEEAAVGQARAGLESVSMAASATSAAKEASTAAKAA
 ASALSTAVVQAKIAERA AKA EAVASDEAKAKAIAAANLAAEASVAAEAALKA EKVAEEA IARAASAKAAA
 RAAAAALASSKEAATASARNAAESEARNEVAVLIAEIDKKSREIDAASSLNARAAAKASSRNVETATIGA
 NINSSKQVVSIPVEIKKFSEPEVSTSWREDEEVTKEKKEHINLNDFDLKS NVF

Xenospira2 (SEQ ID NO:4)

MKIPAI FVTSL LVWGLAEGRVINHESLKTSEDIQGGYSAGIVGDGSDALGSS IENAQKVARAAENVGLNL
 ELGAGARAASVAAAAQAKNTEAAEAGANAALAAAI AKREEAIKASEIANQLLTNAAKAAEATVSATKRAA
 QLTAAAKEATRASAAAAEAATEAQVKANADSIITKRAAIAEAQA AEAQVKA AIARKSAANFLAKAQIAA
 AAESEATKLAAEAVVALTNAEVAVNQARNAQANASTQASMAVRVDSQAANAEAAVAQAETLLVTAEAVA
 AA EAEVANKAATFAKQIVNEKKIHVAKLE

Xenospira3 (SEQ ID NO:6)

MQIPTFVAICLLTSGLVHAGVEEFKSSATEEVISK NLEV DLLKNVDTS AKRRENGAPVLGKNTLQSLEKI
 KTSASVNAKAAAVVKASALALAEAYLRASALSAAASAKAAAALKN AQQAQLNAQEKSLAALKAQSEEEAA
 SARANAATAATQSALERAQASSRLATVAQNVASDLQKRTSTKAAAEAAATLRQLQDAERTKWSANAALV
 SAAAAAAETKTTASSEANAAAKKAAAIASDADGAERSASTE AQSAAKIESVAAAEGSANSASEDSRAAQ
 LEASTAARANVAAAVGDGAI IGLGEEAGAAQ LLAQAKALAEVSSKSENIEDKKF

Xenospira4 (SEQ ID NO:8)

MKIP SILAVSLLIWGLASGAREEVETRDKTSTVVKSEKVEVVAPAKDELKLTSEPIFGRRVGTGASEV
 ASSSGEAI AISLGAGQSAAESQALAAASQSKTAANA AIGASELTNKVAALVAGATGAQARATAASSSALKA
 SLATEEAAEEAEAAVADAKAAAEKAESLAKN LASASARAALSSERANELAQAESAAAAEAQAKTAAAKA
 AEIALKVAEIAVKAEADAAAAA VAAAKARAVADAAAAARA AAVNAIAKAE EESAQAENAAGVLQAAASAA
 AESRAAAAAAATSEAAAEAGPLAGEMKPPHKKWERIPVKKEEWKTSTKEEWKTTNEEWEVK

Xenosin (SEQ ID NO:10)

MKYMLLLLSIFICAHIVCAGVNTTELKKGDELKEESY EKSES KSLKEIKEERASKSKSERLKIREEKREEE
 EKSKSLNLVVREKITKLSSWLKEEKDISPLIEEKNGKGLLGLEDVTDELNIALKSLKEGKKFDTWKFEK
 GSEDVRSLEELDTSVVELLKL IKEGKTDHG AIDLEKNGKVLVDLEKISENILETCGSQKKTVEVVDDKDK
 KWNKESGWKKNLNDLDWKDLDKDKVGGGLLGGLSGLLNSLKSEKGLLGLLNKNQIELLIPLISEIKKKN
 IDFNLFDSVDSVERNLDLKLFTSSVSKVTELLNKGIDITOTILNAKNGDEEDLSGKELKNVKGIFGLIGSL
 KRS LGLENILNLPFKRIPLKI

B)

BBF1 (SEQ ID NO:23)

MKIPALLVTCLYLWGFASAGQSSPLLEIVQGSASATASTAVTARSGLRAGQVAVASQKDATLQADASAAA
 AAAARASADQSASLAQQSASLQSKAAARAKSAEESAAATAKAELQAESIAASASSNAREAAASAKASASA
 MSSAAVQAKLAEKTAKNQALASEEAKLKAAAAASAAAAASAAEAALKAERIAEEAIAKAAA AKAAARAA
 AAALNSAKEAATSSARSAAEAEAKSEVAILISELDKKSREVAASASAKARAAAAASSRNAETAVIGANIN
 VAKEVLAIPIEPKKLPEPELALKEENVAVASSESEVKVETSSEAWSI

BBF2 (SEQ ID NO:25)

MKIPAILVTSL LVWGG LAEGHVVKRD KELKAPALPELLGDGSDTLGASMENGIKVARASQNVGLRTELNA
 AAAAAAAATKQAKDTEAAEAGAAAAIAIAIAKREEAIKASELASKLLTAAAGSSEAAVSATVRAAQLTA
 AASAAKASASASEASAE AQVRANAEANI AKKASAAEAKAAAE AQVKAE LAKKAAAGFLAKARLAASAES
 EATKLAAEAEVALAKARVAVDQSQSAQATATAQAATAVQLQSQAANAEASAVAQAETLLVTAEAVSAAEA
 EAATKATSWGEECHQREKVTFSEDRLNERQDNW

12 / 22

BBF3 (SEQ ID NO:27)

MQIPAI FVTCLLTWGLVHAGSVELGAPKQESVLVEQLLLKNVETSAKRKENGAPKLGESTAAALASTKAT
 AAAEAKASAKVKASALALAEAF LRASAAFAAASAKAAA AVKEATQAQLLAQEKALIALKTQSEQQAASAR
 ADAAAAA AVSALERAQASSRAATTAQDISSDLEKRVATSAAAEAGATLRAEQSAAQSKWSAALAAQTAAA
 AAAIEAKATASSESTAAATSKAAVL TADTSSAEAAAAAEQAQSASRIAGTAATEGSANWASENSRTAQLEA
 SASAKATAAAAVGDGAI IGLARDASAAAQAAAEVKALAEASASLGASEKDKK

BBF4 (SEQ ID NO:29)

MKIP SILAVSLLVWGLASAGKPLIANAQIGKVKTETSSSSSEIETLVSGSQTLVAGSETLASESEALASKS
 EALTSEAEIASVTTKDELILKGEAITGKKLGTGASEVAAASGEAIIATTLGAGQAAAEQAQAAAAQAQKSAA
 AAAANAGESSNSAAALVAAAAAAQ GKAAAAAAATKASLEAADAAEEAESAVALARAASAKAEALASTAA
 AANTRALQAEKSNELAQAEAAAAAEQAQAAAAAAKATQLALKVAETAVKTEADAAAAAVAAAKARAVAD
 AAASRATAVNAIAEAEERDSAQAENTAGVAQAALAAAEAQDSCIGAAATPRHSSSYAWWKL RITSLIVIL
 SPRNRRT

BBSA1 (partial) (SEQ ID NO:30)

GNSESGENWKNGESSES GKNWRNSGSSSES GKNWKNGGSSSES NKHWKNGGSSSES GEKWKNSESSES GKNWK
 NSGSSSES GKNWKNGGSSSES NKHWKSGGSSSES GEKWKNSES GKNKGKSSKSSES WKSNSKNDGSWKSSEE
 SEKWKGDKAVAEDSV SINWADVKEQISNIATSLEKGGNLEAVLKIKKGEKKISSLEEIKEKISVLLKWIQ
 EGKDTSSLLDLKEGSKDIASLKEIKGKILLIVKLVNEGKDTSGLLDLEASGKVILELQSAIEKVLVKSEK
 VTKVSEVSGLVKSKTVSDIKPLQAVIPLILELQKTDINLSTLNKWS TVNVNSIDKERVTKTVPVLLQSMK
 GGEDIQNLLSAKGAKKLGISALDLQAVQ GALGVVGKLSSGGALNSKGLLNLDKGASVLGAGKIGGLIPLP
 KL

BAF1 (SEQ ID NO:41)

MKIPAI IATSLLLWGFASASGPRLLGGRSAASASASASAEASAGGWRKSGASASASAKAGSSNILSRVGA
 SRAAATLVASA AVEAKAGLRAGKATAEEQREALEMLT LSADKNAEARILADDTAVLVQGSAAEQSVAAAK
 TVAVEEESASLDAAAVEAEVAAATSKSSAGQALQSAQTAASALRTSARSALTALKLARLQGAASSNAARM
 MEKALAATQDANAAAQQAMAAESAAAEAAAI AAKQSEARDAGAEAKAAMAALIT AQRNLVQANARAEMA
 SEEAE LDKSRASDAKVN AVARAASKSSIRDELIEIGAEFGKASGEVISTGTRSNGGQDAIATAEASSS
 ASAVGIKKTSGHWGSGKWSRVSKGKGWASSNADADASSSSIIIGGLKRGGLGSEASAAASAEAEASAGTL
 LL

BAF2 (SEQ ID NO:43)

MKIPAILVTSLLAWGLASGRVIESSSSASAQASASAGSRGLLGKRPIGKLEWGKEEKKLEELDEESLNEA
 ALKVGIKNGGLDVAKGAAVLEAAMSDVATLT DQRS LVDLGLGPVANEAEILAEQAATS AQAGAVANSAA
 ERAIAAMEMADRTEYIAALVTTKAAKAEATMAATARATAAASASKISSQESAASAANAANA EAKANAAS
 I IANKANAVLAEEAAVLAATAAKAKESAMKSLSAAQAAAKAQARNAEASAEAQIKLSQARA AVARAAADQ
 AVCSSQAQAASQIQSRASASESAASAQSETNTAAAEAVATADAEAAAQAEAWVMSLKN DLWLHLNMKGEA
 KAEGEAVSISKGHRGGIRSGSISEASAEASSNVSMGGRHGRKDLVSEALAGASAGSSADSL

BAF3 (SEQ ID NO:45)

MKIPAILVTSFLAWGLASGNLLKESKASASASASASASARASGKKNLHVLPLPKKSEHGIVIDKSVFDIKDV
 VLSAVDEINGAPKLGLGWKKVSMGVERAEANAAAAAEALAMIKKIAMARSSAYVQA AWASAQASADALAS
 ARVAQASQEAAAEAKGRAASEALSRAIEASSRADAAAAATLDAMDRTMENARAANA AQTQASGQAENANRS
 AAAILAALLRIAEASALNNEAAVNAAAAAAAASALQAKANAASQATARAAGQASTAAEEAQSAQEAAADKN
 AELTTVMLEKASADQQAASARADYYTASTEAEAAAQASAINALRDGIVVGMGNDAGASAQAMAQVEALAR
 ASEHKALGEKKKGLVWGYGSKGSSSASASASASAEASSRLGKDW

13 / 22

BAF4 (SEQ ID NO:47)

MKIPAILATSLLIWGLVGLVASELESEASAAASAQAEASSSSGRSGKLSASQASASASASASAGSRGGSKGGW
 GQLRRGDVKSEAKSAAAIAVEGAKIGTGIGNTASASAEALSRGLGIGQAAAEQAQAAAGQAEVAAKSCEL
 ADKTTAKAVAMVEAAAEAEIEVANQEVAVKLSTWAAKAARIVEEDSAAVRAAAGKLLLAARAAAAAERR
 ANEESEAANELAQASSAAAAEAEAKANAGREAAAAALAEAAVAIEQEAVILARKAQDARLNAEAAAAA
 AMNARVIASAESEASEDLENRASVARASAAGAAEAKAIATDAGATAEIAAYSWAKKGELINPGPLPKIIS
 VNADLSKSEVEAMKITRGQVQEVKKISTHKGWGWGKEGRSKVSSNASARASASANAAGSLGSKWGRQL
 SASSASADANAEDSOLLKVV

GAF1 (SEQ ID NO:57)

MKIPAIITATLLWGFADASKSYLLGSSASASASASASASAGGSTGGVGVGSVISGGNNIIRGASTTSVT
 LAAAAEAKAALNAGKATVEEQREALQLLTASAEKNAEARS LADDAVLVQGAAEQSVAAAKTVAVEQG
 SNSLDAAAAEAEAAAAASRVSAQQALQAAQTSAAAIQTAAGSALTALKLARKQEAESNNAEQANKALAL
 SRAASAATQRAVAAQNAAAASAASAGAAQAEARNAYAKAKAAIAALTAAQRNYAAAKASASAGSVVAEQD
 AQSRADAENVNAQAARASVRNQEIVEIGAEEFGNASGGVISTGTRSSGGKGVSVTAGAQASASASATS
 SSSSSSGINKGHPRWGHNWGLGSSEASANAEESSASSYSS*

GAF2 (SEQ ID NO:59)

MKIPAIFVTSLWGLASGGVIGPDTSSSSQASASASASASASASSSASIGYNELHKSINAPALAVGVKN
 GGVDVAKGAAVVESAIISDVSTLTDDRTLNLGLAIIGNSAESLARAQASSSASAGAKANALIKQSIAAIEIT
 EKAEYLASIVATKAAKAAEATAAATARATAVAEAAKVSSEQFAAEARAAADAEAKANAASITANKANAVL
 AEAATGLSASAGKAQQSATRALQAARAAAKAQAEILTQKAAQILVLIAEAKAAVSRASADQSVCTSQAQAA
 SQIQSRASAAESAASQAQSEANTIAAEAVARADAEAAQAQAWAESFKRELSSVLEAEANASASASAGAL
 ASGSSSSGASSSADASAGASSYGLGGYRHGGSFSEASAAASAASRAEAA

GAF3 (SEQ ID NO:61)

MKIPAILVTSFLWGLASGGVPKELGTSSISSASASASASASATASSSSKNVHLLPLKSEHGIVIDKSKFN
 IRKVVLSAIDEINGAPNIGLGLKQVSLALAKAQASAQSSAEALAIKKIVALLISAYVRAAEAAARASAE
 ALATVRAAEQAQKIAEAKGRAAAEALSELVEASQKADAAAAGTTDAIERTYQDARAATSAQTKASGEAEN
 ANRNAAATLAAVLSIAKAASGQGGTRAAVDAAAAAAAAAALHAKANAVSQATSKAAAEARVAAEEEAASQ
 ASASASAQLTAQLEEKVSADQQAASASTDTSAIAEAEAAALASTVNAINDGVVIGLGNTASSSAQASAQ
 ASALARAKNARPKIKGWYKIGGATSASASASASASASAQSSSQGLVY

GAF4 (SEQ ID NO:63)

MKIPAILATSLFVWGLVGLVASELVGSDASATASAEASASSSAYGSKYGIGSGAVSGASASASASASASASA
 SSAPAIEGVNVGTGVSNTASASAEALSRGLGIGQAAAEQAQAAAGQAIAAKSCALAAKSTAQAVALEK
 VARAEVDLAESARKATRLSAEAAKAAAEVEKDLVGLRGAAGKLNLAARAGSKAQERANEDSIEANELAQQA
 TAAAGAEAEAKANAAQEAGASALAIQAALNIEQETVKLTRQAQNTLRSENILAAASNARAIASAEAEA
 SSDLNNRANAARSNARAAETRAVATEAASTAEIAAYSSSEKGEITNPGPLPKIVSVTAGLTQNEIAGSG
 AAASASASALASASAGAGAGAGAGAGASAGAGAVAGAGAGAGAGASAGASAGANAGAGASSLLLPQSKLH
 PISRSSASASASAEAEANSSAYA

MalF1 (SEQ ID NO:73)

MAASNKIIIFSFLAIVLLQLATHCSSTAVLISGSAAGASSHNAAGAAAAARAALGASGAAGLGAASGAARR
 NVAVGANGAAAASAAAAARRAGAI GLNGAAGANVAVAGGKKGGAAGLNAGAGASLVSAARRNGALGLN
 GAAGANLAAAGGKKGGAI GLNAGASANVGAAA KKGAI GLNSAASANAAAAA KKGGAIGLNAGASANA
 AAAA KKGSGAVGLNAGASANAAAAA KKGSGAVAANSASANAAAAA QKKAADAANAAASESAAAAA KKG
 AAVAENAAATANAASALRKNALAIASDAAAVRADAAAAA ADDA KANNAASRGSDGLTARANAATLASD
 AARRASNAATAASDAATDRLNAATAASNAATARANAATRADDAAATDADNAASKASDVSAIEADNAARAAD
 ADAIATNRAAEASDAAIAADAAANAADAAAQCNNKVARVSDALALAANAAARGSDAAAEQAQDAVARASD
 AAAAQADGVIAVNGATARDSAIEAAATAGAAQAKAAGRAGAAAAGLRAGAARGAAAGSARGLAGGLAAG
 SNAGIAAGAASGLARGAAAEVCAARIAL

Figure 10

14 / 22

Xenospiral (SEQ ID NO:12)

ATGAAGATTCCAGTATTGCTTGCAACGTGCCTCTACCTTTGCGGATTTGCGTCCGCCGTTTGGAGGGGCCGGGCAAC
TCGTTGCCCCGAGCTCGTGAAAGGTAGCGCATCGGCCACCGCGTCCGACCGCTGTGACCGCTAGATCAGGACTTAGAGCC
GGACAAGTAGCTTTAGCTTCGCAGAAGGATGCCGTACTCCAAGCTCAAGCTGCTGCATCCGCCGCGTCAGAGGCGCGC
GCTGCTGCCGATCTGACGGCTAAACTTAGCCAAGAATCGGCATCAGTGCAATCGCAGGCTGCCGCCAAGGGAAGGAA
ACGGAGGAGGCAGCTGTTGGTCAAGCTAGGGCTGGCCTCGAGTCGGTGTCCATGGCCGCATCAGCCACATCTGCTGCC
AAAGAAGCATCGACCGCCGCCAAGCCGCAGCATCCGCACTATCCACAGCCGTGGTGCAAGCGAAAATAGCTGAGAGG
GCAGCCAAAGCTGAAGCTGTTGCCTCGGACGAAGCCAAGGCCAAGGCGATTGCAGCAGCCAACTTGGCGGCTGAGGCC
AGTGTAGCCGCGAGAAGCAGCTCTCAAGGCCGAGAAAGTGGCCGAAGAAGCCATCGCAAGAGCGGCCTCTGCAAAGGCT
GCCGCAAGAGCTGCTGCTGCCGCTCTAGCCTCCTCGAAGGAAGCAGCCACGGCCAGCGCAAGAAACGCCGCGGAATCC
GAGGCCAGGAACGAAGTAGCTGTATTGATCGCCGAGATTGATAAAAAGAGTAGGGAAATCGACGCAGCCAGTTTCGCTT
AATGCGCGTGCCGCTGCCAAGGCAAGCTCCAGGAACGTAGAAACGGCGACAATCGGGGCCAACATCAACTCTTCGAAA
CAAGTCGTGTCAATTCCAGTGGAATAAAGAAATTCTCGGAGCCGGAAGTGTCAACATCATGGAGAGAAGATGAAGAG
GTTACGAAAGAGAAGAAGGAGCACATAAATCTGAACGACTTCGACTTGAAGAGCAACGTATTT

Xenospira2 (SEQ ID NO:14)

ATGAAGATTCCAGCAATATTCGTACAGTCTCTGCTGGTCTGGGGATTGGCCGAGGGCCGCGTGATTAATCACGAGTCC
CTGAAGACGAGCGAGGATATTCAGGAGGATATTCAGCAGGAATAGTCCGTGATGGATCTGACGCGCTTGGCTCCTCC
ATAGAAAACGCCCAAAAAGTCGCTCGAGCGGCTGAAAACGTGGGCTTGAATCTGGAATTGGGCGCAGGCGCGCGTGCT
GCCAGTGTGCGCTGCTGCCCAGGCCAAAACACAGAGGCTGCGGAAGCAGGAGCAAACGCCGCTCTGGCCGCCGCC
ATTGCCAAACGGGAGGAAGCGATTAAAGCCAGCGAGATAGCAAACCAATTGTTGACCAATGCAGCAAAGCGGCAGAA
GCGACTGTATCGGCAACGAAGAGGGCAGCACAAATTGACGGCTGCAGCGAAAGAAGCAACCAGAGCTTCTGCAGCCGCT
GCTGAAGCTGCTACGGAGGCCAGGTAAAGGCTAACGCCGATTCAATCATCACGAAGAGGGCTGCGATTGCCGAGGCT
CAAGCTGCGGCGGAAGCTCAAGTTAAGGCGGCAATCGCCAGAAAATCGGCAGCGAATTTTTTGGCTAAGGCTCAAATA
GCGGCTGCCGCGGAATCCGAGGCCACGAACTCGCGGCCGAAGCTGTAGTGGCACTAACAAACGCCGAAGTCGCCGTG
AACCAGGCTAGAAACGCACAGGCCAAACGCCTCGACTCAAGCTTCCATGGCTGTTAGGGTAGATTCTCAAGCAGCGAAC
GCTGAAGCAGCCGCTGTAGCGCAAGCCGAAACTCTCTTGTTACGGCAGAAGCTGTGCGAGCTGCGGAGGCTGAGGTT
GCGAACAAAGCCGCCACATTTGCAAACAGATCGTCAACGAGAAGAAAATACATGTAGCAAAGTTGGAA

Xenospira3 (SEQ ID NO:16)

ATGCAGATCCCAACGTTTTGTCGCCATATGCTTGCTCACATCGGGCTTGGTGCACGCAGGCGTCGAGGAATTCAGTCC
TCGGCAACCGAGGAGGTGATCAGCAAAAACCTTAGAAGTCGACCTGTTGAAAAATGTGGACACTAGCGCGAAACGAAGA
GAGAACGGCGCCCCGGTGCTCGGCAAGAACACACTTCAATCCCTGGAGAAGATCAAGACGTCGGCGAGCGTGAATGCC
AAAGCAGCAGCCGTGGTGAAAGCGTCCGCTCTGGCTCTTGCAAGAGGCTATTTGCGAGCGTCCGCATTGTCAGCCGCC
GCTTCAGCCAAGGCAGCCGCCGCCCTGAAAAATGCTCAACAAGCGCAATTAAACGCCCCAGGAAAAGTCTTTGGCCGCG
TTGAAAGCTCAGTCCGAGGAAGAGGCAGCTTCTGCTCGTGCAAACGCAGCAACCGCCGCGACACAGTCGGCACTGGAA
CGCGCTCAAGCCTCCTCCAGGTTAGCAACGGTCCGCCAAAACGTAGCCAGCGACTTGCAGAAACGGACCAGCACCAG
GCCGCGGCTGAAGCCGCTGCCACCCCTCAGACAATTACAGGACGCGGAACGAACGAAATGGAGTGCCAACGCTGCCTTA
GAAGTCTCCGCCGCTGCAGCTGCCGCAGAAACCAAGACCACTGCCTCCTCGGAGGCCGCCAACGCCGCCGCCAAAAG
GCGGCCGCGATAGCTTCTGACGCGGACGGCGCGGAAAGGTGGGCATCTACCGAGGCACAATCAGCTGCGAAGATCGAG
AGTGTGGCAGCCGCCGAGGGATCCGCCAACTCGGCCTCTGAGGATTCCCGGGCCGCTCAATTGGAAGCCTCCACCGCG
GCGAGAGCCAACGTGGCCGCGAGCTGTGCGGGATGGAGCGATTATAGGACTTGGAGAGGAAGCGGGTGCCGCGGCTCAG
TTGCTTGACAGGCGAAGGCATTGGCCGAAGTTAGCTCGAAATCCGAAAATATTGAGGATAAAAAATTTT

Xenospira4 (SEQ ID NO:18)

ATGAAGATCCCATCCATACTCGCGGTTTCCCTGCTGATCTGGGGTTTGGCAAGCGGCGCAAGGGAAGAGGTGGAGACA
CGGGACAAGACCAAGACCTCGACAGTGGTGAAAAGCGAGAAAGTGGAAGTCGTTGCTCCCGCTAAGGATGAACTTAAA
TTAACGAGCGAGCCTATCTTTGGAAGAAGAGTGGGAAGTGGAGCATCCGAGGTGGCATCTAGCAGCGGTGAAGCCATC
GCGATAAGTCTTGGAGCAGGGCAGTCAGCGGCAGAGTCTCAGGCCTTGGCCGCCTCGCAATCCAAAACGGCAGCGAAC
GCCGCCATAGGCGCGAGCGAGCTTACCAACAAAGTTGCTGCTCTAGTTGCTGGCGCGACTGGTGCGCAGGCGAGAGCT
ACGGCCGCCTCCTCGAGCGCGTTGAAGGCCAGCTTGGCGACCGAAGAAGCGGCGGAAGAGGCCGAGGCCGCGTGGCT
GACGCCAAGGCTGCCGCGGAAAAGGCCGAATCCCTGGCGAAAATCTCGCGTCGGCGAGCGCTCGCGCGGCCCTCTCC
TCCGAAAGGGCGAACGAATTGGCTCAAGCTGAGAGCGCTGCAGCGGCCGAGGCGCAGGCCAAGACAGCAGCCGCCGCC
AAAGCAGCGGAATCGCCCTTAAGGTCGCTGAGATAGCGGTGAAGGCGGAAGCGGACGCAGCAGCTGCCGCCGTGGCA
GCTGCAAAGGCAAGAGCCGTGGCAGACGCGGCCGCTGCCCGTGCCGCAGCCGTGAACGCCATCGCCAAGGCGGAAGAG
GAGGCCTCGGCCCAAGCAGAGAACGCCGCCGGTGTTTTGCAAGCAGCCGCCTCCGCCGCGGCGGAATCGCGAGCCGCT
GCAGCTGCCGCCGCTGCTACCTCGGAGGCAGCGGCTGAAGCTGGCCCGTTGGCAGGTGAGATGAAACCACCGCACTGG
AAATGGGAACGGATTCTGTGAAGAAGGAGGAGTGGAACCGTCAACGAAGGAAGAATGGAAAACGACGAATGAAGAA
TGGGAGGTGAAG

15 / 22

Xenosin (SEQ ID NO:20)

ATGAAATACATGCTCTTGTGCTATCTATATTCATCTGTGCACATATTGTATGCGCAGGCGTAAATACAGAATTAAAA
AAAGATGGTGAACATAAGGAAGAGTCTTATGAGAAAAGCGAGTCAAAGAGTTTAAAAGAAATTAAAGAAGAACGTGCT
TCAAAATCAAAAAGTGAACGTTTGAAGATTCTGTAAGAAAAACGCGAAGAGGAAGAAAAATCCAAGAGTCTGAATCTG
GTCGTGGTCAGAGAAAAGATTACCAAACCTTCTTCATGGCTCAAAGAAGAGAAAGATATCAGTCCTCTTTTGAAGAA
AAAAATGGCAAAGGTCTATTGGGTTTGAAGATGTCACGGACGAGTTAAATATCGCTCTTAAATCGTTGAAGGAGGGC
AAAAAGTTTGATACTTGGAATTCGAGAAAGGTAGCGAAGACGTTCTTCTTTGAAGAAGTTGATACGAGCGTCGTT
GAACTTTTAAATTAATAAAGGAAGGAAAACTGACCATGGTGCTATAGATTTGGAGAAGAATGGTAAGGTACTTGTA
GATTTGGAAAAATCTCAGAAAACATACTTGAACTTGTGGATCACAAAAGAAGACTGTGGAAGTTGTAGATGATAAA
GACAAAAAATGGAATAAAGAATCAGGTTGGAAAAAATCTAAATGATCTAGATTGGAAAAAAGATTTAGATAAAGAT
AAAGTTGGTGGCGGTTTGCTTGGCGGTTTAAGTGGCCTCTTAAATAGTTTAAATCAGAAAAAGGTCTTCTAGGTCTT
TTGAATAAGAATCAAATTGAGTTATTAATTCCTTTAATCAGTGAGATAAAAAAGAAAAATATAGATTTTAAATCTCTTC
GATTCTGTTGATTCTGTGCGAAAGAAATTTAGACTTGAACTTTTCAACAAGTTCTGTTTCAAAGTTACTGAATTATTA
AATAAAGGAATCGATATTCAAACAATTTGAATGCGAAAAATGGAGATGAATTCGATTTAAGCGGCAAAGAATTGAAA
AACGTCAAAGGGATATTTGGTTTGATTGGAAGTTTGAAACGCTCATTAGGATTAGAAAATATATTGAACTTACCGTTT
AAACGTATACCTCTGCTTAAATTA

BBF1 (SEQ ID NO:32)

ATGAAGATTCCAGCACTGCTCGTAACGTGCCTCTACCTTTGGGGCTTCGCGTCCGCCGGCCAGAGCTCACCTCTGCTC
GAGATCGTGCGGGTAGCGCGTCGGCCACCGCATCCACCGCTGTGACCGCTAGATCCGGACTTCGTGCCGGTCAGGTA
GCCGTGGCCTCGCAGAAGGATGCCACACTTCAGGCAGATGCCTCAGCGGCCGCCGCGGCCGCTGCACGCGCTTCGCC
GACCAGTCGGCCAGTCTAGCCCAACAGTCGGCGTCTTTGCAGTCCAAAGCTGCCGCCAGAGCAAAATCAGCCGAGGAG
TCAGCGGCAGCTACGGCCAAAGCCGAGTTGCAGGCAGAATCCATTGCTGCATCTGCCAGTTCCAATGCCAGAGAGGCT
GCAGCGTCCGCAAAAGCCTCCGCATCCGCGATGTCATCGGCTGCCGTGCAGGCGAAACTCGCTGAAAAGACGGCCAAG
AATCAAGCTCTGGCTTCCGAAGAAGCCAACTCAAGGCTGCCGCCGCTGCCAGCGCAGCAGCAGCAGCCAGCGCCGCC
GCCGAGGCAGCCCTGAAAGCTGAGAGAATAGCGGAAGAAGCCATCGCCAAGGCGGCCGCTGCCAAAGCAGCCGCCAGA
GCCGCTGCAGCCGCGTTAACTCCGCGAAGGAAGCCGCCACGAGCAGCGCAAGGAGCGCCGCCGAAGCCGAAGCTAAG
AGCGAAGTCGCTATACTGATCAGCGAACTCGACAAGAAGAGCAGGGAAGTCGCCGCTTCGCGCTCCGCCAAGGCACGC
GCTGCTGCTGCGGCTAGCTCCAGAAACGCAGAAACGGCTGTTATCGGAGCTAACATCAATGTGGCCAAAGAGGTCTTG
GCGATTTCCCATCGAGCCAAAGAACTTCCGGAGCCAGAGCTGGCGTTGAAAGAAGAGAATGTGCGCGTTCGCGAGCTCA
GAGAGTGAAGTGAAGGTAGAAACGAGCAGCGAAGCATGGTCAATTTAA

BBF2 (SEQ ID NO:34)

ATGAAGATTCCAGCAATACTGGTTACGTCTCTGCTGGTCTGGGGTGGTCTGGCCGAGGGCCACGTGGTGAAGCGCGAC
AAGGAGCTCAAGGCCCCGGCTTTACCGGAACTACTCGGTGATGGGTCTGACACGCTCGGTGCCTCGATGGAGAACGGG
ATCAAAGTCGCCAGAGCATCGCAGAATGTGGGTCTGAGAACAGAGTTGAATGCAGCCGCGCGGGCTGCAGCCGCTGCT
GCGACCAAGCAGGCCAAAGACACAGAGGCCGCGGAAGCTGGAGCGGCCGCTGCGATTGCCATCGCTATCGCCAAGCGT
GAAGAAGCTATCAAAGCAAGCGAATTAGCCAGCAAGTTGTTGACAGCCGCGGCTGGGTCCAGCGAAGCTGCCGTGTCA
GCGACGGTGAGGGCGGCGCAATTGACGGCCGAGCTAGCGCAGCTGCCAAAGCTTCTGCATCCGCCTCTGAGGCTTCT
GCCGAAGCCCAGGTGAGGGCCAACGCCGAAGCAACATCGCCAAGAAAGCTTCGGCAGCTGAAGCAAAAGCCGAGCC
GAAGCCCAGGTTAAGGCGGAAGCTGCCAAGAAAGCGGCCGCCGCTTTCTTAGCTAAGGCTAGACTAGCGGCCAGCGCC
GAATCCGAGGCCACTAACTCGCAGCCGAAGCTGAAGTAGCACTGGCTAAGGCCAGAGTCGCCGTCGACCAGTCGCAG
AGCGCACAGGCAACCGCTACCGCTCAAGCTGCCACAGCCGTTTCAAGTGCAGTCTCAAGCAGCTAACGCGGAAGCCTCC
GCTGTAGCACAGGCTGAACTCTGCTGGTCACGGCGGAAGCCGCTCTGCGCGGGAAGCCGAAGCCGCGACCAAGCT
ACCAGTTGGGGCGAAGAATGTCATCAACGAGAAAAAGTTACGTTTAGCGAAGATCGATTAAACGAGAGACAAGACAAT
TGGTAG

BBF3 (SEQ ID NO:36)

ATGCAGATCCCAGCGATTTTCGTACGTGCCTGCTCACATGGGGCCTGGTGCACGCAGGTAGCGTGGAAGTCCGGTGCC
CCCAAGCAGGAGTCTGTCTCGTGGAGCAGCTCCTATTGAAGAACGTGGAGACTAGTGCGAAGCGAAAGGAGAACGGC
GCACCGAAACTCGGCGAGAGCACAGCTGCGGCTCTGGCTAGTACCAAGGCAACTGCAGCCGCAGAGGCTAAGGCATCC
GCCAAAGTGAAAGCTTCTGCCTTGGCCCTCGCTGAGGCTTTCTTGCGTGCCTCGGCAGCGTTTGCTGCTGCTTCAGCC
AAAGCTGCTGCCGCTGTAAAGGAAGCAACGCAGGCACAGTTGCTGGCACAGGAGAAGGCTTTGATAGCGTTGAAAAT
CAATCTGAGCAACAAGCTGCCTCTGCTCGCGCGGACGCCGCGGCTGCCGCAGCCGTATCCGCGCTAGAACGCGCCCAG
GCCTCCTCCAGAGCAGCCACGACCGCCCAAGACATCTCCAGCGATCTGGAGAAACGTGTCGCCACCTCAGCCGCTGCT
GAAGCAGGTGCCACCTCAGAGCGGAACAATCCGCCGCGCAATCGAAATGGTCCGCCGCACTGGCCGCCCAACCGCC
GCTGCTGCAGCCGCTATAGAAGCAAAGGCCACCGCTTCTCAGAAAGCACCGCTGCCGCTACTAGTAAGGCCGCCGTG
TTGACCGCTGACACTAGCAGCGCAGAAGCTGCCGCTGCAGCGGAGGCACAATCCGCTTCGCGGATCGCAGGTACAGCA
GCCACCGAGGGATCCGCCAACTGGGCTAGCGAGAACTCGCGTACCGCACAACTGGAAGCTTCGCGCTCAGCGAAGGCC

16 / 22

ACCGCAGCCGCAGCTGTCGGAGATGGAGCTATTATAGGACTTGACGCGGACGCTAGTGCCGCAGCTCAGGCAGCCGCA
GAAGTTAAAGCCTTAGCTGAAGCTAGTGCCAGCTTAGGTGCTTCAGAAAAGGACAAGAAATGA

BBF4 (SEQ ID NO:38)

ATGAAGATTCCATCGATACTCGCGGTGTCCCTGCTGGTTTGGGGTCTGGCCAGCGCAGGCAAACCACTCATTGCCAAT
GCGCAAATAGGGAAGGTCAAGACCGAAACGTCATCGTCTTCAGAGATTGAGACGTTGGTATCAGGAAGCCAGACATTG
GTGGCAGGAAGTGAGACATTGGCTTCAGAAAGCGAGGCATTGGCGTCAAAAAGCGAGGCATTGACGTCAGAAGCCGAG
ATAGCGAGCGTGACAACGAAGGACGAGCTCATACTAAAGGGCGAAGCTATCACTGGAAAGAACTAGGAACCGGGGCG
TCGGAAGTAGCGGCGGCCTCTGGGGAGGCTATCGCAACTACCCTTGGCGCGGGACAAGCTGCAGCAGAGGCACAAGCA
GCCGCCCGCGCGCAAGCAAAATCAGCAGCGGCAGCTGCCGCGAATGCAGGTGAATCCAGCAACAGTGCTGCTGCGTTG
GTTGCTGCTGCAGCTGCAGCACAAGGAAAAGCGGCTGCCGCCGCGCAGCAGCCGCGACGAAGGCTAGCTTAGAGGCCGCA
GACGCTGCTGAGGAAGCTGAGTCGGCCGTGGCCTTGGCTAGGGCTGCCTCCGCAAAGGCGGAAGCGCTCGCATCGACC
GCCGCTGCTGCGAATAACCGTGCTGCTCTCAAGCGGAAAAATCGAACGAGCTGGCGCAAGCTGAGGCTGCAGCCGCC
GCCGAAGCCCAGGCTAAAGCCGCGCTGCTGCCAAGGCAACACAACCTCGCCCTTAAAGTTGCCGAAACTGCGGTGAAA
ACGGAAGCAGATGCAGCAGCTGCCGCCGTTCGCGCCGCAAAAGCCAGAGCAGTCGCAGACGCAGCCGCGTCTCGTGCG
ACCGCAGTGAACGCCATTGCTGAAGCGGAAGAAAGAGACTCTGCACAGGCGGAGAACACCGCTGGTGTAGCACAAGCA
GCGCTCGCTGCTGCGGAAGCACAAGACTCCTGCATCGGCGCTGCCGCGACTCCTAGGCATTGCTCGAGCTATGCATGG
TGGAAGCTTAGGATAACATCCTTGATCGTCATTCTATCGCCACGCAATCGACGTAATA

BBSA1 (partial) (SEQ ID NO:39)

GGAAATTTCGGAAAGCGGCGAAATTTGGAAGAACGGTGAAAGCTCCGAAAGCGGCAAAAATTTGGAGGAACAGCGGAAGC
TCCGAAAGCGGCAAAAATTTGGAAGAATGGCGGAAGCTCAGAAAGCAACAAACATTTGGAAGAACGGTGGAAGCTCGGAA
AGCGGCGAGAAATGGAAAAACAGTGAAAGCTCCGAAAGCGGCAAAAATTTGGAAGAACAGCGGAAGCTCCGAAAGCGGC
AAAAATTTGAAAAACGGCGGAAGCTCGGAAAGCAACAAACATTTGGAAGAGCGGTGGAAGCTCGGAAAGTGGCGAGAAA
TGGAAAAACAGTGAAAGCGGAATAAAGGCAAAAGCTCAAAAAGCAGCGAAAGTTGGAAGAGCAACGAAAACCTCGAAG
AACGACGGCAGCTGGAAGAGCAGTGAAGAATCAGAAAAGTGGAAAGATGGTAAAGCAGTGGCGGAAGACAGCGTTAGT
ATAAACTGGGCAGATGTCAAAGAGCAGATTAGCAACATTGCTACATCCTTAGAAAAGGGTGGTAACCTCGAGGCTGTA
TTGAAAATAAAGAAAGGAGAAAAGAAAATTTCAAGTTTGGAGGAAATCAAGGAGAAAATCTCTGTCTACTGAAATGG
ATTCAAGAAGGCAAAGATACTAGCAGCCTATTAGATTTGAAAGAGGGTAGCAAGGATATTTGCGTCGTTGAAAGAAATC
AAAGGAAAGATCCTTTTGATTGTTAAATTAGTGAACGAAGGGAAAGACACTAGTGGTCTTTTAGATTTAGAAGCGAGT
GGCAAAGTAATTTTAGAATTGCAAAGCGCCATAGAAAAGTTCTCGTAAAGTCAAGAAAGGTAACCAAAGTATCTGAA
GTTTCCGGTTTAGTAAAAGCAAACTGTCTCGGACATAAAACCGCTTCAAGCAGTAATTCCTTTAATCCTTGAATTG
CAAAAAACAGACATTAACCTTAGTACCTTAAACAAGTGGTCCACTGTAAACGTAAATTCTATAGATAAAGAACGCGTC
ACGAAAACGGTTCCAGTGCTCCTTCAATCCATGAAAGGAGGCGAAGATATTCAGAACCTTTTGAGTGCGAAAGGTGCA
AAGAACTTGGCATTAGTGCTTTGGAATTACAGGCAGTTCAGGAGCTCTTGGCGTGGTTGGAAAGCTAAGTTCAGGT
GGTGCCTTGAACCTCAAAGGCTTGTTGAACCTGAAAGACGGCGCTAGTGTGTTAGGTGCAGGAAAAATCGGAGGATTA
ATTCCTTTACCGAAACTTTAAGAGATAGACCGATAAAGGCAGATATACTCTCGGAAGATTTTTTTTGGAAAGTTGAATAG
TCCGCAAAAAAATTATCTCTGATTATTATAATTTAGCCTAAAATATTAAATAAAATGGAGAAATAACGTTGAAATATA
TAAATAA

BAF1 (SEQ ID NO:49)

ATGAAGATCCCAGCGATAATCGCAACGTCCCTTCTCCTCTGGGGTTTCGCCAGCGCCAGCGGGCCGCGCTTACTCGGC
GGCAGATCGGCGCGTCCGCGTCGGCTTCCGCTTCGGCTGAGGCGTCGGCGGGCGGTGAGGAAAGCGGCGCATCC
GCTTCCGCTTCCGCTAAGGCTGGTAGCAGCAACATCCTCAGCCGCGTGGGAGCTTCGAGGGCGGCCGCGACGTTGGTC
GCTTCCGCGCGGTGGAGGCCAAGGCGGGTCTCCGTGCCGGCAAGGCAACCGCCGAGGAGCAGAGGGAGGCTTTGGAA
ATGCTCACCTTGTCGCGCCGACAAGAATGCCGAGGCGCGTATCCTGGCCGACGACACGGCCGTTCTGGTTCAAGGCAGC
GCCGAGGCACAGTCGGTCGCGCCGCGCAAGACCGTCGCGGTGAGGAAGAGTCCGCTTCCTTGGATGCGGCCGCGAGTT
GAAGCGGAGGTGCGAGCCGCCACGTCGAAATCGTCGGCTGGCCAAGCACTCCAGTCCGCACAGACCGCCGCATCTGCT
CTCAGAACTTCCGCCAGGAGCGCCTTGACGGCCCTCAAGCTGGCACGCCTCCAAGGCGCGGCTTCTAGCAACGCTGCC
AGGATGATGGAAGGCGCTGGCCGCCACCCAGGACGCAAATGCCGCCGCCAGCAAGCTATGGCGGCCGAGAGTGCA
GCCGCAGAAGCAGCGGCTATCGCGGCAGCGAAACAATCGGAGGCGAGAGACGCCGGCGCCGAGGCCAAGGCCGCCATG
GCAGCACTCATCACCGCCAGAGGAATCTCGTGACGGCCAATGCCAGGGCGGAAATGGCAAGCGAGGAAGCCGAATTG
GATTCGAAGTCTAGAGCGTCCGACGCCAAGGTGAACGCCGTTGCTCGTGCGGCCTCCAAGTCCAGCATAACGCAGAGAT
GAACTTATCGAGATCGGCGCTGAGTTCGGCAAGGCCAGCGGCGAGGTGATTTCCACCGGCACGCGTTCCAACGGCGGT
CAAGACGCCATCGCCACCGCCGAGGCATCGAGTAGCGCGTCCGCCGTGCGCATCAAGAAAACAAGCGGACACTGGGGG
AGCGGAAAATGGAGTCGTGTCTCAAGGGTAAAGGATGGGCTTCTCGAATGCGGACGCTGACGCCAGCAGCAGCAGC
ATCATCATCGGCGGTCTCAAACGCGGCGGCCTCGGTTGGAAGCCTCTGCGGCAGCTTCCGCAGAAGCGGAAGCTTCC
GCCGGCACACTCCTGCTGTAA

17 / 22

BAF2 (SEQ ID NO:51)

ATGAAGATTCCAGCGATACTCGTGACGTCTCTCCTCGCCTGGGGATTAGCCAGCGGCCGGGTTCATCGAGTCCAGCTCG
TCGGCTTCCGCACAGGCGTCGGCATCGGCCGGCTCGAGAGGCCTGCTCGGTAAACGGCCGATTGGCAAGCTCGAGTGG
GGCAAGGAGGAGAAGAACTCGAAGAACTCGACGAGGAATCGCTCAATGAGGCCGCTCTGAAGGTCGGCATCAAGAAC
GGCGGATTGGATGTCGCGAAGGGCGCGGCAGTCCTCGAGGCAGCGATGAGCGACGTCGCGACCCTTACGGATCAGCGT
TCTCTTGTGGATCTCGGTCTCGGCCCGGTGCGGAACGAGGCCGAGATCCTGGCGGAGGCGCAGGCCGCCACGAGCGCC
CAAGCTGGCGCTGTCGCTAATAGCGCCGCGGAGCGTGCGATCGCGGCGATGGAGATGGCCGACAGAACCGAATATATT
GCGGCACTTGTCAACCACAAAGCCGCCAAAGCTGCCGAGGCCACTATGGCCGCTACTGCCCGTGCCACCGCCGCCGCC
TCAGCCTCCAAGATATCCAGTCAGGAATCAGCCGCATCGGCCGCTAACGCCGCCAACGCCGAAGCCAAGGCCAACGCC
GCTTCCATAATCGCTAACAAAGGCGAACGCCGTCCTGGCTGAGGCCGCCGCCGCTACTCGCAGCCACTGCTGCCAAGGCC
AAGGAATCGGCGATGAAATCGCTTAGCGCCGCTCAGGCCGCCGCCAAGGCACAAGCCAGGAACGCCGAGGCCCTCCGCC
GAAGCTCAGATCAAACCTTTCCAGGCCAGGgCCGCCGTGGCACGCGCTGCAGCCGATCAGGCCGTCTGTTCTCTCCaG
GCTCAGGCCGCAAGTCAGATACAATCGAGGGCATCCGCATCCGAATCCGCGGCATCGGCACAATCAGAGACCAACACC
GCCGCGGCCGAAGCGGTGCGCACCGCTGACGCCGAAGCGGCCGCGCAAGCTGAAGCGTGGGTTCATGTCGCTGAAGAAC
GATCTGTGGCTGCATCTCAACATGAAGGGTGAGGCCAAGGCCGAAGGCGAGGCCGTTTCGATCAGCAAAGGACATCGC
GGCGGTATCAGGTGGGCAGCATCTCGGAAGCCAGCGCCGAGGCAAGCAGCAACGTTTCCATGGGCGGACGTTCATGGA
CGGAAGGACCTCGTCTCTGAAGCGTTAGCGGGAGCATCAGCGGGCAGCAGTGCCGACTCCCTTTGA

BAF3 (SEQ ID NO:53)

ATGAAGATACCAGCGATACTCGTGACGTCCCTCCTCGCCTGGGGACTGGCCAGCGGGAATCTCCTTAAGGAGTCGAAA
GCTTCCGCGTCCGCGTCCGCGTCCGCTTCCGCGAGGGCCAGCGGCAAGAAGAATCTTCACGTGTTGCCATTACCGAAG
AAAAGCGAGCATGGCATCGTGATCGACAAGTCGGTGTTGACATCAAGGATGTAGTGCTGAGCGCGGTGACGAGATC
AACGGCGCCCCGAAACTCGGCCTGGGATGGAAGAAGGTCAGCATGGGGGTGGAGCGCGCCGAGGCCAACGCAGCCGCT
GCCGCCGAGGCATTGGCGATGATCAAGAAGATTGCCATGGCCCGCAGCAGTGCTACGTCCAGGCGGCCTGGGCATCG
GCCCAGGCATCAGCTGACGCATTGGCTAGCGCCAGGGTGCGCACAGGCGTCTCAGGAGGCTGCGGAGGCCAAAGGGTAGA
GCGGCTTCCGAGGCGCTCTCCAGAGCCATCGAAGCATCCTCGCGAGCCGATGCGGCAGCCGCTGCGACGCTGGACGCG
ATGGACCGCACCATGGAGAACGCGAGGGCGGCAAAATGCCGCGCAAACGCAGGCCAGCGGCCAAGCTGAGAACGCAAAT
CGCAGCGCTGCTGCCATCCTCGCAGCTCTGCTACGTATCGCGGAGGCATCCGCGTTGAACAACGAGGCCGCGGTCAAC
GCGGCCGCGGCCGCGAGCCGCGAGCGTCTGCCCTTCAGGCCAAGGCTAACGCGGCTTCTCAAGCAACCGCCAGAGCCGCA
GGACAGGCGTCGACGGCCGCCGAAGAGGCGCAATCCGCCCAAGAAGCCGCCGATAAGAACGCGGAGCTGACCACGGTC
ATGCTCGAAAAGGCTAGTGCTGATCAACAGGCGGCATCCGCTAGGGCTGACTACTACACCGCCTCAACCGAGGCCGAA
GCCGCTGCACAGGCGTCTGCTATCAACGCACTCAGGGACGGAATAGTTGTGCGGAaTGGGAAATGACGCTGGCGCATCG
GCCAAGCGATGGCACAGGTAGAAGCTCTCGCTCGCGCCAGCGAGCACAAGGCGTTAGGCGAGAAGAAGAAGGGCCTG
GTTTGGGGCTACGGAAGCAAGGGCAGTAGCTCCGCCAGCGCATCCGCCAGCGCCTCCGCCGAAGCATCCTCGAGACTC
GGAAAGGACTGGTAG

BAF4 (SEQ ID NO:55)

ATGAAGATTCCAGCGATACTTGCGACGTCCCTCCTCATCTGGGGTCTTGTCGGCGCCAGCGAGCTCGAATCGGAAGCG
AGTGCGGCGGCGTCTGCGCAAGCGGAAGCGTCCTCGTCTGGTTCGCTCCGGCAAACGTGTCCGCGTCTCAGGCTTCCGCC
AGCGCGTCCGCCAGCGCGTCAGCCGGCAGCAGAGGTGGCAGCAAAGGTGGCTGGGGCCAGCTCCGCCGTGGTGATGTT
AAGAGCGAGGCGAAGAGCGCCGCCGCGATCGCGGTGGAAGGAGCTAAAATCGGCACCGGAATCGGAAATACCGCGTCC
GCATCCGCGGAGGCGCTCTCACGAGGACTCGGCATCGGACAGGCGGCCGCGGAGGCGCAAGCCGCAGCCGCAGGTGAG
GCAGAGGTGCGCCGCGAAATCGTGCGAACTTGCCGACAAGACCACCGCCAAAGCGGTGCGCATGGTTCGAAGCGGCAGCC
GAGGCCGAAATCGAGGTGGCCAATCAGGAGGTGCGAGCCGTCAAATTATCGACTTGGGCCGCTAAAGCAGCAAGGATA
GTCGAGGAAGACAGCGCCGCCGTGAGGGCGGCTGCCGGCAAATTGCTTTTGGCCGCGAGAGCTGCCGCCGCCGCCGAG
AGACGCGCCAACGAGGAATCCGAGGCGGCCAACGAACCTTGCTCAAGCGTCATCTGCCGCTGCCGCCGAGGCCGAAGCC
AAAGCGAACGCCGCCGCTGAGGCCGCTGCCGCTGCCCTTGCTATCGCCGAGGCCGCCGCTCGCCATCGAACAAGAAGCC
GTCATTTTGGCTCGCAAGGCACAAGATGCCCGTTTGAATGCTGAAGCCGAGCCGCCGCTGCGATGAACGCCCGTGTC
ATCGCTTCCGCCGAATCCGAGGCCAGTGAAGATCTGGAGAATCGCGCTAGTGTGGCGCGTGCCAGTGCGGCCGGTGCC
GCTGAGGCAAAGGCTATCGCCACCGATGCCGGCGCCACTGCCGAGATCGCGGCCTACAGTTGGGCCAAGAAGGGCGAA
CTGATCAACCCCGGCCCGTTGCCGAAGATCATCAGCGTCAACGCCGATCTGTCCAAGAGCGAGGTGAGGCCATGAAG
ATCACCCGGGGTCAAGTACAGGAAGTCAAGAAAATCAGCACTCACAAAGGTGGCTGGGGATGGGGAAAGGAAGGAAGG
TCGAAGGTATCTTCCAACGCTAGTGCCAGAGCTAGTGCCAGCGCCAATGCAGCCGCCGCTAGCCTCGGCAGCAAATGG
GGAAGACAACATCCGCATCATCCGCGTGGGTGACGCCAACGCCGAAGCCGACAGCCAGTTGCTGAAAGTGTGGTGA

GAF1 (SEQ ID NO:65)

ATGAAGATCCCAGCGATAATCGCAACGACCCTCCTTCTCTGGGGTTTCGCCGACGCCAGCAAGTCGTACCTCTTAGGC
TCATCCGCGTCTGCTTCCGCTTCCGCTTCCGCCTCGGCATCAGCGGGAGGAAGCACCGGCCGCGTCCGCGTCCGATCT
GTAATATCCGGTGGCAACAACATCATCAGAGGAGCTTCGACCACATCCGTGACATTGGCAGCCGCCGCGCAGCGGAGGCC

18 / 22

AAGGCAGCTCTGAATGCTGGAAAAGCGACTGTCGAAGAGCAAAGGGAAGCGTTACAGTTGCTCACCGCGTCCGCTGAA
AAAAACGCCGAGGCGCGTTCCTTGGCCGACGATGCGGCCGTTCTAGTTCAGGGTGCCGCTGAGGCGCAATCGGTCGCC
gCCGCGAAGACGGTCGCGGTGAGCAAGGATCCAACTCTCTGGATGCAGCTGCAGCCGAAGCGGAAGCCGCCGCCGCC
GCATCCAGGGTATCGGCCCAGCAGGCACTCCAGGCCGCGCAGACCTCCGCCGCCGCTATTCAAACCGCTGCCGGTAGC
GCCCTGACGGCTCTCAAATTGGCACGCAAACAGGAAGCGGAATCCAATAATGCCGCCGAACAGGCAAATAAAGCATTG
GCCTTAAGTCGCGCAGCCAGCGCTGCCACTCAACGAGCCGTGGCAGCTCAGAACGCGGCTGCCGCATCAGCGGCTTCG
GCTGGAGCCGCACAAGCTGAGGCAAGGAACGCCCTACGCCAAAGCCAAAGCAGCGATAGCTGCTCTTACGGCCGCCCAA
AGAAATTACGCCGCGGCCAAGGCTAGCGCAAGCGCGGGTAGCGTGGTGGCCGAACAAGATGCTCAATCTAGAGCGGCC
GATGCCGAGGTGAACGCCGTTGCCCAAGCCGCTGCCCGAGCCAGCGTTTCGCAATCAGGAGATCGTTGAAATCGGCGCG
GAATTCGGCAACGCCAGCGGCGGAGTGATCTCGACCGGCACACGTTCTTCCGGAGGCAAGGGTGTCTCCGTTACCGCT
GGAGCTCAGGCTAGCGCGTCCGCTTCCGCGACCTCCTCCTCCTCCTCCTCCTCCGGCATCAACAAAGGACATCCCAGA
TGGGGGCACAATTGGGGTTTAGGTTCTTCGGAAGCGTCAGCAAACGCTGAAGCCGAAAGCAGCGCTTCCTCTTATTCA
TCTTAA

GAF2 (SEQ ID NO:67)

ATGAAGATTCCAGCGATATTCGTGACGTCTCTGCTCGCCTGGGGACTCGCCAGCGGCGGAGTCATAGGTCCCGACACG
TCCTCATCGTCCCAGGCATCGGCATCGGCATCGGCGTCAGCATCGGCGTCGGCATCATCGTCCGCATCGATCGGTTAC
AACGAACTCCATAAATCGATCAATGCGCCCGCCTTGGCGGTTCGGCGTCAAGAACGGCGGAGTGGATGTGCGCAAGGGC
GCGGCCGTTGTGCAaTCAGCGATATCCGACGTATCGACTCTAACCGATGATCGTACGTTGAACGGTCTCGCTATCATC
GGGAATAGCGCCGAGAGTCTGGCAAGAGCACAGGCTTCCTCGAGCGCCAGCGCCGGCGCAAAGCCAATGCTCTCATC
AAACAATCGATAGCGGCTATAGAGATCACCGAAAAGGCAGAGTACCTTGCGTCGATCGTCGCCACCAAGGCAGCGAAG
GCCGCCGAGGCCACAGCGGCCGCGACCGCTCGCGCCACTGCCGTGCGCGAGGCTGCCAAGGTTTCAGCGAGCAATTC
GCGGCCGAGGCACGCGCGGCCGCCGACGCCGAAGCCAAGGCCAACGCCGCTTCATCATCGCCAACAAAGCGAACGCC
GTCCTCGCGGAGGCAGCCACCGGACTTAGCGCCAGCGCTGGCAAAGCCCAACAATCGGCGACCAGGGCGTTGCAAGCC
GCACGAGCTGCCGCTAAGGCTCAAGCCGAACTTACCCAGAAAGCCGCTCAAATCTTAGTCCTCATTGCTGAAGCCAAA
GCCGCCGTGAGCCGAGCAAGCGCCGATCAATCCGTCTGTACGTCCCAGGCACAAGCCGCCAGTCAGATTCAATCGAGA
GCCTCCGCGGCCGAATCCGCGGCATCGGCTCAATCGGAAGCCAACACCATTCGCGGCCGAGGCGGTTCGCTAGAGCTGAC
GCCGAGGCGGCCAGTCAAGCTCAAGCGTGGGCCGAATCCTTCAAACGCGAACTCTCGAGTGTGTTTTGGAGGCCGAG
GCCAATGCCTCGGCTAGTGCCTCGGCTGGTGGCCAGTGGTAGCAGCAGCTCGGGCGCGAGTTCCAGCGCGGAT
GCCAGCGCCGAGCGAGCAGCTATGGATCCTTGGGCGGATATCGACACGGCGGAAGCTTCAGCGAGGCATCGGCAGCC
GCGTCAGCGGCCAGTCGCGCCGAGGCTGCGTAA

GAF3 (SEQ ID NO:69)

ATGAAGATTCCAGCGATACTCGTGACGTCCCTTCCTCGCCTGGGGACTGGCCAGCGGGGGTGTCCCTAAAGAGTTGGGA
ACTTCCATTTCTTCCGCGTCCGCATCCGCATCCGCATCCGCATCCGCGACCGCGTCCTCCAGTAGCAAGAATGTTTAC
TTATTACCATTGAAAAGCGAGCATGGCATCGTAATTGACAAGTCAAATTCACATCAGAAAGGTAGTGTTGAGCGCA
ATCGATGAGATCAACGGCGCGCCCAACATCGGTCTGGGATTGAAACAGGTCAAGTTTGGCGCTCGCAAAGCCCAGGCT
AGTGCTCAATCGAGCGCCGAGGCATTGGCAATCATCAAGAAAATCGTCGCGCTCCTCATCTCGGCCTACGTCAGAGCA
GCCGAGGCCGCGGCTCGAGCATCCGCCGAAGCTTTAGCTACCGTTAGGGCTGCGGAACAAGCGCAAAAAATTGCTGAA
GCGAAGGGTAGAGCGGCTGCTGAGGCGCTCTCCGAGTTAGTCGAGGCGTCCCAGAAGGCCGATGCGGCGGCCGCGGGA
ACGACGGACGCGATCGAACGCACCTACCAGGATGCCAGAGCGGCCACTTCCGCACAGACCAAGGCCAGCGGCCGAAGCC
GAGAATGCTAATCGCAATGCTGCCGCCACCCTCGCGGCGGTCTTGAGCATCGCTAAGGCCGCTCCGGTCAAGGAGGC
ACTCGAGCCGCTGTCGATGCAGCTGCTGCCGCTGCCGCCGAGCCGCTCTGCATGCTAAAGCTAACGCGGTTTTGCAA
GCTACCAGCAAAGCAGCCGCTGAAGCTAGAGTCGCGGCTGAGGAGGCAGCATCCGCCCAGGCATCCGCCTCAGCAAGC
GCACAGCTGACCGCACAAATTAGAGGAGAAAGTCAGCGCCGATCAACAAGCAGCCTCCGCCAGTACTGATACCTCCGCT
GCTATAGCCGAGGCTGAAGCTGCCGCGTTAGCGTCCACCGTCAACGCGATCAACGACGGAGTGGTCATCGGATTAGGA
AATACCGCCAGTTCTTCTGCCCAAGCTTCCGCACAGGCCAGTGCTCTCGCTCGCGCAAAAAATGCGCGCCCTAAAATA
AAGGGCTGGTACAAAATCGGAGGCGGACTTCCGCTTCTGCAAGCGCATCGGCCAGCGCTTCCGCCCAGTCATCCTCG
CAAGGACTGGTATACTAG

GAF4 (SEQ ID NO:71)

ATGAAGATTCCAGCGATACTTGGCAGCTCCCTTTTCGTCTGGGGTCTTGTGCGGCCAGCGAACTCGTCGGATCGGAC
GCGAGCGCGACGGCATCTGCTGAAGCGTCAGCATCGTCATCCGCATACGGTAGCAAGTATGGTATTTGGTAGTGGTGCT
GTCTCCGGTGATCAGCCAGCGCCTCTGCCAGCGCGTCTGCTAGCGCATCAGCCAGCAGTGCTCCCGCGATCGAAGGA
GTAAACGTTGGCACCGGAGTCAGTAACACCGCTTCCGCGTCCGCAGAAGCTCTCTCCCGTGGACTCGGCATCGGACaA
GCGGCTGCCGAAGCGCAAGCCGCTGCCGCTGGCCAAGCGGCGATCGCTGCGAAATCGTGCGCGCTAGCGGCCAAGAGC
ACCGCTCAAGCGGTTGCCCTGGTTGAGAAAGTGGCCCGCGCCGAGGTAGATCTGGCCGAAAGCGCGAGAAAGGCTACA
AGATTATCGGCAGAAGCAGCCAAGGCAGCGGCGGAAGTCGAGAAGGACCTCGTCGGTCTGAGAGGGGCTGCCGGTAAA
CTGAATCTGGCTGCGAGAGCCGGTTCTAAAGCCCAAGAAGCGGCCAACGAAGACTCTATAGAGGCTAACGAAGTTGCC
CAAGCAACGGCCGCCGCCGCTGCCGAGGCTGAAGCCAAGGCGAATGCCGCCCAGGAGGCAGGCGCCTCCGCTTTGGCC

19 / 22

ATCGCCCAAGCCGCCCTTAACATCGAGCAAGAGACTGTTAAATTGACCCGCCAGGCCAGAATACTCGTCTCAGATCT
GAAAATATTCTCGCCGCGGCCAGCAATGCCCCGCGCCATCGCTTCCGCTGAGGCCGAGGCCAGTAGTGATTTGAATAAT
CGTGCGAATGCAGCGCGTTCCAATGCCCCGAGCTGCTGCCGAGACCAGAGCCGTAGCTACCGAAGCCGCTTCTACCGCC
GAGATCGCAGCTTATAGTTTCATCCGAGAAAGGCGagATCACCAATCCCGGTCTCTGCCCAAGATCGTCAGTGTTACC
GCAGGTCTGACCCagAACGAAATAGCGGGATCAGGAGCGGCCGCTAGTGCTAGTGCCAGTGCTCTTGCCAGTGCCAGT
GCCGGTGCCGGTGCCGGTGCCAGGTGCAGGTGCAGGAGCCGGTGCAAGTGCCAGGAGCCGGTGCCAGTTGCAGGTgCAGGAGCCGGT
GCAGGAGCCGGTGCTAGTGCCGGAGCGAGTGCCGGAGCGAATGCCGGTGCCGGTGCCAGCAGTTTACTCTTGCCGCAG
AGTAAACTCCATCCAATCTCCAGGTCTTCCGCCTCTGCCTCCGCTTCCGCCGAGGCCGAAGCTAACAGTTTCGGCGTAT
GCGTAA

MalF1 (SEQ ID NO:75)

ATGGCAGCGTCGAACAAAATCATCTTCAGCTTTTTAGCTATTGTTCTATTACAACCTGCCACACACTGTTTCATCAACA
GCTGTATTGATTTCTGGTTCGGCTGCTGGTGCTTCTCACACAATGCTGCTGGTGCCAGCTGCAGCAGCCAGAGCTGCC
TTAGGCGCTTCTGGGGCTGCAGGTTTAGGTGCTGCATCTGGTGCTGCAAGAAGAAACGTAGCAGTTGGTGCTAACGGT
GCCGCCGCCGCTAGTGCTGCAGCTGCAGCTGCCAGACGAGCTGGCGCTATTGGCCTAAATGGAGCAGCTGGAGCTAAT
GTAGCTGTCGCTGGTGGCAAAAAGGAGGTGCTGCTGGATTAAATGCTGGCGCTGGTGCTTCTTTAGTATCTGCAGCT
GCAAGACGAAATGGAGCCCTTGGACTTAACGGTGCCAGCTGGAGCAAATCTCGCAGCAGCTGGTGGCAAAAAGGAGGT
GCTATTGGATTAAACGCTGGAGCATCAGCCAATGTTGGTGCCGCTGCTGCCAAGAAAAATGGAGCCATAGGACTTAAC
TCAGCTGCTTCAGCTAATGCTGCCGCTGCCGCTGCTAAAAAGGTGGAGCCATTGGATTGAATGCTGGAGCTTCAGCA
AATGCTGCTGCTGCCGCTGCCAAGAAGAGTGAGCTGTTGGATTAAATGCTGGAGCTTCTGCTAACGCTGCTGCTGCT
GCTGCCAAGAAAAGTGAGCTGTTGCTGCCAATTCGCTGCTTCAGCAAATGCAGCTGCTGCTGCACAAAAGAAAGCC
GCTGCTGATGCCGCAAATGCTGCTGCTTCTGAAAGTGCTGCTGCTGCTGCAGCCAAGAAAGCCGCCGCTGTTGCTGAA
AATGCAGCTGCCACCGCCAATGCCGCTTCAGCTTTACGTAAAAATGCATTAGCCATTGCCAGTGATGCAGCAGCTGTC
CGTGCTGATGCCGCTGCCGCCGCCGCTGACGATGCTGCTAAAGCTAACAACGCTGCTTCCCGTGGAAGTGATGGTTTA
ACTGCCCCGCGCAATGCCGCCACTTTAGCCAGTGATGCTGCCCGTAGAGCTAGCAATGCAGCAACAGCTGCCAGCGAT
GCTGCCACTGACCGATTGAACGCCGCCACCGCTGCTAGCAACGCTGCCACTGCTCGTGCAAATGCCGCCACACGTGCC
GATGATGCCGCCACTGATGCCGACAATGCTGCTTCAAAGGCCAGTGATGTATCAGCTATTGAAGCCGACAACGCTGCA
CGAGCTGCTGATGCTGATGCTATCGCTACCAACCGTGCCGCTGAAGCAAGCGATGCTGCTGCTATTGCCGCTGATGCC
GCTGCCAATGCTGCTGATGCCGCTGCCAATGTAATAACAAAGTTGCCCGAGTAAGTGATGCCCTTAGCTCTCGCCGCT
AATGCTGCTGCCGAGGATCTGATGCCGCCGCTGAAGCTCAAGATGCTGTTGCCAGAGCAAGTGACGCTGCCGCTGCC
CAAGCTGATGGTGTTGCCATTGCCGTAAATGGAGCTACTGCGAGAGACTCAGCAATTGAAGCCGCTGCTACTGCTGGA
GCTGCCCAAGCTAAAGCCGCTGGACGTGCTGGAGCTGCTGCAGCTGGTTTAAAGAGCTGGTGCCGCTAGAGGTGCTGCC
GCTGGTAGTGCCCGCGGTCTAGCTGGAGGATTAGCTGCAGGTTCCAATGCTGGAATCGCGGCTGGTGCCAGCTTCTGGA
TTAGCAAGAGGCGCAGCTGCTGAAGTTTGCGCAGCTAGAATAGCATTGTAA

Figure 11

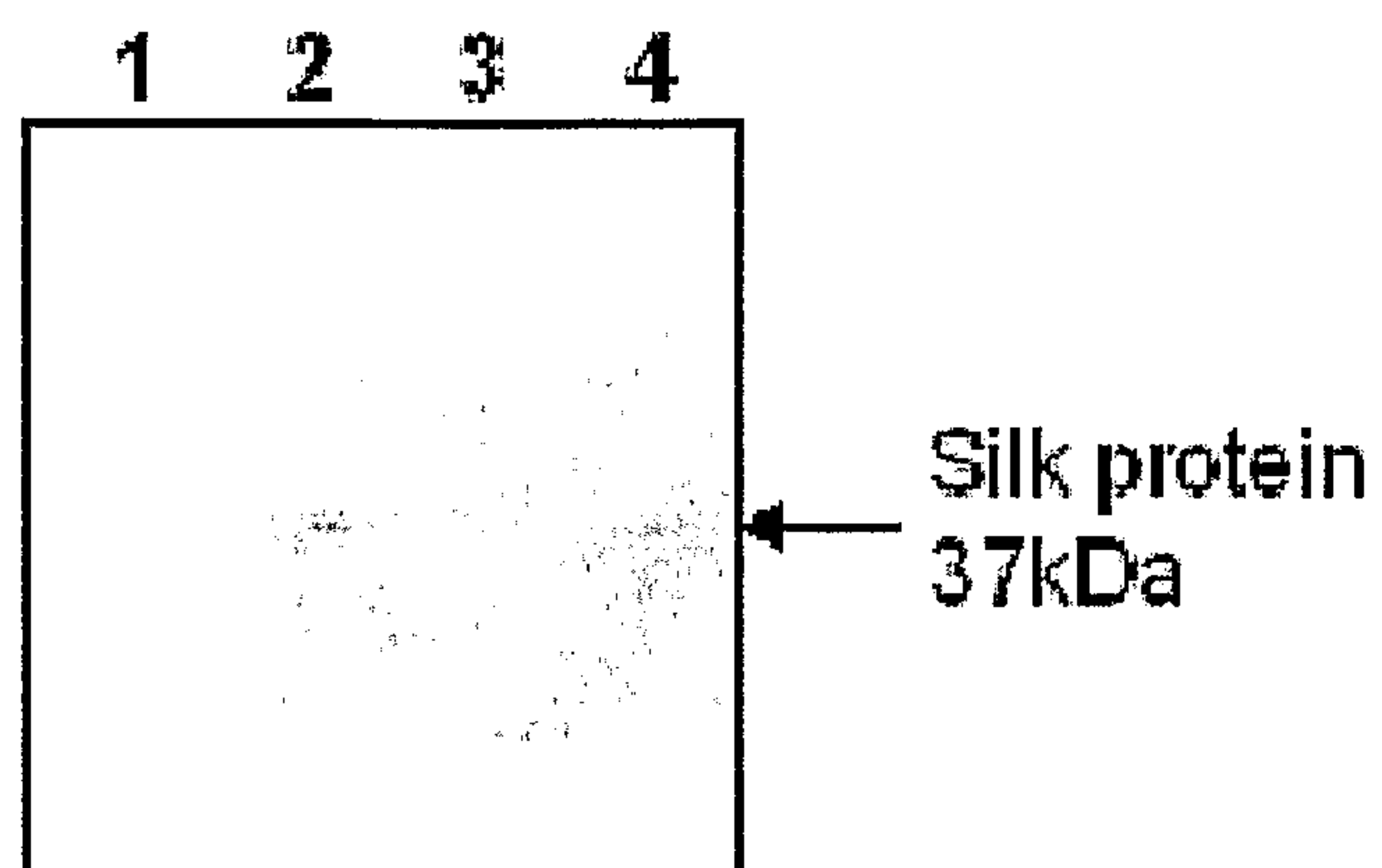
20 / 22

Xenosin - Inclusion of intron

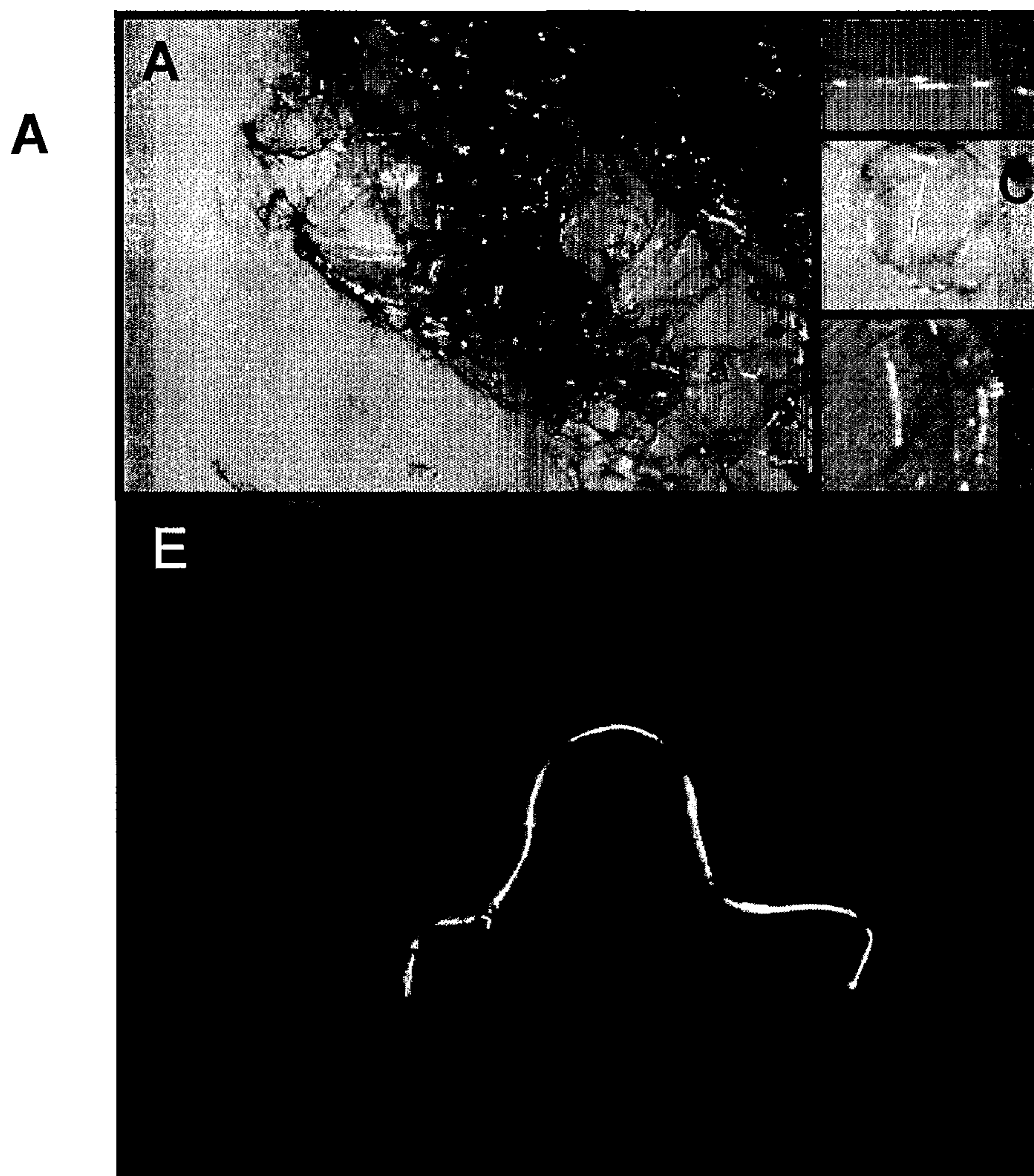
ATGAAATACATGCTCTTGTTGCTATCTATATTCATCTGTGCACATATTGTATGCGCAGGCGTAAATACAG
AATTAAAAAAGATGGTGAACATAAGGAAGAGTCTTATGAGAAAAGCGAGTCAAAGAGTTTAAAAGAAAT
TAAAGAAGAACGTGCTTCAAAATCAAAAAGTGAACGTTTGAAGATTCGTGAAGGTAATTCGTGAGATTCA
AGATTCAAATCAATTAAATTTGAAAATTATGAAAGTAGTATTGTTAAATTATAAGATAGAAGATTTTATC
TAAAAAATAATAAATTAAGCTTTTTGTATTTTTGGATATTGTAGATATTTTAAATATAGAATTCTTATAA
AGTTAAAAAATATTTTATAAATTAAACAACTTTTTATTATTTTTATGATCTAAAAATTAAAAATTTCAAG
TTAAAGTTCAAATTAAAAATTTGTAAAAAATATGGAAAAACATAAAAAATTGAATTTGTTGTAATTTAAA
AAGGATTTTTATTATTTATTGATTAAATTATGAATATAAGTTCGAAAAATCCTAAATATTAATGTTTTAAA
TTTTAATTCTTAACAAAATATATTTAATTTAATTCTTAACAAAGATACATTTAAAGAATTTTCGCAAATTT
AAAAATTAGGTTTTTAATTTTAAGAATCAAATGGTAAAAAACATTTTAAATTTGAAATATATAAAAGTAA
ATCTTTTAATCGACAAACGGATGAATTTATTGATTAGAAAAACGCGAAGAGGAAGAAAAATCCAAGAGTC
TGAATCTGGTCGTGGTCAGAGAAAAGATTACCAAACCTTCTTCATGGCTCAAAGAAGAGAAAGATATCAG
TCCTCTTTTGAAGAAAAAAATGGCAAAGGTCTATTGGGTTTGAAGATGTCACGGACGAGTTAAATATC
GCTCTTAAATCGTTGAAGGAGGGCAAAAAGTTTGATACTTGGAAATTCGAGAAAGGTAGCGAAGACGTTT
GTTCTTTGAAGAACTTGATACGAGCGTCGTTGAACTTTTAAATTAATAAAGGAAGGAAAAACTGACCA
TGGTGCTATAGATTTGGAGAAGAATGGTAAGGTACTTGTAGATTTGGAAAAAATCTCAGAAAACATACTT
GAACTTGTGGATCACAAAAGAAGACTGTGGAAGTTGTAGATGATAAAGACAAAAAATGGAATAAAGAAT
CAGGTTGGAAAAAAATCTAAATGATCTAGATTGGAAAAAAGATTTAGATAAAGATAAAGTTGGTGGCGG
TTTGCTTGGCGGTTTTAAGTGGCCTCTTAAATAGTTTAAAATCAGAAAAAGGTCTTCTAGGTCTTTTGAAT
AAGAATCAAATTGAGTTATTAATTCCTTTAATCAGTGAGATAAAAAAGAAAAATATAGATTTTAAATCTCT
TCGATTCTGTTGATTCTGTGAAAGAAATTTAGACTTGAACTTTTCACAAGTTCTGTTTCAAAGTTAC
TGAATTATTAAATAAAGGAATCGATATTCAAACAATTTTGAATGCGAAAAATGGAGATGAATTCGATTTA
AGCGGCAAAGAATTGAAAAACGTCAAAGGGATATTTGGTTTGATTGGAAGTTTGAAACGCTCATTAGGAT
TAGAAAATATATTGAACTTACCGTTTAAACGTATACCTCTGCTTAAATTA

Figure 12

21 / 22

**Figure 13**

22 / 22

**Figure 14**