Compounds and methods for the production of suckerin and uses thereof

Cross-Reference to Related Application

[0001] This application claims the benefit of priority of United States of America Provisional Patent Application No. 61/715,068 filed October 17, 2012, the contents of which being hereby incorporated by reference in its entirety for all purposes.

Field

[0002] The invention relates to new robust proteins and polymers, their chemical designs and methods of production and use thereof.

Background

[0003] The plastics industry alone accounts for about 7% of all petroleum used per annum, amounting to billions of dollars spent on non-renewable resources that clearly have a negative impact on the environment. Therefore there is a need to develope industrially useful materials including textiles, bio-compatible materials and "smart" materials with tailored mechanical properties and innumerable additional functionalities that manifest at the atomic, molecular, supra-molecular, meso and macroscopic scales.

[0004] There are many groups and companies around the world uniquely devoted to engineering high-value materials based on proteins. An excellent case study comes from the high-performance adhesive proteins used by marine mussels to adhere to their substrates[1]. These proteins exhibit unrivalled adhesion in an aqueous milieu, dramatically outperforming the other best known glues. Protein based materials are also gaining significant traction in the field of biomedical engineering where they are being used for controlled drug release and as substrates for tissue engineering and repair. Key advantages of proteins include 1) biocompatibility; 2) ability to assemble proteins, often through multiple scales of structural hierarchy, to achieve desired nano and microstructures and 3) diversity in material states and shapes-for example films, fibers, hydrogels, foams, nano-scale templates, nano-spheres and microspheres etc. can all be engineered into materials with tailored structure and mechanical properties. In the push towards these advanced "green techno-materials" silkworm and spider silk protein-based materials have gained significant headway[2,3]. The list of applications and high-profile publications and patents on the subject is ever growing and far reaching. As part of the effort to understand and manipulate silk into useful materials, groups from around the world have studied the primary amino acid sequence, structure, processing mechanisms and mechanical properties of natural silk spinning animals including silkworms and spiders[4-11]. The strong industrial push towards the production of artificial silk is best demonstrated by the efforts of several groups focusing on the genes and proteins[12-18] and the methods for processing and spinning reconstituted or recombinantly derived silk fibroins[19-23]. Each of these report spun artificial spider dragline silk fibers from recombinant

fibroins, but unfortunately, the mechanical properties of these fibers generally remain inferior to those observed in nature, for example[24]. One of the main reasons for this shortcoming is that the spinning mechanism employed by the spider has yet to be adequately mimicked. Therefore, there are several obstacles that that must be overcome before the production of artificial silk on an industrial scale becomes reality.

[0005] First, the production of artificial spider silk on an industrial scale will require a cost effective recombinant silk fibroin protein source. Native silk genes are extremely large and repetitive and, as a result, they can be unstable in bacteria or yeast because of the occurrence of undesired recombination events. Excessively large transcripts may also misfold in expression hosts, a process that can result in translational pausing and the truncation of recombinant products[25]. In addition, native spider silk genes contain heavy codon usage biases that render them incompatible with many expression systems[25]. Mammalian expression systems appear to have less trouble with spider silk genes, but these systems are more costly than microbial expression systems. More recently, the drive toward the production of large quantities of recombinant fibroins has focused on the use of transgenic organisms such as plants[26-30], silkworms[31,32], and goats[24,33]. Xia et. al.[34] recently demonstrated that by modifying the tRNA pool in *E. coli* it was possible to express a large spider silk protein of 285KDa that was close to the molecular weight of native spider silk proteins. The authors were able to spin their proteins into fibers with mechanical properties rivaling those of native silks. The data clearly demonstrates that silk properties depend directly on molecular weight and that silks spun from proteins of lower molecular weight were inferior. While this recent work offers significant potential, it also highlights a second major problem with the creation of biomimetic silk, the fact that fibroins have a tendency to aggregate into unwanted precipitates during processing. In fact efficient spinning was only achieved using the extremely harsh (and expensive) hexafluoroisopropanol that would be impractical to scale up. Nexia Biotechnologies Inc., have demonstrated the ability to spin a fragment of recombinant dragline silk fibroin from the aqueous phase[24]. However the basis of this achievement remains unclear. While these results suggest that a method for the production of artificial silk from the aqueous phase may be at hand, the mechanical properties of these materials remain inferior to those measured from native dragline silk. The main reason for this shortcoming appears to be that a precise level of control over fibroin folding and chain alignment is not yet attainable. In the spider, the fibroins self-assemble through several scales of structural hierarchy, involving the sequential development of at least three liquid-crystalline mesophases within the silk-gland[11,35-37]. Control over fibroin folding, self-assembly and the global molecular orientation of fibroin dopes within the silk gland involves very precise micro-environments and elongational flow forces along the

length of the glandular lumen that facilitate polymer chain alignment and the spinning of the fibroin dope into a solid state beta-sheet re-enforced fiber. During spinning, silk proteins are folded and packed into colloidal and/or liquid crystal assemblies and their conversion into solid state fibers requires a combination of draw elongation[38,39], pH and salt gradients along the spinning gland[38] and very precise molecular designs of the N- and C-terminal domains that are sensitive to shear and micro-environment and that are believed to facilitate processing and polymerization into a beta-sheet re-enforced polymer network[40-43]. Clearly, the mimetics of such a mechanism will require a combination of significant advances in the field in order to translate these concepts into industrial processes.

[0006] Finally, as desirable as the mechanical properties of spider silk are in the dry state, dragline undergoes a process known as "supercontraction" when it is hydrated, where it shrinks to about 50% of its original length and becomes rubbery, where glassy structure becomes mobile and some of the beta-sheet structure that re-enforces the fiber melts and re-arranges[9,44,45]. Clearly supercontraction must be avoided if artificial spider silks are to be used for industrial applications. Despite these potential caveats, it is important to consider that the mechanical properties of native silk fibers are truly exceptional and in many cases may not be required for many applications. In fact the silk field has witnessed a dramatic expansion in the development of silk-based materials where extreme strength and toughness are not absolutely required. Currently the major challenges in the silk field for many applications are 1) industrial scale-up and production of large quantities of pure recombinant protein in a cost effective manner, a topic whose challenges have been addressed thoroughly by[25]; 2) preventing aggregation and/or solubilizing the protein for processing using inexpensive and/or environmentally friendly conditions; 3) controlling/tailoring the supra-molecular organization of the materials in order to tailor properties.

[0007] Efforts to engineer other new materials inspired by biological structures are limited by the lack of genomic data from many model organisms. Although knowledge of the primary sequence is key, several challenges remain with identifying target protein sequences. In particular, the genomes of most model organisms studied in biomimetics research have not been sequenced, and the proteins that constitute these materials are notoriously difficult to solubilize and characterize. A comprehensive analysis through traditional cDNA cloning approaches can take years, substantially limiting the scope and depth of information that can be extracted. Therefore, a comprehensive view of molecular design and natural manufacturing processes that ranges from genotype to mechanical phenotype is usually not directly accessible.

[0008] The Decapodiformes are a superorder of the class cephalopoda that include

over 300 species of squid and cuttlefish. The tentacles and arms of all Decapodiformes species are lined with hooks, spines and crown shaped "teeth" that are used in predation, grappling and object manipulation (Figure 1A). Squid and Cuttlefish exhibit impressive predatory prowess that involves a rapid tenticular strike. Their attack efficiency depends intimately on their ability to ensnare prey using their strong suckers which are lined with crowns of sharp, teeth-like structures. While initially thought to be comprised entirely of chitin, more recent work by Miserez[46,47] indicates that these robust structures are assembled almost entirely from protein. These structures are unique to Decapodiformes and are collectively referred to as Sucker Ring Teeth (SRT).

Summary

[0009] A first aspect of the invention includes an isolated polypeptide comprising; or consisting of:

a. the amino acid sequence set forth in any one of SEQ ID Nos. 2, 21, 23, 24 or 25; or

b. a variant of the amino acid sequence of (a) that shares at least 40 %, , preferably at least 50 %, more preferably at least 60 %, 80% or 90% sequence identity with any one of the amino acid sequences of (a) over its entire length

c. a fragment of the amino acid sequence of (a) or (b), the fragment having a length of at least 50, preferably at least 100, more preferably at least 150, most preferably at least 200, 400, or 600 amino acids;

wherein the polypeptide is at least partially in beta sheet conformation.

[0010] Another aspect of the invention includes an isolated nucleic acid molecule comprising a nucleic acid sequence encoding the polypeptide as described herein.

[0011] Another aspect of the invention includes a host cell comprising the expression construct as described herein able to express the polypeptide of the invention. The host cell may be any suitable host cell for recombinant production. The cell may be a prokariotic cell or a eukariotic cell. There are many cell based systems known in the art. Alternatively the polypeptide may be expressed by cell free translation.

[0012] Another aspect of the invention includes a material comprising the polypeptide as described herein. Preferably the material maintains the beta sheet structure. In various embodiments the material is selected from the group: a fibre, filament, a film, a foam, a nano fibre, a nano sphere, a nano particle, a liquid crystal mesogen, a colloid, a copolymer, or a block copolymer.

[0013] Another aspect of the invention includes a method of making the material comprising the polypeptide as described herein comprising the steps of: (a) crushing the polypeptide; (b)

heating and/ or using a solvent to dissolve the crushed polypeptide into a liquid; and (c) forming the material from the liquid.In various embodiments forming the material comprises spinning the liquid into a fibre. In various other embodiments forming the material comprises placing the liquid into a mold; and removing the mold after the liquid solidifies.

[0014] Another aspect of the invention includes a tissue scaffold comprising the polypeptide as described herein.

[0015] Another aspect of the invention includes a multimer of the polypeptides as described herein.

[0016] Other aspects of the invention would be apparent to a person skilled in the art with reference to the following drawings and discription of various non-limiting embodiments.

Brief description of the drawings

[0017] The drawings are not necessarily drawn to scale, emphasis instead generally being placed upon illustrating the principles of various embodiments. In the following description, various embodiments of the invention are described with reference to the following exemplary drawings.

[0018] **Figure 1:** (A) Low-magnification SEM micrograph image of a single Squid Sucker Ring Tooth (SRT) from *Dosidicus gigas*. (B) high-magnification SEM image of fracture surface showing the nanotubular architiechture of SRT cross section through (A) as indicated.

[0019] **Figure 2:** Integrating RNA-seq, proteomics and material science. RNA-seq of the tissue of interest is performed and the transcriptome is assembled *de novo*. In parallel, comprehensive proteomics of the protein building blocks is conducted, as well as materials and biophysical characterization. Transcriptome data are integrated with one or a combination of different proteomic techniques (Edman sequencing, amino acid analysis and mass spectrometry analysis) to identify the relevant transcripts. The integrated approach allows for the rapid generation of genomic to mechanical phenotype maps, eventually providing key information for subsequent structural analysis and biotechnological applications that mimic the properties of the starting materials, using for instance recombinant protein expression or synthetic peptide mimics.

[0020] **Figure 3:** Optical micrograph of a microtomed cross-section surface of a SRT and elastic modulus mapping of the sections obtained by nanoindentation.

[0021] **Figure 4:** Isolation and characterisation of Suckerin protein **(a)** SDS Page of the crude SRT extract. The most abundant protein at ~39kDa was targeted for this study. **(b)** Primary sequence modularity of Suckerin, illustrating extreme modularity, with alternating Ala/His and Gly/Tyr rich domains. These domains are composed of smaller repeats with some homology to dragline silks.

[0022] **Figure 5: (A)** Raman spectrum of a microtomed section compared against a *Bombyx mori* silk standard, showing matching of the Amide I band centered at 1666cm-1, which is attributed to β-sheet conformation. The lower intensity suggests that the β-sheets are less oriented than in silk fibers. **(B)** Raman spectra in cross-polarized modes confirm that the β-sheet domains in SRT are randomly oriented in the cross-sectional plane.

[0023] **Figure 6: (A)** styalized drawing of the random orientation of β-sheets in SRT. **(B)** Evaluation of photo-cross-linking efficiency with recombinant protein (rec)Suckerin-39. **(a)** Effect of APS concentration on cross-linking: SDS-PAGE of 1 mg/ml (rec)Suckerin-39 with Ru salt at various APS concentrations. Lanes 1 and 2 demonstrate little to no protein migration indicating effective cross-linking. Cross-linked intermediates are visible in Lanes 3-6. At lower APS only monomers were observed. **(b)** SDS-PAGE of 1mg/ml (rec)Suckerin-39 at various light exposure times. **(c)** MALDI-TOF analysis of cross-linking at various light exposure times. Bottom spectrum: (rec)Suckerin-39 control sample. At 5 sec (middle spectrum), oligomeric intermediates are detected; at 2 min (top spectrum) no signal is observed indicative of effective cross-linking.**(C)** the mean distances between prolines for all three species, reflecting the mean length of the [M1] domains.

[0024] **Figure 7:** Recombinant suckerin-39 protein expression. Lanes 7 and 8 show recombinant suckerin at ~ 39 kDa.

[0025] **Figure 8:** Engineering of recombinant Suckerin-39. **(a)** FTIR of (rec)Suckerin-39 films illustrating the β-sheet dominated content. **(b)** MALDI-TOF analysis of photo cross-linked (rec)Suckerin-39. Oligomers are observed under intermediate cross-linking conditions (middle spectrum). In the fully cross-linked state, all peaks disappear (top). **(c)** Elastic modulus of (rec)Suckerin-39 films before and after cross-linking under dry and wet conditions; moduli of dragline silk are shown for comparison. **(d)** *H* vs. *E* Ashby plot of engineering materials and comparison with crosslinked (rec)Suckerin-39 (wet). Wear resistance is predicted to scale with $H3/E2$, which can be represented by guidelines with slope *2/3* on the chart. In this representation, best-performing materials against wear lie along the upper-most guidelines. **(e)** (rec)Suckerin-39 as a matrix for cell culture, showing hMSCs viability after 2 weeks by live/dead staining and variation in cell morphology as a function of protein concentration.

[0026] **Figure 9.** Fourier Transform infrared Spectroscopy of (A) Suckerin-based film and (B) Suckerin based fiber. The data show that the transformed Suckerin-based material contains a significant percentage of beta-sheet structure. Fiber/Film wavenumber; Side Chain 1598/1595; Beta-Sheet 1614/1613; Beta-Sheet 1629/1629; Random coil Film 1644; alpha helix 1644/1653; turn 1673/1669; beta-sheet 1689/1689.

[0027] **Figure 10.** depiction of the top 20 common peptide domains observed in known

Suckerins. Counts of each module are normalized to total lengths of all Suckerins within each species. (A) The top 20 *D.gigas* modules sorted based on 3-mer and 5-mer frequency. For comparisons, the frequency for *S.lessoniana* and *S.esculenta* are shown. (B) The top 20 modules for all 3 species sorted based on 4-mer frequency. (C) The top 20 modules for all 3 species sorted based on 8-mer frequency.

[0028] **Figure 11. (A)** SDS PAGE of SRT proteins from 3 distantly related Decapodiformes. Lane 1: Molecular weight marker. Lane 3, 4 and 5 are, *D.gigas*, *S. lessoniana* and *S. esculenta* total SRT protein extracts respectively. **(B)** 2D isoelectric focusing gel of *D.gigas* (A and B); *S. lessoniana* (C and D) *and S. esculenta* (E and F) SRT proteins with pH 7-10 focusing strips.

[0029] Figure 12. Phylogenetic relationships of all known Suckerin proteins. Left: Neighbor Joining model based phylogenetic relationships of all known Suckerin proteins from *D. gigas* (Dgig), *S. lessoniana* (Sless), and *S. esculenta* (Sesc) and bootstrap scores listed at branch nodes. Right: Large scale modular architecture of Suckerin proteins from the sequences. (SEQ ID 2 = Dgig-9 in CLADE 1); (SEQ ID 3 = Dgig-25 in CLADE 3); (SEQ ID 21= Dgig-20 CLADE 2); (SEQ ID 23 = Dgig-22 in CLADE 4); (SEQ ID 24= Dgig-23 CLADE 6); (SEQ ID 25 = Dgig-24 CLADE 5).

[0030] **Figure 13.** Mechanical Properties and structural arrangement of SRT tip cross-section under various micro-environments **(A)** Nanomechanics of *D. gigas* SRT tip cross sections in air (■); water (●); 100 % ethanol (♦); 0.02 M Urea (▲:); 0.2 M Urea (◊); 2M Urea (►) **(B)** Corresponding Micro-raman spectroscopy of *D. gigas* sucker SRT section subjected to the same micro-environments described in (C). Arrow and dotted lines indicate 1236 cm^{-1} peaks that correspond to beta-sheet structure. 1236 cm^{-1} peaks that correspond to beta-sheet structure. Curves from top to bottom: Curve 1 Dry; Curve 2 Dry after Urea; Curve 3, 17 minutes; 4, 35 minutes; 5, 48 minutes and 6, 90 minutes exposure to 0.2 M Urea respectively.

Detailed Description

[0031] Here we describe entirely novel proteins that assemble into silk-like beta-sheet re-enforced materials and that offer multiple advantages for circumventing the major pitfalls of silk engineering. The novel SRT proteins (Suckerins) were identified from RNA-seq databases that were generated from the Sucker tissue of three distantly related Decapodiformes species. Similarity in amino acid composition, primary amino acid sequence and modular sequence designs and phylogenetic analysis suggested that these proteins are encoded by a gene family.This description relates to new robust polymers produced by all Decapodiformes species, their chemical designs and methods of production and use thereof.

[0032] The Sucker ring teeth (SRT) exhibit a distinct nano-tubular architecture (Figure

1B) and function under multiple loading regimes including bending, torsion and compression. The actual proteins of SRT and the nucleic acids that encode them have never been reported before the priority date.

[0033]. Using Next Generation Sequencing combined with proteomics tools, we have identified an entirely novel family of genes that encode a new family of modular SRT proteins unique to Decapodiformes. The proteins have been named Suckerins.

[0034] Accordingly, a first aspect of the invention includes an isolated polypeptide comprising; or consisting of:

a. the amino acid sequence set forth in any one of SEQ ID Nos. 2, 21, 23, 24 or 25; or
b. a variant of the amino acid sequence of (a) that shares at least 40 %, , preferably at least 50 %, more preferably at least 60 %, 80% or 90% sequence identity with any one of the amino acid sequences of (a) over its entire length
c. a fragment of the amino acid sequence of (a) or (b), the fragment having a length of at least 50, preferably at least 100, more preferably at least 150, most preferably at least 200, 400, or 600 amino acids;

wherein the polypeptide is at least partially in beta sheet conformation.

[0035] The term "isolated polypeptide" refers to a polymer of amino acids and its equivalent and does not refer to a specific length of the product; thus, peptides, oligopeptides and proteins are included within the definition of a polypeptide. Isolated polypeptides are separated from other cellular components with which it may naturally occur including cellular debris or are synthesized using known methods such as liquid or solid phase synthesis preferably by SPPS using either Fmoc or Boc. The resulting polypeptides are preferably 70, 80 or 90% pure, preferably at least 95 or 98% pure polypeptide containing less than 30% contaminants, preferably less than 20 or 10% and most preferably less than 5 or 2% contaminants that cannot be identified as the polypeptide. In various embodiments the polypeptides may be linked together to form multimers. This term also does not refer to, or exclude modifications of the polypeptide, for example, glycosylates, acetylations, phosphorylations, and the like. Included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, natural amino acids, etc.), polypeptides with substituted linkages as well as other modifications known in the art, both naturally and non-naturally occurring. Polypeptides of the present invention are isolated, synthesized or engineered peptides and may have between about 70 and about 600 amino acids that form a structural protein. In various embodiments the poly peptides may have between about 50 to 500, preferably between about 100 to 400, more preferably

between about 150 to 300, most preferably between about 200 to 280 amino acids.

[0036] The term "at least partially in beta sheet conformation" refers to a protein at least capable of forming an architecture whereby at least two backbones of 3 or more amino acids are connected laterally via hydrogen bonds. The protein may include other structural architecture provided at least part of the polypeptide is able have at least two backbones of 3 or more amino acids are connected laterally via hydrogen bonds. Preferably, at least 20% or at least 40, 50, 60, 70, 80 or 90% of the polypeptide is able to form into a beta sheet. Beta strands may be formed of typically 3 to 10 amino acids connected laterally by at least two or three backbones via hydrogen bonds, generally forming a pleated sheet. The hydrogen bonds need not be perfect and may exhibit localized disruptions. Generally the hydrogen bonds lie roughly in the plane of the beta sheet. The bonding pattern may be parallel, antiparallel or a mixture of both parallel and antiparallel bonding. The pleats may form in an isotropic formation or an anisotropic formation and may be anywhere in the range of 5 to 8 angstroms apart. In preferred embodiments the beta sheet structure may be isotropic pleats of 5 to 6 angstroms.In preferred embodiments the polypeptides conform into beta sheets under dry or hydrated treatments via localized hydrogen bonds that play a role in the mechanical properties of the suckerin polypeptides.

[0037] The term "percent sequence identity" is taken to include an amino acid sequence which is at least 40, 50, 60, 70, 80 or 90% identical, preferably at least 95 or 98% identical at the amino acid level over at least 10, 20, 50, 100, 200 or 300 amino acids with the amino acid sequences within each alternative sequence. In particular, percent sequence identity should typically be considered with respect to those regions of the sequence known to be essential for the structure of the protein rather than non-essential neighboring sequences. Preferred polypeptides of the invention comprise a contiguous sequence having greater than 40, 60 or 70% identity, more preferably greater than 80 or 90% identity, to one or more of amino acids set forth in any one of SEQ ID Nos. 2 to 53 within each alternative sequence.

[0038] The term "percent Sequence homology" as used herein in relation to amino acid sequences, means that a residue of a given molecule is identical or homologous or similar to that at the corresponding position of a reference molecule. Similarity or homology means that the residue has similar characteristics as the one in the corresponding position of the reference molecule but is not identical. One example for such similarity is a "conservative amino acid replacement", where one amino acid is replaced by another with similar properties. Examples of conservative substitutions are the replacements among the members of the following groups: 1) alanine, serine, and threonine; 2) aspartic acid and glutamic acid; 3) asparagine and glutamine; 4) arginine and lysine; 5) isoleucine, leucine, methionine, and valine; and 6) phenylalanine, tyrosine, and tryptophan. In some

embodiments, the term "percent sequence homology" is taken to have the same meaning as above in relationship to polypeptides that are isolated from the sucker ring teeth of organisms that share descent from a common decipodiforme ancestor.

[0039] Calculation of the % sequence identity or % sequence homology preferably firstly requires the production of an optimal alignment, taking into consideration gap penalties. A suitable computer program for carrying out such an alignment is the GCG Wisconsin Bestfit package and others known in the art. Examples of other software that can perform sequence comparisons include, but are not limited to, the BLAST package and the GENEWORKS suite of comparison tools. Both BLAST and FASTA are available for offline and online searching.

[0040] Although the final % sequence identity or % sequence homology can be measured in terms of identity, the alignment process itself is typically not based on an all-or-nothing pair comparison. Instead, a scaled similarity score matrix is generally used that assigns scores to each pairwise comparison based on chemical similarity or evolutionary distance. An example of such a matrix commonly used is the BLOSUM62 matrix - the default matrix for the BLAST suite of programs. GCG Wisconsin programs generally use either the public default values or a custom symbol comparison table if supplied (see user manual for further details). It is preferred to use the public default values for the GCG package, or in the case of other software, the default matrix, such as BLOSUM62.

[0041] The term "fragment" includes a length of at least 20 amino acids; a length of at least 50, preferably at least 100, more preferably at least 150, most preferably at least 200 amino acids sequences set forth in any one of SEQ ID Nos. 2, 21, 23 24 or 25; or the variant having at least 70, at least 80, at least 90, at least 95, at least 96, at least 97, at least 98 or at least 99 % sequence identity with the amino acid sequence set forth in in any one of SEQ ID Nos. 2, 21, 23 24 or 25; or the variant over the stretch of at least 20 amino acids. In preferred embodiments the fragments occur over domains, particularly over consensus domains or peptides whereby the fragment of the amino acid sequence of (a) that shares at least 90 %, preferably at least 95 %, more preferably at least 99 % sequence identity or sequence homology with any one of the amino acid consensus domains or peptides described herein.

[0042] The polypeptide sequences SEQ ID Nos 2, 21, 23 24 or 25 represent clades 1-6. A clade is a subgroup of polypeptides within the suckerin polypeptides that is closely related to polypeptides within the same clade. Preferably the polypeptides of Clades 1 to 6 are represented by a specific sequence whereby any amino acid sequence of that shares at least 40 %, preferably at least 50 %, more preferably at least 60 % sequence identity or sequence homology with the representative amino acid sequences over its entire length,

falls within the same clade, wherein the representative amino acid sequence of clade 1 comprises the sequence set forth in SEQ ID NO. 2, the representative amino acid sequence of clade 2 comprises the sequence set forth in SEQ ID NO. 21, the representative amino acid sequence of clade 3 comprises the sequence set forth in SEQ ID NO. 3, the representative amino acid sequence of clade 4 comprises the sequence set forth in SEQ ID NO. 23, the representative amino acid sequence of clade 5 comprises the sequence set forth in SEQ ID NO. 25, the representative amino acid sequence of clade 6 comprises the sequence set forth in SEQ ID NO. 24.

[0043] Preferably the polypeptide remains soluble in water and/or under mildly acidic conditions. Preferably mildly acidic conditions have a pH in the range of about 4 to 7, or 4 to 6 or 4 to 5. Alternatively, mildly acidic conditions include any acid having a pKa of above 4. The solubility under non-toxic contitions has the advantage that the polypeptides can be processed in large volumes without producing toxic waste.

[0044] In various embodiments the polypeptide comprises at least one domain 1 comprising or consisting of an amino acid sequence selected from the group consisting of HH, TT, SS, AV, TTH, THH, IAAL, SY, VTHHAP, VVLLAAF, HTTHHA, AATVSHTTHHA, FPY, THT, HVT, HHP, VSH and TVS. Preferably the at least one domain 1 comprising amino acid sequences having one or more Alanine, Valine, Threonine, Serine or Histidine.

[0045] In various enbodiments the polypeptide comprises at least one glycine rich domain 2.

[0046] In various enbodiments the glycine rich domain 2 comprises at least one peptide sequence selected from GGLYG, GGLYG, GGYG, GLYGG, YGIG, GIG and GYG.

[0047] In preferred embodiments the glycine rich domain may also have a high number of Tyrosine and Leucine amino acids. Domain 2 is a glycine rich [M2] module that may be ~ 3 -90 residues long or 10-100 residues long and is itself often assembled from smaller tripeptides tetrapeptides and pentapeptides modules or referred to herein as sub-domains with consensus sequences. In various embodiments at least one glycine rich domain 2 comprises at least one sub-domain 3 comprising amino acid sequence $GGX_1X_2X_3Y$ (SEQ ID NO:1) wherein $X_1$is nothing, L, F or V; wherein X_2 is nothing, G, or F; and wherein X_3 is nothing, G or A. In various preferred embodiments the X_1 of the sub-domain 3 is Leucine. In various preferred embodiments the sub-domain 3 has the amino acid sequence GGLY. In various preferred embodiments the at least one glycine rich domain 2 may further comprise at least one sub-domain 4 comprising amino acid sequence GGY. The sub-domains may only occur once within the domain 2 or the sub-domains may be repeated several times within domain 2.

[0048] In various embodiments the amino acid sequence comprises, at least one

consensus sequence selected from sequences set forth in SEQ ID NOS. 105, 106, 107, 108, 109, 110.

[0049] In various embodiments the polypeptides of clade 1 comprises a group of sequences comprising at least one domain 1 including the peptide THH and at least one domain 2 including the peptide GGLYG, These polypeptides may include any one of amino acid sequence set forth in SEQ ID NOS. 2, 7, 8, 10, 12, 13, 15, 17, 22 and 38 having a consensus sequence set forth in SEQ ID NO. 105 ($XGX_1XXGGYGX_2X$ wherein X is any amino acid, X_1 is W or G and X_2 is L or A). The polypeptises of clade 2 comprises a group of sequences comprising at least one domain 1 including the peptide IAAL and at least one domain 2 including the peptide GGLYG, These polypeptides may include any one of amino acid sequence set forth in SEQ ID NOS. 21, 30, 42, 43 and 52, having a consensus sequence set forth in SEQ ID NO. 106 ($AALXCQXX_1AX_2X_3XnPXX_4XGXX_5X_6X_7GXYG$ wherein X is any amino acid, $X_1$is A or E, X_2 is A or I, X_3 is L or I, X_4 is G or V, X_5 is P or G, X_6 is F or L, X_7 is G or N and n is 90-200). The polypeptides of clade 3 comprises a group of sequences comprising at least one domain 1 including the peptide SY and at least one domain 2 including the peptide GGYG, These polypeptides may include any one of amino acid sequence set forth in SEQ ID NOS. 3, 5, 19, 26, 35, 47 and 50, having a consensus sequence set forth in SEQ ID NO. 107 ($X_1XGGX_2GXnX_3G$ wherein X is any amino acid, $X_1$is G or L, X_2 is Y or G, X_3 is G or V, and n is 2-15). The polypeptides of clade 4 comprises a group of sequences comprising at least one domain 1 including the peptide VTHHAP, or VVLLAAF and at least one domain 2 including the peptide GLYGG. These polypeptides may include any one of amino acid sequence set forth in SEQ ID NOS. 23, 32 and 44 having a consensus sequence set forth in SEQ ID NO. 108 ($LIFVVLLAAFGX_1AX_2X_3EXnVTHHAPYX_4XGXnGLYGG$ wherein X is any amino acid, $X_1$is L or F, X_2 is C or Y, X_3 is G or S, X_4 is T or G and n is 2-100). The polypeptide of clade 5 comprises a group of sequences comprising at least one domain 1 including the peptide FPY, THT, HVT, HHP and at least one domain 2 including the peptide YGIG or GIG, These polypeptides may include any one of amino acid sequence set forth in SEQ ID NOS. 25, 46 and 49 having a consensus sequence set forth in SEQ ID NO. 109 ($GX_1VSX_2QXPFXX_3X_4PXYGIGXnGAX_5XHGF$ wherein X is any amino acid, $X_1$is L or I, X_2 is A or G, X_3 is A or P, X_4 is G or F, X_5 is F or Y and n is 10-15). The polypeptides of clade 6 comprises a group of sequences comprising at least one domain 1 including the peptide TVS and at least one domain 2 including the peptide GYG, These polypeptides may include any one of amino acid sequence set forth in SEQ ID NOS. 24, 37, 45, 51 and 53, having a consensus sequence set forth in SEQ ID NO. 110 ($XGXX_1XnSX_2XTVSH$ wherein X is any amino acid, $X_1$is Y or F, X_2 is V or I, and n is 2-10).

[0050] In various embodiments the polypeptide comprises an amino acid sequence set forth in any one of SEQ ID Nos. 2 to 53; or (b) a variant of the amino acid sequence of (a) that shares at least 40 %, preferably at least 50 %, more preferably at least 60 % sequence identity or sequence homology with the amino acid sequence of the amino acid sequence set forth in any one of SEQ ID Nos. 2 to 53 over its entire length.

[0051] Preferably the percent sequence identity should typically be considered with respect to those domains or regions of the sequence known to be essential for the structure of the protein rather than non-essential neighboring sequences. Preferably the polypeptide comprises at least one domain 1 comprising amino acid sequences having one or more Alanine, Valine, Threonine, Serine or Histidine and forms the beta-sheet domains that are occupied by A, V, T S, and H rich sequences. This domain is designated as the [M1] module. In various embodiments domain 1 comprises amino acid sequence HH, or TT, or SS or AV. or AA, or, TV. In various embodiments domain 1 is flanked by prolines at either side of the domain or at least at one side of the domain. Without being limited by any theories it is postulated that in various embodiments the prolines form the edges of the beta-sheet. In various preferred embodiments the domain 1 has the amino acid sequence THH or TTH. or VSH. In various embodiments the domain 1 has the amino acid sequence HTTHHA. In various embodiments the domain 1 has the amino acid sequence AATVSHTTHHA.

[0052] Fragments and derivatives of full length polypeptides, particularly include fragments or derivatives having substantially the same biological structure and mechanics involved with the formation of the beta sheet. A particularly preferred polypeptide consists of amino acids with domain 1 and 2 mentioned herein. There may be repeats, substitutions and deletions within the domains meaning the domains can greatly vary in length.

[0053] In various preferred embodiments the polypeptide may further comprise at least one proline residue. The proline residues may form at the edges of the beta sheets, as such they are often either before and/or after domain 1. In various embodiments the proline residues flank either side of the domain 1.

[0054] In various embodiments the polypeptide comprises a peptide unit from N- to C-terminus: Proline-Domain 1-Domain 2,.

[0055] In various embodiments the polypeptide comprises a peptide unit from N to C terminus, Proline-Domain 1-Proline-Domain 2.

[0056] In various embodiments the domain 2 comprises from N to C terminus, (sub-Domain 3-sub-Domain 4) n, wherein n is independantly 1 to 4. In various embodiments the domain 2 comprises comprises from N to C terminus, (sub-Domain 3-sub-Domain 4-sub-Domain 3)n, wherein n is independantly 1 to 3.

[0057] In various embodiments any of the described peptide units may be repeated

from 2 to 13 times. There may be other amino acid residues between each of the domains or each of the peptide units mentioned herein provided the polypeptide is able to form a beta sheet structure.

[0058] The polypeptide may be isolated and purified from a species in the Decapodiformes super order.Decapodiforme is a superorder of cephalopoda having 10 limbs and sucker ring teeth within the suckers on there limbs. The identified orders are Bathyteuthoidea, Spirulidea, idiosepidea, Sepolidea, Oegeopsidea and Myopsidea. There are about 95 identified genera and about 450 species. The polypeptide may be isolated and purified from any of these 450 species.

[0059] Another aspect of the invention includes an isolated nucleic acid molecule comprising a nucleic acid sequence encoding the polypeptide as described herein.

[0060] The term "isolated nucleic acid" as used herein refers to any nucleic acid molecule in any possible configuration, such as single stranded, double stranded or a combination thereof. Isolated nucleic acids include for instance DNA molecules, RNA molecules, analogues of the DNA or RNA generated using nucleotide analogues or using nucleic acid chemistry, locked nucleic acid molecules (LNA), peptide nucleic acid molecules (PNA) and tecto-RNA molecules. DNA or RNA may be of genomic or synthetic origin and may be single or double stranded. Such nucleic acid can be e.g. mRNA, cRNA, synthetic RNA, genomic DNA, cDNA, synthetic DNA, a copolymer of DNA and RNA, oligonucleotides, etc. Any nucleic acid capable of expressing the polypeptides of the invention including the domains in a host cell would be suitable. Isolated nucleic acid are separated from other cellular components with which it may naturally occur including cellular debris or are synthesized using known methods The resulting nucleic acids are preferably 70, 80 or 90% pure, preferably at least 95 or 98% pure nucleic acid containing less than 30% contaminants, preferably less than 20 or 10% and most preferably less than 5 or 2% contaminants that cannot be identified as the nucleic acid as described herein.

[0061] Many nucleotide analogues are known and can be used in the isolated nucleic acid of the invention. A nucleotide analogue is a nucleotide containing a modification at for instance the base, sugar, or phosphate moieties. As an illustrative example, a substitution of 2'-OH residues of siRNA with 2'F, 2'O-Me or 2'H residues is known to improve the *in vivo* stability of the respective RNA. Modifications at the base moiety include natural and synthetic modifications of A, C, G, and T/U, different purine or pyrimidine bases, such as uracil-5-yl, hypoxanthin-9-yl, and 2-aminoadenin-9-yl, as well as non-purine or non-pyrimidine nucleotide bases. Other nucleotide analogues serve as universal bases. Universal bases include 3-nitropyrrole and 5-nitroindole. Universal bases are able to form a base pair with any other base. Base modifications often can be combined with for example a sugar

modification, such as for instance 2'-O-methoxyethyl, e.g. to achieve unique properties such as increased duplex stability.

[0062] In a preferred embodiment the isolated nucleic acid molecule comprises any one of the sequences set forth in SEQ ID Nos. 54 to 100. Preferably the isolated nucleic acid molecule is optomised for expression in a given host.

[0063] The term "optomised for expression " as used herein, refers to the addition of regulatory elements. This could include promoter or enhancher sequences as known in the art. promoter or enhancher sequences vary from organism to organism, but are well known to persons skilled in the art for different organisms. The promoter selected is preferably a strong promotor capable of high levels of transcription to drive rapid polypeptide expression. Generally viral promoters have these criteria as these are required for efficient viral propagation, and they frequently induce much higher levels of transcription than eukaryotic promoters by using mechanisms to control and recruit host transcription machinery. Moreover, viral promoters tend to be far more compact and hence easier to manipulate and accommodate into gene vectors. Human cytomegalovirus major immediate early gene promoter (hCMV) would by suitable. Other suitable viral promoters include the simian virus 40 (SV40), Rous sarcoma virus long terminal repeat (RSV-LTR), Moloney murine leukaemia virus (MoMLV) LTR, and other retroviral LTR promoters. Any other suitable promotor known in the art capable of high levels of of transcription would be suitable. The isolated nucleic acid of the invention is operably linked to the promoter, and preferably lies 3' to the promoter sequence, more preferably lies directly adjacent to the promoter sequence.

[0064] In a prefered embodiment the nucleic acid sequence is comprised in an expression construct preferably a vector as is known in the art. The term "vector" relates to a single or double-stranded circular nucleic acid molecule that can be transfected into cells and replicated within or independently of a cell genome. A circular double-stranded nucleic acid molecule can be cut and thereby linearized upon treatment with restriction enzymes. An assortment of nucleic acid vectors, restriction enzymes, and the knowledge of the nucleotide sequences cut by restriction enzymes are readily available to those skilled in the art. A nucleic acid molecule encoding a polypeptide can be inserted into a vector by cutting the vector with restriction enzymes and ligating the pieces together. Preferably the vector is a plasmid.

[0065] Another aspect of the invention includes a host cell comprising the expression construct as described herein able to express the polypeptide of the invention. The host cell may be any suitable host cell for recombinant production. The cell may be a prokariotic cell or a eukariotic cell. There are many cell based systems known in the art. Alternatively the polypeptide may be expressed by cell free translation. The host cell may include a

multicellular host system of a plant or an animal. Transgenic models are known in the art.

[0066] Another aspect of the invention includes a material comprising the polypeptide as described herein. Preferably the material maintains the beta sheet structure. In various embodiments the material is selected from the group: a fibre, filament, a film, a foam, a nano fibre, a nano sphere, a nano particle, a liquid crystal mesogen, a colloid, a copolymer, or a block copolymer.

[0067] Another aspect of the invention includes a method of making the material comprising the polypeptide as described herein comprising the steps of: (a) crushing the polypeptide; (b) heating and/ or using a solvent to dissolve the crushed polypeptide into a liquid; and (c) forming the material from the liquid.In various embodiments forming the material comprises spinning the liquid into a fibre. In various other embodiments forming the material comprises placing the liquid into a mold; and removing the mold after the liquid solidifies.

[0068] Another aspect of the invention includes a tissue scaffold comprising the polypeptide as described herein.

[0069] Another aspect of the invention includes a multimer of the polypeptides as described herein.

[0070] The present invention is further illustrated by the following examples. However, it should be understood, that the invention is not limited to the exemplified embodiments.

Examples of preferred embodiments

Tissue Collection

Jumbo squid were caught off the east coast of the Baja Peninsula La Paz, Mexico. *Sepioteuthis lessoniana* squid were caught off the east coast of Singapore, Changi. *Sepia esculenta* were bought in a pet shop in Singapore. These represent three distantly related Decapodiformes species 1) jumbo squid (*Dosidicus gigas,* Order Oegeopsida); 2) green eyed reef squid (*Sepioteuthis lessoniana,* Order Myopsida) and 3) the golden cuttle fish (*Sepia esculenta:* Order Sepiidae). Live animals were sacrificed, dissected and materials and SRT and sucker tissues harvested immediately. Tissues were stored in RNA-later (Qiagen).

Macro structure of sucker ring teeth

[0071] The protein-based SRTs exhibit a precise nanotubular architecture (Fig. 1b) and feature micro-mechanical properties matching those of strong synthetic polymers[47], with elastic moduli in the range of 6–8GPa in the dry state (Fig 3) and 2-4 GPa when hydrated Notably, these robust properties are achieved in the absence of covalent bonds between the constituent proteins, which are soluble in 2M urea/weak acidic conditions implying that the cooperation of weak interactions are responsible for holding the entire structure together into a robust supramolecular network.

Raman spectroscopy

[0072] Microtomed SRT cross-sections were probed with a confocal Raman CRM200 microscope (WITec) equipped with a diode-pumped 785 nm near infrared laser (Toptica Photonics), a water immersion objective (Nikon, ×60, numerical aperture of 1) and a P-500 piezo-scanner (Physik Instrumente). ScanCtrlSpectroscopyPlus software (WITec) was used for measurement and image processing. Protein amide I chemical images (with 500nm steps and 0.4s integration time) were obtained using a sum filter and integrating the 1600-1700cm^{-1} spectral range.

Isolation and purification of a new class of proteins

[0073] RNA-seq uses deep sequencing to survey the transcriptome in an unbiased and comprehensive manner. Full-length transcripts can be assembled *de novo* from short sequence reads generated from massively parallel sequencing, and this approach has had a profound impact in a number of areas of research. However, to our knowledge its broad integrative potential for biomimetic engineering has not yet been demonstrated. Transcriptomic data sets were integrated with proteomics data in three ways to identify relevant transcripts. First, where Edman protein sequence was available, we identified assembled transcript sequences by sequence alignment. Second, if only amino acid composition was available, we calculated amino acid profiles for each transcript, and identified the transcript using the most similar profiles in terms of molecular weight, amino acid composition and isoelectric point values. Third, we used the translated transcriptome database as a reference protein database for mass spectrometry analysis.

[0074] Using Next Generation Sequencing combined with proteomics tools, we have identified an entirely novel family of genes that encode a family of modular SRT proteins unique to decapodiformes of the cephalopods. The proteins have been named Suckerins. The complete genes (Gene SEQ ID 53-100) and protein (Protein SEQ ID 2-52) sequences (in corresponding order) for Suckerins identified from the Jumbo Squid *Dosidicus gigas;* Green Eyed Reef squid *Sepioteuthis lessoniana*; and the Golden Cuttle fish *Sepia esculenta*. are provided.

[0075] Integrating the primary protein sequence with Raman micro-spectroscopy, *x*-ray scattering and mechanical data at various length scales allowed us to rapidly generate genotype to mechanical phenotype maps for each of these model biological materials (Fig. 2).

RNA-Seq Tanscriptome Library Preparation

[0076] Total RNA was extracted with a Qiagen RNeasy mini kit. 2-10µg Total RNA per tissue was used in the construction of three separate RNA-Seq libraries. DEPC water was added to total RNA to a final 50µl volume and incubated at 65°C for 5min. Poly-A mRNA was

enriched with oligo dT beads (Invitrogen) and used for constructing strand-specific paired-end libraries (ScriptSeq™ mRNA-Seq library kit v1, Epicenter, Illumina). Phusion PCR polymerase (Thermo Scientific) was used for the final library amplification (12 cycles). PCR cleanup was performed with the MinElute PCR purification kit (Qiagen). Library quantity and quality was assessed with an Agilent 21000 Bioanalyzer.

Library Sequencing

[0077] Each library was diluted to 8pM and clusters were generated on paired-end-read flow cells on an Illumina cBot. Each library was sequenced on a separate lane and 2x76 bp paired-end-reads were collected.

Transcriptome Assembly

[0078] Raw reads were converted to fastq format using the Illumina's Offline Base Caller (v1.6). *De novo* transcript assembly was performed with the Trinity software suite using standard parameters on a computational cluster. The final butterfly output predicted transcript files were used for subsequent analysis.

Protein annotation

[0079] Open reading frames were predicted using custom Perl scripts. Predicted transcripts were quantified by RSEM software. Predicted protein sequences were first searched using USEARCH against NR (Genbank) and Pfam protein databases to identify homologous sequences. Custom Perl scripts were written to match predicted proteins against proteomic data including: *(i)* De novo LC MS/MS peptides, *(ii)* MW, *(iii)* Edman sequences, and *(iv)* amino acid composition.

LC MS/MS

[0080] 5μl of digested peptides were separated with a Nano C18 HPLC column (Agilent 1200) and data were collected with a QTOF 6520 (Agilent). The *m/z* range for the parent ion was 350 to 1200Da, and 100–800Da for the fragment ions. Data were acquired with MassHunter Acquisition B.040 and processed with MassHunter Qualitative Analysis B.040 (Agilent). PEAKS Studio 5.3 was used for DeNovo sequence tag creation, and searches against our created transcriptome libraries were conducted using Spider and Peaks search routines.

X-RAY diffraction

[0081] Wide-Angle X-Ray Scattering (WAXS) and Small angle scattering (SAX) were collected using a synchrotron source.

2D electrophoresis.

[0082] SRT from the tentacles of *D. gigas*, S. *Lessoniana* and *S. esculenta* were pulverized separately in liquid nitrogen and re-suspended in Bio-Rad rehydration buffer with the appropriate ampholytes at 2mg/ml concentration. 125 ul of each sample was incubated

overnight with BioRad 3-10 and 7-10 Ready strips in preparation for isoelectric focussing. Samples were focused using a 0-4000V ramp. The strips were then subjected to reduction and alkylation steps according to manufacturers' instructions and then the second dimension was run with a low initial voltage of 10V/cm followed by 200V for 30 minutes. Mini-Gels were stained with Sypro Ruby stain according to manufacturers' instructions

[0083] SRT from the tentacles were pulverized in liquid nitrogen, homogenized in 5% acetic acid/2M Urea, and subjected to SDS-PAGE. Bands were transferred to a PVDF membrane, stained with Commassie blue and subjected to Edman sequencing (Iowa State Protein Facility). Suckerin was also isolated by Size-Exclusion (SE) HPLC (Agilent 1260 Infinity LC) and subjected to chymotrypsin digestion and LC-MS/MS. *De novo* peptides determined using PEAKS software were screened against our transcriptome database.

[0084] A list of candidate transcripts was generated by matching the overall amino acid composition of the SRT against the translated predicted transcripts. We initially focused our attention on the most abundant protein present in the sucker crown (Fig. 4a). An *N*-terminal Edman sequence of the isolated protein was mapped to one candidate sequence, which was also among the most highly expressed transcripts in our transcriptome library (Table 1). The gene contained highly repetitive sequences resulting in fragmentation of the transcript assembly. Therefore we used RACE-PCR to identify the full-length transcript. Total RNA was subjected to RACE-PCR using Invitrogen's Generacer Kit. PCR primers were designed based on Trinity assembled transcripts and KOD extreme Taq-polymerase was used for PCR (Merck Milipore). Products were sub-cloned into pCR2.1 by TA cloning, subjected to Sanger sequencing and alignments performed with ClustalW.

[0085] The identified gene/protein was a highly modular, glycine-, histidine- and tyrosine-rich 39 kDa protein with partial homology to silk[4,5] and oyster shell matrix proteins[55]. However, the protein is clearly distinct from these or any other proteins, supporting the view that we have identified a structural protein, named Suckerin that had previously not been described. The identity of Suckerin was corroborated by LC-MS/MS with 80% sequence coverage (data not shown). Here, the repetitive nature of the protein sequence informed our decision to use the more promiscuous chymotrypsin protease for protein digestion for subsequent LC-MS/MS analysis that provided ~ 80% coverage of the protein sequence confirming the identity and sequence design of this protein

Table 1. Transcript expression levels determined by RSEM. Transcript abundance estimation from RNA-SEQ reads mapped to Trinity assembled transcripts. ***Expected Count***: the estimated number of reads that are derived from the transcript. ***Tau***: the estimated fraction of transcripts made up by this transcript. ***Transcripts per million***: estimated number of transcripts per million transcript molecules. ***Abundance Rank***: the relative rank of the

transcript in the entire transcriptome sorted by Tau.

Dosidicus gigas Sucker (149,005 predicted transcripts)					
Transcript ID	**Expected Read Counts**	**Tau (Estimated fraction of transcripts)**	**Transcripts per million (Tau x 10^6)**	**Abundance Rank**	**Annotation**
comp3_c1	1,888,587	0.028133893	28133.89	2	Myosin Regulatory Light Chain
comp14_c0	152,983	0.02743	27432.24	3	Suckerin-39
comp15_c0	292,197	0.01860	18597.67	5	Suckerin-39
comp4_c0	1,918,297	0.016067085	16067.09	7	Tropomyosin
comp0_c0	9,655,271	0.015045237	15045.24	9	Myosin Heavy Chain

[0086] The combined use of proteomic tools with RNA-seq–generated transcriptome databases, the sequence identification, structure function analysis and biomimetic/protein engineering was rapidly sped up. One of the main advantages of the approach is that it provides a permanent database of all expressed genes in the target tissue that can be rapidly used for further searches. This approach facilitates the rapid characterization of novel structural proteins and provides insights into natural processing mechanisms of uncharacterized biological materials within a matter of weeks, providing a strong foundation for the elucidation of complex structural and functional dynamics.

Amino acid analysis

[0087] Tissues and purified proteins were hydrolyzed in vacuo in 6M HCl/5%phenol overnight, washed on a vacuum evaporator with methanol (SpeedVac), and amino acid composition was analyzed on a post-column derivatization, Ninydrin-based S433 Sykam analyzer.

Evolutionary Relationships of the Suckerin Gene/Protein Family Members.

[0088] The Suckerins are clearly an entirely novel class of proteins and blast searches of each of the known members provides low homology with all other known proteins, with values ranging from 20-39% for the closest matching unrelated structural proteins. All of the Suckerins we have identified to date exhibit variable degrees of identity and divergence. However, they are clearly members of the same or related family of proteins based on the fact that 1) they are rich in Glycine, Alanine, Histidine and/or Tyrosine 2) they are usually

highly modular at multiple scales of hierarchy. Suckerins in general are assembled from large modules (defined as [L]) ~ 20-40 amino acids long that are reiterated ~1-13 times within the entire full length native protein. [L] modules often contain at least 1 and often all three types of smaller modules defined as [M1], [M2] and [M3] where [M1] is usually A, V, T, H or S Rich, [M2] and [M3] which are G rich; 3) the proteins are present in the SRT structure but are not observed in any other known tissue characterized to date. By way of example, which is in no way intended to limit the scope of the current invention, the conserved modular nature of two Suckerins from the 3 diverse decapodiformes when compared and contrasted have similar domains but still have a level of variation. Sequence alignments of selected Suckerins from the three species show some conservation of both N-terminal signal peptides and extensive conservation of repetitive sequences. Taken together, this provides strong evidence that all of the known Suckerins reported herein are members of a gene family.

[0089] Our sequence data on the Suckerins was obtained from three distantly related squid species. In fact, according to current squid cladistics and systematic efforts by others, these species diverged very early on in the history of decapodiformes evolution, having shared a common ancestor dating back as far as 350 million years ago [48,49] and references therein. This common ancestor is believed to have given rise to all known extant squid species. This information provides strong support for the view that, while they have yet to be sequenced, SRT from all decapodiformes are assembled from related proteins expressed from members of the Suckerin gene family described herein. Therefore similar domain structures are considered to describe all members of the Suckerins identified and those yet to be identified that are clearly related evolutionarily and uses thereof.

[0090] Similarity at the genetic level, in amino acid composition, signal peptide sequence, primary protein sequence and modular design at multiple amino sequence length scales support the view that the Suckerins are encoded by members of a gene family. Using the Maximum Joining (NJ) model a statistically reasonable unrooted tree was generated (Fig. 12). Based on potential caveats involved in the use of NJ methods for the phylogenetic analysis of protein families, we make only the most conservative observations based on this tree. First, the NJ tree is statistically well supported, indicating that all of the Suckerin proteins we have identified are evolutionarily related. Second, the genes appear to have diverged by both gene duplication and speciation events. Third, the proteins cluster into 6 distinct clades suggesting that the Suckerin genes comprise a multigene superfamilly. Fourth, five of the six clades contain proteins from all three of the species studied. These species shared a common ancestor ~354 MYa suggesting that the Suckerin gene family is ancient, with Suckerin gene(s) likely arising at least in the Devonian period. Understanding the

molecular mechanisms that underlie the evolution of modular proteins remains a major challenge and an elucidation of these mechanisms as they relate to the Suckerin protein family will likely provide unique insights into the biomechanical and ecological roles of these proteins. Based on the observed variations in size and the presence or absence of small and large modules across the Suckerin proteins it appears likely that genetic divergence has involved gene duplication, segmental expansion/deletion, and slippage of DNA polymerase, non-reciprocal homologous crossing-over and/or gene conversion events observed in genes encoding highly modular proteins. As is the case with spider silk proteins, Suckerins are encoded by guanine:cytosine rich sequences that are known recombination hot-spots. The occurrence of these regions, combined with high degree of homologous modular sequences may have influenced the divergence in the Suckerin gene family. Finally, in addition to molecular drive mechanisms, it will, in the future, be key to unravel the relative contribution the selective pressures acting at the level of the mechanical properties of the SRT.

Phylogentic analysis.

[0091] Full-length Suckerin protein sequences were aligned using ClustalW as implemented in BioEdit sequence alignment editor. We used three different methods of phylogenetic estimation – Neighbor Joining (NJ), Maximum Parsimony (MP) and Maximum Likelihood (ML). Since bootstrap support values were generally low for most clades in the MP and ML trees, we have presented only the NJ tree for the present analysis (Fig. 12). Homology searches against the NCBI non-redundant database were unable to pick up sequences related to Suckerins for use as an out-group. Thus, the tree presented here is an unrooted tree. The NJ tree was generated using MEGA5 using the p-distance method, pairwise-deletion of gaps and 100 bootstrap replicates for node support. Most of the nodes in the NJ tree showed a high bootstrap support value. Based on the topology of the NJ tree, the Suckerin sequences could be assigned to six distinct phylogenetic clades. This suggests that the common ancestor of *D. gigas*, *S. Lessoniana* and *S. esculenta* possessed at least six suckerin genes.

[0092] The protein modular architectures of the suckerin sequences are shown juxtaposed with the NJ tree to clarify the evolutionary history of this protein family.

Structure of Native SRT

[0093] Our current data indicates that the Suckerin proteins are encoded by a large gene family that is of ancient origin. The large scale modular architecture of the Suckerins likely facilitates their intra and inter-molecular assembly into polymer networks consisting of amorphous and/or semi-amorphous structure stabilized by optimally small, nano-confined beta-sheets. The observed distinctions in molecular weights and large scale modular architectures of the Suckerins may offer the potential for the modulation of network

properties including cross-link density and intervening amorphous chain lengths, which would directly influence the mechanical properties of the SRT. It is interesting to speculate that these factors may play a role in the creation of the observed gradients in SRT mechanical properties that likely-prevent modulus mismatch failure at the tissue-SRT interface. The hierarchical design and mechanical behavior of native SRT provides a unique model system for the development of biomimetic engineering strategies and the Suckerin gene family provides an expanded tool-box for the engineering of a range of tailored high-performance biomimetic materials.

[0094] We conducted SDS-PAGE and 2D electrophoresis on three distantly related Decapodiformes species 1) jumbo squid (*Dosidicus gigas,* Order Oegeopsida); 2) green eyed reef squid (*Sepioteuthis lessoniana,* Order Myopsida) and 3) the golden cuttle fish (*Sepia esculenta:* Order Sepiidae). Figures 11A & B show that, for all three species examined, SRT are assembled from mixture proteins whose molecular weights range from ~8 KDa to ~ 60 KDa and whose pIs fall in a relatively confined window between 7 and 10. The data also indicated that *D. gigas* SRT contain a larger repertoire of proteins than those of *S. lessoniana* and *S. esculenta.* Interestingly, the amino acid composition of Suckerin-39 closely matched the global amino acid composition of the entire *D. gigas* SRT suggesting possible similarities between proteins, at least within *D. gigas.*

[0095] To understand the molecular bases for the mechanical properties of native SRT we first examined the protein conformations of SRT slices using micro-RAMAN spectroscopy. Interestingly the SRT exhibits the presence of beta-sheet-structure that is also observed in silks (Figure 7d & e). Polarized micro-raman spectroscopy supports the view that the distribution of beta-sheets is isotropic, an arrangement that would be useful for a material that is subject to a range of loading regimes. In silico predictions using JPRED suggest that the beta-sheet domains are occupied by A, V, T S, rich sequences in the [M1] module. The data indicates that the SRT maintains silk-like supra-molecular organization while being assembled from an entirely novel set of proteins, the Suckerins. The SRT beta sheets likely stabilize and re-enforce the network and may offer superior strength based on nano-confinement principles [50].

[0096] An exhaustive search of all protein databases such as swissprot failed to provide statistically significant hits, supporting the view that the Suckerins represent a novel class of structural proteins. The Suckerins generally have a ~20-24 amino acid signal peptide with ~ 40%-90 % similarity between all of the proteins identified. The full length Suckerin proteins exhibit similarity in amino acid composition, with a heavy biases Glycine, Tyrosine, Leucine, Alanine, Threonine, Histidine, Serine and Valine (Figure 4b). The complete primary amino acid sequences are listed in SEQ ID NOS.: 2-52 and color coded modular designs of the

Suckerins are shown in Figs. 4b. Most of the Suckerins identified contain sets of commonly occurring small peptide modules, including GGY and GGLY. GGY peptides are also present other structural proteins including the spider silk protein family(4, 5), silk worm fibroins (58), the shell matrix proteins Shematrins (55) and Prismalin-14 (59), and crocodile skin β-keratins (*60*) suggesting convergent evolution of this peptide design.

[0097] The current invention relates to the novelty of nucleic acids encoding the Suckerin proteins and the novelty in amino acid primary sequence and modular designs observed in all known Suckerins, natural and synthetic variants thereof and evolutionary related Suckerin genes and proteins. Here, the molecular design of the proteins is described in detail.

[0098] In general the Suckerins range in molecular weight form ~ 5 KDa to 60 KDa. The [L] modules are reiterated consecutively (~ between 3 and 13 times) depending on the size of the [L] and [M] modules and the overall size of the protein. [L] modules usually occur throughout majority of the protein, with the exception of the peptide signal sequence. As mentioned above, the large-scale [L] modules themselves comprise smaller [M] modules. In the clearest cases [L] modules are assembled from clustered domains and/or sequential organization of [M1], [M2]. Figure 4b provides a clear example of the modular architecture of selected Suckerin from *D. gigas, S. lessoniana* and *S. esculenta*. [M1] and [M2] modules often appear in sequence and can be preceded and/or proceeded by proline residues. Clearly defined hexapeptides that occur within [M1] and [M2] and/or that span [M1] and [M2] are listed in Table 2. The glycine rich [M2] module can be ~ 0-90 residues long and is itself often assembled from smaller sub-domains of tripeptides tetrapeptides and pentapeptides also shown in Table 2 Tri, tetra and pentapeptides can also sometimes be rearranged within the glycine rich [M2] domain. In some of the Suckerin proteins the strict use of the small glycine rich peptides listed in Table 2 is not adhered to and the sequence, while biased in glycine residues contains few if any clearly definable tri, tetra and pentapeptides. Deviations from the bipartite [M1][M2] design of the [L] modules can also be observed. However all the proteins examined so far have at least one sub-domain within the [M2] domain 2 comprising amino acid sequence GGXX'Y (SEQ ID NO:1), wherein X is L, F or V; and wherein X' is missing, G, GG, FG, or GA; for example Suckerin (SEQ ID NOS 2-52). Nevertheless, some proteins identified to be present in the SRT represent a protein that has diverged significantly from the common ancestor of the original Suckerin, or another related Suckerin gene/protein family of similar design and function and there is less consensus in domain 1 however at least one domain 1 amino acid sequence HH, TT, SS, or AV is present in all known suckerin sequences.

[0099] The proteins are encoded by genes (SEQ ID NOS: 53-100) producing modular

proteins can also be emphasized by examination of SEQ ID NOS 1-52 where the use and arrangement of large and small scale modules is unique and not observed in any other known proteins. A clear example, which is not intended to limit the scope of design and/or use of Suckerin based materials, is the occurrence of domain organization of the large module with precise placement of prolines, where [L]=P[M1]P[M2] which is observed on many occasions throughout the known Suckerins (SEQ ID NOS 2-53). Note that Proline can be present or absent depending on the module and/or Suckerin protein considered. Only amino acid sequence set forth in SEQ ID NO: 20 has no proline in the protein sequence. These modular designs are not observed in any other protein and may play a significant role in the structure and function of the material as will be described below. The current invention relates to peptides, protein modules, full length proteins and copolymers based on the novel amino acid motifs of Suckerins large and small, as defined above, including point mutations and insertion or deletion of [L] or [M] modules.

[00100] A summary of commonly occurring small domain peptide modules is presented in Table 2 and depicted in Figure 10. Medium sized modules are also evident in the Suckerins, where we use [M1] to denote a ~6-15 amino acid long A, V, T, S, H rich domain modules and [M2] to denote the ensemble of repetitive and non-repetitive Glycine rich domain sequences. [M1] and [M2] modules often occur in tandem are often flanked by proline residues. The combined Pro[M1]Pro[M2] unit (which is an example large module [L]) ranges in length from ~10 to 80 amino acids is re-iterated 3-15 times in the different proteins. Similarities and distinctions between Suckerins are also observed upon examination of their remarkable large scale modular architectures (Figs 4b and 12). The large scale modularity of suckerin highlights the fact that some of these proteins exhibit extreme conservation of large scale module design, while others only marginally do. It is also notable that different Suckerins exhibit differences in glycine-rich module length, a feature that may have a direct impact on SRT mechanics.

[00101] Without being bound by any theory proline is unique among the 20 protein-forming amino acids in that the amine nitrogen is bound to not one but two alkyl groups, thus making it a secondary amine this coupled with the distinctive cyclic structure of proline's side chain locking its ϕ backbone at a dihedral angle at approximately −60°, giving proline an exceptional conformational rigidity compared to other amino acids. Thus, when proline is bound as an amide in a peptide bond, its nitrogen is not bound to any hydrogen, meaning it cannot act as a hydrogen bond donor, but can be a hydrogen bond acceptor. The distinct side chain/amine group interactions allow proline to aid in the formation of beta turns and minimizes aggregation of proteins.

Table 2. Summary of peptides found in [L] modules of Suckerin proteins. Hexapeptides that appear more than twice in all known Suckerin are listed. Hexapeptides can appear in domain 1[M1], domain 2 [M2] and/or domain 3 that can span M2. Domain 2 and 3 [M2] peptides include tri, tetra and pentapeptides that occur more than 10 times in all known Suckerins are listed.

	Hexapeptides
M1	VSHTTH,AVSHTT,TVSHTT,ATVSHT,AATVSH,TAVSHT,ATAVSH,AATAVS,GAVSH T,STVSHG,VSTVSH,AVIALS,AAVSHT,VSRTTH,SVSTVS,AAAVSH,AAATVS,TSVS RT,SVSRTT,VSHVTH,SSSGFV,GAATVS,TSSGSY,HTSSGS,SSSSGF,VIALSS,ISS SSG,ATSVSR,SNYASA,AATSVS,SVTHHA,LVSLAV,TATSTT,SSVTHH,ALVSLA,SL AVVL,FSALVS,TSNSNS,TSAYHA,SALVSL,VSLAVV,AVVLGV,TVSHGV,TTMFSA,A ATTVS,TSTTST,LSYTSA,TSTSNS,SYTSAY,AVSHVT,VLSYTS,TTSTSN,YGAAVS, GAAAVS,GAAVSH,YGAAAV,STSNSN,ATSTTS,STTSTS,ALSSCT,VVLLAA,VSQTT H,GATVSH,ASVSTV,TSVSQT,ATTVSR,VTHTTH,CTLAVV,AASILT,TVSRTT,TTAV SH,RSVSTV,SCTLAV,ATTAVS,AVLAIS,SSCTLA,TSSVTH,VSHVSH,TTVSRT,AAT TYR,HTVTHT,THTVTH,AATVGY,HSVSTV,TVSHVS,ATVGYS,TAVSHV,TVTHTT,A TTSVS,VTSSVT,GHSVST,SVSTLS,TVSTVS,TTTKTV,SSISRV,ISTVSH,SISRVS,VS TVHR,VTVTHH,TVHTVS,VKTVTH,TVTGLV,HTVSHV,KTTSSV,VSHTAH,TVSQTT, STTTKT,SISTVS,TSVKTT,VSIIAA,SSVSHV,QRSVST,ISSSAK,AAASVT,SVRHTV,A AVSHV,VKTVVH,VRHTVS,TSSVHH,TKTVTV,KTVTVT,RVSHTA,SHVTHS,TTSVHH ,VTSSSV,GATVHT,STVTHT,AASVST,AVTSSS,TTHGVA,GISSSA,MAAAVL,VVSHV T,QVTSSV,VATAVS,SINTVS,TSVSHT,YGATAV,VSTLSH,DSSSYS,AVMALS,SSSA KG,SRVSHT,ALDSSS,TGSSVS,AASSVS,AVATYR,HTVSTV,TVTHHA,TSVHHT,SS SYSH,RHTVST,VATYRV,ASINTV,TTSVSQ,ATTVSH,TVTVTH,TTKTVT,AAAVLF,V SHAGA,QTSHSV,AALDSS,SVSHTT,AISTVG,AASTVT,TGLVSA,SHTAHS,GSSVST ,TSSSVT,AHVSQS,SSVSTL,VHTVSH,AAAAVS,SSSVTH,STSVSQ,GSTTTK,TVTH SN,GASVST,TVSHGS,TRSVST,VSRTAL,SVKTTS,ATTIFA,AATVSQ,AAASVS,SSS ISR,LTTTSV,VGASVS,VSQTSH,GQSVST,SVHHTT,GATAVS,GTSVKT,YVSRTA,S VQTVS,YASAIA,TSSSIS,ATVSQT,SVSTVH,TTTSVH,ASVTHG,QAISTV,ASTVTH,S VSQTS,GSSIST,QSVSTV,SSISTV,VKTTSS,TVSHAG,ASTSVS,VTGLVS,VGQSVS, TTVSHT,AASVTH,TTSSVH,ATVHTV,SVSQTT,STVSHA,SQTSHS

	TTHHAP,VSHTTH,HTTHHA,SHTTHH,THHAPY,HHAPYG,HHAPLG,THHAPL,GYGLHH,GLHHGG,YGLHHG,LHHGGL,HHGGLY,VTHHAP,HGLLHG,GHGLLH,RTTHHA,VSHVTH,SRTTHH,TGGHGH,SVTHHA,GHGHGG,VSHGVH,SSVTHH,GGHGHG,HANHVG,HGHGGY,YHANHV,AYHANH,SHVTHH,HVTHHA,HTTHGV,SHTTHG,THHAPA,HHAPAA,VTHTTH,SHVSHG,VSHVSH,THTTHG,HTVTHT,THTVTH,TTHGVH,THGVHH,HGVHSP,SHGIGH,THHAPV,SHGVHS,HHGPYG,TTHHAG,SVHHHV,SHGSHY,HGVSHP,THHAGY,HHLPAA,HAGAHP,SSVHHH,HGVHHP,VSHGLH,HHAPVY,SHGVHP,VTVTHH,SQTTHH,HVSHGP,SHTAHS,HGVHHL,HGIHAP,VSHGIH,HTVSHV,HTAHSP,QTTHHA,VGHGIH,SHDVHQ,VSHTAH,GHGIHP,VHTVSH,HHHVVP,HSVPHI,HTTHGI,HTVSHG,GHHGPY,HHTTHG,HHAPTY,HDVHQP,TVHHPA,THHAPT,HGVAHP,RGHGAH,GVHHLP,THGFHP,TSSVHH,VTHGFH,SHAGAH,THGIHH,HGFHPA,SHGISH,SVHHTT,SHVTHS,TTSVHH,GGHHGP,VSHGSH,VHHTTH,HHAGYG,HGIHPT,HAVHRV,HGVHPS,VHHLPA,SGGHHG,HHVVPS,GHAVHR,VHTVHH,SHGVSH,GSGGHH,GHTVSH,VSHDVH,HAPVAH,VHHHVV,HGIGHP,SHSVPH,IGHAVH,HYPMGH,HGSHYP,TVTHHA,TSVHHT,HGAHVS,HTTHGG,SHGIHA,GHGAHV,THGVAH,HVTHSI,HGISHA,TVTHHP		
M2	**Tri-peptide**	**Tetra-peptide**	**Penta-peptide**

	GYG,GGY,YGG,GGL,YGL,GLG,GLY,LYG,LGG,GFG,HGG,YGF,YGY,PYG,FGG,GYP,LGY,LGA,GLH,AYG,GAY,GLA,PLG,QGG,GQG,GHG,GLL,FGY,GGV,AGY,LHG,YGV,YGA,YGQ,PVG,GAG,VGG,GVG,AHG,GGH,SHG,SYG,GAV,HHG,LAG,PGA,LGL,YGS,LGF,GGF,GGS,GSY,FPG,HGL,GAA,VYG,YGH,GYL,GVS,YPG,VGY,GGA,THG,SSG,GWG,GVY,GYS,GVH,HGV,GGW,GSG,SGY,GGG,LGV,SGG,GIG,GGI,YLG,TYG,NQG,TGG,LLG,VSG,GAP,VLG,GVL,HGI,GYH,GLS,FGL,GST,PGY,GLV,GVF,GAT,GSS,PAG,GAH,PGV,GYY,VGA,GHY	GGLY,GGYG,YGGY,GYGL,GYGG,GLYG,LYGG,YGLG,YGGL,GLGG,LGGY,HGGL,YGFG,YGYG,GFGG,GLGA,LGYG,GGLG,GAYG,GQGG,GFGY,GLGY,LHGG,FGGL,AGYG,GYGY,YGQG,QGGY,GYGQ,LGGL,GYGA,GYGF,AHGG,GLAG,GYGV,FGGY,GGHG,LGFG,SYGG,GYGS,VGGY,HGGY,HHGG,GLGL,YGGH,LGLG,GHGL,GGYL,YGGW,GGLL,GVYG,YGGV,GYLG,GVGG,PYGG,GSYG,GYGH,YGVG,GHGG,GVSG,SGYG,TYGG,VSGG,GAGY,LLGG,YLGG,GYPG,VYGG,VGYG,GSGY,GGVY,GGWG	

[00102] The large-scale modular architecture of Suckerin (Fig. 4b) exhibits two alternating domains containing: *(i)* Ala/His rich motifs ~13 residues long often flanked by proline residues, and *(ii)* Gly/Tyr rich sequences ~2-70 residues long comprised of silk-like GGY repeats, suggesting convergent evolution of these motifs. The Ala-His rich motifs are reminiscent of poly-Ala motifs known to assemble into the reinforcing β-sheet based nano-crystals in spider dragline silk[50]. Micro-Raman spectroscopy (Fig. 5a) on ultra-microtomed SRT cross-sections confirmed a clear β-sheet band across the entire ring with the position of the Amide I band matching that of *Bombyx mori* silk[56]. Polarized Raman spectra (Fig. 5b) showed that, in contrast to spider dragline and silkworm silks, the β-sheet domains are randomly oriented. This isotropic distribution (Fig. 6A), is consistent with the predatory function of SRT that must resist complex compressive, shear and torsional load regimes, with β-sheet domains providing the load-bearing function.

[00103] In view of the extreme potential for robust tailorable materials and protein-based materials re-enforced by beta-sheets, the identification and characterization of the chemical and physical properties described above provide evidence that the Suckerin genes and proteins are useful new reagents having utility in a range material forms and applications. *Suckerin-based films, fibers, foams, scaffolds, micro-spheres, nanoparticles and other forms* can be engineered from synthetic, recombinant or naturally derived Suckerin peptides, proteins and homologues thereof.

[00104] Here we further summarize the proteins encoded by the Suckerins and related variants thereof. The equations are not intended to summarize all Suckerin proteins and potential mimetic variants thereof but rather to facilitate use and are in no way designed to limit the scope of the use of Suckerin-based proteins motifs, modules, fragments and mimetic versions thereof.

[00105] In one simplified version [L]=Large Suckerin module 10-40 residues long comprising Proline and Domains [M] modules. An exemplary, but non limiting list of small peptides is observed in Suckerins and provided in Table 2. [M1]= A, V, T S H rich modules 1-15 residues long; [M2] = a glycine rich module containing the following: A tripeptide module found in any Suckerin protein or any close variant thereof with > 66% identity to natural Suckerin tripeptide, named [sM3] or [sM4]. Sub-Domain 3 [sM3] and sub-Domain 4 [sM4] sits within the [M2] domain. Tetrapeptide module or variant thereof with > 75 % identity to natural observed Suckerin and a pentapeptide or mimetic variant thereof with >60% identity to natural observed Suckerin. A summary of the sequences found in known Suckerins can be found in Table 2.

Example 1. Suckerin (SEQ ID NO: 2) Large Module 4:

[L]= P[M1] P [M2]

[00106] Each of the Suckerin Large modules can be summarized in a similar manner. Use of individual [L] modules, reiterated [L] (for example but not limited to $[L]_{1\text{-}100}$) or [M] modules (for example, but not limited to $[M1]_{1\text{-}100}$, $[M2]_{1\text{-}100}$) and variable arrangement of modules and peptides (listed in Table 2) are available for Suckerin-based materials engineering. Mimetic proteins based on these modules are also considered part of the current invention. It is emphasized that Suckerin proteins are continuous molecules and the description of the genes and proteins in terms of modular designs is by no way intended to limit the use of the proteins and modules in tandem and/or in overlapping sequences and/or that omit or insert individual amino acids, modules or non-repetitive sequence. They are merely provided as a generalized framework for the construction of Suckerin based

materials. The simplified representation of an [L] module in Example 1 is presented to facilitate the design and use of the current invention for the scope of potential engineered nucleic acids and proteins based on the essential components that define the Suckerin gene and protein family and/or evolutionarily related gene and protein families and mimetic genes and proteins based on these.

Mechanical Properties

[00107] The SRT in bulk exhibits exceptional mechanical properties. Furthermore, as is the case with silk, SRT proteins have the potential to be sculpted into a wide variety of materials with an extreme range of mechanical properties. Properties can be tailored by 1) designing Suckerin based proteins with chosen amount of large and small scale modules, to tailor the proportion of beta sheets and other secondary and tertiary structures; 2) Processing micro-environment can be used to modulate chain alignment, structure and mechanical properties; 3) Physical processing, including but not limited to shear, draw and elongational flow can be used to modify chain alignment, beta-sheet crystallization and mechanical properties. The native SRT also exhibits unique nano-tubular architecture. This fact supports the view that it is feasible to assemble Suckerins into a range of useful meso-scale structures.

[00108] Based on these observations and on the similarities in sequence design of Suckerins with Spidroins and silkworm fibroins we asked whether the SRT also contained 3-dimentional silk-like nano-crystals where inter-digitation of residues in the 200 crystal plane can contribute to mechanical properties of the nano-structures and thus the material. We therefore deployed synchrotron-based wide angle x-ray diffraction (WAX) on SRT cross-section sections which revealed a diffuse pattern with line scan analysis providing the signature 4.7 Å inter-strand spacing characteristic of beta-sheets. However, the pattern exhibits significant peak broadening and clear equatorial reflections are absent, indicating that SRT do not contain silk-like 3D nano-crystals and suggesting that SRT beta-sheet strands are likely smaller that 5 nm. While the absence of inter-chain packing is predicted to result in a decrease in local rupture forces, small beta-sheet domains have been shown by simulation to exhibit superior properties, where hydrogen bonds exhibit co-opperativity under mechanical deformation. Moreover, confined beta-sheets can exhibit stick-slip behavior, enhancing toughness and enabling self-healing behavior where hydrogen bonds are recovered following deformation. Based on this information it will be of interest to determine how primary Suckerin sequence design at once enables beta-sheet formation while placing limits on the dimensions and packing of the beta-sheets. Sequences in the [M1] region bear similarity with the beta-sheet forming poly-Alanine sequences of dragline silks and VT motifs recently shown to adopt beta-sheet structure in orb-spider viscid silk. However, the [M1]

motifs are not as repetitive as Spidroin and silk worm sequences, suggesting that irregularity in side chain chemistry in the Suckerins may reduce the potential for inter-sheet crystal packing. Examination of the majority of the Suckerin proteins clearly shows the rigorous positional conservation of prolines in many cases. Proline is a know beta-sheet disruptor that may also place limits on beta-sheet size and directly influence the mechanical properties of the SRT. At this stage we can also not rule out the possibility that the [M2] domains contain order. For example the glycine rich regions of dragline spidroins occupy amorphous domains that, in some cases contain local structure. In addition, glycine rich sequences can have the capacity to adopt amyloid like conformations.

Solubility of Suckerin-based Proteins

[00109] The large scale modular architecture we observe across many of the Suckerins appears well suited for intra and inter-molecular assembly into beta-sheet cross-linked 3-dimentional polymer networks. To further dissect the role Suckerin structure plays in determining the mechanical properties of the SRT, we developed complementary nano-indentation and micro-raman spectroscopy experiments using a set of micro-environmental conditions designed to target and disrupt hydrophobic interactions and hydrogen bonds. Fig. 13A is a plot of Modulus vs. Time obtained by nano-indentation of a section of *D. gigas* SRT tip under dry, hydrated, ethanol and Urea treatments. In the dry and hydrated states the moduli remain linear. Ethanol treatment results in a modest decay in modulus, which is contrary to the behavior of silkworm and spider silk fibers whose moduli increase in the presence of polar solvents.

[00110] Urea treatments were used to target and disrupt hydrogen bonds in SRT sections. In this case, the samples exhibit significant loss in modulus with decay plateaus that correlate with Urea concentration (Fig. 13A). At the highest concentration of Urea used (2M) the final modulus of the SRT was 20 MPa which is dramatically lower than both dry and hydrated conditions. These data provide the first direct evidence of the key role for hydrogen bonds in SRT mechanics and function. We then used micro-raman spectroscopy under a parallel set of micro-environmental conditions to evaluate the changes in secondary structures that occur during these treatments (Fig. 13B). While ethanol treatment yielded no significant peak shifts, Urea treatment leads to a distinct in shift in the Amide III 1236 peak that corresponds directly to a loss of beta-sheet structure and to the decrease in modulus observed by nano-indentation. Together, these data indicate that the hydrogen bonds localized to beta-sheets play a key role in the mechanical properties of the SRT.

[00111] In contrast to silk proteins, Suckerins can be purified and remain soluble under mild acid conditions, while the solubilization of silkworm cocoons and spider threads often requires high salt concentrations and solvents such as Hexafluoroisoproponal that are

hazardous to handle even in small volumes. We have also determined that high concentrations of Suckerin protein remain soluble in low concentrations of Urea (1-2M) and/or under mildly acidic conditions. Facile control of solubility offers unrivalled ability to concentrate and process this material into an extreme range of materials.

[00112] Urea treatments were used to target and disrupt hydrogen bonds in SRT sections. In this case, the samples exhibit significant loss in modulus with decay plateaus that correlate with Urea concentration (Fig. 13A). At the highest concentration of Urea used (2M) the final modulus of the SRT was 20 MPa which is dramatically lower than both dry and hydrated conditions. These data provide the first direct evidence of the key role for hydrogen bonds in SRT mechanics and function. We then used micro-raman spectroscopy under a parallel set of micro-environmental conditions to evaluate the changes in secondary structures that occur during these treatments (Fig. 13B). While ethanol treatment yielded no significant peak shifts, Urea treatment leads to a distinct shift in the Amide III 1236 peak that corresponds directly to a loss of beta-sheet structure and to the decrease in modulus observed by nano-indentation. Together, these data indicate that the hydrogen bonds localized to beta-sheets play a key role in the mechanical properties of the SRT.

Designing synthetic suckerin proteins and uses thereof.

[00113] The following methods for designing synthetic variants of proteins based on Suckerins are described. This section emphasizes that natural Suckerin proteins from any Decapodiformes, natural protein variants thereof, synthesized or engineered suckerins can be use in making sukerin based materials. In general Suckerins can be comprised of a series of repeats with or without local amino acid residue differences that still maintain the essence of the overall large and/or small scale design of the natural proteins observed in Decapodiformes. Genetic engineering or other methods can be used to create chimeras of known Suckerin modules or chimeric genes/proteins that contain non-repetitive sequences while retaining the essence of the large and/or small scale molecular design of Suckerin. Modifications can be done to tune mechanical properties and/or to confer new functionality or processing ability. Novel mechanical properties can include but are not limited to decrease or increase extensibility, failure strength, toughness, compressibility among others.

Engineering Suckerin-Based proteins to include different functionality

[00114] Functionalization of Suckerin-based materials in any and all material forms into materials with desired functionalities is anticipated. Functionalization by the incorporation of the amino acids described herein using genetic engineering or synthetic chemistry is one way in which the Suckerin based material can be endowed with additional functionality. The modules may reduce or confer structure by the introduction of for example, but not limited to, random coils, alpha helices, coiled coils, beta-turns and/or additional beta-sheet forming

domains/sequences to the Suckerin. In another embodiment, modules are included that confer cell binding, adhesive domains for example fibronectin domains. In another embodiment the additional modules enhance solubility, improve spinability, and/or improve the ability to assemble and process the material. Incorporated peptide sequences can be used to provide enzymatic cleavage cites to tailor the biodegradability of the material. Peptides released by cleavage or any other means may be biologically active or cytotoxic depending on design and intended use. In one embodiment cell signaling properties are conferred to the Suckerin-based material.

[00115] Amino acid side chain modification offers extreme potential for the functionalization of Suckerin based proteins and materials described in the Claims. Suckerins contain amino acids that can be converted, by chemical, enzymatic and any other chosen mode(s) of processing into side chains that confer new functions. The Suckerin-based materials may be any material that incorporates the amino acids described herein with individual, multiple and/or combinations of amino acid side chain modifications that confer functionality. These modifications may find use in areas including, but not limited to 1) industrial processing; 2) the incorporation and/or cross linking of additional atoms, molecules, peptides and polymers; 3) materials with novel biophysical functions, including but not limited to tailored materials with adhesive and/or anti-adhesive properties; materials that provide specific cell signaling cues.

[00116] There exist a wide variety of known amino acid side chain modification that occur in biology and these are generally referred to as post-translational modifications. Examples of potential amino acid side-chain modification to the Suckerin-based materials include but are not limited to the following: 1) acetylation/deacetylation; 2) alkylation; 3) amidation; 4) biotinylation; 6) citrullination; 7) formylation; 8) gamma carboxylation; 9) glutamylation of glutamic acid; 10) glycosylation; 11) glycation; 12) glycylation; 13) heme moiety may be covalently attached; 14) hydroxylation; 15) iodination; 16) isoprenylation; 17) lipoylation including: prenylation, and GPI anchor formation such as myristoylation farnesylation and geranylgeranylation; 18) methylation/demythalation of arginine or lysine; 19) nucleotide addition or derivatrives thereof may be covalently attached (ADP, ribosylation, Flavin); 20) nitrosylation; 21) oxidation; 22) palmitoylation; 23) pegylation; 24) phosphatidylinositol may be covalently attached; 25) phospopentetheinylation; phosphorylation, the addition of a phosphate group to serine, threonine, tyrosine and/or histidine; 26) polysialylation (addition of polysialylic acid to NCAM); 27) pyroglutamte formation; 28) tRNA mediated addition of amino acids such agenylation; 29) sulfation, for example the addition of a sulphate group to tyrosine; 30) selenoylation; 31) Tyrosine hydroxylation into adhesive Dihydroxyphenylalanine side chains.

Expression of recombinant Suckerin proteins.

[00117] Expression of recombinant engineered Suckerin genes can be inserted into vectors for subcloning, clone propagation, protein expression and other manipulations common to those skilled in the art of molecular biology and its related techniques.

[00118] While there are many challenges and subtleties involved in the generation of recombinant silk proteins, we have successfully expressed full length 39 Kilo Dalton Suckerin (SEQ ID 2) and purified it using standard IMAC chromatography under denaturing conditions (Figure 7). These results demonstrate the feasibility of expressing and purifying the Suckerins in a straightforward manner.

[00119] Suckerins and variants thereof can be produced using *in vitro* translation systems, by recombinant protein expression using any from a range of prokaryotic or eukaryotic hosts/ cell lines. These host cells include but are not limited to bacterial cells, fungal cells, insect cells, mammalian cells and algal and other plant cells. The proteins can also be produced in transgenic organisms for example in plants, silkworms, mammals or any other suitable host. Depending on the application the product may be Suckerin protein or variants and/or fragments thereof according to the Claims. These proteins may be synthesized in the same or different hosts as desired. Host cells and organisms expressing and/or over expressing desired Suckerin proteins, protein modules and/or variants and combinations thereof provide valuable reagents for many applications including but not limited to the production of nano-fibers, filaments, fibers, films, foams, bulk material and other materials listed but not limited to those presented in the Claims Section of this document.

[00120] To translate the SRT molecular design into engineering applications, we expressed SEQ ID NO. 61 and purified recombinant protein based on SEQ ID NO.: 2 named (rec)Suckerin-39 and processed the purified proteins into structural and functional materials. Full-length Suckerin-39 were readily expressed without the issues commonly faced with the expression of large silk proteins in bacteria (such as gene instability and translational pausing), and (rec)Suckerin protein-based materials could be processed from mildly acidic aqueous solutions. Dry (rec)Suckerin-39 exhibited a β-sheet enriched content by ATR-FTIR (estimated to more than 50%, Fig. 8a), suggesting that we have engineered a supramolecular network that mimics the native SRT. In the dry state, (rec)Suckerin-39 films exhibited moduli comparable to the native SRT (~ 7.5 GPa, Fig. 8c), whereas under hydrated conditions the modulus decreased to 5–8 MPa (Fig. 8c), which is considerably lower than the native hydrated material (4GPa) but similar to hydrated spider dragline silk (10 MPa)[9].

[00121] This low wet modulus suggested a window of opportunity for tuning the stiffness,

which we hypothesized could be modulated by targeting the abundant tyrosine residues via the introduction of di-tyrosine covalent cross-linking. We used a ruthenium-based photo-optical cross-linking approach [61] and established optical cross-linking conditions using SDS-PAGE (Fig. 6B) and MALDI-TOF analysis (Fig. 8b). These conditions were then used to cross-link the (rec)Suckerin-39 films. The elastic modulus of cross-linked recombinant films increased to 8–9 GPa in dry conditions. Notably, the modulus remained similar under hydrated conditions with values ranging from 6–7 GPa (Fig. 8c). This represents one of the the highest elastic modulus of any recombinant protein-based material under hydrated conditions —the modulus is higher than that of hydrated dragline silk by 2 to 3 orders of magnitude. The results also indicate an efficient route for tuning the mechanical properties of Suckerin-based materials over a very broad range.

[00122] Comparing the modulus and hardness (*H*) against engineering materials on a classical Ashby plot (Fig. 8d), the cross-linked films are stiffer and harder than any engineered polymer (including PMMA, PEEK or Polyamides), and are also predicted to exhibit a superior wear resistance than synthetic structural polymers.

Expression and purification of recombinant (rec)Suckerin-39 protein

[00123] The full-length Suckerin-39 open reading frame was amplified by RT-PCR from the 3' RACE library using primers designed based on the 5' sequence of the Suckerin39 gene which was confirmed by Sanger sequencing. The 5' primer was specific to Suckerin-39 (5'-TGAAGGAGTAGAAAGTAGTCTCC-3' SEQ ID NO. 101) while the 3' primer used was the GeneRacer 3' primer (5'- GCTGTCAACGATACGCTACGTAACG-3' SEQ ID NO. 102). The PCR product was cloned into the pCR2.1 vector by TA cloning. The sequence omitting the signal peptide sequence was then amplified by PCR using the following primers:
Suckerin-39-FWD primer:5'-AAAAAAGCTAGCATTTTGCCAGCGGCAACATCTG-3' (SEQ ID NO. 103;
Suckerin-39-REV primer: 5'-AAAAAACTCGAGTTAGTGGAGGAGACCATATCCAC-3' (SEQ ID NO. 104).

[00124] The PCR fragment was then cloned into the NheI/XhoI site of the pET23 vector (Novagen). A stop codon was added at the 3' end of the gene to prevent expression of the c-terminal HIS tag encoded in the expression vector, resulting in a protein that exactly mirrors the native suckerin-39 sequence. The sequence verified construct was transformed into BL21 DE3 cells and maintained with carbenicillin selection.

[00125] Test expression revealed that SRT-39 was expressed as inclusion bodies and we used this method to generate relatively pure protein fractions. A single colony was inoculated and grown overnight in 20ml of LB with carbenicillin (100µg/ml). The next day, the overnight culture was inoculated into a 1L shake flask culture with carbenicillin (100µg/ml) at

a 1:100 dilution. The culture was grown at 37°C until an OD_{600} of 0.4-0.6. IPTG was added to a final concentration of 0.5mM and protein induction was performed at 37 ° C for 4 hours. The cells were pelleted at 10,000 rpm for 15 minutes and washed twice with 10mL of 20mM Tris pH8 and stored at -20 ° C.

The cell pellet was resuspended in 50ml of Lysis buffer (50mM Tris pH 7.4, 200mM NaCl, 1mM PMSF) and lysed using a high pressure homogenizer (Microfluidics Corp). Inclusion bodies and cellular debris were pelleted at 19,000 rpm for 1 hour at 4°C. The pellet was washed twice with urea wash buffer (100mM Tris pH7.4, 5mM EDTA, 2M urea, 2% (v/v) Triton X-100, 5mM DTT) and twice with wash buffer (100mM Tris pH7.4, 5mM EDTA, 5mM DTT). Inclusion bodies were pelleted by centrifugation at 5,000g for 15 min at 4°C in between washes. The inclusion body pellet was re-solubilized in 5% or 20% acetic acid, snap frozen with liquid nitrogen and lyophilized overnight. Recombinant proteins of ~ 90-95% purity were obtained in this manner and yields were estimated at 10 mg/L.

MALDI-TOF

[00126] Proteins mixed with sinapinic acid dissolved in a mixture of 50/50 Q-water/ACN with 0.1%TFA and dried prior to irradiation. Experiments were conducted on a Kratos Axima TOF^2 (Kratos-Shimadzu Biotech) equipped with an N_2 laser (337nm, 4ns pulse width). An accelerating voltage of 20kV was used, spectra were recorded in linear mode by averaging at least 100 laser shots at a power of 120 system units.

Peptide synthesis

[00127] Suckerins and variants thereof can be assembled by peptide synthesis either manually or for example, with the aid of commonly available peptide synthesizers.

[00128] Naturally occurring Suckerin proteins exhibit some imperfectly repetitive structure. In some cases imperfections are likely a consequence of the process by which the Suckerin protein genes evolved rather than a requirement for the local and supra-molecular organization of the SRT. Therefore local imperfections will not dramatically effect properties, nucleic acid sequences are engineered which encode clean Suckerin proteins, each of which encodes a polypeptide having direct repeats of naturally occurring and/or engineered variants of [L], [M1],[M2] and [M3] and/or sequences bearing at least 60% similarity to these. Alternatively, nucleic acid sequences may include sequences from different Suckerins, [L], [M1], [M2] and [M3] peptides to form a "copolymer" Suckerin protein.

Formation of Suckerin-based material

[00129] Suckerin proteins exhibit solubility in mild acetic conditions and can also be heated and deformed to process them into any complex shape. Suckerin-based solutions, gels, fibers, melts, liquid-crystals and any other chosen state of matter are processed further by manipulation by methods including but not limited to micro-environment, addition of water,

salts, elements of the periodic table, solvent, heating, cooling, magnetism, mechanical stress, mechanical strain, pressure, compression, torsion and/or shear for example.

[00130] Suckerin proteins, peptides, copolymers and variants and combinations thereof are folded into structures containing beta-sheet secondary structure.

[00131] Suckerin-based solutions, gels, fibers, melts, liquid-crystals and any other chosen state of matter are processed further by manipulation by methods including but not limited to microenvironment, addition or removal of water, solvent, addition of elements from the period table, mechanical stress, magnetism, mechanical strain, pressure, compression, torsion and/or shear.

[00132] We envisage a broad range of applications for SRT-based materials, which could rival natural and engineered silks in fields such as photonics and sensing devices, or as tissue scaffolds with tailored properties. They could also be exploited as biocompatible films for food and drug packaging, and as cost-effective encapsulants that shield drugs against thermal degradation during transportation and storage. Their tandem repeat design with high Tyr content is also deemed ideal to direct gold nanoparticles growth[61] through the reducing activity of Tyr residues, which are useful in sensing and diagnostic applications.

Suckerin powder

[00133] SRT were pulverized into a powder in liquid nitrogen using a mortar and pestle. A dry powder containing the protein resulted. The powder can be used for further processing to make any of the products listed herein.

Fabrication, visualization, and characterization of SRT

[00134] Hydrated and/or Heated SRT powder can be formed into many different patterns

Spun fiber material

[00135] As another illustration of their processability, melted SRT-proteins were readily spun into fibers. After drawing the liquid SRT proteins into fibers, we observed slight shifts and changes in relative FTIR peak intensities, indicating in this case significant re-arrangement of SRT protein secondary structure. However, a clear β-sheet signature was still observed for the fibers, supporting the view that they maintain a silk-like supramolecular organization. Silks have proven to be suitable for an extreme range of both textile and high-technology applications[2,3]. However, a large majority of these materials require harsh processing conditions to solubilize and process the fibroin or spidroin proteins. Our data clearly demonstrate that these silk-like materials can be sculpted and spun by facile aqueous-based processing.

Heat-drawn spinning of SRT proteins into fibers

[00136] Natural, recombinant and synthetic Suckerin-based materials, including fibers are of utility in which other natural fibers are considered and used. For example Suckerin-

based materials and/or fibers can be mixed with various plastics, resins and any other compound to prepare reinforced plastic, resin and/or composite products. Given that the natural and synthetic Suckerins can be effectively melted, they can be processed together or in parallel with thermal plastics and other materials at temperatures up to the degradation temperature of the peptide backbone ~280 degrees.

Fabrication of recombinant (rec)Suckerin-39 films and nano-surfaces

[00137] (Rec)Suckerin-39 was found to be soluble in 5% acetic acid at 20mg/ml. These conditions were considered suitable as they represented mild, eco-friendly aqueous-based conditions for the assembly of solid-state Suckerin based films from solution. Films were generated by the addition of 10 μl of the protein solution into custom 3D printed mini-wells (2mm diameter, 2mm height) secured to a glass microscope slide base. The samples were allowed to dry for 12 hours resulting in ~5 µm thick films. A second layer was then introduced on top of the first layer in the same manner to create a film that was thick enough to eliminate surface effects in subsequent nano-indentation experiments. Tested films were approximately ~7-10 µm thick. The test chamber humidity was ~ 55% RH.

Photo-cross linking of (rec)Suckerin-39 proteins

[00138] (Rec)Suckerin-39 protein cross-linking conditions were first established using 1mg/ml preparations of rec-Suckerin-39 in 5% acetic acid with 0.2 mM $Ru(bpy)_3Cl_2$ (Sigma Aldrich) and a range of Ammonium persulfate (APS) concentrations (2.5 mM to 0 mM) in 20 μl reaction volumes. Following the introduction of the reactants, the samples were immediately exposed to white light from a 500W halogen lamp for 2 minutes. Samples were then mixed with Laemmli loading buffer, boiled for 10 minutes and subjected to SDS-PAGE. Heavily cross-linked samples (2.5 and 1.0 mM APS samples) displayed little to no mobility on SDS-PAGE gels compared to samples tested with lower concentrations of APS (Fig. 6Ba). By titrating APS concentration we identified conditions where the proteins were cross-linked into lower molecular weight oligomers (Fig. 6Ba). We used MALDI-TOF to confirm the presence of these cross-linking intermediates where dimers, trimers and tetramers were observed using 0.6 mM APS, 0.2 mM $Ru(bpy)_3Cl_2$ and 2 minutes light exposure (Fig. 8b). At higher APS concentrations (rec)Suckerin-39 is not observed by MALDI supporting the view that the protein is cross-linked into very high molecular weight assemblies. We also performed parallel experiments to evaluate the effect of light exposure time on cross-linking using 1mg/ml protein in 5% acetic acid with 2.5 mM APS and 0.2 mM $Ru(bpy)_3Cl_2$. Here, intermediates are observed after a light exposure duration of 5 seconds. Beyond exposure durations greater than 10 seconds, SDS PAGE and MALDI indicate that (rec) Suckerin-39 is cross-linked into very high molecular weight oligomers (Figs. 6Bb-c). Together the data support the view that the Di-Tyrosine cross-linking method is highly efficient for our Tyrosine

rich proteins and suggests that reaction conditions can be used to tune the cross-link density and mechanical properties of Suckerin based materials.

Photo-cross-linking of rec-Suckerin films

[00139] Cross-linked (rec)Suckerin-39 films were generated by the following method. 20 mg/ml (rec)Suckerin-39 in 5% acetic acid with 2.5 mM APS and 0.2 mM $Ru(bpy)_3Cl_2$ in 20 µl reaction volumes were used. 10 µl of the sample was introduced into the custom 3D printed test platform described above and allowed to dry in the dark for 2 hours. Partially dried films were then exposed to light from a 500 W halogen lamp for 3 minutes and allowed to dry for another 2 hours, followed by another round of 3 minutes exposure to light. A second layer was then introduced in the same manner to ensure that the film was thick enough to eliminate potential substrate effects during nano-indentation. The final processed films were approximately 5-10 µm thick.

Nanoindentation of native and recombinant Suckerin-based materials

[00140] Ultra-microtomed cross sections of Sucker Ring Teeth (SRT) were probed by Nanoindentation using a TriboScan 950 (Hysitron), following previously described procedures. The engineered films were probed in various locations using a cube corner fluid cell tip at maximum loads of 50 μN, chosen such as to eliminate stiffness substrate effects. A total of 50 indents were carried out at a loading rate of 5 μN/sec using 5 s holding time before unloading. Water droplets were then introduced within the mini-well and the film was hydrated for 30 minutes prior to indentation. The films were tested under fully hydrated conditions in three separate regions of the sample at maximum loads of 10 μN, for a total of 50 indents. Between each set of indents in one region, water was re-introduced within the mini-well in order ensure fully hydrated conditions. Identical experimental set-up was employed for cross-linked films. Because of the stiffer nature of the films even in hydrated conditions, maximum loads of 50 μN were kept identical for both dry and the wet measurements. All indentation curves were analyzed using the classic Oliver-Pharr analysis.

[00141] Suckerin-based materials engineered in any shape and/or physical state, as described above can be designed to provide an extreme range of desired functions by way of modification of the materials themselves and/or the addition/incorporation of additional functional atoms, molecules and/or materials. Modifications may include 1) amino acid side chain modifications; 2) the addition of metals, peptides, proteins and/or any pharmacological agent including the addition (via embedding, adsorption, cross-linking etc) of low molecular weight, biological and/or synthetic compounds, RNA, DNA, molecules, carbohydrates inorganic compounds etc.; 3) the addition of cells (at any stage of differentiation, including stem cells, predifferentitated stem cells and/or induced pleuripotent stem cells), and/or combinations of cell types; Below, we provide several examples of the potential design and

utility of functionalized Suckerin-based materials.

Addition/binding/inclusion of elements of the periodic table:

[00142] The combination of elements of the periodic table with Suckerin-based materials offers enormous potential in a wide range of industrial and biomedical applications. For example, the incorporation of copper into other biological proteins is already known to result in the creation of light weight, wear resistant materials, such as those observed in blood worm's teeth[51]. The incorporation of metals may also enable electron transport and/or conductivity and/or influence oxidation/reduction pathways in the aqueous milieu are also considered. The Suckerin-based materials may also be combined with radio-active materials that may find use in for the targeting, for example, of wide range of cancer cells. In all cases the physico-chemical properties of the Suckerin-based material can be use to enable binding/release of the elements and to provide mechanical stability for the bound elements in a wide range of industrial and/or biomedical niches.

Suckerin-Based materials for controlled drug release:

[00143] The primary sequence design, secondary, tertiary, quaternary structure and/or nano, meso, micro and macro scale structures imparted from the physical and/ or chemical processing of these Suckerin-based materials, in the novel embodiments described below can be used to provide functional control over the incorporation and/or physical adsorption of pharmacologically active entities (hereafter referred to as "agents") and the subsequent specio-temporal control of their release. Agents can be coated, attached and/or incorporated on or into materials for drug delivery, alone or in combination include but are not limited to the following: 1) Growth factors; 2) Single and/or double stranded DNA and/or oligonucleotides; 3) RNA; 4) Sugars; 5) Lipids; 6) Proteolipids; 7) any class of prescription or over the counter drug; 8) Peptides; 9) Cytokines; 10) Chemokines; 11) Glycoproteins; 12) Organic and/or inorganic atoms (nano, meso, micro and macro scale metallic materials; 13) Ions; 14) Salts, 15) Metals; 16) Radioactive Isotopes; 17) Antibodies; 18) Viruses; 19) Vulnerary agents; 20) Hemostatic agents; 21) Antibiotics; 22) Antithelmintics; 23) anti-fungal agents; 24) hormones; 25) anti-inflammatory agents; 26) ATP; 27) enzymes; 28) protein domain module(s) and combinations thereof; 29) Neurotransmitter(s); 30) ion channel and porin related proteins and molecules.

[00144] It is well documented that beta-structure can dictate the diffusion and release kinetics of an extreme range of agents from beta-sheet containing silk proteins. These materials are being designed to test the release kinetics and biological activity of model compounds/ molecules with diverse functions, with the eventual goal of using a chosen agent (by way of illustration a drug or protein etc.), or combination of agents to effect cellular homeostasis, cell signaling, differentiation and/or the physiological status of targeted cells or

tissues. The fact that Suckerin beta-sheet content can be tailored indicates that these materials will find utility in sustained and controlled drug delivery. These materials will provide not only tailored release kinetics but they can also be engineered to maintain specific mechanical properties and offer biocompatibility.

[00145] Suckerin-based materials may be any size scale, thickness, physical state and morphology. These may included gels, electrogels, liquid crystals, hydrogels, thin films, films, nano-fibers, fibrils, fibers, (including tissue scaffolds), nano-spheres and other nano-scale shapes, micro-spheres and all macroscopic materials. The agent loaded materials may be used alone, or in combination with any other agent(s) without limitation.

Suckerin based materials as sensors/switches and/or actuators.

[00146] The field of nano-technology offers huge potential for the engineering of molecular scale switches, sensors and actuators etc. that could be used to assess local micro-environments, measure cellular status and perform mechanical functions and/or electrical functions at the nano-scale. However, the materials currently in use, namely carbon-based nano-tubes, nano-crystalline materials and synthetic materials have their own limitations in terms of engineering flexibility and biocompatibility. Given that Suckerins can be sculpted into a wide range of materials, they clearly have enormous potential as substrates and structures that will form the basis for the addition of switching, sensing functions for example. By way of example, the beta-sheet content of material can be monitored, (using antibodies, congo-red binding, thioflavin-T binding, raman spectroscopy and/or any other technique) which can provide direct read-outs of the stability of the material, it's bioresorption rates and, by mathematical inference with polymer network theory, it's mechanical status. In another embodiment the Suckerin material is modified to include sensing structures, compounds, chemicals and materials.

Suckerin based materials with cells and tissues:

[00147] When processed into the appropriate physical states and structures many synthetic materials offer the potential to act as scaffolds/ adhesion substrates for cell and tissue growth and in vivo tissue repair. While these materials offer significant potential they exhibit limitations in engineering flexibility, biocompatibility and biodegradability that proteins may not. Therefore there has been a recent surge in interest of creating engineered biomimetic scaffold that offer greater control of these key aspects. For example there is an extensive body of literature on the use of collagens and fibronectin in tissue repair. Each of these materials has advantages and disadvantages, depending on the cell/tissues used and the repair strategy being used. Suckerin-based materials and mimetics thereof represent a novel class of potential cellular substrates/ tissue scaffolds that could offer engineering flexibility, biocompatibility, direct control over the biomechanical properties of the substrate

and tailored biodegradability to provide substrates that are 1) robust 2) have precisely tailored physical properties 3) provide finely tuned surfaces that influence cell cellular homeostasis and cell signaling 4) have tailored degradation kinetics. Examples of cell types that can be combined with Suckerin-based materials include but are not limited to 1) Neurons 2) Glia 3) any class of stem cells and stem cell lineages 4) muscle cells 5) bone cells 6) pancreas 7) hematopoetic cells 8) liver cells 9) cardiac myocytes 10) cartilaginous cells; 11) bone 12) tendon 13) cells derived from or comprising artery or venous tissue 14) skin cells 15) reproductive cells 16) cells conferring immunity 17) genetically engineered cells 18) plant cells 19) bacterial cells 20) yeast cells 21) algal cells 22) oligodendrocytes etc.

Evaluation of cell culture compatibility of (rec)Suckerin-39 films

[00148] As an additional proof-of-concept of the versatility of (rec)Suckerin-39, we examined the viability and proliferation of human mesenchymal stem cells (hMSC) grown on (rec)Suckerin-39 films.

[00149] Human mesenchymal stem cells (hMSCs, Lonza PT-2501) were cultured in Minimum Essential Medium (MEM)- (Invitrogen 12571-063) with 10% fetal bovine serum (FBS) and 100 U/mL penicillin and 100 µg/mL streptomycin at 37°C in a 5% CO_2 humidified atmosphere. Lyophilized (rec)Suckerin39 was dissolved in 50% acetic acid to 10 mg/ml and diluted where needed in 50% acetic acid. 50 µl of each protein solution was deposited onto a 14mm glass coverslip and dried under air flow in a fume hood. The protein film was then crosslinked with a 1% glutaraldehyde solution, rinsed thrice with distilled water and air dried. The film-covered cover slip was then placed in a 24-well tissue culture plate and UV sterilized. 3.75×10^4 hMSCs per well were seeded on the prepared (rec)Suckerin-39 films and cultured for 16 days with frequent media change. For live/dead staining, the cells were treated with 2 µM fluorescein diacetate and 2.5 µg/mL propidium iodide for 0.5 h at 37°C. Thereafter, the hMSCs were imaged using a Nikon fluorescence microscope (TE2000-U).

[00150] The hMSCs were cultured for more than 2 weeks and displayed high cell viability for all conditions demonstrating that (rec)Suckerin-39 is biocompatible and can support cell growth (Fig. 8f). Notably, we observed that hMSCs grown on films generated from higher concentrations of (rec)Suckerin-39 displayed increasingly branched and spindle-shaped cell morphologies, which may be related to cell-lineage specification of hMSCs as a result of variations in matrix elasticity at different protein concentrations.

Uses of Suckerin based materials

[00151] Mechanically and functionally useful nano-fibers; meso and macro-scale fibers, films, thin films nano- meso and micro and macro-scale fibrous mats, materials with tailored porosity, liquid-crystals, gels, nano-spheres, microspheres, foams, hydrogels, electrogels. Fibers, woven fabrics, high-tech clothing, rope, sails, fishing line parachute, wings on arial

devices, flexible tie downs for electrical components, sutures biomedical implants, tissue scaffold. Biomedical applications include but are not limited to, tissue repair scaffold, tissue engineering scaffold, reconstructive eye surgery ear surgery implant. Sutures used in surgical procedures including eye surgery, reconstructive nerve surgery, tympanic membrane surgery, vascular closure, bowel surgery, cosmetic surgery, and central nervous system surgery. Natural and synthetic Suckerin-based fibers may also be of utility in the generation of antibiotic impregnated sutures and implant material and matrix material for bone, connective tissue and soft tissue. Implants and matrix material for reconstruction may be impregnated with growth factors, differentiation factors, and/or cell attractants to facilitate incorporation of exogenous material and optimize repair. Sensors, switches actuators insulators and above mentioned devices in this category. Drug delivery devices including nano-particle, microspheres, membranes gels foams Cosmetic components and fillers, composite materials. The materials of the current invention may also be incorporated into bandages, surgical and dental wound packing material, diapers and catamenial devices, and the like

[00152] By "comprising" it is meant including, but not limited to, whatever follows the word "comprising". Thus, use of the term "comprising" indicates that the listed elements are required or mandatory, but that other elements are optional and may or may not be present.

[00153] By "consisting of" is meant including, and limited to, whatever follows the phrase "consisting of". Thus, the phrase "consisting of" indicates that the listed elements are required or mandatory, and that no other elements may be present.

[00154] The inventions illustratively described herein may suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for example, the terms "comprising", "including", "containing", etc. shall be read expansively and without limitation. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the inventions embodied therein herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention.

[00155] By "about" in relation to a given numerical value, such as for temperature and period of time, it is meant to include numerical values within 10% of the specified value.

[00156] The invention has been described broadly and generically herein. Each of the

narrower species and sub-generic groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

[00157] Other embodiments are within the following claims and non- limiting examples. In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group.

References

1 Lee, P. L., Messersmith, P. B., Israelachvili, J. N. & Waite, J. H. Mussel-Inspired Adhesives and Coatings. *Annual Reviews of Materials Research* **41**, 99-132 (2011).

2 Omenetto, F. G. & Kaplan, D. New Opportunities for an Ancient Material. *Science* **329**, 528-531 (2010).

3 Tao, H., Kaplan, D. L. & Omenetto, F. G. Silk materials--a road to sustainable high technology. *Adv Mater* **24**, 2824-2837, doi:10.1002/adma.201104477 (2012).

4 Guerette, P., Gizinger, D., Weber, B. & Gosline, J. M. Silk Properties Determined by Gland-Specific Expression of a Spider Fibroin Gene Family. *Science* **272**, 112-115 (1996).

5 Gatesy, J., Hayashi, C. Y., Motriuk, D., Woods, J. & Lewis, R. Extreme Diversity, Conservation, and Convergence of Spider Silk Fibroin Sequences. *Science* **291**, 2603-2605 (2001).

6 Hayashi, C. Y. & Lewis, R. V. Molecular architecture and evolution of a modular spider silk protein gene. *Science* **287**, 1477-1479 (2000).

7 Hayashi, C. Y. & Lewis, R. V. Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks. *J Mol Biol* **275**, 773-784, doi:10.1006/jmbi.1997.1478 (1998).

8 Ayoub, N. A., Garb, J. E., Tinghitella, R. M., Collin, M. A. & Hayashi, C. Y. Blueprint for a high-performance biomaterial: full-length spider dragline silk genes. *PLoS One* **2**, e514, doi:10.1371/journal.pone.0000514 (2007).

9 Gosline, J. M., Guerette, P. A., Ortlepp, C. S. & Savage, K. N. The Mechanical Design of Spider Silks: From Fibroins Sequence to Mechanical Function. *Journal of Experimental Biology* **202**, 3295– 3303 (1999).

10 Holland, C., Vollrath, F., Ryan, A. J. & Mykhaylyk, O. O. Silk and synthetic polymers: reconciling 100 degrees of separation. *Adv Mater* **24**, 105-109, 104, doi:10.1002/adma.201103664 (2012).

11 Vollrath, F. & Knight, D. P. Liquid Crystalline Spinning of Spider Silk. *Nature* **410**, 541-548 (2001).

12 Lewis, R. V. & Hayashi, C. Y. Extremely elastic spider silk prtein and DNA coding thereof. United States patent 5994099 (1999).

13 Lewis, R. V., Hayashi, C. Y., Gatesy, J. E. & Motriuk, D. Spider silk protein encoding nucleic acids, polypeptides, andtibodies and methods of use thereof. US patent 7,521,228 (2009).

14 Scheibel, T. & Huemmerich, D. Proteins of natural origin and materials made therefrom. 8097583 (2012).

15 Scheibel, T., Huemmerich, D. & Ackerschott, C. Recombinant spider silk proteins. United States patent 8034897 (2011).

16 Scheibel, T., Huemmerich, D. & C., A. Recombinant spider silk proteins. United States patent 7951908 (2011).

17 Fahnestock, S. R. Recombinantly produced spider silk. United States patent 6268169 (2001).

18 Johansson, J., Hjalm, G., Stark, M., Engstrom, W. & Hedhammar, M. Spider silk proteins and methods for producing silk proteins. United States patent 8173772 (2012).

19 Kaplan, D. L., Jin, H. J., Rutledge, G. & Fridrikh, S. Silk biomaterials and methods of use thereof. United States patent 7674882 (2010).

20 Kaplan, D. L., Kim, U.-J., Park, J. H. & Jin, H. J. Concentrated aqueous silk fibroin solution and use thereof. (2009).

21 Lock, R. L. Process for making silk fibroin fibers. United States patent 5252285 (1993).

22 Mello, C., Arcidiacono, S. & Butler, M. M. Methods for the purification and aqueous fiber spinning of spider silks and other structural proteins. United States patent 7335739 (2008).

23 Islam, S. *et al.* Methods and apparatus for spinning spider silk protein. United Stated patent 7057023 (2006).

24 Lazaris, A. *et al.* Spider silk fibers spun from soluble recombinant silk produced in mammalian cells. *Science* **295**, 472-476, doi:10.1126/science.1065780 (2002).

25 Fahnestock, S. R., Yao, Z. & Bedzyk, L. A. Microbial production of spider silk proteins. *J Biotechnol* **74**, 105-119 (2000).

26 Scheller, J., Guhrs, K. H., Grosse, F. & Conrad, U. Production of spider silk proteins in tobacco and potato. *Nat Biotechnol* **19**, 573-577, doi:10.1038/89335 (2001).

27 Yang, J. Production of silk-like proteins in plants. United States patent 6965060

(2005).

28 Yang, J. Production of silk-like proteins in plants. United States patent 6608242 (2003).

29 Yang, J., Barr, L. A., Fahnestock, S. R. & Liu, Z. B. High yield recombinant silk-like protein production in transgenic plants through protein targeting. *Transgenic Res* **14**, 313-324 (2005).

30 Menassa, R. *et al.* Spider dragline silk proteins in transgenic tobacco leaves: accumulation and field production. *Plant Biotechnol J* **2**, 431-438, doi:10.1111/j.1467-7652.2004.00087.x (2004).

31 Yamao, M. *et al.* Gene targeting in the silkworm by use of a baculovirus. *Genes Dev* **13**, 511-516 (1999).

32 Teule, F. *et al.* Silkworms transformed with chimeric silkworm/spider silk genes spin composite silk fibers with improved mechanical properties. *Proc Natl Acad Sci U S A* **109**, 923-928, doi:10.1073/pnas.1109420109 (2012).

33 Karatzas, C. N., Turner, J. D. & Lazaris-Karatzas, A. Production of biofilament in transgenic animals. United States patent 7157615 (2007).

34 Xia, X. X. *et al.* Native-sized recombinant spider silk protein produced in metabolically engineered Escherichia coli results in a strong fiber. *Proc Natl Acad Sci U S A* **107**, 14059-14063, doi:10.1073/pnas.1003366107 (2010).

35 Kerkam, K., Viney, C., Kaplan, D. & Lombardi, S. Liquid crystallinity of natural silk secretions. *Nature* **349**, 596-598 (1991).

36 Willcox, J. P., Gido, S. P., Muller, W. & Kaplan, D. Evidence of cholesteric liquid crystalline phase in natural silk spinning process. *Macromolecules* **29**, 5106-5110 (1996).

37 Knight, D. & Vollrath, F. Hexagonal columnar liquid crystal in the cells secreting spider silk. *Tissue Cell* **31**, 617-620, doi:10.1054/tice.1999.0076 (1999).

38 Vollrath, F., Madsen, B. & Shao, Z. The effect of spinning conditions on the mechanics of a spider's dragline silk. *Proc Biol Sci* **268**, 2339-2346, doi:10.1098/rspb.2001.1590 (2001).

39 Vollrath, F. & Knight, D. P. Liquid crystalline spinning of spider silk. *Nature* **410**, 541-548, doi:10.1038/35069000 (2001).

40 Hagn, F. *et al.* A conserved spider silk domain acts as a molecular switch that controls fibre assembly. *Nature* **465**, 239-242, doi:10.1038/nature08936 (2010).

41 Hagn, F., Thamm, C., Scheibel, T. & Kessler, H. pH-dependent dimerization and salt-dependent stabilization of the N-terminal domain of spider dragline silk--implications for fiber formation. *Angew Chem Int Ed Engl* **50**, 310-313,

doi:10.1002/anie.201003795 (2011).

42 Rammensee, S., Slotta, S., Scheibel, T. & Bausch, A. R. Assembly Mechanism of Recombinant Spider Silk Proteins. *Proc Natl Acad Sci U S A* **105**, 6590-6595 (2008).

43 Askarieh, G. *et al.* Self-assembly of spider silk proteins is controlled by a pH-sensitive relay. *Nature* **465**, 236-238, doi:10.1038/nature08962 (2010).

44 Guan, J., Vollrath, F. & Porter, D. Two mechanisms for supercontraction in Nephila spider dragline silk. *Biomacromolecules* **12**, 4030-4035, doi:10.1021/bm201032v (2011).

45 van Beek, J. D., Kummerlen, J., Vollrath, F. & Meier, B. H. Supercontracted spider dragline silk: a solid-state NMR study of the local structure. *Int J Biol Macromol* **24**, 173-178 (1999).

46 Miserez, A., Weaver, J. C., Kisailus, D. & Birkedal, H. in *MRS Spring meeting 2009.* (Materials Resaerch Society).

47 Miserez, A. *et al.* Microstructural and Biochemical Characterization of the Nano-porous Sucker Rings from Dosidicus gigas. *Adv Mater* **21**, 401-406 (2009).

48 Strugnell, J., Jackson, J., Drummond, A. J. & Cooper, A. Divergence time estimates for major cephalopod groups: evidence from multiple genes. *Cladistics* **22**, 89-96 (2006).

49 Lindgren, A. R., Pankey, M. S., Hochberg, F. G. & Oakley, T. H. A multi-gene phylogeny of Cephalopoda supports convergent morphological evolution in association with multiple habitat shifts in the marine environment. *BMC Evol Biol* **12**, 129, doi:10.1186/1471-2148-12-129 (2012).

50 Keten, S., Xu, Z., Ihle, B. & Buehler, M. J. Nanoconfinement Controls Stiffness, Strength and Mechanical Toughness of Beta-Sheet Crystals in Silk. *Nat Mater* **vol. 9**, 359-367 (2010).

51 Lichtenegger, H. C., Schöberl, T., Bartl, M. H., Waite, H. & Stucky, G. D. High Abrasion Resistance with Sparse Mineralization : Copper Biomineral in Worm Jaws. *Science* **298**, 389-392 (2002).

52 Grabherr, M. G. *et al.* Full-Length Transcriptome Assembly from RNA-Seq Data Without a Reference Genome. *Nat Biotechnol* **29**, 644-652 (2011).

53 Li, B. & Dewey, C. N. RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome. *BMC Bioinformatics* **12**, 323, doi:10.1186/1471-2105-12-323 (2011).

54 Edgar, R. C. Search and Custering Orders of Magnitude Faster than BLAST. *Bioinformatics (Oxford, England)* **26**, 2460-2461, doi:10.1093/bioinformatics/btq461 (2010).

55 Yano, M., Nagai, K., Morimoto, K. & Miyamoto, K. Shematrin: A Family of Glycine-Rich Structural Proteins in the Shell of the Pearl Oyster Pinctada fucata. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **144**, 254 – 262 (2006).

56. Rousseau, M.-E., Lefevre, T., Beaulieu, L., Asakura, T. & Pezolet, M. Study of Protein Conformation and Orientation in Silkworm and Spider Silk Fibers Using Raman Microspectroscopy. *Biomacromolecules* **5**, 2247-2257 (2004).

57. Fancy, D.A. & Kodadek, T. Chemistry for the Analysis of Protein-Protein Interactions: Rapid and Efficient Cross-Linking Triggered by Long Wavelength Light. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 6020-6024 (1999).

58. C. Z. Zhou *et al.*, Fine organization of Bombyx mori fibroin heavy chain gene. *Nucleic Acids Res* **28**, 2413 (Jun 15, 2000).

59. M. Suzuki *et al.*, Characterization of Prismalin-14, a novel matrix protein from the prismatic layer of the Japanese pearl oyster (Pinctada fucata). *Biochem J* **382**, 205 (Aug 15, 2004).

60. L. Dalla Valle *et al.*, Beta-keratins of the crocodilian epidermis: composition, structure, and phylogenetic relationships. *J Exp Zool B Mol Dev Evol* **312**, 42 (Jan 15, 2009).

61. P. A. Guerette et al., (2013) Accelerating the design of biomimetic materials by integrating RNA-seq with proteomics and materials science. Nat Biotechnol, 31:908-915

Claims:

1. An isolated polypeptide comprising or consisting of:
    a. the amino acid sequence set forth in any one of SEQ ID Nos. 2, 21, 23, 24 or 25; or
    b. a variant of the amino acid sequence of (a) that shares at least 40 %, , preferably at least 50 %, more preferably at least 60, 80 or 90 % sequence identity with any one of the amino acid sequences of (a) over its entire length
    c. a fragment of the amino acid sequence of (a) or (b), the fragment having a length of at least 50, preferably at least 100, more preferably at least 150, most preferably at least 200, 400, or 600 amino acids;

    wherein the polypeptide is at least partially in beta sheet conformation.

2. The polypeptide of claim 1, wherein the polypeptide is soluble in water and/or under mildly acidic conditions.

3. The polypeptide of claim 1 or 2, wherein the polypeptide comprises at least one domain 1 comprising or consisting of an amino acid sequence selected from the group consisting of HH, TT, SS, AV, TTH, THH, IAAL, SY, VTHHAP, VVLLAAF, HTTHHA, AATVSHTTHHA, FPY, THT, HVT, HHP, VSH and TVS.

4. The polypeptide of any one of claim 1-3, wherein the polypeptide comprises at least one glycine rich domain 2.

5. The polypeptide of claim 4, wherein the glycine rich domain 2 comprises an amino acid sequence selected from the group consisting of GGLYG, GGLYG, GGYG, GLYGG, YGIG, GIG and GYG.

6. The polypeptide of claim 4 or 5, wherein the at least one glycine rich domain 2 comprises at least one sub-domain 3 comprising the amino acid sequence $GGX_1X_2X_3Y$ (SEQ ID NO:1)
    wherein X_1 is nothing, L, F or V;
    wherein X_2 is nothing, G, or F; and
    wherein X_3 is nothing, G or A;

7. The polypeptide of claim 6, wherein X_1 of sub-domain 3 is Leucine.

8. The polypeptide of claim 7, wherein sub-domain 3 has the amino acid sequence GGLY.

9. The polypeptide of any one of claims 4-8, wherein the at least one glycine rich domain 2 further comprises at least one sub-domain 4 comprising the amino acid sequence GGY.

10. The polypeptide of any one of claim 1-5, wherein the polypeptide comprises at least one amino acid consensus sequence selected from the sequences set forth in SEQ ID Nos. 105, 106, 107, 108, 109 and 110.

11. The polypeptide of any one of claims 1-6, wherein the polypeptide comprises
    a. the amino acid sequence set forth in any one of SEQ ID Nos. 2 to 53; or
    b. a variant of the amino acid sequence of (a) that shares at least 40 %, preferably at least 50 %, more preferably at least 60 % sequence identity with the amino acid sequence of the amino acid sequence set forth in any one of SEQ ID Nos. 2 to 53 over its entire length.

12. The polypeptide of any one of claims 4-11 comprising a peptide unit from N- to C-terminus: Proline-Domain 1-Domain 2.

13. The polypeptide of any one of claims 4-11 comprising a peptide unit from N to C terminus, Proline-Domain 1-Proline-Domain 2.

14. The polypeptide of claim 12 or 13, wherein domain 2 comprises from N to C terminus, (sub-Domain 3-sub-domain 4)$_n$; wherein n is independently 1 to 4.

15. The polypeptide of claim 12 or 13 wherein domain 2 comprises from N to C terminus, (sub-Domain 3-sub-Domain 4-sub-Domain 3)$_n$; wherein n is independently 1 to 3 .

16. The polypeptide of any one of claims 12-15 wherein the peptide unit is repeated from 2 to 13 times.

17. The polypeptide of any one of claims 1 to 16 isolated and purified from a species in the Decapodiformes super order.

18. An isolated nucleic acid molecule comprising a nucleic acid sequence encoding the polypeptide of any one of claims 1 to 17.

19. The isolated nucleic acid molecule of claim 18, wherein the nucleic acid sequence comprises any one of the sequences set forth in SEQ ID Nos. 54 to 100.

20. The isolated nucleic acid molecule of claim 18 or 19, wherein the nucleic acid sequence is comprised in an expression construct preferably a vector.

21. A host cell comprising the expression construct of claim 20, wherein the cell is capable of expressing the polypeptide of any one of claims 1-17.

22. A composition or material comprising the polypeptide of any one of claims 1-17.

23. The material of claim 22, wherein the material is selected from the group: a fibre, a filament, a film, a foam, a nano fibre, a nano sphere, a nano particle, a liquid crystal mesogen, a tissue scaffold, a colloid, a copolymer, or a block copolymer.

24. A method of making the material of claim 22 or 23 comprising the steps of:
crushing the polypeptide of any one of claims 1-17;
heating the crushed polypeptide into a liquid; and
forming the material from the liquid.

25. The method of claim 24, wherein forming the material comprises spinning the liquid into a fibre.

26. The method of claim 24, wherein forming the material comprises placing the liquid into a mold; and removing the mold after the liquid solidifies.

27. A tissue scaffold comprising the polypeptide of any one of claims 1-17.

28. A multimer of the polypeptides of any one of claims 1-17.