



## OPEN

## A new class of animal collagen masquerading as an insect silk

SUBJECT AREAS:  
BIOINSPIRED MATERIALS  
BIOMIMETICS  
BIOMATERIALS-PROTEINS  
PROTEIN FOLDINGTara D. Sutherland<sup>1</sup>, Yong Y. Peng<sup>2</sup>, Holly E. Trueman<sup>1</sup>, Sarah Weisman<sup>1</sup>, Shoko Okada<sup>1</sup>, Andrew A. Walker<sup>1,3</sup>, Alagacone Sriskantha<sup>1</sup>, Jacinta F. White<sup>2</sup>, Mickey G. Huson<sup>4</sup>, Jerome A. Werkmeister<sup>2</sup>, Veronica Glattauer<sup>2</sup>, Violet Stoichevska<sup>2</sup>, Stephen T. Mudie<sup>5</sup>, Victoria S. Haritos<sup>1</sup> & John A. M. Ramshaw<sup>2</sup><sup>1</sup>CSIRO Ecosystem Sciences, Clunies Ross Street, Acton, ACT, 2601, Australia, <sup>2</sup>CSIRO Materials Science and Engineering, Bayview Avenue, Clayton, VIC, 3169, Australia, <sup>3</sup>Research School of Biology, Australian National University, Canberra, ACT, 0200, Australia, <sup>4</sup>CSIRO Materials Science and Engineering, Waurin Ponds, Geelong, VIC, 3216, Australia, <sup>5</sup>Australian Synchrotron, Blackburn Road, Clayton, VIC, 3168, Australia.Received  
6 June 2013Accepted  
17 September 2013Published  
4 October 2013Correspondence and  
requests for materials  
should be addressed toT.D.S. (Tara.  
Sutherland@CSIRO.  
au)

Collagen is ubiquitous throughout the animal kingdom, where it comprises some 28 diverse molecules that form the extracellular matrix within organisms. In the 1960s, an extracorporeal animal collagen that forms the cocoon of a small group of hymenopteran insects was postulated. Here we categorically demonstrate that the larvae of a sawfly species produce silk from three small collagen proteins. The native proteins do not contain hydroxyproline, a post translational modification normally considered characteristic of animal collagens. The function of the proteins as silks explains their unusual collagen features. Recombinant proteins could be produced in standard bacterial expression systems and assembled into stable collagen molecules, opening the door to manufacture a new class of artificial collagen materials.

Collagens are structural proteins that are abundant and ubiquitous in the connective tissue of all metazoan animals. Animal collagens comprise some 28 diverse molecules that are characterised by a rod-like structure where polyproline II-like polypeptide chains are super-coiled about a common axis to form a triple-helix<sup>1</sup>. Since the close packing between the chains is such that only glycine (Gly) can fit in the centre of the triple-helix, collagen-forming proteins contain characteristic tripeptide repeats (Gly-Xaa-Yaa)<sub>n</sub>, where Xaa and Yaa can be any residue. The structure is stabilized by a high content (~20%) of the imino acids proline (Pro) and hydroxyproline (Hyp).

Whilst it is commonly regarded that collagens function within animals, the existence of an extracorporeal animal collagen was postulated in the 1960s when K. M. Rudall described the X-ray diffraction pattern of some hymenopteran insect cocoons as collagen-like<sup>2,3</sup>. Rudall noted that “during the brilliant summer of 1959 there were plagues of gooseberry sawfly (*Nematus ribesii*)” and this enabled cocoons or silk fibres drawn from the salivary gland to be obtained in sufficient quantity for study<sup>2</sup>. The diffraction patterns suggest twisted cables of collagen molecules with dimensions of 3 nm diameter. Amino acid analyses found high Gly (33.6%) and Pro (10.0%) content consistent with collagens, as well as high Ala (12.2%) characteristic of silks. Of most interest was that Hyp, which is seen as a characteristic of animal collagens was absent<sup>4</sup>.

Other described insect collagens are apparently homologous to molecules found in mammals. For example, a collagen from *Drosophila melanogaster*<sup>5</sup> is similar in structure to the type IV collagen found from hydra to human<sup>6</sup>. Type IV collagens contain a central domain of about 1200 residues containing the characteristic (Gly-Xaa-Yaa)<sub>n</sub> collagen repeat with high levels of Pro and Hyp, but with many interruptions (up to 25) from inserts of up to 30 amino acids. The collagen domain is flanked by large non-collagenous domains. Cysteine residues in the non-collagenous domains crosslink the proteins to generate the network structure of the extracellular matrix. Insect collagens with similarities in composition and length (~280–290 nm) to the predominant interstitial type I and type II animal collagens have been isolated, for example, from locust<sup>7</sup> and cockroach<sup>8</sup>. The mature interstitial collagen proteins contain just over 1000 residues, which primarily consist of uninterrupted Gly-Xaa-Yaa repeats containing around 12% Pro and 10% hydroxyproline<sup>7,8</sup>. Rather than networks, these proteins assemble into fibrils and fibres.

Here, we examine the cocoon silk of the willow sawfly, *Nematus oligospilus*, a hymenopteran species closely related to the sawfly species described by Rudall<sup>2,3</sup>. The native silk structure and composition were investigated using wide angle X-ray scattering and amino acid analysis. A combined proteomic-transcriptomic approach was used to identify the silk genes. The silk proteins were expressed in *Escherichia coli* and the recombinant molecules were demonstrated to be protease resistant and have circular dichroism spectra characteristic of collagen.



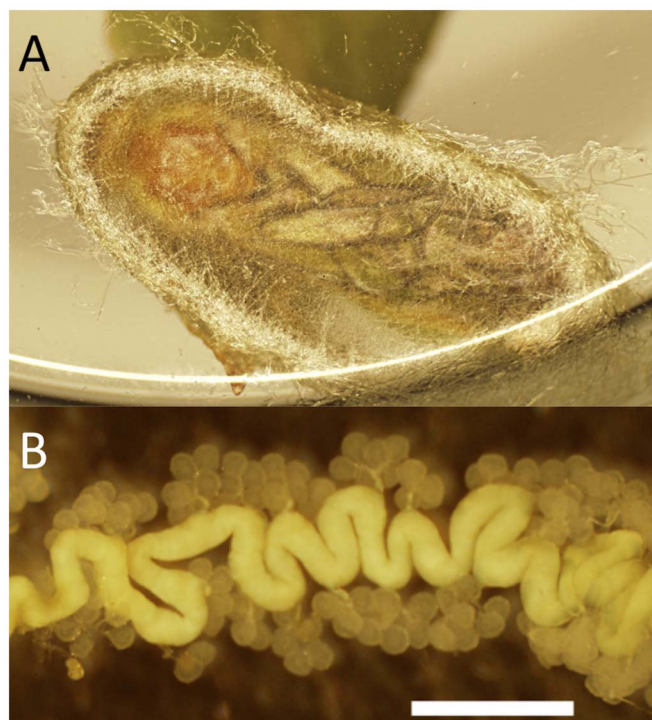
## Results

**X-ray diffraction patterns of silk.** Diffraction studies using wide-angle X-ray scattering (WAXS) on *N. oligospilus* cocoon silk showed a 0.286 nm meridional reflection peak (Fig. 1A) characteristic of the axial distance between residues along strands of the collagen triple helix<sup>9</sup>. This feature was similar to that observed by Rudall for *N. ribesii* silks<sup>2,3</sup>. Other observed peaks from the *N. oligospilus* cocoon silk were consistent with a structural repeat with 4.59 nm spacing in the axial direction (Fig. 1B) and therefore were attributed to various harmonics of long axial repeats of a collagen structure. Equatorial spots at 1.2 nm, attributed to lateral spacing between triple helices by Rudall, were not observed in the willow sawfly WAXS and peaks at 0.326 and 0.382 nm were not observed in the gooseberry sawfly (Figure 1), implying differences in structural arrangement of the collagen molecules within the two silks. Previously, Rudall had attributed a meridional arc observed at 0.465 nm to protein in a cross- $\beta$  structure, although no other characteristic peaks of cross- $\beta$  structure were detected.

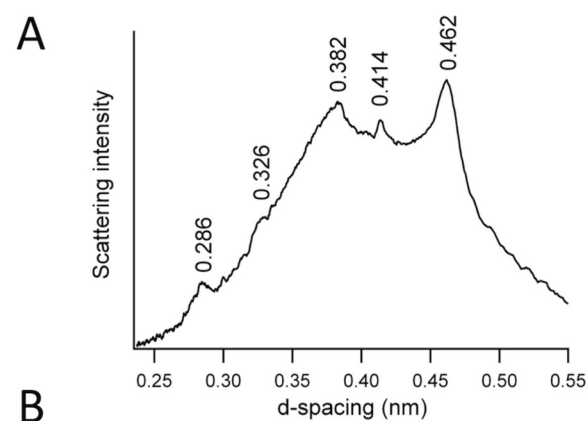
**Hydroxyproline (Hyp) analysis of cocoon silk.** Amino acid analysis of the willow sawfly silk did not detect any Hyp, consistent with Rudall's results from gooseberry sawfly cocoons in the 1960s<sup>2,3</sup>.

**Native silk production.** The sawfly silk proteins are produced at high concentrations in the labial gland, a dedicated silk gland. The silk gland comprises a convoluted tubular structure of about 12–15 mm in length when not stretched, onto which a series of nodular structures are attached (Fig. 2). Secretory cells within these nodules produce the silk proteins, which are then secreted into the gland lumen<sup>10</sup>. Haematoxylin and eosin staining of the gland shows the protein organized within the secretory cells into tactoids (Supplementary Fig. 2).

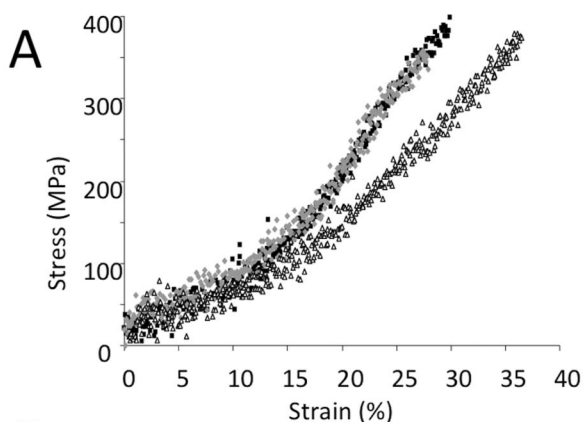
**Mechanical and chemical properties of native silk.** Mature silk fibres ( $2.2 \pm 0.4 \mu\text{m}$  diameter) were obtained directly from larvae that had commenced spinning their cocoon. The mechanical properties of the silk fibres (Fig. 3) exceeded that of mammalian



**Figure 2** | Images of silk and silk gland of sawflies. (A): Silken cocoon with encased pupa. (B): Silk gland showing numerous nodules comprising large secretory cells attached to the lumen through delivery ducts. Scale bar 1 mm.



**Figure 1** | Wide angle X-ray scattering (WAXS) data from *N. oligospilus* cocoons. (A): WAXS spectrum (average of five spots). The individual 1-D WAXS spectra are shown in Supplementary Figure 1. (B): Proposed assignments of WAXS peaks from the Willow sawfly and comparison with peaks observed in Gooseberry sawfly (from Rudall)<sup>2</sup>.



**Figure 3** | Mechanical properties of cocoon silk. (A): example stress-strain curves of three *N. oligospilus* silken fibres. (B): a summary of all data for measured fibres.



collagens, with breaking stresses of  $322 \pm 68$  MPa and breaking strains of  $34 \pm 4\%$ , compared to 100–120 MPa breaking stress and 13% strain at break reported for other animal collagens<sup>11</sup>. A further comparison of the mechanical properties of collagen and other biological fibres to various synthetic materials can be found in Gosline et al.<sup>12</sup>.

The sawfly cocoons could readily be dissolved at room temperature in common protein denaturants such as 8 M urea or 6 M guanidine hydrochloride. Following denaturing polyacrylamide gel electrophoresis analysis, solutions of solubilized cocoons resolved as three discrete proteins predicted to be 53, 47 and 32 kDa (Fig. 4A, native silk). Under reducing conditions the protein pattern was not altered, indicating that disulfide bonds were not present between the proteins.

**Identification of silk protein sequences.** The silk proteins were identified after matching the mass of tryptic peptides from the three proteins from the cocoon (Fig. 4A) to predicted tryptic peptide masses from proteins encoded by three cDNA populations identified in a silk gland cDNA library isolated from late final instar larvae (53 kDa: 9 peptides; 47 kDa: 9 peptides, and 32 kDa: 3 peptides; Fig. 5). Over 70% of the sequences obtained from the sawfly silk gland cDNA library encoded the three silk proteins. The measured tryptic fragments did not match any other proteins in the Genbank database. The primary amino acid sequence of the proteins is shown in Figure 5 and described in the discussion.

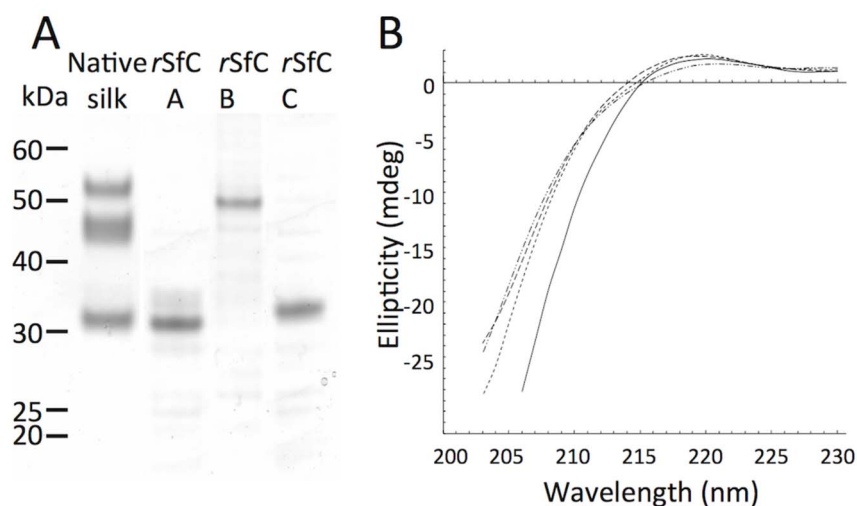
**Expression and analysis of silk proteins.** The ability of the sawfly silk proteins to fold into collagens was confirmed after recombinant expression of the proteins. The small size and absence of Hyp in the silk proteins were conducive to their recombinant expression in standard *E. coli* fermentation systems (Fig. 4A, rSfC A-C). The recombinant proteins were expressed from two different vectors (pET and pCold) and in neither case did they migrate according to their predicted molecular weight or alongside their native equivalents. Whilst it is common for collagen proteins to migrate inconsistently with protein markers, the inconsistency with the native proteins suggests either the proteins are not fully unfolding in the guanidine hydrochloride solubilisation treatment or that post-translational modifications are occurring in the native system. The recombinant molecules were digested with pepsin at 20°C, a technique commonly used to isolate and purify collagen molecules.

Consistent with the collagen structure, a substantial part of each of the recombinantly-produced sawfly silk proteins was resistant to pepsin digestion, whereas the entire molecule was protease-sensitive at higher temperatures. The circular dichroism spectra of the pepsin-resistant fragments confirmed the proteins were folded into collagen triple-helices (Fig. 4B), with each spectrum showing characteristic collagen positive ellipticity around 220 nm<sup>13</sup>.

## Discussion

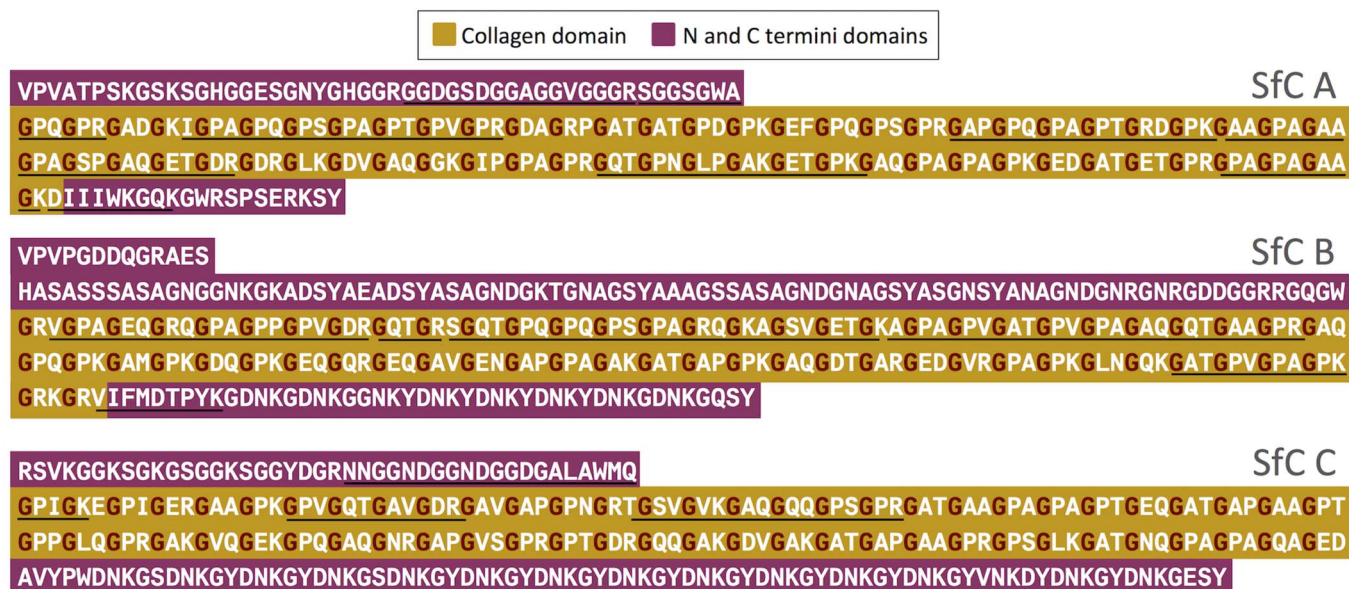
The present study examines the silk (Fig. 2) of the willow sawfly, *Nematus oligospilus*, a hymenopteran species restricted to willows in temperate areas of Australia. As with all silks, the cocoon is produced from a concentrated protein solution that the insect accumulates in a dedicated silk gland. In the willow sawfly, the silk gland is derived from the larval labial gland, a common adaptation in insects<sup>10</sup>. In contrast to homologous glands where the silk proteins are produced in the anterior region and then accumulate in a posterior lumen, the lumen of the sawfly silk gland is ringed with secretory units along its entire length (Fig. 2), and the silk proteins are organized into higher level structures within the secretory units (Supplementary Fig. 1). The cocoon is made of silk-like micro-scale fibres (Fig. 2) and the X-ray diffraction pattern from the fibres is characteristic of a collagen triple helix (Fig. 1). However, the fibres dissolve readily in protein denaturants, a property not common to either collagens or silks (Fig. 4A). Despite their comparatively low chemical stability, the mechanical properties of the silk fibres exceeds that of mammalian collagens (Fig. 3), with breaking stresses and breaking strains around three times higher than that reported for other animal collagens<sup>11</sup>. Collectively, these findings suggest the sawfly produces unusual silk using novel collagens.

The solubilized cocoon contained three silk proteins (Fig. 4A) whose primary sequence was identified from cDNA isolated from a silk gland cDNA library (Fig. 5). Analysis of the protein sequences conclusively identified them as collagen-forming proteins: each contained a central sequence block of 60–62 contiguous Gly-Xaa-Yaa repeats that comprised the majority of the protein sequence. Interestingly, the length of the collagen domains (180 to 186 residues) correspond to an axial length of 515 to 532 Å, similar to the axial period of 550 Å observed in negatively stained tactoids from gooseberry sawfly by Rudall (1967). The collagen structure in the silk proteins was further confirmed after the three proteins were individually expressed in *E. coli* (Fig. 4A) and purified recombinant



**Figure 4 | Sawfly silk proteins.** (A): SDS-PAGE of recombinant sawfly silk proteins alongside native silk proteins with protein marker ladder shown on left. (B): Circular dichroism spectra of recombinant sawfly cocoon collagens after pepsin treatment showing characteristic collagen maxima at about 220 nm, SfC A (---), SfC B (- · - · -) and SfC C (·····). The solid line (—) shows a representative collagen spectra for comparison. The spectra for SfC B (- · - · -) and SfC C (·····) are overlapping.





**Figure 5 | Architecture and amino acid sequences of the willow sawfly silk proteins.** Tryptic peptides identified by mass spectrometry from proteins in cocoon silk are underlined. The first position of each Gly-Xaa-Yaa repeat is shown in red.

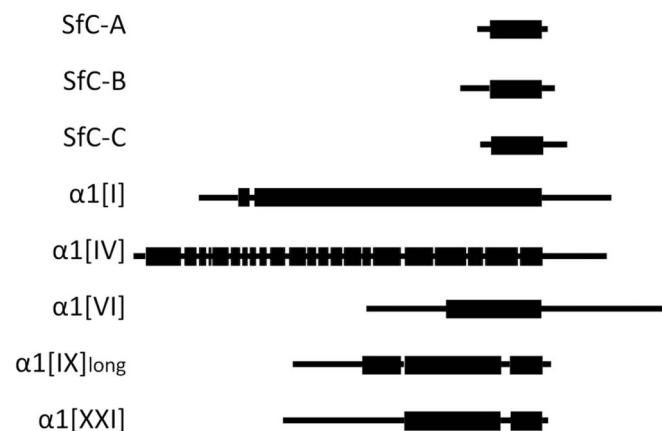
molecules were demonstrated to be protease-resistant with circular dichroism spectra consistent with the collagen structure (Fig. 4B). Comparison of the primary sequence, size and molecular organization (Fig. 6) of the sawfly silk proteins to animal collagen proteins available in the Genbank database found the silk collagens were only similar to each other and not to other described collagens, suggesting they evolved independently of other collagens and thereby constituting a new class of animal collagens. We termed the sawfly silk sequences SfC A–C (Sawfly Collagen A, B or C).

The vast majority (98%) of the triplets in the sawfly silk proteins are commonly found in other animal collagens, which only contain a few of the 400 possible combinations at any significant level<sup>14</sup>. Only 2% of the triplets, Gly-Glu-Phe and Gly-Lys-Ile in SfC A, Gly-Ala-Met in SfC B and Gly-Asn-Gln in SfC C, are considered rare<sup>14</sup>. However, the distribution and composition of imino acids in the

sawfly silks is highly unusual for animal collagens, in that there are no hydroxyprolines, a post-translational modification of proline considered characteristic. In other animal collagens, the X-position in the primary sequence Gly-Xaa-Yaa triplet is commonly proline and the Y-position is frequently hydroxyproline. The overall high level of imino acids serves to correctly position the protein backbone for collagen folding<sup>15</sup>, and Hyp in the Y-position is associated with thermal stability<sup>16</sup>. A lack of hydroxyproline however, is not unique. Recently a new group of collagen-like proteins have been described from bacteria which also lack hydroxyproline<sup>17</sup>. A range of these proteins have been characterised as recombinant products<sup>18–20</sup>. This work has shown that these collagens were stable at around mammalian body temperature, 35°C–38°C, despite the lack of hydroxyproline, with intra-molecular ion pair formation being partly responsible for this stability<sup>21,22</sup>.

In addition to the absence of Hyp, the proline distribution in the sawfly silk proteins is biased, occurring predominantly in the X-position (53%, 39% and 38% for SfC A–C, respectively) and rarely in the Y-position (6%, 4% and 8%, respectively). In other animal collagens, proline occupies around 28% of the X-position and Hyp occupies about 38%<sup>23</sup>. However, the total proportion of imino acids in the sawfly silk proteins is similar to that found in other animal collagens, comprising around 16% of the residues in the collagen-forming domains, consistent with the role for this residue in assisting the folding of the protein backbone into the collagen structure.

Mechanisms to increase thermal stability in the sawfly silks, in the absence of Hyp in the Yaa position, include the presence of particular triplets that enhance collagen stability. Most notably, SfC A–C have 7, 4 and 6 arginine residues in the Yaa position, respectively. Arginine in the Y-position confers triple-helix stability similar to Hyp<sup>24</sup>. However, the ecology of the animal and biology of silk production make the requirement for thermal stability in the collagen molecules less stringent than the requirement in mammals. Sawflies are poikilothermic and therefore only require the collagen to be stable at environmental temperatures, rather than mammalian body temperature. Furthermore, the biophysical environment of the silk proteins both during silk fabrication and post-fabrication serve to enhance the collagen molecules stability. Unlike other collagens, the silk proteins are maintained in a highly concentrated silk protein solution in the silk gland. The thermal stability of collagen rises as water is removed, such as when aggregates form<sup>25,26</sup>, and therefore concentration and



**Figure 6 | Comparison of the molecular architecture collagen-forming silk proteins SfC A–C to a selection of other animal collagens.** These include an interstitial collagen  $\alpha 1$ [I]; a network forming collagen,  $\alpha 1$ [IV]; a beaded filament-forming collagen,  $\alpha 1$ [VI]; and 2 FACIT collagens of different sizes,  $\alpha 1$ [IX] and  $\alpha 1$ [XXI]. Lines indicate comparative protein length and blocks indicate regions of collagen forming sequence within the chains. The published sequences for the collagens from the insect *Drosophila melanogaster* are highly homologous to the type IV sequences.



aggregation of the collagen molecules in the silk gland probably serves to enhance the molecule's thermal stability. After silk fibre fabrication, the collagen is dehydrated and the molecules will be stable at any normal environmental temperature.

The collagen domains of the sawfly silk proteins are flanked by variable length glycine-rich, non-collagenous domains of 49, 103 and 41 (N-termini) and 21, 46 and 84 (C-termini) residues for the SfC A, B and C chains, respectively. Repetitive motifs occurred in the C-terminal domain of the SfC B and SfC C proteins: eight repeats of the tetrapeptide Gly/Tyr-Asp-Asn-Lys in SfC B, and 14 repeats of the pentapeptide Gly-Tyr-Asp-Asn-Lys in SfC C. With the exception of the C-terminal domain of SfC C, which showed 76% similarity over 83 residues to a Gly-Tyr-Asp-Asn-Lys repeat in an uncharacterized protein from the fungus *Arthroderma otae*, none of the terminal domains showed sequence similarities to known proteins. Generally, animal collagens contain N- and C-terminal peptides that are essential for folding of the triple-helix and correct fibrillogenesis but are then proteolytically removed and are not present in the mature material<sup>27</sup>. Several peptides corresponding to the N- and C-termini of the sawfly silk proteins were identified by liquid chromatography-tandem mass spectroscopy of proteins solubilized from cocoons (Fig. 5) indicating that most of these regions, if not all are present in cocoons. These regions may contribute to the mechanical properties of the fibres.

The data presented in this paper have demonstrated the presence of the collagen triple-helix structural motif in an insect silk, confirming the initial proposal by Rudall in the 1960s<sup>2</sup>. Collagen is an established biomedical material and many tons of animal-derived collagen are used annually in pharmaceutical and medical applications. However, there have been concerns over batch-to-batch variability and the transmission of diseases in the materials, leading to a preference for development of a recombinant material that can be produced under controlled conditions<sup>28</sup>. Replication of other animal collagens in recombinant systems requires co-expression of enzymes capable of converting a number of the proline residues in the protein to Hyp<sup>29,30</sup> and, to date, such systems have not allowed large-scale production of the material<sup>31</sup>. The discovery of relatively small collagen proteins that fold into collagen triple-helices but do not contain hydroxyproline, that can be fabricated into materials from concentrated protein solutions, and that form strong fibres without covalent cross-linking provides opportunities for collagen materials to be prepared as recombinant products in *E. coli*. The collagen silk proteins therefore constitute a new class of animal collagens ideally suited to the development of innovative collagen-based biomedical materials.

## Methods

**Collection and culture of insects.** Willow sawfly larvae (*Nematus oligospilus*) were collected from willow trees (*Salix sp.*) around Lake Burley Griffin (Canberra, Australia) during the summers of 2007–2012. Larvae were maintained in the laboratory on a diet of fresh willow leaves.

**Amino acid analysis.** After spinning the cocoons, the sawflies were removed from the cocoons before metamorphosis was complete and cocoon parts not in contact with leaves or other substrate were selected for analysis. Six cocoons were washed in distilled water for 30 min three times and dried. The hydroxyproline composition of the washed silk was determined in duplicate after 24 h gas phase hydrolysis at 110°C using the Waters AccQTag Ultra chemistry at the Australian Proteome Analysis Facility Ltd (Macquarie University, Sydney). The limit of sensitivity in this analysis was estimated as 0.1 pmol total, corresponding to about 1 in 5000 residues.

**Wide angle X-ray scattering (WAXS).** Diffraction patterns were collected at the SAXS/WAXS beam-line of the Australian Synchrotron. The beam-line was operated at a beam energy of 18 keV (wavelength of 0.6888 Å) with a sample to detector distance of 550 mm, yielding a q-range of approximately 0.08–2.6 Å<sup>-1</sup>. Use of a pixel counting detector (Pilatus-1M, Dectris, Baden, Switzerland) and evacuated flight tube achieved a reasonable signal to noise ratio from the weakly scattering sample. Data reduction and azimuthal integration, to produce 1D profiles, were achieved using scatterBrain (Australian Synchrotron).

**Protein solubilization and gel electrophoresis.** Sawfly cocoons were incubated in 6 M guanidine HCl with or without 5% v/v 2-mercaptoethanol for 30 min at 95°C,

after which no solid material remained. The guanidine HCl was removed after proteins were precipitated by the addition of nine volumes of cold 100% ethanol, incubated at -20°C for 1 hr, micro-centrifuged at maximum speed (15000 g) for 15 min at 4°C, the pellet washed with 90% cold ethanol, the pellet dried and the proteins resuspended in SDS-PAGE running buffer as required. Solubilized protein solution was run on a Nu-PAGE 4–12% Bis-Tris protein gel (Life Technologies) and stained with Coomassie Blue R-250.

**Mechanical testing.** When disturbed, final instar larvae in the process of spinning would rapidly move away from stimuli leaving a single fibre that could be collected for mechanical testing. Fibres were mounted across a 2 mm gap on paper frames, fixed at either end with epoxy glue, and examined on an optical microscope to determine their exact gauge length and diameter (and to examine and discard any samples with defects before mechanical testing). Tensile measurements were carried out on an Instron Tensile Tester model 4501 at a rate of 2.5 mm.min<sup>-1</sup>. Tests were conducted in air at 21°C and 65% relative humidity. Data from fibres that broke at the mounting points were excluded.

**Silk protein gene discovery.** The DNA sequence corresponding to the silk protein genes were obtained from cDNA isolated from a cDNA library constructed from silk glands of four *N. oligospilus* larvae according to methods described previously<sup>32</sup>. Briefly, 29.1 µg total RNA was isolated using the RNAqueous4PCR kit (Life Technologies), from which mRNA was isolated using the Micro-FastTrack™ 2.0 mRNA Isolation kit (Life Technologies). This mRNA was used to construct a cDNA library using the CloneMiner™ cDNA kit (Life Technologies). The cDNA library comprised approximately 2.9 × 10<sup>7</sup> colony forming units, with an average insert size of 1.1 kb. From this library, 50 randomly chosen clones were sent for sequence analysis. Three genes represented by three distinct groups of sequences containing repeating (Gly-Xaa-Yaa)<sub>n</sub> sequences were found in 20 of the clones and these were termed SFC A, B and C (deposited in Genbank as KF534808, KF534809 and KF534807). Other sequences contained vector only sequences with the exception of five, which were found as single occurrences, with possible identities determined from database searches as peptidase, esterase, actin, RNA helicase and translation elongation factor. SFC A and SFC B were represented by two sequences differing by 8 and 6 single nucleotide polymorphisms respectively, suggesting the presence of two separate alleles. No variations were seen within the SFC C group of sequences.

**Mass spectrometry.** The protein bands on SDS-PAGE gels were cut out with a razor and analyzed as described previously<sup>32</sup>. The proteins were digested with trypsin, and the resultant peptides were analysed by liquid chromatography-tandem mass spectrometry on an Agilent LC/MSD Trap XCT spectrometer. Agilent SpectrumMill software was used to compare the peptides to both the sequences obtained from a cDNA library of the sawfly silk gland and to the entire NCBI database of protein sequences.

**Protein expression and purification.** Full length cDNA clones from each *N. oligospilus* collagen types (SfC A–C) were PCR amplified using oligonucleotides designed against the 5' and 3' cDNA sequences and with appropriate restriction enzyme sites for transfer into the pET14b vector (Novagen) behind the His tag. The amplicons were purified, digested with the restriction enzymes, cloned into pET14b vectors (Novagen) and the sequences verified by DNA sequencing of the inserts. Constructs with the correct sequence were used for expression in *E. coli* Rossetta 2 (DE3) cells (Novagen) using Overnight express media (Novagen) at 20°C for 48 hrs. Alternatively, sequences were modified at either end using the Quickchange Site-directed Mutagenesis Kit (Agilent Technologies) according to the manufacturer's instructions, to allow transfer into the pDONR 222 vector (Invitrogen). In order to increase expression of SFC C, the V-domain from *S. pyogenes* ScI2 gene, which is a registration and triple-helix promoting sequence, was inserted at the N-terminal of the construct. Sequences were then inserted into pCold expression vectors (Takara Bio Inc., Japan), then transformed into competent *E. coli* BL21 cells for expression as required. Transformed cells were grown on auto-induction media (Novagen) at 20°C or 2YT media (16 g.L<sup>-1</sup> Trypton; 5 g.L<sup>-1</sup> yeast extract, 5 g.L<sup>-1</sup> NaCl) containing 0.1 mg/ml ampicillin at 37°C in shaker flasks for 7 h, cooled to 25°C, induced with 1 mM IPTG, grown for a further 10 h, and then a further 16 h at 15°C. The cells were harvested by centrifugation (3000 g for 30 min).

Cells were lysed using Bugbuster and expression levels was assessed using SDS PAGE. For circular dichroism, cells were lysed by sonication in 40 mM sodium phosphate buffer pH 8.0, and the cell lysate clarified by centrifugation (20,000 g for 40 min) and the clear supernatant retained. The expressed silk proteins were purified by absorbing the clarified lysates on an IMAC HyperCel™ column (Pall), with elution by stepwise increments up to 500 mM imidazole, adjusted to pH 8.0 with HCl. Cross-flow filtration was used to lower the salt content and to concentrate the protein solution. Final purification was by gel permeation chromatography on a HiPrep Sephacryl™ S-200 column (GE Healthcare). The individual triple-helical collagen segments were prepared by digestion of the proteins with 0.1 mg/ml pepsin in 50 mM acetic acid. Purity of all products was assessed by SDS-PAGE.

**Circular dichroism spectroscopy.** Circular dichroism spectra were collected for pepsin treated silk protein samples in 50 mM acetic acid using 1 mm path length cells in a JASCO J-815 instrument.

1. Traub, W., Yonath, A. & Segal, D. M. On the molecular structure of collagen. *Nature* **221**, 914–917 (1969).



2. Rudall, K. M. Silk and other cocoon proteins. *Comparative Biochemistry* **4B**, 397–433 (1962).
3. Rudall, K. M. Comparative biology and biochemistry of collagen In *Treatise on Collagen* ed BS Gould Chapter 2 **2A**, 83–137. London and New York: Academic (1967).
4. Rudall, K. M. & Kenchington, W. Arthropod silks: The problem of fibrous proteins in animal tissues. *Ann. Rev. Entomol.* **16**, 73–96 (1971).
5. Yasothornsrikul, S., Davis, W. J., Cramer, G., Kimbrell, D. A. & Dearolf, C. R. *viking*: identification and characterization of a second type IV collagen in *Drosophila*. *Gene* **1**, 17–25 (1997).
6. Fowler, S. J., Jose, S., Zhang, X., Deutzmann, R., Sarras, M. P. & Boot-Handford, R. P. Characterization of hydra type IV collagen. Type IV collagen is essential for head regeneration and its expression is up-regulated upon exposure to glucose. *J. Biol. Chem.* **15**, 39589–99 (2000).
7. Ashhurst, D. E. & Bailey, A. J. Locust collagen: morphological and biochemical characterization. *Eur. J. Biochem.* **103**, 75–83 (1980).
8. Francois, J., Herbage, D. & Junqua, S. Cockroach collagen: isolation, biochemical and biophysical characterization. *Eur. J. Biochem.* **112**, 389–96 (1980).
9. Fraser, R. D. B., MacRae, T. P. & Suzuki, E. Chain conformation in the collagen molecule. *J. Mol. Biol.* **129**, 463–481 (1979).
10. Sutherland, T. D., Young, J. H., Weisman, S., Hayashi, C. Y. & Merritt, D. J. Insect silk: One name, many materials. *Annu. Rev. Entomol.* **55**, 171–188 (2010).
11. Alexander, R. McN. & Bennet-Clark, H. C. Storage of elastic strain energy in muscle and other tissues. *Nature* **265**, 144–117 (1977).
12. Gosline, J. M., Guerette, P. A., Ortlepp, C. S. & Savage, K. N. The mechanical design of spider silks: from fibroin sequence to mechanical function. *J. Experim. Biol.* **202**, 3295–3303 (1999).
13. Mohs, A., Silva, T., Yoshida, T., Amin, R., Lukomski, S. *et al.* Mechanism of stabilization of a bacterial collagen triple helix in the absence of hydroxyproline. *J. Biol. Chem.* **282**, 29757–29765 (2007).
14. Ramshaw, J. A. M., Shah, N. K. & Brodsky, B. Gly-X-Y tripeptide frequencies in collagen: a context for host-guest triple-helical peptides. *J. Struct. Biol.* **122**, 86–91 (1998).
15. Cram, D. J. The design of molecular hosts, guests, and their complexes. *Science* **240**, 760–767 (1988).
16. Rosenbloom, J., Harsch, M. & Jimenez, S. Hydroxyproline content determines the denaturation temperature of chick tendon collagen. *Arch. Biochem. Biophys.* **158**, 478–484 (1973).
17. Rasmussen, M., Jacobsson, M. & Björck, L. Genome-based identification and analysis of collagen-related structural motifs in bacterial and viral proteins. *J. Biol. Chem.* **278**, 32313–32316 (2003).
18. Xu, Y., Keene, D. R., Bujnicki, J. M., Höök, M. & Lukomski, S. Streptococcal Scl1 and Scl2 proteins form collagen-like triple helices. *J. Biol. Chem.* **277**, 27312–27318 (2002).
19. Mohs, A., Silva, T., Yoshida, T., Amin, R., Lukomski, S., Inouye, M. & Brodsky, B. Mechanism of stabilization of a bacterial collagen triple helix in the absence of hydroxyproline. *J. Biol. Chem.* **282**, 29757–29765 (2007).
20. Xu, C., Yu, Z., Inouye, M., Brodsky, B. & Mirochnitchenko, O. Expanding the family of collagen proteins: recombinant bacterial collagens of varying composition form triple-helices of similar stability. *Biomacromolecules* **11**, 348–356 (2010).
21. Persikov, A. V., Ramshaw, J. A. M., Kirkpatrick, A. & Brodsky, B. Peptide investigations of pairwise interactions in the collagen triple-helix. *J. Mol. Biol.* **316**, 385–394 (2002).
22. Persikov, A. V., Ramshaw, J. A. M. & Brodsky, B. Prediction of collagen stability from amino acid sequence. *J. Biol. Chem.* **280**, 19343–19349 (2005).
23. Kadler, K. Extracellular matrix. 1: Fibril-forming collagens. *Protein Profile* **1**, 519–638 (1994).
24. Yang, W., Chan, V. C., Kirkpatrick, A., Ramshaw, J. A. M. & Brodsky, B. Gly-Pro-Arg confers stability similar to Gly-Pro-Hyp in the collagen triple-helix of host-guest peptides. *J. Biol. Chem.* **272**, 28837–28840 (1997).
25. Miles, C. A. & Ghelashvili, M. Polymer-in-a-box mechanism for the thermal stabilization of collagen molecules in fibres. *Biophys. J.* **76**, 3243–3252 (1999).
26. Mathews, M. *Connective tissue*. (Springer-Verlag, Berlin, 1975).
27. Kadler, K. E., Hojima, Y. & Prockop, D. J. Assembly of collagen fibrils *de novo* by cleavage of the type I pC-collagen with procollagen C-proteinase. *J. Biol. Chem.* **262**, 15696–701 (1987).
28. Werkmeister, J. A. & Ramshaw, J. A. M. Recombinant protein scaffolds for tissue engineering. *Biomed. Mater.* **7**, 012002 (2012).
29. Ricard-Blum, S. The collagen family. *Perspect. Biol.* **3**, a004978 (2011).
30. Brodsky, B. & Ramshaw, J. A. M. The collagen triple-helix structure. *Matrix Biol.* **15**, 545–554 (1997).
31. Myllyharju, J. Recombinant collagen trimers from insect cells and yeast. *Methods Mol. Biol.* **522**, 51–62 (2009).
32. Sutherland, T. D. *et al.* A highly divergent gene cluster in honeybees encodes a novel silk family. *Genome Res.* **16**, 1414–1421 (2006).

## Acknowledgments

We thank Nicole Phair-Sorensen for assistance with mechanical tests and Stephen Trowell for critical reading of the manuscript. This work was funded by CSIRO. The study was facilitated by access to the Australian Proteome Analysis Facility supported under the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS). Part of the research was undertaken at the SAXS/WAXS beam-line at the Australian Synchrotron, Victoria, Australia. Supplementary Information is linked to the online version of the paper.

## Author contributions

T.D.S. and V.S.H. conceived the project. T.D.S., Y.Y.P., S.W., M.H., J.A.W., V.G., and J.A.M.R. designed experiments. T.D.S., M.H. and J.A.M.R. supervised experiments. H.E.T., A.S., S.O., Y.Y.P., S.W., A.A.W., J.F.W., M.H., V.G., S.T.M., and V.S. performed experiments. All authors were involved in analysis of the data from their experimental areas. T.D.S. and J.A.M.R. wrote the manuscript, with input to specific sections from the other authors. All authors reviewed the manuscript.

## Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Sutherland, T. D. *et al.* A new class of animal collagen masquerading as an insect silk. *Sci. Rep.* **3**, 2864; DOI:10.1038/srep02864 (2013).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported license. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>