PROCEEDINGS B

rspb.royalsocietypublishing.org

Review

Cite this article: Walker AA, Holland C, Sutherland TD. 2015 More than one way to spin a crystallite: multiple trajectories through liquid crystallinity to solid silk. *Proc. R. Soc. B* **282**: 20150259.
http://dx.doi.org/10.1098/rspb.2015.0259

Received: 4 February 2015
Accepted: 11 May 2015

Subject Areas:
biomaterials, structural biology, evolution

Keywords:
silk, liquid crystal, coiled coil, collagen, cross-β-sheet, mesophase

Authors for correspondence:
Andrew A. Walker
e-mail: a.walker@uq.edu.au
Tara D. Sutherland
e-mail: tara.sutherland@csiro.au

†Current address: Institute for Molecular Bioscience, The University of Queensland, St Lucia, Queensland 4072, Australia.

THE ROYAL SOCIETY PUBLISHING

More than one way to spin a crystallite: multiple trajectories through liquid crystallinity to solid silk

Andrew A. Walker[1,2,†], Chris Holland[3] and Tara D. Sutherland[2]

[1]Research School of Biology, Australian National University, Canberra 0200, Australia
[2]Food and Nutrition, CSIRO, Canberra 2600, Australia
[3]Department of Materials Science and Engineering, The University of Sheffield, Sheffield S1 3JD, UK

Arthropods face several key challenges in processing concentrated feedstocks of proteins (silk dope) into solid, semi-crystalline silk fibres. Strikingly, independently evolved lineages of silk-producing organisms have converged on the use of liquid crystal intermediates (mesophases) to reduce the viscosity of silk dope and assist the formation of supramolecular structure. However, the exact nature of the liquid-crystal-forming-units (mesogens) in silk dope, and the relationship between liquid crystallinity, protein structure and silk processing is yet to be fully elucidated. In this review, we focus on emerging differences in this area between the canonical silks containing extended-β-sheets made by silkworms and spiders, and 'non-canonical' silks made by other insect taxa in which the final crystallites are coiled-coils, collagen helices or cross-β-sheets. We compared the amino acid sequences and processing of natural, regenerated and recombinant silk proteins, finding that canonical and non-canonical silk proteins show marked differences in length, architecture, amino acid content and protein folding. Canonical silk proteins are long, flexible in solution and amphipathic; these features allow them both to form large, micelle-like mesogens in solution, and to transition to a crystallite-containing form due to mechanical deformation near the liquid–solid transition. By contrast, non-canonical silk proteins are short and have rod or lath-like structures that are well suited to act both as mesogens and as crystallites *without* a major intervening phase transition. Given many non-canonical silk proteins can be produced at high yield in *E. coli*, and that mesophase formation is a versatile way to direct numerous kinds of supramolecular structure, further elucidation of the natural processing of non-canonical silk proteins may to lead to new developments in the production of advanced protein materials.

1. Introduction: silk spinning and liquid crystallinity

(a) Silk comprises multiple diverse materials produced by distinct groups of arthropods

Silk production occurs in a diverse range of arthropods including crustaceans, mites, centipedes and insects, and it is now recognized that silk production has evolved independently more than 23 times within the arthropods [1]. For the purposes of this review, we define silks as protein materials that are converted (spun) from a highly concentrated liquid feedstock (dope), and which undergo a liquid-to-solid phase transition concurrently with being mechanically drawn from the silk gland into the external air (pultrusion).

Silk materials are semi-crystalline polymers consisting of ordered domains containing crystallites with precise secondary structures and high H-bonding density, interspersed with disordered amorphous domains with a lower H-bonding density [2,3]. Of particular note, as it serves as a basis for our sub-classification, the structures of the ordered crystallites may be β-sheets, α-helical coiled-coils, collagen helices or the polyglycine II structure [1,4]. Further structural variation derives from crystallite orientation, a feature distinguishing the extended-β-sheet

© 2015 The Author(s) Published by the Royal Society. All rights reserved.

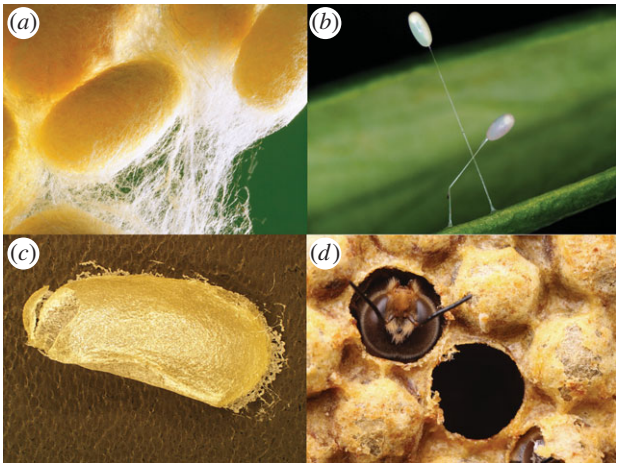

Figure 1. Canonical and non-canonical silks. (*a*) Silkworm (*B. mori*) cocoon fibres, a canonical silk. (*b–d*) Non-canonical silks produced by other insect species. (*b*) Lacewing (*M. signata*) egg-stalk silk, photograph by Holly Trueman. (*c*) Sawfly (*Nematus oligospilus*) cocoon silk. (*d*) Honeybee (*Apis mellifera*) silk and wax on cell caps of a hive, photograph by Alex Wild.

structure (amide backbones parallel to fibre axis) from the cross-β-sheet structure (amide backbones perpendicular to fibre axis). The canonical silks such as the cocoon fibres of silkworms (*Bombyx mori*; Lepidoptera) and dragline silks of orb-spiders (Araneidae) contain crystallites of the extended-β-sheet type; an adaptation that has also arisen independently in raspy crickets (Orthoptera: Gryllacrididae) [5] and some sawflies and parasitic wasps (Hymenoptera) [4]. Silks containing crystallites of other types, referred to here collectively as *non-canonical silks*, are produced by at least nine groups of insects [1,4]. In this review, we compare canonical silks made by the silkworm and orb-spiders with the non-canonical silks made by three insect taxa (figure 1): the aculeates (ants, bees and wasps; Hymenoptera: Aprocrita), sawflies of the tribe Nematini (Hymenoptera: Tenthredinidae) and lacewings (Neuroptera: Chrysopidae).

X-ray scattering experiments on aculeate silk reveal longitudinal *d*-spacings of 0.51 nm, indicating that crystallites are of the coiled-coil type and orientated parallel to the fibre axis, and lateral spacings of approximately 2.5 nm corresponding to lateral packing between superhelices [4,6]. These data suggest superhelices share a common orientation but not ordered in a lattice-like way with respect to adjacent superhelices. Hence, the 'crystallites' in aculeate silk are probably best understood as single-coiled-coil superhelices. Sawflies in the tribe Nematini make pupal cocoons in which another superhelical structure, the collagen triple helix, is oriented on average parallel to the fibre axis [4]. The third non-canonical silk we consider in this review are egg-stalks made by mother green lacewings (*M. signata*), which have the cross-β-sheet structure [7].

Understanding the relationship between silk protein amino acid sequence, processing and the hierarchical structure of the final material spans the fields of physiology, protein folding, molecular evolution and materials science. We will argue that there are compelling reasons to believe that the process of spinning canonical and non-canonical silks is substantially different, constituting two separate general strategies and hence 'fitness peaks' [8] in the adaptive landscape of the system as a whole—including protein structure and length, gland morphology and the material's mechanical properties before, during and after spinning. Understanding the full range of silk processing modes used by organisms is invaluable in guiding the endeavours for a long-standing biotechnological challenge: true biomimetic manufacture of silks and faithful replication of a silk fibre's mechanical properties [9].

(b) Principles of silk spinning by arthropods

While the key challenges of silk processing are shared across disparate silk-producing taxa, we note that common challenges may be overcome via disparate strategies. We summarize the main challenges in silk processing under three headings.

(i) Overcoming flow viscosity

To facilitate transition to a solid material, silk dope is a highly concentrated solution of proteins, frequently 20–40% of dry weight [10]. Initial rheological measurements on freshly extracted spider and silkworm dopes found both to be viscoelastic materials, similar to a high molecular weight polymer acting as a weak gel, with a high viscosity in the KPa · s range [10–12]. Taken in isolation and from a polymer processing perspective, this material would present these animals with a considerable challenge to process into a micrometre-sized fibre as the forces required to flow the material down the silk duct are prohibitively high [13]. However, microscopic analysis of the silk spinning apparatus in both animals provided a solution; unspun canonical silks appear to be liquid crystals [14–16]. In the silk ducts of silkworm silk glands and orb-spider major ampullate glands, liquid crystalline textures indicating the presence of nematic mesophases (see §1c) are observed, which persist in the dope until disappearing shortly before the spinneret [14]. The presence of liquid crystallinity in silk dopes reduces friction between molecules, lowering viscosity, which is vital as it reduces the energetic requirements of the animal producing the fibre, a key factor governing the evolution of the silk production process. [17]. While it is well documented that silk dopes undergo shear thinning [11,18], a feature that is shared with other liquid crystals under flow [17], this is not evidence *per se* that silk mesophases entail more efficient processing. However, recent use of shear-induced polarization light imaging (SIPLI) [19] demonstrated that the mechanical work input required to create fibrillar structures in silk is three orders of magnitude less compared with a non-liquid crystalline synthetic polymer melt (high-density polyethylene) [20].

Like canonical silks, non-canonical silks are produced from liquid crystalline feedstocks. In final instar honeybee larvae, silk glands contain birefringent structures called tactoids (characteristic structures formed when mesophases exist in equilibrium with isotropic phases [13]) surrounded by an isotropic fluid [21–23]. Consistent with this notion, tactoids are observed to form first at the periphery of the gland lumen, with their tips terminating at the surface of cuboidal cells where silk proteins are secreted. Later, tactoids are observed within the entire gland lumen [21,23]. In the silk glands of bumblebees and hornets, silk dope has been observed to form a pattern called 'fibrous bars' which we suggest is actually the product of fusion of tactoids into a cohesive mesophase spanning the silk gland. Honeybee tactoids show a regular banding pattern with periodicity of 500 nm, while the 'width' of the fibrous bars is 1000–1600 nm. In both cases, the banding pattern is likely to correspond to a repeating aspect of the mesophase structure such as the pitch of a chiral nematic (see §1c).

(ii) Control of solubility and aggregation

Silk must be capable of solidifying as it leaves the body of the organism, but premature solidification inside the silk gland must be avoided. As mesophases form at high mesogen concentration but still retain liquid properties, mesophase formation is probably the major mechanism by which solubility is maintained and aggregation avoided. In silkworms and spiders, solubility of silk proteins is also maintained during storage by modulation of dope pH, ionic composition and water content [24]. As silk dope flows anteriorly towards the spinneret along an increasingly narrow duct, the pH is lowered, cation concentrations increase, and in the case of Lepidopteran silks (i.e. *B. mori*) a sheath of sericin proteins is thought to have the effect of dehydrating the dope. Each of these changes primes the dope for aggregation and the liquid → solid transition.

In many silks of both the canonical and non-canonical types, covalent cross-linking upon extrusion, either owing to oxidative cross-linking through cysteine residues or due to enzymatic tanning, may assist aggregation and solidification [7,25]. However, covalent cross-linking is not a requisite feature of fabrication of either canonical or non-canonical silk proteins, as demonstrated by the potential for silkworm, hornet and sawfly silks to be dissolved in chaotropic solutions without reducing agents [26–28].

(iii) Formation and orientation of crystallites

In the model system of the silkworm, proteins in silk dope have a structure rich in β-turns with substantial conformational flexibility (the 'silk I' structure) [29]. Rheological studies combined with various spectroscopic, microscopic and scattering techniques suggest mesophase intermediates are also involved in the development of molecular and supramolecular structure in silkworm glands [14]. On the nanoscale, Rheo-IR studies have indicated that silk proteins will align in response to flow and that alignment occurs prior to silk I → silk II conversion [30]. On the microscale, confocal rheology has demonstrated the development of fibrillar structures oriented along the shearing direction [31] and on the macroscale SIPLI has shown flow-induced birefringence [20]. However, care must be taken to assign specific structures responsible for the birefringence patterns observed using visible light microscopy either in the fibre or the gland itself [16,32] as the liquid crystalline texture of the dope disappears prior to the silk I → silk II conversion. Therefore, the structural unit (mesogen) is unlikely to comprise solely β-sheet crystallites [14] (see §3a).

The process of converting silk I to the final silk II structure (extended-β-sheet crystallites surrounded by amorphous chains) is achieved by mechanically deforming the dope as it flows down the duct prior to its emergence from the spinneret. Mechanical deformation causes molecular extension of fibroin proteins due to shearing forces from friction caused at the duct wall, extensional forces due to the drawing of the material, and elongational flow. Molecular extension brings adjacent chains into close apposition, promoting dehydration, the formation of intermolecular hydrogen bonds and aggregation, and hence the silk I → silk II conversion yielding aligned β-sheet crystallites [14,33]. Thus, the directionally applied mechanical deformation both contributes to the formation of crystallite secondary structure and the orientation of crystallites with respect to the fibre axis.

Crystallite formation and orientation in non-canonical silk processing can be inferred to be quite different compared with processing of silk by silkworms and spiders. In canonical silks, β-strands within crystallites are on average parallel with the fibre axis and therefore fully extended in this direction. By contrast, protein chains in non-canonical silks present as α-helices, collagen helices and cross-β-sheet crystallites lie at an angle to the fibre axis and may be further extended by applying tension to the fibre to form extended-β-sheets [4,34]. Therefore, the crystallites are unlikely to be formed in response to molecular extension during silk processing. Consistent with this reasoning, recombinant versions of aculeate hymenopteran and sawfly silk proteins form mature secondary structures similar to crystallites in solution, without the application of mechanical force [28,35] (see §2b). In addition, as crystallites are not formed due to forces acting parallel to the fibre axis during the liquid-to-solid transition, an alternative process must direct their orientation within the non-canonical silk fibres. We will argue that liquid crystalline states are ideally placed to direct this process, though in a fundamentally different manner to what occurs in canonical silk processing.

(c) In search of a mesogen: the mechanism of mesophase formation by silk proteins

Liquid crystals (mesophases) are states of matter having the properties of both liquids and crystalline solids. For example, molecules in close proximity may share alignment (orientational order) and/or a lattice-like arrangement (positional order) while retaining the ability to flow like a liquid. The first known liquid crystals were preparations of elongated organic molecules in which mesophase formation could be induced by temperature changes (thermotropic mesophases). The key feature that causes some molecules to be mesogenic and some not is their shape. A length/width ratio (axial ratio) *above approximately five* is sufficient for mesogenicity, provided the molecule can be concentrated; the higher the axial ratio of a molecule, the lower the concentration required to induce a mesophase [36]. Mesophase formation is a spontaneous event that occurs because the loss of orientational entropy associated with molecular alignment is outweighed by a gain in positional entropy [37]. In the simplest case, entropic minimization results in a *nematic* phase where the molecules share local orientational order without having positional order; the addition of positional order yields a *smectic* phase. Finally, chiral molecules may form layers at a preferred angle to underlying layers, producing chiral nematic and chiral smectic phases.

Lyotropic mesophases are more complex states of matter than thermotropic mesophases and form due to the interaction of two different components. A typical example is the formation of lipid bilayers—a type of liquid crystal—as a simple consequence of mixing amphiphilic molecules such as phospholipids and water. Apart from temperature, the key variable controlling the structure of the mesophase formed is the relative concentration of the two components. Phases dominated by vesicles, micelles, hexagonal columns, lamellae, cubic and inverse phases are formed with successively increasing concentrations of phospholipid [38]. Importantly for our discussion of mesogen structure in canonical silk processing (§3a), anisotropic lyotropic structures such as hexagonal columns and lamellae are subject to the same entropic considerations discussed above for small molecules, and thus may form large-scale versions of the nematic and smectic phases.

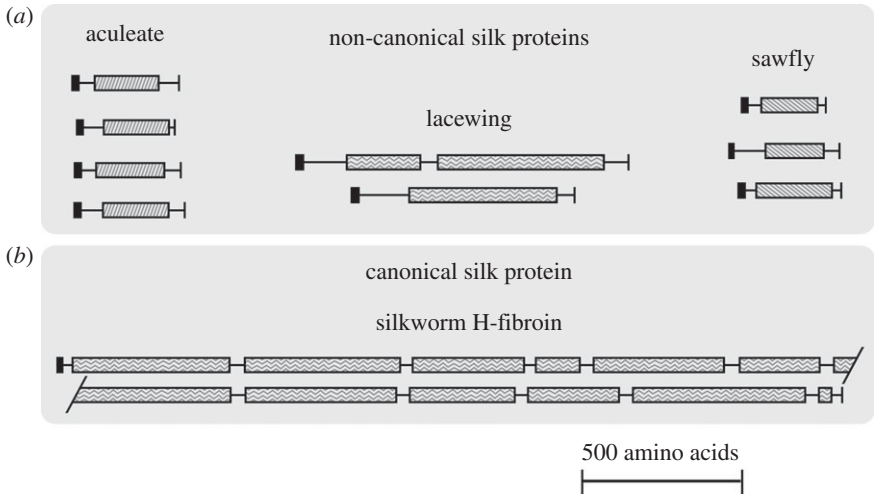

Figure 2. Comparison of amino acid sequences of proteins that form canonical and non-canonical silks. (*a*) Amino acid sequences of non-canonical silk proteins from aculeates, sawflies and lacewings showing convergence to short sequences with high-repeat regularity. (*b*) Canonical silk proteins such as silkworm H-fibroin (pictured) are typically very long with 'spacer' regions.

Like lipids, the structures formed by proteins depend intimately on the hydrophobic effect. However, the large size of proteins, their chemical complexity and potential for highly selective interactions arguably yields an even broader scope for mesogenic behaviour. Demonstrated polypeptide mesogens include silk proteins [16,39], elongated virus capsules [40], F-actin tubules [41], interstitial collagens [42] and α-helix-forming polypeptides such as poly-γ-benzyl-L-glutamate (PBLG; [43]). Among this list, the size of mesogens and the method of mesogen formation vary widely. For example, the 1500 nm F-actin filaments studied by Suzuki *et al*. [41] are orders of magnitude larger than PBLG molecules studied by Robinson [43]. High axial ratios can be achieved either through the use of intrinsically rod-like structures (collagen, PBLG) or end-to-end polymerization of multiple small globular proteins (tobacco mosaic virus (TMV), F-actin). The exact nature of mesogens in silk dope is unknown, but much can be inferred by analysis of silk protein amino acid sequences, studies on natural silk spinning *in vivo*, and the behaviour of regenerated and recombinant silk proteins.

2. Comparison of canonical and non-canonical silk proteins

(a) Length, architecture and repetition

Comparison of non-canonical and canonical silk protein sequences reveals marked differences in length, architecture and sequence repetition (figure 2). The main structural protein in silkworm silk is heavy-chain fibroin (H-fibroin, 350 kDa), which is linked covalently to light-chain fibroin (L-fibroin, 26 kDa) and non-covalently to the glycoprotein fibrohexamerin (P25, 30 kDa) to form a complex with stoichiometry 6 : 6 : 1 [3,44]. Orb-spider dragline silk is composed of two major proteins, MaSp1 and MaSp2, which have molecular weights in the range 250–500 kDa [3]. Thus, canonical extended-β-sheet-forming silk proteins are typically larger than 250 kDa. By contrast, non-canonical silk proteins from aculeates, sawflies and adult lacewings range between 22 and 86 kDa in size. For the superhelical silk proteins (those that form coiled-coils and collagens), protein length and in particular superhelical domain length appears tightly constrained. For example, aculeate silk is made up of four coiled-coil-forming proteins, silk fibroins 1–4. While the molecular weight of these proteins ranges between 29 and 45 kDa, the length of the coiled-coil domain is always approximately 30 heptads [45]. Sawfly silk is made up of three fibroins, SfCollA-C, that range between 22 and 32 kDa but each contain either 78 or 79 collagen tripeptides [28]. The lacewing egg-stalk proteins MalXB1 and MalXB2 are larger and less constrained in length compared with the superhelical silk proteins, being 86 kDa (50 repetitive motifs) and 54 kDa (29 repetitive motifs), respectively [7].

Analysis of non-canonical silk protein amino acid sequences suggests they have a common elongated tertiary structure. Whereas 'spacer' regions are a characteristic feature of canonical silk proteins [2,24], non-canonical silk proteins usually show no (or at most one) interruptions to the register of the central repetitive domain (figure 2). For example, the coiled-coil prediction algorithm MARCOIL [46] predicts a coiled-coil domain of approximately 210 residues in aculeate silk proteins, probably uninterrupted by heptad irregularities such as stutters or stammers [45]; the 78–79 (Xaa–Yaa–Gly) repeats in sawfly collagen silk proteins occur consecutively without interruption; and the cross-β-sheet-forming MalXB1 and MalXB2 contain one and zero interruptions within their repetitive domains, respectively. This high repeat continuity combined with the inherently rod-like nature of coiled-coils and collagen helices, or the lath-like nature of β-hairpin ribbons, is a formula to produce highly elongated proteins in the folded state. We estimate that the central repetitive domain of each type of non-canonical silk protein is between 25 and 70 nm long in the folded state, and has an axial ratio between 14 and 40 (table 1). Thus, non-canonical silk proteins have all the necessary features required of a mesogen (i.e. axial ratio and varying hydrophilicity), and folded proteins are likely to form mesophases spontaneously when appropriately concentrated.

(b) Amino acid composition and protein folding

The propensity of amino acid residues towards particular secondary structures, as determined from solved protein structures [47], is the basis of secondary structure prediction algorithms such as GOR4 [48]. Residue folding propensities

Table 1. Dimensions of protein and polypeptide mesogens. (n.d., not determined.)

protein	predicted geometry	estimated length of mesogen (nm)	estimated width of mesogen (nm)	axial ratio	predicted threshold concentration for mesophase (v/v)[a] (%)	observed threshold concentration for mesophase (v/v) (%)	references
TMV particle	rod-like	300	18	16.7	42	2–10	[40]
F-actin tubules	rod-like	1500	7	214	3.7	8	[41]
collagen	rod-like	60	1.5	40	19	10	[42]
PBLG	rod-like	29	1.5	20	37	14	[43]
aculeate fibroin	rod-like	30.6	2.1	14.6	48	n.d.	[45]
sawfly SfColl	rod-like	67.8	1.5	37.7	20	n.d.	[28]
lacewing MalXB1	lath-like	27.8	2.8 × 0.54	16.7[b]	42	n.d.	[7]
H-fibroin	n.d.	n.d.	n.d.	approximately 5	n.d.	less than 30	[39]

[a]Using Onsager's relationship.
[b]Using average of width dimensions.

are derived from the structures of globular proteins in aqueous environments, and are therefore a better guide to the structure of proteins in silk dope than in the final solid silk. Interestingly, canonical silk proteins are usually poor in the classic β-sheet-forming residues Val, Ile, Tyr, Phe, Cys and Trp. Instead, the small residues Gly, Ala and Ser are dominant. Consistent with this discrepancy, canonical silk proteins do not fold into β-sheet-rich structures spontaneously in solution. Reconstituted silkworm silk, before exposure to mechanical shear or solvents such as MeOH, exists in the β-turn-rich silk I structure [29,49]. Owing to its amino acid composition, GOR4 predicts silkworm H-fibroin to take predominantly a random coil conformation (85.5%) with minor β-sheet (9.4%) and α-helical (5.1%) components. This can primarily be explained as reflecting the primary importance of extrinsic factors, especially mechanical deformation, for the folding of canonical silk proteins into β-sheet-rich structures.

By contrast, the residues with the highest propensities to form α-helices and coiled-coils (Glu, Lys, Leu, Arg and Ala; [47,50]) collectively make up 58–64% of the honeybee fibroins AmelF1-F4 [25]. The GOR4 algorithm predicts 70–90% of each protein to fold into α-helices, and circular dichroism experiments using recombinantly expressed proteins confirm that honeybee fibroins fold into native-like coiled-coil structures in solution [51]. Similarly, sawfly collagen fibroins have primary structural features sufficient to induce the collagen structure in recombinant silk proteins in solution [28]: the high number of collagen tripeptide repeats (Xaa–Yaa–Gly) results in an overall composition of 27–34% Gly and 8–16% Pro. Proline occurs at 38–53% of Xaa positions, orienting the protein backbone favourably for collagen folding [52]. Despite (unusually for an animal collagen) lacking hydroxyproline, the presence of hydroxylysine in the Yaa position is likely to increase thermal stability of the collagen triple helix. Lacewing egg-stalk proteins in dried silk dope show a highly ordered biaxially orientated arrangement of β-sheets [22], suggesting cross-β-sheet ribbons also form spontaneously in solution despite GOR4 predicting a predominantly random coil structure (80–95%). We conclude that there is a fundamental difference in protein folding between the non-canonical silk proteins and the canonical silk proteins of silkworms and spiders. Notably, folded mesogenic proteins in the silk glands of aculeates, sawflies and lacewings are likely to have essentially the same structure as the crystallites present in solid silk. Thus, we propose that the elongated tertiary structure of non-canonical silk proteins has convergently arisen in multiple insect groups owing to its ability to act efficiently both as a mesogen in silk dope, and as a crystallite in the solid silk.

3. Multiple trajectories through liquid crystallinity to solid silk

(a) What is the structure of the canonical silk I mesogen?

In silk glands, the protein structures that constitute mesogenic units have not been clearly characterized. However, taking again the silkworm as a model, we can place various limits on what its structure might be. For example, as the H-fibroin protein makes up the majority of the silk, and the mesogen must occupy a large volume of the dope [37], it is likely that the mesogen contains all or part of the H-fibroin protein. Possibly, the mesogenic unit may correspond to the 6 : 6 : 1 complex formed by the three main silk proteins [44], to a single H-fibroin chain or part thereof, or to an end-to-end aggregation of multiple 6 : 6 : 1 complexes [39]. At the level of secondary structure, we can rule out the possibility that the mesogen consists of β-sheets, as mesophase occurs along the length of gland duct but disappears before the silk I → silk II structural transition near the silk press [13,14]. Instead, the major secondary structure present in the mesogen is likely to be the β-turn-rich silk I.

We propose the flexible and amphiphilic nature of H-fibroin in the silk I structure, consisting of long stretches of hydrophobic repeats alternating with hydrophilic linkers [24,33], is highly suited to the formation of micelle-like lyotropic mesophases. Viney [39] argued persuasively that the low level of birefringence exhibited by mesophasic silk dope is unlikely to be due to *orientation birefringence* resulting from the long-range orientation of polarizable bonds but instead represents *form*

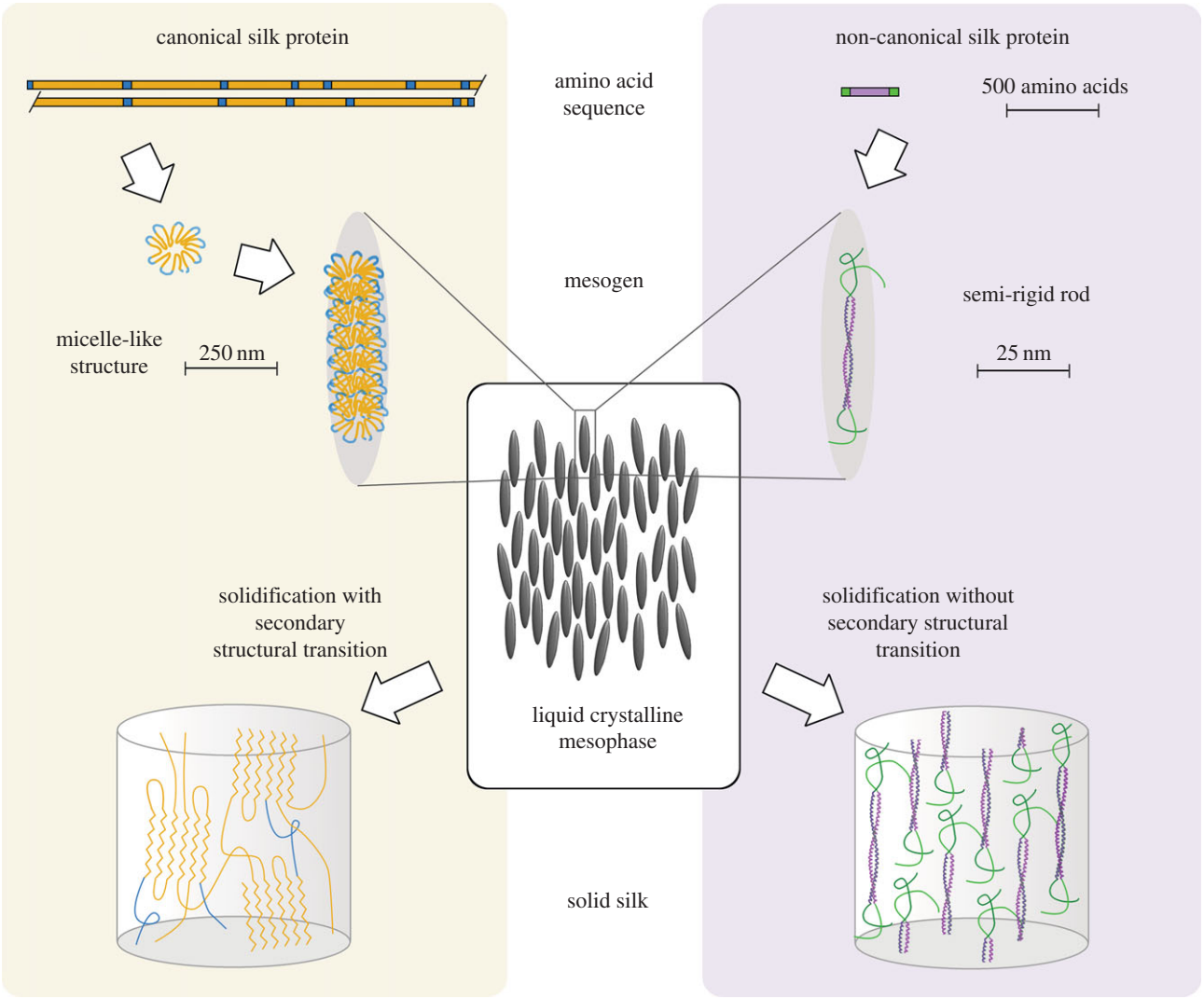

Figure 3. Comparison of liquid crystalline processing of canonical and non-canonical silk protein. Canonical silk proteins from silkworms and spiders (left) assemble into large, micelle-like mesogens by virtue of their flexibility and amphiphilicity, while non-canonical silk proteins (right) fold into comparatively small rod or lath-like mesogens with well-defined secondary structure. Liquid crystalline intermediates reduce flow viscosity, assist solubility and participate in the formation of supramolecular structure. Solidification occurs concurrently with a structural transition to the final extended-β-sheet structure for canonical silk proteins, whereas non-canonical silk proteins do not change structure markedly during solidification.

birefringence, originating from the anisotropic distribution of two coexisting phases which are themselves isotropic. In the case of H-fibroin, these two phases may be best described as: (i) a hydrophobic phase consisting of the dominant repeats of the silk protein; and (ii) a hydrophilic phase consisting of water, ions and the hydrophilic linkers of the silk protein. Anisotropy may arise if conditions inherently favour anisotropic lyotropic structures such as hexagonal columns, or alternatively if isotropic structures such as micelles aggregate through head-to-tail polymerization as has been observed under conditions of shear [53]. Either way, the mesogen probably resembles micelles observed in native and reconstituted silk [33,53] with the single added feature of elongation (figure 3). Analysis of silk protein sequences and physical characterization of silk dope and fibres suggests this model may be generalized to other groups that produce silk with extended-β-sheet crystallites, such as spiders [2] and raspy crickets [5].

(b) Multi-purpose mesogenic and crystalline protein structures in non-canonical silks

In contrast to the mechanism by which canonical silk proteins form mesogens, the best candidate for mesogens formed by non-canonical silk proteins corresponds to a single-coiled-coil superhelix, collagen triple helix or cross-β-sheet ribbon (figure 3). Non-canonical silk proteins, with a central semi-rigid superhelix and more flexible domains at each terminal, appear ideally suited to form liquid crystalline mesophases. Applying standard molecular dimensions to silk proteins produced by aculeates, sawflies and mother lacewings, the rigid portion of each molecule is predicted to have an axial ratio between 14 and 40 (table 1). According to Onsager's relationship [37], the corresponding concentration thresholds for mesophase formation fall between 19 and 50% of total volume (though experimentally measured thresholds where they are available are substantially lower). As silk proteins are known to accumulate to 30–40% of dry weight in some silk glands [54], it is reasonable to suppose non-canonical silk proteins reach these concentration thresholds.

The major conclusion of our analysis of non-canonical silk proteins (§2) is that the convergently evolved architecture of non-canonical silk proteins probably arose due to selection pressure for proteins that can act both as mesogens in the liquid phase and crystallites in the solid phase, without a major structural transition. Non-canonical silk processing thus contrasts with canonical silk processing, in which

proteins are specialized to transition between discrete mesogenic (silk I) and crystalline (silk II) forms. As each processing mode entails numerous adaptations in protein amino acid sequence and gland structure, it is likely that effective 'fitness troughs' [8] prevent their interconversion during evolution, and that they therefore represent distinct fitness peaks in the adaptive landscape of silk production.

(c) Towards an understanding of non-canonical silk processing

The potential of non-canonical silk proteins to act both as mesogens and crystallites suggests a unique silk processing pathway. It has long been recognized that materials such as insect cuticles, plant cell walls and interstitial collagens can be considered solid analogues of mesophases [38]. The same is true of non-canonical silks, but we propose it is true in the more direct sense that they are the actual product of solidifying liquid crystalline phases. For example, in aculeate and sawfly silk, protein superhelices are arranged roughly parallel to the fibre axis, but without positional orientation with respect to adjacent superhelices [4], i.e. they are solid analogues of nematic mesophases.

Based on the analysis above, we are now in a position to propose a tentative model for the process of silk spinning using non-canonical silk proteins (figure 3). Firstly, proteins folded into rod- or lath-shaped molecules accumulate in silk gland lumen. Above a critical concentration, the molecules transition into a mesophase, producing local alignment of molecules. Near the spinneret where the gland walls are close together, the average orientation of the mesogens undergoes flow-induced alignment in the direction of the spinneret, which is parallel to the eventual fibre axis. Non-canonical silk proteins are positioned closely together in mesophase, interacting via non-covalent bonds with water and solute molecules, and with adjacent proteins. As the lumen is drawn from the gland to the external air, sufficient bonds are present between adjacent proteins to overcome capillary break-up. This mechanism is consistent with the wetting properties of fibre bundles where the spreading parameter is negative, meaning bundles of fibres are naturally preferred [55]. Hence, the fibre is stable over the timescale required for it to dehydrate and for protein–protein bonds to replace protein–water bonds. After solidification, the fibre is complete and has molecular orientation that is continuous with the orientation in the preceding liquid crystal phase.

We note that this kind of fabrication is similar to the fabrication of materials such as dogfish egg-sacks [56] and mantis ootheca [57]. In both cases, concentrated protein solutions progress through distinct mesophases before solidifying as analogues of lamellar and smectic phases. Moreover, the proteins present in these materials show numerous similarities at the amino acid sequence level to non-canonical silk proteins. However, it has not previously been appreciated that fabrication of this type is a feasible route for producing solid, cylindrical, micrometre-scale fibres on demand.

(d) New artificial protein materials through biomimetic self-assembly of silk mesogens

The ability to produce recombinant versions of non-canonical silk proteins has generated interest in their use for the creation of artificial protein materials [28,35,58–60]. However, attempts to date make use of non-natural processing steps, such as solubilization in hexafluoroacetone [58] or detergent micelles [35,51,60], and solidification using methanol baths [35,60]. It is not clear what kind of liquid crystal phenomena, if any, occur during these types of material fabrication. Interestingly, precedent does exist for the formation of artificial fibres using folded, rod-shaped proteins in mesophase. Over the past two decades, systems have been developed in which the mesogenic properties of elongated virus capsules direct the self-assembly of hierarchically structured materials [61]. For example, an engineered ZnS-binding TMV was induced by Lee and co-workers to form smectic phases that were then solidified, allowing the production of a highly ordered material with regularly spaced metallic particles [62].

Further advantages may be gained on these already sophisticated fabrication systems if it is possible to harness the natural and straightforward mesogenic behaviour that we suggest characterizes non-canonical silk proteins. For example, aculeate silk proteins are around 30 nm long (cf. 300 nm for the TMV) and hence might be used to introduce structure at finer scales compared with virion mesogens. In addition, non-canonical silk proteins are easily engineered, express at high levels and are easily refolded into native-like conformations [28,35,60]. However, the greatest advantage of non-canonical silk proteins over alternative protein mesogens, such as virus particles, is that liquid crystalline processing into solid materials is their natural function. Accordingly, their primary sequences are likely to contain numerous adaptations predisposing them to act as mesogens in the liquid state, and strong and flexible structural proteins in the solid state.

4. Conclusion

Silk proteins that form crystallites with structures other than extended-β-sheets have convergently evolved features that distinguish them from canonical silk proteins made by silkworms and spiders. Non-canonical silk proteins are characterized by relatively small size (less than 86 kDa) and high-repeat continuity within the central repetitive domain. This primary structure yields a rod-like or lath-like tertiary structure 23–70 nm in length, with more flexible domains at each end. Importantly, this structure is likely to be formed spontaneously in solution without reliance on extrinsic factors.

We propose the evolutionary convergence of this rod/lath-like structure in non-canonical silk proteins is the result of selection for proteins that can act efficiently both as mesogens in the liquid state and crystallites in the solid state, without an intervening transition in secondary or tertiary structure. This lies in stark contrast to canonical silk proteins, which are highly specialized to both forms, mesogens and crystallites, but which undergo extensive changes at the level of protein structure to transition between the two forms. We propose a simple qualitative model of silk fabrication using non-canonical silk proteins that further suggests the molecular orientation in the solid silk is continuous with molecular orientation in the liquid crystal state, i.e. that non-canonical silks are fabricated by solidification of mesophases. Recombinant non-canonical silk proteins from aculeates and sawflies, and structural analogues of lacewing egg-stalk proteins, have been used to make biomimetic materials but processing so far only poorly mimics natural spinning

and hence is unlikely to capture the full potential of these proteins for assembling hierarchical structures. We conclude that further investigation of liquid crystalline mesophases formed by recombinant non-canonical silk proteins is likely to yield further advances towards the creation of sophisticated protein materials.

Authors' contributions. A.A.W. conceived the central idea of the manuscript, which was further developed and analysed by A.A.W., T.D.S. and C.H. All authors wrote and approved the final document.

Competing interests. We declare we have no competing interests.

Funding. This study was enabled by funding support from the CSIRO and from the Australian National University.

References

1. Sutherland TD, Young JH, Weisman S, Hayashi CY, Merritt DJ. 2010 Insect silk: one name, many materials. *Annu. Rev. Entomol.* **55**, 171–188. (doi:10.1146/annurev-ento-112408-085401)
2. Bini E, Knight DP, Kaplan DL. 2004 Mapping domain structures in silks from insects and spiders related to protein assembly. *J. Mol. Biol.* **335**, 27–40. (doi:10.1016/j.jmb.2003.10.043)
3. Fu C, Shao Z, Vollrath F. 2009 Animal silks: their structures, properties and artificial production. *Chem. Commun.* **2009**, 6515–6529. (doi:10.1039/B911049F)
4. Rudall KM. 1962 Silk and other cocoon proteins. In *Comparative biochemistry* (eds M Florkin, HS Mason), pp. 397–433. New York, NY: Academic Press.
5. Walker AA, Weisman S, Church JS, Merritt DJ, Mudie ST, Sutherland TD. 2012 Silk from crickets: a new twist on spinning. *PLoS ONE* **7**, e30408. (doi:10.1371/journal.pone.0030408)
6. Kameda T, Nemoto T, Ogawa T, Tosaka M, Kurata H, Schaper AK. 2014 Evidence of α-helical coiled coils and β-sheets in hornet silk. *J. Struct. Biol.* **185**, 303–308. (doi:10.1016/j.jsb.2013.12.005)
7. Weisman S, Okada S, Mudie ST, Huson MG, Trueman HE, Sriskantha A, Haritos VS, Sutherland TD. 2009 Fifty years later: the sequence, structure and function of lacewing cross-β silk. *J. Struct. Biol.* **168**, 467–475. (doi:10.1016/j.jsb.2009.07.002)
8. Kauffman S. 1987 Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.* **128**, 11–45. (doi:10.1016/S0022-5193(87)80029-2)
9. Tokareva O, Michalczechen-Lacerda VA, Rech EL, Kaplan DL. 2013 Recombinant DNA production of spider silk proteins. *Microb. Biotechnol.* **6**, 651–663. (doi:10.1111/1751-7915.12081)
10. Laity PR, Gilks SE, Holland C. 2015 Rheological behaviour of native silk feedstocks. *Polymer* **67**, 28–39. (doi:10.1016/j.polymer.2015.04.049)
11. Holland C, Terry AE, Porter D, Vollrath F. 2006 Comparing the rheology of native spider and silkworm spinning dope. *Nat. Mater.* **5**, 870–874. (doi:10.1038/nmat1762)
12. Kojic N, Bico J, Clasen C, McKinley GH. 2006 *Ex vivo* rheology of spider silk. *J. Exp. Biol.* **209**, 4355–4362. (doi:10.1242/jeb.02516)
13. Rey AD, Herrera-Valencia EE. 2012 Liquid crystal models of biological materials and silk spinning. *Biopolymers* **97**, 374–396. (doi:10.1002/bip.21723)
14. Asakura T, Umemura K, Nakazawa Y, Hirose H, Higham J, Knight DP. 2007 Some observations on the structure and function of the spinning apparatus in the silkworm *Bombyx mori*. *Biomacromolecules* **8**, 175–181. (doi:10.1021/bm060874z)
15. Knight DP, Vollrath F. 1999 Liquid crystals and flow elongation in a spider's silk production line. *Proc. R. Soc. Lond. B* **266**, 519–523. (doi:10.1098/rspb.1999.0667)
16. Vollrath F, Knight DP. 2001 Liquid crystalline spinning of spider silk. *Nature* **410**, 541–548. (doi:10.1038/35069000)
17. Donald AM, Windle AH, Hanna S. 1992 *Liquid crystalline polymers*, 2nd edn. Cambridge, UK: Cambridge University Press.
18. Holland C, Terry AE, Porter D, Vollrath F. 2007 Natural and unnatural silks. *Polymer* **48**, 3388–3392. (doi:10.1016/j.polymer.2007.04.019)
19. Mykhaylyk OO. 2010 Time-resolved polarized light imaging of sheared materials: application to polymer crystallization. *Soft Matter* **6**, 4430–4440. (doi:10.1039/c0sm00332h)
20. Holland C, Vollrath F, Ryan AJ, Mykhaylyk OO. 2012 Silk and synthetic polymers: reconciling 100 degrees of separation. *Adv. Mater.* **24**, 105–109. (doi:10.1002/adma.201103664)
21. Flower NE, Kenchington W. 1967 Studies on insect fibrous proteins: the larval silk of *Apis*, *Bombus* and *Vespa*. *J. R. Micr. Soc.* **86** 297–310. (doi:10.1111/j.1365-2818.1967.tb00589.x)
22. Lucas F, Rudall KM. 1968 Extracellular fibrous proteins: the silks. In *Compr biochem* (eds M Florkin, EH Stotz), pp. 475–558. Amsterdam, The Netherlands: Elsevier.
23. Silva-Zacarin EC, Silva De Moraes RL, Taboga SR. 2003 Silk formation mechanisms in the larval salivary glands of *Apis mellifera* (Hymenoptera: Apidae). *J. Biosci.* **28**, 753–764. (doi:10.1007/BF02708436)
24. Foo C, Wong P, Bini E, Hensman J, Knight DP, Lewis RV, Kaplan DL. 2006 Role of pH and charge on silk protein assembly in insects and spiders. *Appl. Phys. A Mater. Sci.* **82**, 223–233. (doi:10.1007/s00339-005-3426-7)
25. Sutherland TD, Campbell PM, Weisman S, Trueman HE, Sriskantha A, Wanjura WJ, Haritos VS. 2006 A highly divergent gene cluster in honey bees encodes a novel silk family. *Genome Res.* **16**, 1414–1421. (doi:10.1101/gr.5052606)
26. Chen X, Knight DP, Shao Z, Vollrath F. 2001 Regenerated *Bombyx* silk solutions studied with rheometry and FTIR. *Polymer* **42**, 9969–9974. (doi:10.1016/S0032-3861(01)00541-9)
27. Kameda T, Kojima K, Zhang Q, Sezutsu H, Teramoto H, Kuwana Y, Tamada Y. 2008 Hornet silk proteins in the cocoons produced by different *Vespa* species inhabiting Japan. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **151**, 221–224. (doi:10.1016/j.cbpb.2008.07.004)
28. Sutherland TD *et al*. 2013 A new class of animal collagen masquerading as an insect silk. *Sci. Rep.* **3**, 2864. (doi:10.1038/srep02864)
29. Asakura T, Yamane T, Nakazawa Y, Kameda T, Ando K. 2001 Structure of *Bombyx mori* silk fibroin before spinning in solid state studied with wide angle X-ray scattering and ^{13}C cross-polarization/magic angle spinning NMR. *Biopolymers* **58**, 521–525. (doi:10.1002/1097-0282(20010415)58:5<521::AID-BIP1027>3.0.CO;2-T)
30. Boulet-Audet M, Terry AE, Vollrath F, Holland C. 2014 Silk protein aggregation kinetics revealed by Rheo-IR. *Acta Biomater.* **10**, 776–784. (doi:10.1016/j.actbio.2013.10.032)
31. Holland C, Urbach JS, Blair DL. 2012 Direct visualization of shear dependent silk fibrillogenesis. *Soft Matter* **8**, 2590–2594. (doi:10.1039/C2SM06886A)
32. Holland C, O'Neil K, Vollrath F, Dicko C. 2012 Distinct structural and optical regimes in natural silk spinning. *Biopolymers* **97**, 368–373. (doi:10.1002/bip.22022)
33. Jin HJ, Kaplan DL. 2003 Mechanism of silk processing in insects and spiders. *Nature* **424**, 1057–1061. (doi:10.1038/nature01809)
34. Bauer F, Bertinetti L, Masic A, Scheibel T. 2012 Dependence of mechanical properties of lacewing egg stalks on relative humidity. *Biomacromolecules* **13**, 3730–3735. (doi:10.1021/bm301199d)
35. Sutherland TD, Church JS, Hu X, Huson MG, Kaplan DL, Weisman S. 2011 Single honeybee silk protein mimics properties of multi-protein silk. *PLoS ONE* **6**, e16489. (doi:10.1371/journal.pone.0016489)
36. Viney C. 1997 Liquid crystalline phase behaviour of proteins and polypeptides. In *Protein materials* (eds K McGrath, DL Kaplan), pp. 281–311. Boston, MA: Birkhauser.
37. Onsager L. 1949 The effects of shape on the interaction of colloidal particles. *Ann. NY Acad. Sci.* **51**, 627–659. (doi:10.1111/j.1749-6632.1949.tb27296.x)
38. Donald AM, Windle AH, Hanna S. 2006 *Liquid crystalline polymers*, 2nd edn. Cambridge, UK: Cambridge University Press.
39. Viney C. 1997 Natural silks: archetypal supramolecular assembly of polymer fibres. *Supramol. Sci.* **4**, 75–81. (doi:10.1016/S0968-5677(96)00059-4)
40. Fraden S, Maret G. 1993 Angular correlations and the isotropic-nematic phase transition in

suspensions of tobacco mosaic virus. *Phys. Rev. E* **48**, 2816–2838. (doi:10.1103/PhysRevLett.63.2068)

41. Suzuki A, Maeda T, Ito T. 1991 Formation of liquid crystalline phase of actin filament solutions and its dependence on filament length as studied by optical birefringence. *Biophys. J.* **59**, 25–30. (doi:10.1016/S0006-3495(91)82194-4)
42. Giraud-Guille MM. 1989 Liquid crystalline phases of sonicated type I collagen. *Biol. Cell* **67**, 97–101. (doi:10.1111/j.1768-322X.1989.tb03014.x)
43. Robinson C. 1961 Liquid crystalline structure in polypeptide solutions. *Tetrahedron* **13**, 219–234. (doi:10.1039/DF9582500029)
44. Inoue S, Tanaka K, Arisaka F, Kimura S, Ohtomo K, Mizuno S. 2000 Silk fibroin of *Bombyx mori* is secreted, assembling a high molecular mass elementary unit consisting of H-chain, L-chain, and P25, with a 6:6:1 molar ratio. *J. Biol. Chem.* **275**, 40 517–40 528. (doi:10.1074/jbc.M006897200)
45. Sutherland TD, Weisman S, Trueman HE, Sriskantha A, Trueman JW, Haritos VS. 2007 Conservation of essential design features in coiled coil silks. *Mol. Biol. Evol.* **24**, 2424–2432. (doi:10.1093/molbev/msm171)
46. Delorenzi M, Speed T. 2002 An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* **18**, 617–625. (doi:10.1093/bioinformatics/18.4.617)
47. Chou PY, Fasman GD. 1974 Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins. *Biochemistry* **13**, 211–222. (doi:10.1021/bi00699a001)
48. Garnier J, Gibrat J-F, Robson B. 1996 GOR secondary structure prediction method version IV. *Methods Enzymol.* **266**, 540–553. (doi:10.1016/S0076-6879(96)66034-0)
49. Zhang C, Song D, Lu Q, Hu X, Kaplan DL, Zhu H. 2012 Flexibility regeneration of silk fibroin *in vitro*. *Biomacromolecules* **13**, 2148–2153. (doi:10.1021/bm300541g)
50. Gromiha MM, Parry DAD. 2004 Characteristic features of amino acid residues in coiled-coil protein structures. *Biophys. Chem.* **111**, 95–103. (doi:10.1016/j.bpc.2004.05.001)
51. Walker AA, Warden AC, Trueman HE, Weisman S, Sutherland TD. 2013 Micellar refolding of coiled-coil honeybee silk proteins. *J. Mater. Chem. B* **1**, 3644–3651. (doi:10.1039/C3TB20611D)
52. Bachmann A, Kiefhaber T, Boudko S, Engel J, Bachinger HP. 2005 Collagen triple-helix formation in all-trans chains proceeds by a nucleation/growth mechanism with a purely entropic barrier. *Proc. Natl Acad. Sci. USA* **102**, 13 897–13 902. (doi:10.1073/pnas.0505141102)
53. Greving I, Cai M, Vollrath F, Schniepp HC. 2012 Shear-induced self-assembly of native silk proteins into fibrils studied by atomic force microscopy. *Biomacromolecules* **13**, 676–682. (doi:10.1021/bm201509b)
54. Akai H. 1983 The structure and ultrastructure of the silk gland. *Experientia* **39**, 443–449. (doi:10.1007/BF01965158)
55. Rey AD, Golmohammadi M, Valencia EEH. 2011 A model for mesophase wetting thresholds of sheets, fibers and fiber bundles. *Soft Matter* **7**, 5002–5009. (doi:10.1039/c1sm05113j)
56. Knight DP, Feng D, Stewart M, King E. 1993 Changes in macromolecular organization in collagen assemblies during secretion in the nidamental gland and formation of the egg capsule wall in the dogfish *Scyliorhinus canicula*. *Phil. Trans. R. Soc. Lond. B* **341**, 419–436. (doi:10.1098/rstb.1993.0125)
57. Walker AA, Weisman S, Kameda T, Sutherland T. 2012 Natural templates for coiled-coil biomaterials from praying mantis egg-cases. *Biomacromolecules* **13**, 4264–4272. (doi:10.1021/bm301570v)
58. Bauer F, Scheibel T. 2012 Artificial egg stalks made of a recombinantly produced lacewing silk protein. *Angew. Chem. Int. Ed.* **51**, 6521–6524. (doi:10.1002/anie.201200591)
59. Kambe Y, Sutherland TD, Kameda T. 2014 Recombinant production and film properties of full-length hornet silk proteins. *Acta Biomater.* **10**, 3590–3598. (doi:10.1016/j.actbio.2014.05.013)
60. Weisman S, Haritos VS, Church JS, Huson MG, Mudie ST, Rodgers AJW, Dumsday GJ, Sutherland TD. 2011 Honeybee silk: recombinant protein production, assembly and fiber spinning. *Biomaterials* **31**, 2695–2700. (doi:10.1016/j.biomaterials.2009.12.021)
61. Flynn CE, Lee SW, Peelle BR, Belcher A. 2003 Viruses as vehicles for growth, organization and assembly of materials. *Acta Mater.* **51**, 5867–5880. (doi:10.1016/j.actamat.2003.08.031)
62. Lee SW, Mao C, Flynn C, Belcher A. 2002 Ordering of quantum dots using genetically engineered viruses. *Science* **296**, 892–895. (doi:10.1126/science.1068054)