Silk Road Revealed: Mechanism of silk fibre formation in *Bombyx mori*

R.O. Moreno-Tortolero,[ab*] Y. Luo,[c] F. Parmeggiani, [ac] N. Skaer,[d] R. Walker,[d] L. Serpell,[e] C. Holland,[f] S.A. Davis[a*]

(a) School of Chemistry, University of Bristol (UoB) Cantock's Cl, Bristol BS8 1TS, UK; (b) Max Planck-Bristol Centre for Minimal Biology, School of Chemistry, University of Bristol, Bristol, United Kingdom; (C) School of Biochemistry, UoB University Walk, Bristol BS8 1TD, UK; (d) Orthox LTD Milton Park, 66 Innovation Dr, Milton, Abingdon OX14 4RT, UK; (e) Sussex Neuroscience, School of Life Sciences, University of Sussex Falmer Brighton, BN1 9RH, UK; (f) Dept. Materials Science and Engineering, University of Sheffield Mappin St, Sheffield City Centre, Sheffield S1 3JD, UK.

*e-mail: ro.morenotortolero@bristol.ac.uk, s.a.davis@bristol.ac.uk

Abstract

Silk fibroin (SF) is one of the most exploited and studied natural materials, yet the complex sequence of dynamic transitions required to transform from a liquid state to a solid fibre remain unresolved. We found that fibroin heavy chain (FibH) from *Bombyx mori* (*B. mori*), a multidomain protein with highly repetitive domains, folds to form β-solenoid structures. The single fibroin molecule undergoes conformational transformations from a globular form to an extended solenoid driven by pH reduction. Following this, the N-terminal domain (NTD) drives higher-order oligomerisation beyond the proposed biological tetrameric unit when in the extended state.[1,2] The proposed supramolecular structures explain the complex rheological behaviour observed for silk and the liquid crystalline textures observed within the *B. mori* gland. Overall, our results demonstrate that silk assembly is a highly organised process, where a mild reduction in pH promotes molecular reorganisation and higher-order oligomerisation of the N-terminal domain into structures that precede fibre formation. Moreover, we suggest that NTD is at least partially hydrolysed and lost during the standard regeneration/reconstitution processing, explaining the historical differences between native and regenerated feedstocks. We believe our results expand our understanding of the liquid-to-solid transformation of fibroin molecules, as well as indicate to other researchers a clear strategy to determine the quality of their feedstock, which can be translated into improved materials fabrication.

Main

Lepidopteran silks are non-woven composite materials made predominantly of two protein families, fibroins which form twin reinforcement fibres (brins), and sericins, a glue-like matrix fibre coating. This coating has historically been removed to enable unravelling for textile applications as well as in preparation for biomedical use under the guise of biocompatibility, however this degumming process typically induces considerable collateral hydrolysis.[3]

Within the larval insect, fibroin is stored as a liquid, sustaining a triggered transformation to a very stable solid material, the silk fibre. This transformation starts within the silk gland, which is a highly specialised organ.[4] However, the mechanism of this transformation at the molecular level remains largely unknown. Whilst there are some commonly agreed principles, including the contribution of pH and shear stress/strain in triggering assembly and promoting structural conformational changes,[5,6] there is less consensus regarding the conformational changes at the molecular level. In particular, the molecular structure of the liquid state silk, often called silk-I, is highly debated.[7–11] Silk-I used to be called α-silk, as initially incorrectly assumed to include α-helical folds, to distinguish between this

conformational polymorph and the better characterised Silk-II, which is known to be rich in antiparallel β-sheet structure.[12] Others classify silk-I as random coil or an intrinsically disordered state. However, discrete reflections in X-ray diffraction suggest otherwise,[13,14] although intrinsically disordered domains are likely to be present.

The most accepted Silk I model is a rippled sheet or a type-II β-turn rich structure.[15] However this model requires intermolecular hydrogen bonding networks to persist,[16] failing to explain the presence of an otherwise identical conformation in dilute solutions and in the solid state measured by NMR, as well as small fibrillar structures observed herein and previously. Modelling of consensus sequences has indicated type-II β-turn and β helix as potentially alternative structures, with a right-handed β-helix showing lower relative free-energy values than the corresponding type-II β-turn.[16]

Similarly, debate also surrounds the larger scale nature of silks' mesogen that produces the range of distinct liquid crystalline optical textures observed *in vivo* between the anterior section of the middle gland and the spinneret.[17] Silk goes from an isotropic texture in the posterior and middle gland sections to a series of complex transitioning optical textures from the anterior section of the gland towards the spinneret.[18] At the start of the anterior section of the silk gland, a cellular optical texture is observed, which transforms to an isotropic phase prior to reverting to a fully nematic phase before the spinneret.[18] The emergence of the cellular optical texture has been attributed to epitaxial anchoring of the mesogen under confined flow,[19] but this does not adequately account for the subsequent transition to a nematic texture, under flow, as the tube diameter in the gland decreases towards the spinneret. Curiously, at a similar position to the cellular optical texture, evidence of cholesteric order have been observed using electron microscopy.[20] More importantly, despite the evidence of supramolecular order, a transition from Silk I to Silk II polymorph only occurs later near the spinneret.[18,20]

Fibroin, both within the gland and in the fibre, is composed of three different proteins, fibroin heavy chain (FibH), fibroin light chain (FibL) and fibrohexamerin or P25 of molecular weights 392, 28 and 25 KDa, respectively. These form a complex known as the elementary unit with molar ratios of 6:6:1, wherein FibH and FibL form a heterodimer stabilized by a single disulphide linkage located at the C-terminal domain of FibH and the interaction between these and P25 is non-covalent.[21,22] However, the structural details of this system are not well known, except for the crystal structure of the N-terminal domain of FibH. Interestingly, the roles of FibL and P25 are unclear.

While the exact mechanism of silk fiber formation in silkworms remains uncertain, it is generally accepted that only FibH is essential for fibre formation, with FibL and P25 being auxiliary in the secretion process;[23–25] some silk-producing moths and other related species lack these latter proteins entirely.[26–28] Accordingly, due to its dominant mass contribution, FibH is thought to be the main protein responsible for the properties of fibroin across different length-scales, from the NMR chemical shifts,[29] to the aforementioned diffraction patterns of Silk-I and Silk-II.[30]

In this study, we propose that at the molecular level Silk-I actually corresponds to a folded fibrillar conformation of FibH. Furthermore, we rationalise how physicochemical triggers such as pH, metal ions, and stress/strain induce conformational changes in this mesogen that result in concomitant supramolecular assembly or reorganization consistent with the observed optical textures. The abrupt loss of the cellular texture derives from disruption of the preceding cholesteric phase, and is a key intermediate to facilitate the minimum energy transition to the ordered nematic phase, as previously proposed.[18,20]

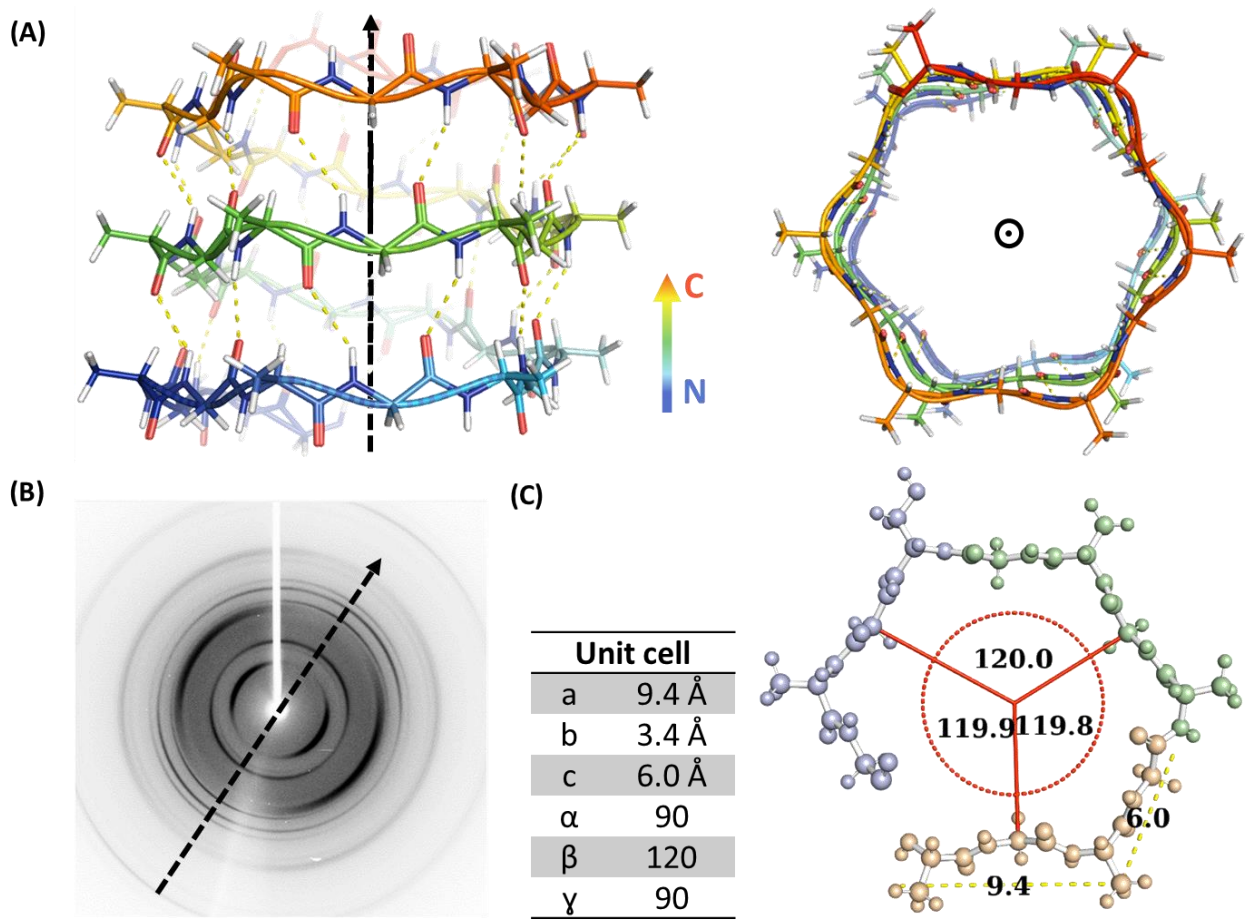
To investigate this hypothesis, we employed a comprehensive approach combining both computational simulations and experimental

methods, encompassing various length scales ranging from individual molecules to macroscopic properties. Our methodology involved the utilization of a fibroin solution achieved by dissolving silk fibres in concentrated LiBr solutions, known as regeneration, resulting in a material referred to as native-like silk fibroin (NLSF). This NLSF exhibits numerous characteristics like those of the native silk fibroin (NSF), including apparent mass and rheological properties. It is important to differentiate NLSF from standard regenerated silk fibroin (RSF), which, although it can be obtained in larger quantities than NSF, fails to accurately replicate the rheological and mass properties of the native system.

Molecular model creation and validation

FibH is a highly repetitive protein, with long low complexity domains (see SD.1) that can be described mostly by a hexapeptide of the form (GAGAGX), where X can be any of the residues A, S, Y or V in that order of abundance. Using the Alphafold prediction algorithm, implemented via ColabFold,[31–33] we found that these repetitive domains are predicted to adopt a β-solenoid conformation. Albeit the predictions showed low confidence, and relatively high predicted Local Distance Difference Test (pLDDT) in average of about 50, it was noteworthy that all generated models maintained the solenoidal topology (SD.2). Briefly, the β-solenoid fold results from a supersecondary folding of strands as depicted in Figure 1.A, although in this case the abundance of glycine and the absence of a tightly packed core lead to motifs structurally closer to a polyproline-II or polyglycine-II configuration, which can be frequently assigned as disordered by CD spectroscopy.[34] Similar β-helical configurations have been suggested before, mostly via computational approaches, but have not reached mainstream acceptance.[10,16] The fold is stabilised by inter-strand hydrogen bonds running along the solenoid axis.

Dilute solutions of NLSF were cast and left to dry slowly, to obtain the expected Silk-I diffraction pattern, with the films being characterised by fibre X-ray diffraction experiments (see Fibre X-ray diffraction in Materials and Methods in ESI). Since the predicted model of the protein is fibrillar, we hypothesise the molecule would be preferentially oriented parallel to the surface of the film.[35] The experimentally acquired X-ray diffraction (Figure 1.B) shows a high degree of orientation and is consistent both with previous Silk-I data,[14,36] and our newly proposed solenoid model. Here, we propose a trigonal unit cell containing a single curved strand, as depicted in Figure 1.C; unit cell parameters are a: 9.4, b: 3.4 and c: 6 Å. We note that the proposed unit cell is similar to that proposed for the PolyGly-II polymorph,[37] known to resemble Silk-I.[15] A detailed discussion of these results is provided in SD.3.



Unit cell	
a	9.4 Å
b	3.4 Å
c	6.0 Å
α	90
β	120
γ	90

Figure 1. (A) β-solenoid molecular model obtained from fibroin heavy chain, observed perpendicular to its screw axis, left, and parallel to it, right. (B) Experimental Silk-I diffraction pattern obtained from drop-casted film showing the proposed fibre axis. (C) Molecular detail of a β-solenoid domains, with the estimated hexagonal geometry and unit cell representation.

Furthermore, the measured NMR chemical shifts (^{1}H, ^{13}C and ^{15}N) were very close to those reported elsewhere,[38,39] and predicted shifts from the solenoid models (see table S3 in SI). In addition, we note that when predicting the torsion angles for reduced motifs in FibH, two possible populations for G residues are possible within generic GX motif (see Figure S16). One of these (φ, ϕ = 77°, 10°) was used to derive the proposed type-II β-turn model for Silk-I.[29] However, the second population of torsion angles (ca. φ, ϕ = -80°, 180°) was not used to constrain the modelling and matches closely those found in the solenoidal model predicted by ColabFold as well as previously proposed angles from density functional theory (DFT) calculations.[11,40] This conformation implies that in the transformation from Silk-I to Silk-II, G residues would not need to go through sterically hindered angles (see Figure S17),[41] thus reducing the overall energetic cost of the Silk-I to Silk-II transformation. Further to these observations, NOESY experiments suggest Y residues close to the methyl sidechain of A which we believe could be an inter-strand (within the solenoid) stabilising interaction. There are numerous Y residues on the predicted models, with the aromatic group hovering over methyl groups from A residues on contiguous strands (see Figure S13). For further discussion on NMR data, we point the readers to SD.4.

pH Effect: rheological effect of NTD

It has largely been recognised that in native spinning both shear stress and pH changes are responsible for the transformation of the soluble fibroin into the solid fibre.[5,42] Rheological characterisation of native feedstock

has provided snapshots into the strain-induced assembly. In most studies the silk dope is extracted from the posterior segment of the middle gland, where the pH is near neutral.[43,44] However, direct evidence of the structural changes induced by pH and their effects on rheological properties are missing.

Here we used NLSF samples obtained at concentrations of 60 to 80 mg/mL, lower than those found inside the gland (190-300 mg/mL),[44] buffered at different pH values to replicate the pH gradient within the silk gland, going from pH 8 in the posterior section down to 6.2 in the anterior section.[45] We observed that the material goes through a reversible gel-to-sol transition when buffered above pH 7, with the sample at pH 7 separating into two clear phases, gel-like and liquid-like. Single phases were observed for samples either below or above pH 7. Using NLSF at pH 7, under oscillatory rheology testing we were able to replicate many previous features seen in NSF (i.e. observation of crossover), with the exception of a lower overall modulus and viscosity and a lack of a single relaxation mode in the terminal region, both of which we attribute to a lower concentration and increased polydispersity of the NLSF against NSF studies (See SD.5).[44] Furthermore, we were able to uncover a reversible pH driven sol-gel transition occurring at a critical pH of 7, similar to the pH-dependent oligomerisation of NTD.[1] Figure 2.A shows two master curves demonstrating the effect of pH on the obtained material; master curves were created from the data shown in Figures S20 and the applied shifts are presented in Table S4. Decreasing the pH further turns the gel opaque irreversibly, providing evidence of larger aggregates forming. However, no evidence suggests that the pH drops below 6 within the gland.

For samples below pH 7, the observed shear modulus showed an almost linear dependency with pH, increasing as the pH decreased. This suggests a more stable molecular network is formed as the pH decreases, evidenced by increased elastic modulus. Crucially, NLSF at pH 7 (liquid-like) showed curves like those reported for the extracted liquid SF from the gland, suggested to be aquamelts, due to their similarity to typical Maxwellian polymer melts.[46] Notably, the cross-over frequency seemed to be shifted towards higher frequencies as pH increased, indicating a shortening in the lifespan of the interaction. Again, at pH lower than 7, the observation of gel phases indicates long-lived and increased stability of the formed structures. In all cases, given that the lineal viscoelastic limit (LVE) of the material was not exceeded, samples remained clear, and repeated cycles showed similar behaviour.

Similarly, when measuring viscosity against shear rate for samples at pH 7, the material undergoes shear-thinning until the shear viscosity increases, preceded by a normal force increase, resulting in fibrous aggregates akin to NSF (see Figure S18).[47] Interestingly, under similar experimental conditions, this transition seems to be hindered at higher pH values (something also seen for native silks and our samples in Lithium Bromide),[48] in congruence with the dominant liquid character and the decreased stability of the interaction that gives rise to the slowest relaxation mode.[49]

The proximity of the critical pH observed here with that reported for NTD (below pH 7),[1] strongly hints at this domain being involved in the observed structural transitions. Furthermore, the transition is also reflected in an earlier onset of aggregation observed at pH lower than 7 when temperature ramps are conducted on CD and DLS, see SD.6 and SD.7, respectively. As a consequence, we hypothesise that the absence of NTD would render the material pH-insensitive within the range (pH 8-6) and such a rheological transition would not be seen.

As a means of comparison and in support of our use of NLSF as a model system for NSF, RSF was also studied for its rheological behaviour and response to pH changes. It is well known that silk fibroin behaves very differently after standard regeneration/reconstitution (RSF).[50,51] For example, when shear is applied to NSF extracted directly from the posterior part of the middle

gland, higher-order assembly into nanofibrils is observed, contrary to reconstituted SF at a similar pH of 7.[5,52,53] Similarly, reversible sol-gel transitions have been observed only for NSF when changing pH,[42] while our RSF is insensitive to pH in the studied range. It is possible to promote assembly into nanofibrillar structures in RSF, but only when the pH is decreased to near the theoretical isoelectric point (PI) of FibH (ca. 4.7),[54] well below the physiological pH of the gland.[50,52,55]

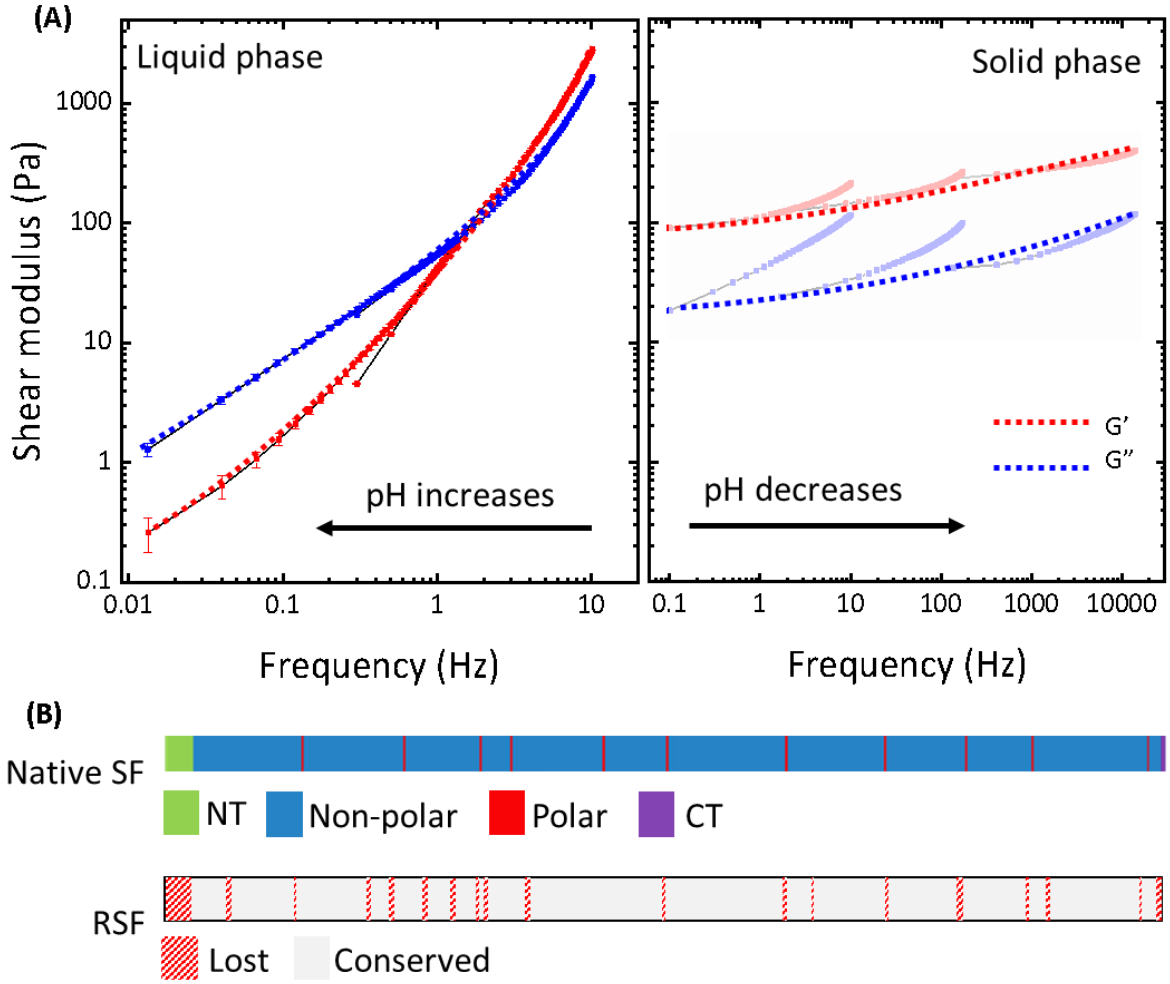


Figure 2. Master curves created using oscillatory shear data of the liquid and solid-like samples shown on the left and right, respectively (A). Carton showing the multidomain architecture of the primary structure of FibH (top), followed by an aligned cartoon representing the total sequence coverage from the hydrolysate peptides recovered from RSF (B)

Many authors argue that the molecular weight reduction in RSF is solely responsible for these observations.[56] However, we considered this reduction insufficient to explain the significant differences in literature found viscosities, or the complete insensitivity to pH (between pH 8 and 6) that we observed. Reported viscosities obtained from NSF are far superior to those reported for RSF even at similar weighted concentrations (between 500-4000 Pa.s for SF and between 0.1-1 Pa.s RSF).[44,57] Such a significant difference - 4 orders of magnitude- is not substantiated by the molecular weight difference, which on average only falls by approximately half after 30 mins of boiling during degumming.[58] This reduction in MW would be translated in a proportional reduction of ~0.088, close to one order of magnitude drop in viscosity (3 orders of magnitude away from the experimental difference). More details on the calculations and discussion behind this point can be found in SD.8.

Thus, we hypothesised that RSF is not just becoming lower MW but is also likely losing its NTD, which as noted previously, is highly implicated in the improved rheological properties of NLSF and NSF. To prove this hypothesis, we conducted proteomics analysis on the peptide fragments (0.1-6 kDa) obtained directly from the liquid solutions of both RSF and NLSF. It was noteworthy that although substantial material was obtained from RSF, no material was recovered from NLSF. Together, the SDS-PAGE results obtained from NLSF and this observation would indicate a near-native MW and the likely presence of all its functional domains. On the other hand, LC-MS/MS obtained from the RSF fragments found a total of 113 peptides matched the provided sequences (FibH, FibL and P25). With 35 matching FibH for 10.17 % coverage, 52 peptides FibL for a 96.56 % coverage, and 26 matched P25 with total coverage of 89.09 % (see Figure S 27, A, B and C). For brevity, the following discussion will be focused on FibH, given the predominant role of this protein in the system. However, the high coverage of both FibL and P25 indicates a high degree of hydrolysis in RSF.[58] Of the 35 detected peptides from FibH, 22 belong to the first 150 amino acids; 2 of these have a 100 % confidence match, with other 3 high confidence matches with 81, 78, and 60 % of confidence. The higher number of matched fragments and their higher confidence alludes to preferential cleavages at the NTD (Figure S24, D and E summarise these results). Although the results do not allow us to unambiguously determine NTD as the sole determinant for the transitions, as p25 and FibL could play a role in assembly, the high degree of homology across all *Dytrisia,*[1,59,60] and the absence of the latter two proteins in Saturnids[26,61,62] reinforces the idea that this domain is critical in driving assembly.

Beyond this observation, the complete sequence coverage plot indicates that most of the matched peptides correspond to the more hydrophilic segments of the chain (see Figure S28 A), i.e. hydrolysis during degumming in Na_2Co_3 is limited by accessibility. Furthermore, besides G, the most abundant residue in FibH, S and T are particularly enriched at the terminal positions suggesting possible intramolecular nucleophilic attack under the alkaline conditions. Other amino acids such as N, K, and E were found at the terminal position of the peptides. Curiously and regardless of its low abundancy, N (only 20 in FibH) was found prominently at the C-terminal of the fragments, with S and T more so at the N-terminal positions (see sequence logo in Figure S28.B). Overall, during degumming induced scission is likely led mainly by S residues, as suggested before,[63] but other hydrolysis mechanisms are also occurring. The hydrophilic spacers and termini are richer in polar amino acids, often considered better nucleophiles in aqueous conditions. In particular, residues R, H, and K are only found in the terminal domains, and E and D both in the spacers and terminals (see Figure S 28).

Mechanism of fibre formation

Despite the significant rheological differences observed with pH changes for NLSF, we did not observe clear differences in the NMR and CD data in the studied pH range, suggesting that most of the secondary structures of the protein remain unchanged. To better understand the system, we then conducted pH-controlled TEM at pH 8 and 6, above and below the identified transition. Here we observed that the protein appeared as globules with an apparent size of about 24 ±8 nm at pH 8, but elongated fibrillar structures were observed at pH 6 (SD.10). Although these fibrillar structures were slightly thicker than the expected solenoid (6 ± 1 nm against ca. 3 nm), acquiring higher resolution images proved difficult given the high sensitivity of the samples to beam damage. The observation of no change in NMR and CD data with pH, but such dramatic morphological change observed in TEM suggests that only small subdomains withing FibH might be changing (such as NTD and the flexible linkers), allowing for extension of the solenoid, which seems to be contained within the globules, suiting the observation in DLS and offering a conciliating

explanation between the micelle and liquid crystalline models.[64,65] Beyond the morphological change, we observed the presence of supramolecular structures resembling bottle-brush structures like those observed before from *Samia ricini* (*Saturniid*) fibroin,[66] and recently in NSF.[67] Based on these observations and the apparent role NTD has on the reversible sol-gel transition, we propose that FibH can form high-order oligomers driven by head (NTD) interactions, making the core of the supramolecular fibre (SD.11).

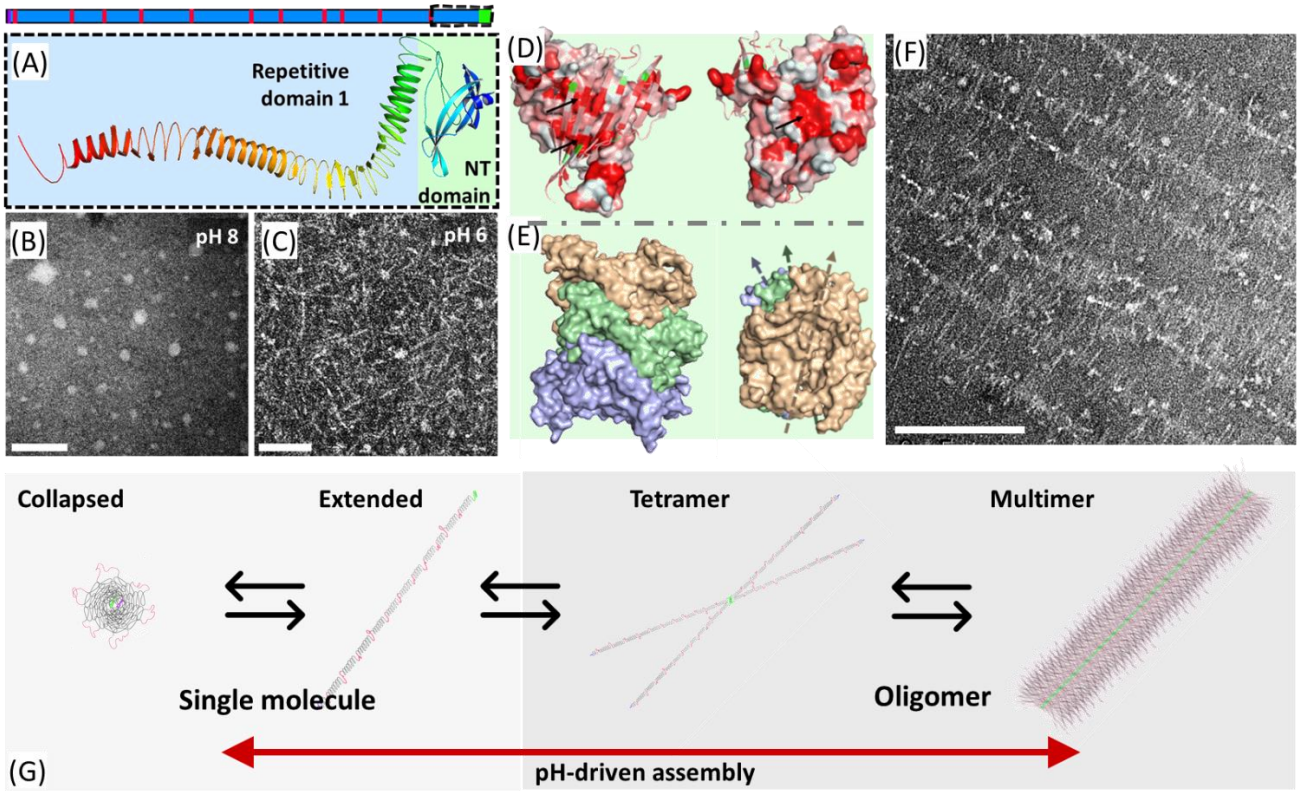


Figure 3. (A) Cartoon of Fibroin heavy Chain with 12 repetitive blocks (blue), 11 flexible linkers (red), NTD (green) and CTD (purple), under this a molecular cartoon of one of the folded models corresponding to NTD and first repetitive domain in a beta-solenoid type of fold. (B) Negative stained TEM of NLSF at pH 8. (C) Negative stained TEM of NLSF at pH 6. (D) Surface representation of the proposed biological unit of the NTD tetramer coloured by hydrophobicity with top dimer represented as a cartoon (left) and as seen from the back (right) with black arrows pointing at hydrophobic patches. (E) Proposed NTD tetramer stacking shows 3 tetrameric units stacked from the side (left) and the top (right) with arrows indication the twist of the stacking. (F) Negative stained TEM of NLSF at pH 6 showing supramolecular brush-like structures formed via NTD stacking. (G) Proposed pH driven self-assembly of the protein, from globular unimer, extended solenoid, dimer/tetramers and multimeric brush-like fibril (left to right). Scale bars 100 nm for B and C, and 500 nm for F.

The structure of the N-terminal domain solved by X-ray crystallography has a saddle shape, with highly hydrophobic patches on the surfaces, on one side only covered by a small helix linked to the rest of the structure by a flexible spacer (Fig. 3. D). It was noted before that this helix could flip out of the hydrophobic space. We propose that NTD dimeric units could stack based on our hypothesis, leaving the rest of the multidomain protein body protruding laterally. These structures impart great ordering by creating supramolecular brush-like fibres (Fig. 3. E, F) and might be responsible for the cholesteric textures observed in the past.[20] Hence, we conducted docking simulations from NTD crystal units at pH 8, 7 and 6 to verify that these can undergo pH-dependent oligomerisation. Our results suggest that NTD likely forms high-order oligomers, with the interactions being strongly favoured at pH below 7 (see Figure S33, B-E). Similarly, we noted that solenoid structures would be left in position for lateral interactions to emerge upon

oligomerisation. Docking simulations of solenoid units verified that such interactions were possible and are equally favourable parallelly or antiparallelly (see SD.12).

The last observations and the fact that no evidence of the bottle-brush structures can be found within the fibre (SD.13) suggest that these structures exist temporarily. However, they are likely to fulfil a role in the assembly pathway, given the conservation of the NTD domain across all species. We believe that NTD-driven oligomerisation could facilitate controlled phase separation, thus aiding in the dehydration of the dope and fostering lateral interaction of the solenoid units. Upon acidification, these structures align with the flow generating the cholesteric textures. Here, extensive lateral interaction would form a network. Thus, in these first steps, the system is characterised by two main interactions, NTD-NTD and solenoid-solenoid. As the silk duct diameter is reduced, a critical stress is reached, and NTD interactions are potentially disrupted, leaving a network dominated entirely by lateral solenoid interactions. The order imparted by the NTD oligomers is lost, and thus the network suffers a reorganisation, with the solenoid axis aligning with the flow and giving rise to the recently observed fractal network, both in NSF and high-quality RSF.[64] Figure 4 summarises the proposed fibre formation model, with cartoons and experimental microscopy data showing snapshots of the observed structures and assigned to the observed optical textures.

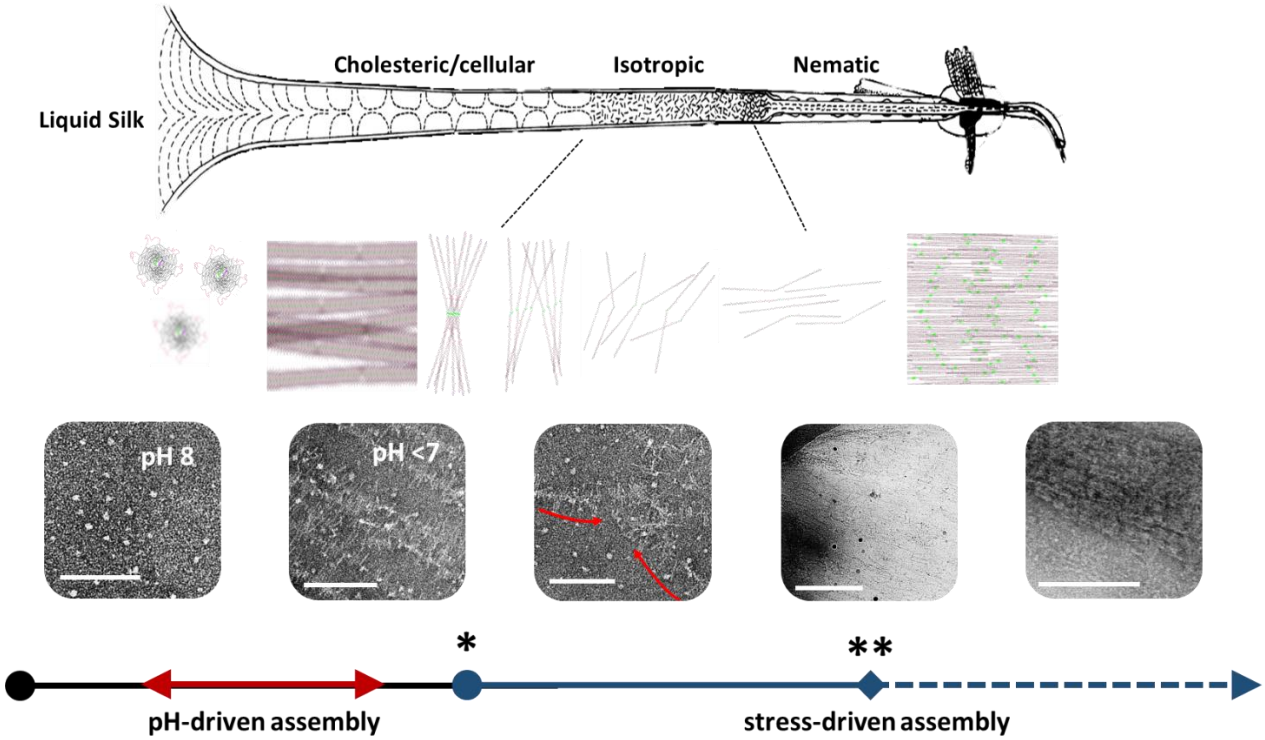


Figure 4. Cartoon with the schematic representation of the proposed self-assembly pathway for fibroin. Illustration with optical textures was adapted from the literature.[14] In the first instance, fibroin at pH greater than 7 exists as a globular unit. At pH lower than 7 extend and form supramolecular brush-like fibrillar structures that would align with the direction of the flow and originate a cholesteric texture. At critical stress, the supramolecular structures break, leaving free isotopically ordered anisotropic oligomers of lower order that align and deform under flow to generate the solid silk fibre after the silk press.

The described transition would explain the transformation of cholesteric-isotropic-nematic observed before.[18] Moreover, the last proposed organisation leaves the solenoid axis, and thus the hydrogen bonding network, parallel with the direction of elongation and ideally placed for unfolding and stretching of the backbone. These hypotheses are supported by observing two maxima in yield stress in stress-controlled experiments on gel-like material at pH 7; similar

stress maxima were recently observed when pulling fibres from other native-like fibroin solutions.[68] The first yield occurs at relatively low strains and appears narrower than the second. Such a double yield behaviour suggests the disruption of two different interactions, which we believe correspond to the initial disruption of NTD interactions, with narrower energy distribution, and secondly, a stress-induced denaturation (hydrogen bond breakage) with subsequent backbone stretching. Further discussion in SD.14. The described process leads to the extension of the backbone and β-sheet formation with nano-fishnet architecture proposed before.[69] Based on this hypothesis, we would predict that if the stress is kept low enough during fibre formation, one would potentially be able to obtain Silk-I based fibres (or a combination of polymorphs).

Conclusions

This study employed a comprehensive array of techniques to enhance our understanding of the Silk-I structural polymorph of silk fibroin. Through our investigations, we made significant advancements by uncovering a pH-induced reversible sol-gel transition, which we subsequently correlated with morphological and structural alterations of the protein at the molecular level.

We propose that FibH from *B. mori* is a multidomain protein with twelve low complexity domains with the ability to fold into β-solenoids. We accredit the Silk-I diffraction pattern to these domains, which are linked by flexible spacers and flanked by two terminal domains, of which NTD is the main driver of assembly. As pH drops below 7, NTD oligomerises forming large bottle-brush supramolecular fibres. The solenoid units, anchored within the construct are enabled to form stable lateral interactions driven by hydrophobic Ala-rich surfaces. Stress in the first instance disrupts the initial bottle-brush structures, leaving behind a fractal network of solenoids interacting laterally. As stress increases, the fold is denatured and the backbone of the polypeptide is stretched in the direction of the flow, driving the assembly into a nanofishnet molecular architecture with β-sheet crystallites as nodes.[69,70] On the other hand, our study on standard RSF showed that these have largely lost NTD, FibL and P25. Thus, it contains mainly a polydisperse mixture of the low complexity domains of the protein. In the absence of NTD, these fragments are devoid of the pH switch, and therefore lack the degree of preorganisation, albeit still retaining most of the Silk-I features.

We believe the β-solenoid structure to be incidental in *Bombyx mori*, as other fibrillar structures are compatible with the proposed assembly mechanism, with only NTD being essential in driving a pH-driven assembly, as demonstrated in *Samia ricini,* where the repetitive domains are rich in polyA motifs that fold in α-helixes,[71,72] yet still form the bottle-brush structures.[66] Beyond the consequences for insect silk, our work also brings into question the understanding of low complexity proteins as intrinsically disordered, especially those rich in G. Our work paves the way for the utilization of native-like silk fibroins (NLSF), which have been widely disregarded in previous literature. Through our findings, we provide a convincing rationale for the historical disparities between native and RSF, as well as unparalleled understanding of the fibroin assembly process, from a liquid state to the remarkable structural material that constitutes silk fibre.

Acknowledgements

This work was supported by various funding sources: the Wellcome Trust [086906/Z/08/Z], [100917/Z/13/Z]; The EIC Accelerator [947454]; The NIHR [NIHR i4i Invention for Innovation award II-LB-0417-20005]; R.O.M.T. received support from the EPSRC National Productivity Investment Fund (NPIF; EP/R51245/XF) and the EPSRC Doctoral Prize Fellowship at the University of Bristol (EP/W524414/1). F.P. received support from an EPSRC early career fellowship (EP/S017542/1). Y.L. received support from the China Scholarship Council

during his PhD program. The authors would like to express their gratitude to the following individuals and facilities for their contributions and assistance: J.C. Eloi from the Chemistry Imaging Facility (funded by the University of Bristol and EPSRC, EP/K035746/1 and EP/M028216/1), J. Mantell from the Wolfson Bioimaging Facility, U. Borucu from the GW4 Cryo-EM Facility (funded by the Wellcome Trust, 202904/Z/16/Z and 206181/Z/17/Z, and BBSRC, BB/R000484/1) for their help and insightful contribution in TEM; M. Crump, and specially C. Williams for their assistance and discussions on protein NMR, and access to the 700MHz spectrometer (funded by BrisSynBio, a BBSRC/EPSRC Synthetic Biology Research Centre, BB/L01386X/1, and BBSRC Alert 20, BB/V019163/1); C. Arthur for contributions and assistance with LC-MS/MS; R. Cruz-Samperio for help with SDS-PAGE and discussions; P. Laity for contributions and thoughtful discussions related to this manuscript. The views expressed in this work are those of the author(s) and do not necessarily reflect those of the NIHR, the Department of Health and Social Care or any of their funding bodies.

Author Contributions

R.O.M.T. conceived and designed the idea presented in this study and conducted the experimental work. Y.L. performed docking simulations under the supervision of F.P. who also contributed to the development of ideas. N.S. and R.W. provided valuable input in conceptualizing ideas and acquiring funding. L.S. contributed to the acquisition and analysis of fibre X-ray diffraction data, as well as the conceptualization of ideas. C.H. played a role in the conceptualization of ideas, experimental design, and data analysis. S.A.D. provided continuous supervision to R.O.M.T, contributed to experimental design, conceptualisation of ideas and acquired funding. All authors participated in revising the manuscript, and R.O.M.T. and S.A.D. wrote and edited the manuscript. The results were discussed among all authors, and they all contributed to the final version of the manuscript.

Competing interests

The authors declare no competing interests.

References

1. He, Y.-X. *et al.* N-Terminal Domain of Bombyx mori Fibroin Mediates the Assembly of Silk in Response to pH Decrease. *J. Mol. Biol.* **418**, 197–207 (2012).
2. Hao, Z. *et al.* New insight into the mechanism of in vivo fibroin self-assembly and secretion in the silkworm, Bombyx mori. *Int. J. Biol. Macromol.* **169**, 473–479 (2021).
3. Partlow, B. P. *et al.* Silk Fibroin Degradation Related to Rheological and Mechanical Properties. *Macromol. Biosci.* **16**, 666–675 (2016).
4. Ma, Y. *et al.* A single-cell transcriptomic atlas characterizes the silk-producing organ in the silkworm. *Nat. Commun. 2022 131* **13**, 1–16 (2022).
5. Laity, P. R. & Holland, C. The rheology behind stress-induced solidification in native silk feedstocks. *Int. J. Mol. Sci.* **17**, (2016).
6. Foo, C. W. P. *et al.* Role of pH and charge on silk protein assembly in insects and spiders. *Appl. Phys. A* **82**, 223–233 (2006).
7. Asakura, T. *et al.* A repeated β-turn structure in poly(Ala-Gly) as a model for silk I of Bombyx mori silk fibroin studied with two-dimensional spin-diffusion NMR under off magic angle spinning and rotational echo double resonance. *J. Mol. Biol.* **306**, 291–305 (2001).
8. Hu, X., Kaplan, D. & Cebe, P. Dynamic Protein and Water Relationships during b-Sheet Formation. *Macromolecules* **41**, 3939–3948 (2008).
9. Asakura, T. Structure of Silk I (Bombyx mori Silk Fibroin before Spinning) -Type II β-

Turn, Not α-Helix-. *Mol. 2021, Vol. 26, Page 3706* **26**, 3706 (2021).
10. Zhao, B., Cohen Stuart, M. A. & Hall, C. K. Dock 'n roll: folding of a silk-inspired polypeptide into an amyloid-like beta solenoid. *Soft Matter* **12**, 3721–3729 (2016).
11. Haskew, M. J., Deacon, B., Yong, C. W., Hardy, J. G. & Murphy, S. T. Atomistic Simulation of Water Incorporation and Mobility in Bombyx mori Silk Fibroin. *ACS Omega* **6**, 35494–35504 (2021).
12. Marsh, R. E., Corey, R. B. & Pauling, L. An investigation of the structure of silk fibroin. *Biochim. Biophys. Acta* **16**, 1–34 (1955).
13. Kratky, O., Schauenstein, E. & Sekora, A. An Unstable Lattice in Silk Fibroin. *Nat. 1950 1654191* **165**, 319–320 (1950).
14. He, S.-J., Valluzzi, R. & Gido, S. P. Silk I structure in Bombyx mori silk foams. *Int. J. Biol. Macromol.* **24**, 187–195 (1999).
15. Lotz, B. Rippled Sheets: The Early Polyglycine Days and Recent Developments in Nylons. *ChemBioChem* e202100658 (2022) doi:10.1002/CBIC.202100658.
16. Lazo, N. D. & Downing, D. T. Crystalline regions of Bombyx mori silk fibroin may exhibit β-turn and β-helix conformations. *Macromolecules* **32**, 4700–4705 (1999).
17. Walker, A. A., Holland, C. & Sutherland, T. D. More than one way to spin a crystallite: multiple trajectories through liquid crystallinity to solid silk. *Proceedings. Biol. Sci.* **282**, 20150259 (2015).
18. Asakura, T. *et al.* Some Observations on the Structure and Function of the Spinning Apparatus in the Silkworm Bombyx mori. *Biomacromolecules* **8**, 175–181 (2006).
19. Rey, A. D. Liquid crystal models of biological materials and processes. *Soft Matter* **6**, 3402–3429 (2010).
20. Willcox, P. J., Gido, S. P., Muller, W. & Kaplan, D. L. Evidence of a Cholesteric Liquid Crystalline Phase in Natural Silk Spinning Processes. *Macromolecules* **29**, 5106–5110 (1996).
21. Inoue, S. *et al.* 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**, 40517–40528 (2000).
22. Inoue, S. *et al.* Assembly of the silk fibroin elementary unit in endoplasmic reticulum and a role of L-chain for protection of α1,2-mannose residues in N-linked oligosaccharide chains of fibrohexamerin/P25. *Eur. J. Biochem.* **271**, 356–366 (2004).
23. Zabelina, V. *et al.* Mutation in Bombyx mori fibrohexamerin (P25) gene causes reorganization of rough endoplasmic reticulum in posterior silk gland cells and alters morphology of fibroin secretory globules in the silk gland lumen. *Insect Biochem. Mol. Biol.* **135**, 103607 (2021).
24. Tanaka, K., Inoue, S. & Mizuno, S. Hydrophobic interaction of P25, containing Asn-linked oligosaccharide chains, with the H-L complex of silk fibroin produced by Bombyx mori. *Insect Biochem. Mol. Biol.* **29**, 269–276 (1999).
25. Takahashi, T. *et al.* The physiological accumulation of mutant fibroin light chains induces an unfolded protein response in the posterior silk gland of the Sericin cocoon Nd-sD strain of silkworm Bombyx mori. *J. Insect Biotechnol. Sericology* **86**, 105–112 (2017).
26. Tanaka, K. & Mizuno, S. Homologues of fibroin L-chain and P25 of Bombyx mori are present in Dendrolimus spectabilis and Papilio xuthus but not detectable in Antheraea yamamai. *Insect Biochem. Mol. Biol.* **31**, 665–677 (2001).
27. Li, J. Y. *et al.* Expression profiling and

regulation of genes related to silkworm posterior silk gland development and fibroin synthesis. *J. Proteome Res.* **10**, 3551–3564 (2011).
28. Yonemura, N., Mita, K., Tamura, T. & Sehnal, F. Conservation of Silk Genes in Trichoptera and Lepidoptera. *J. Mol. Evol.* **68**, 641–653 (2009).
29. Suzuki, Y., Yamazaki, T., Aoki, A., Shindo, H. & Asakura, T. NMR study of the structures of repeated sequences, GAGXGA (X = S, Y, V), in bombyx mori liquid silk. *Biomacromolecules* **15**, 104–112 (2014).
30. Takahashi, Y., Gehoh, M. & Yuzuriha, K. Structure refinement and diffuse streak scattering of silk (Bombyx mori). *Int. J. Biol. Macromol.* **24**, 127–138 (1999).
31. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods 2022* 1–4 (2022) doi:10.1038/s41592-022-01488-1.
32. Mirdita, M. *et al.* ColabFold - Making protein folding accessible to all. *bioRxiv* 2021.08.15.456425 (2021) doi:10.1101/2021.08.15.456425.
33. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nat. 2021 5967873* **596**, 583–589 (2021).
34. Lopes, J. L. S., Miles, A. J., Whitmore, L. & Wallace, B. A. Distinct circular dichroism spectroscopic signatures of polyproline II and unordered secondary structures: Applications in secondary structure analyses. *Protein Sci.* **23**, 1765–1772 (2014).
35. Fändrich, M. & Dobson, C. M. The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *EMBO J.* **21**, 5682–5690 (2002).
36. Okuyama, K. *et al.* Analysis of Silk I structure by X-ray and electron diffraction methods. *J. Sericultural Sci. Japan* **57**, 23–30 (1988).
37. Crick, F. H. C. & Rich, A. Structure of Polyglycine II. *Nat. 1955 1764486* **176**, 780–781 (1955).
38. Asakura, T., Ohgo, K., Komatsu, K., Kanenari, M. & Okuyama, K. Refinement of Repeated β-turn Structure for Silk I Conformation of Bombyx mori Silk Fibroin Using 13C Solid-State NMR and X-ray Diffraction Methods. *Macromolecules* **38**, 7397–7403 (2005).
39. Suzuki, Y., Yamazaki, T., Aoki, A., Shindo, H. & Asakura, T. NMR study of the structures of repeated sequences, GAGXGA (X = S, Y, V), in bombyx mori liquid silk. *Biomacromolecules* **15**, 104–112 (2014).
40. Zhou, P., Li, G., Shao, Z., Pan, X. & Yu, T. Structure of Bombyx mori silk fibroin based on the DFT chemical shift calculation. *J. Phys. Chem. B* **105**, 12469–12476 (2001).
41. Ho, B. K. & Brasseur, R. The Ramachandran plots of glycine and pre-proline. *BMC Struct. Biol.* **5**, 1–11 (2005).
42. Terry, A. E., Knight, D. P., Porter, D. & Vollrath, F. pH induced changes in the rheology of silk fibroin solution from the middle division of Bombyx mori silkworm. *Biomacromolecules* **5**, 768–772 (2004).
43. Boulet-Audet, M., Holland, C., Gheysens, T. & Vollrath, F. Dry-Spun Silk Produces Native-Like Fibroin Solutions. *Biomacromolecules* **17**, 3198–3204 (2016).
44. Laity, P. R., Gilks, S. E. & Holland, C. Rheological behaviour of native silk feedstocks. *Polymer (Guildf).* **67**, 28–39 (2015).
45. Domigan, L. J. *et al.* Carbonic anhydrase generates a pH gradient in Bombyx mori silk glands. *Insect Biochem. Mol. Biol.* **65**, 100–6 (2015).
46. Holland, C. *et al.* Silk and Synthetic Polymers: Reconciling 100 Degrees of Separation. *Adv. Mater.* **24**, 105–109 (2012).
47. Mu, X., Fitzpatrick, V. & Kaplan, D. L. *From Silk Spinning to 3D Printing: Polymer*

Manufacturing using Directed Hierarchical Molecular Assembly. *Advanced Healthcare Materials* vol. 9 1901552 (Wiley-VCH Verlag, 2020).
48. Koeppel, A., Laity, P. R. & Holland, C. The influence of metal ions on native silk rheology. *Acta Biomater.* **117**, 204–212 (2020).
49. Koeppel, A. & Holland, C. Progress and Trends in Artificial Silk Spinning: A Systematic Review. *ACS Biomater. Sci. Eng.* **3**, 226–237 (2017).
50. Koebley, S. R. *et al.* Silk reconstitution disrupts fibroin self-assembly. *Biomacromolecules* **16**, 2796–2804 (2015).
51. Vollrath, F., Porter, D. & Holland, C. There are many more lessons still to be learned from spider silks. *Soft Matter* **7**, 9595–9600 (2011).
52. Greving, I., Cai, M., Vollrath, F. & Schniepp, H. C. Shear-Induced Self-Assembly of Native Silk Proteins into Fibrils Studied by Atomic Force Microscopy. *Biomacromolecules* **13**, 676–682 (2012).
53. Sparkes, J. & Holland, C. The rheological properties of native sericin. *Acta Biomater.* **69**, 234–242 (2018).
54. Chen, P. *et al.* pH-Triggered transition of silk fibroin from spherical micelles to nanofibrils in water. *Macromol. Res.* **16**, 539–543 (2008).
55. Holland, C., Urbach, J. S. & Blair, D. L. Direct visualization of shear dependent silk fibrillogenesis. *Soft Matter* **8**, 2590–2594 (2012).
56. Aoki, M., Masuda, Y., Ishikawa, K. & Tamada, Y. Fractionation of Regenerated Silk Fibroin and Characterization of the Fractions. *Molecules* **26**, (2021).
57. Frydrych, M., Greenhalgh, A. & Vollrath, F. Artificial spinning of natural silk threads. *Sci. Rep.* **9**, 1–10 (2019).
58. Wray, L. S. *et al.* Effect of processing on silk-based biomaterials: Reproducibility and biocompatibility. *J. Biomed. Mater. Res. Part B Appl. Biomater.* **99B**, 89–101 (2011).
59. Zhou, C. Z. *et al.* Silk fibroin: structural implications of a remarkable amino acid sequence. *Proteins* **44**, 119–122 (2001).
60. Bini, E., Knight, D. P. & Kaplan, D. L. Mapping Domain Structures in Silks from Insects and Spiders Related to Protein Assembly. *J. Mol. Biol.* **335**, 27–40 (2004).
61. Collin, M. A., Mita, K., Sehnal, F. & Hayashi, C. Y. Molecular evolution of lepidopteran silk proteins: Insights from the ghost moth, hepialus californicus. *J. Mol. Evol.* **70**, 519–529 (2010).
62. Kludkiewicz, B. *et al.* The expansion of genes encoding soluble silk components in the greater wax moth, Galleria mellonella. *Insect Biochem. Mol. Biol.* **106**, 28–38 (2019).
63. Partlow, B. P. *et al.* Silk Fibroin Degradation Related to Rheological and Mechanical Properties. *Macromol. Biosci.* **16**, 666–675 (2016).
64. Yang, S., Zhao, C., Yang, Y., Ren, J. & Ling, S. The Fractal Network Structure of Silk Fibroin Molecules and Its Effect on Spinning of Silkworm Silk. *ACS Nano* (2023) doi:10.1021/ACSNANO.3C00105/ASSET/IMAGES/LARGE/NN3C00105_0010.JPEG.
65. Eliaz, D. *et al.* Micro and nano-scale compartments guide the structural transition of silk protein monomers into silk fibers. *Nat. Commun. 2022 131* **13**, 1–12 (2022).
66. Inoue, S. I. *et al.* Nanostructure of Natural Fibrous Protein: In Vitro Nanofabric Formation of Samia cynthia ricini Wild Silk Fibroin by Self-Assembling. *Nano Lett.* **3**, 1329–1332 (2003).
67. Song, K. *et al.* Silkworm spinning: the programmed self-assembly from natural silk fibroin to superfibre. *bioRxiv* 2021.03.08.434386 (2021)

doi:10.1101/2021.03.08.434386.
68. Wan, Q. *et al.* Mesoscale structure development reveals when a silkworm silk is spun. *Nat. Commun. 2021 121* **12**, 1–8 (2021).
69. Liu, R. *et al.* “Nano-Fishnet” Structure Making Silk Fibers Tougher. *Adv. Funct. Mater.* **26**, 5534–5541 (2016).
70. Kim, Y. *et al.* Nano-fishnet formation of silk controlled by Arginine density. *Acta Biomater.* **128**, 201–208 (2021).
71. Nakazawa, Y. & Asakura, T. Heterogeneous exchange behavior of Samia cynthia ricini silk fibroin during helix–coil transition studied with 13C NMR. *FEBS Lett.* **529**, 188–192 (2002).
72. Yao, J., Nakazawa, Y. & Asakura, T. Structures of Bombyx mori and Samia cynthia ricini silk fibroins studied with solid-state NMR. *Biomacromolecules* **5**, 680–688 (2004).