

<https://doi.org/10.1038/s42003-024-06474-1>

Molecular organization of fibroin heavy chain and mechanism of fibre formation in *Bombyx mori*



Rafael O. Moreno-Tortolero^{1,2}✉, Yijie Luo^{1,3}, Fabio Parmeggiani^{1,3,4}, Nick Skaer⁵, Robert Walker⁵, Louise C. Serpell⁶, Chris Holland⁷ & Sean A. Davis¹✉

Fibroins' transition from liquid to solid is fundamental to spinning and underpins the impressive native properties of silk. Herein, we establish a fibroin heavy chain fold for the Silk-I polymorph, which could be relevant for other similar proteins, and explains mechanistically the liquid-to-solid transition of this silk, driven by pH reduction and flow stress. Combining spectroscopy and modelling we propose that the liquid Silk-I fibroin heavy chain (FibH) from the silkworm, *Bombyx mori*, adopts a newly reported β -solenoid structure. Similarly, using rheology we propose that FibH N-terminal domain (NTD) templates reversible higher-order oligomerization driven by pH reduction. Our integrated approach bridges the gap in understanding FibH structure and provides insight into the spatial and temporal hierarchical self-assembly across length scales. Our findings elucidate the complex rheological behaviour of Silk-I, solutions and gels, and the observed liquid crystalline textures within the silk gland. We also find that the NTD undergoes hydrolysis during standard regeneration, explaining key differences between native and regenerated silk feedstocks. In general, in this study we emphasize the unique characteristics of native and native-like silks, offering a fresh perspective on our fundamental understanding of silk-fibre production and applications.

The major protein component of silk, fibroin, is stored in arthropods as a liquid gel before undergoing a triggered conversion via spinning, into a highly stable solid fibre. This transformation starts within the silk gland, a highly specialised sack-shaped organ where not only the silk proteins are produced but where tight control over pH and metal ion concentrations is exerted^{1–5}. However, there is a gap in our knowledge surrounding the molecular level events during this transformation. There are some commonly agreed principles, including the contribution of pH and flow stress/strain, energy input/work⁶, in triggering assembly and promoting structural conformational changes^{7,8}. Though, there is less consensus regarding protein conformational changes. In particular, the molecular structure of the liquid state silk, often called silk-I, is heavily debated^{9–13} and is the focus of our research.

For silks' model organism *Bombyx mori*, fibroin, both within the gland and in the fibre, is composed of three different proteins, fibroin heavy chain (FibH), fibroin light chain (FibL) and a glycoprotein, fibrohexamerin (P25)

of molecular masses 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^{14,15}. However, the structural details of this system are not well known, save for the crystal structure of the N-terminal domain (NTD) of FibH. It is generally accepted that only FibH is essential for fibre formation, with FibL and P25 being auxiliary in the secretion process^{16–18}, with some silk-producing moths and other related species lacking these latter proteins entirely^{19–21}. 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 shifts²², to the diffraction patterns of Silk-I (liquid state) and Silk-II (fibre)²³.

Silk-I has previously been termed α -silk, as it was erroneously 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

¹School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK. ²Max Planck-Bristol Centre for Minimal Biology, University of Bristol, Bristol BS8 1TS, UK. ³School of Biochemistry, University of Bristol, University Walk, Bristol BS8 1TD, UK. ⁴School of Pharmacy and Pharmaceutical Sciences, Cardiff University, Redwood Building, King Edward VII Ave, Cardiff CF10 3NB, UK. ⁵Orthox Ltd, Milton Park, 66 Innovation Drive, Abingdon OX14 4RQ, UK. ⁶Sussex Neuroscience, School of Life Sciences, University of Sussex, Brighton BN1 9QG, UK. ⁷Department of Materials Science and Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, UK. ✉e-mail: ro.morenotortolero@bristol.ac.uk; s.a.davis@bristol.ac.uk

antiparallel β -sheet structures²⁴. The exact structure of silk-I remains ambiguous, and it is often classified as either a random coil or an intrinsically disordered state^{25,26}. However, discrete reflections in X-ray diffraction data suggest neither of these classifications is fully adequate^{14,15,27}. Over time, several models have been proposed to account for the diffraction patterns (summarised in Table S1), however, all of these suffer from limitations inherent to the technical difficulties in acquiring aligned fibre X-ray diffraction patterns or complete datasets from electron diffraction. Most of the proposed unit cells so far belong to low-symmetry systems, perhaps due to the initial required assumptions necessary to arrive at structural models given the limited data available. Currently, the most accepted Silk I model is a type-II β -turn rich structure²⁸, similar to the crankshaft model proposed by others before^{29,30}. However, this model requires extensive intermolecular hydrogen bonding networks to persist³¹, failing to account for the observation of similar chemical shifts both in dilute solutions and in the solid state measured by NMR, as well as small fibrillar structures observed herein and previously by electron microscopy³².

Similarly, modelling of consensus sequences has indicated a right-handed β -helix showing lower relative free-energy values than the corresponding type-II β -turn³¹. Moreover, the observation of a range of distinct liquid crystalline textures (patterns) in vivo at the start of anterior section of the gland and the spinneret remains unexplained³³, with the mesogenic structures being unidentified. 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 spinneret³⁴. The emergence of the cellular optical texture has been attributed to epitaxial anchoring of rod-like mesogens under confinement³⁵. Yet, this model 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 has been observed using electron microscopy³⁶. More importantly, despite the evidence of supramolecular order, a transition from the Silk I to Silk II polymorph only occurs later near the spinneret^{34,36}.

In this study, we propose that at the molecular level, Silk-I corresponds to a folded fibrillar conformation of FibH. Furthermore, we rationalise how physicochemical triggers such as pH, metal ions, and stress/strain induce

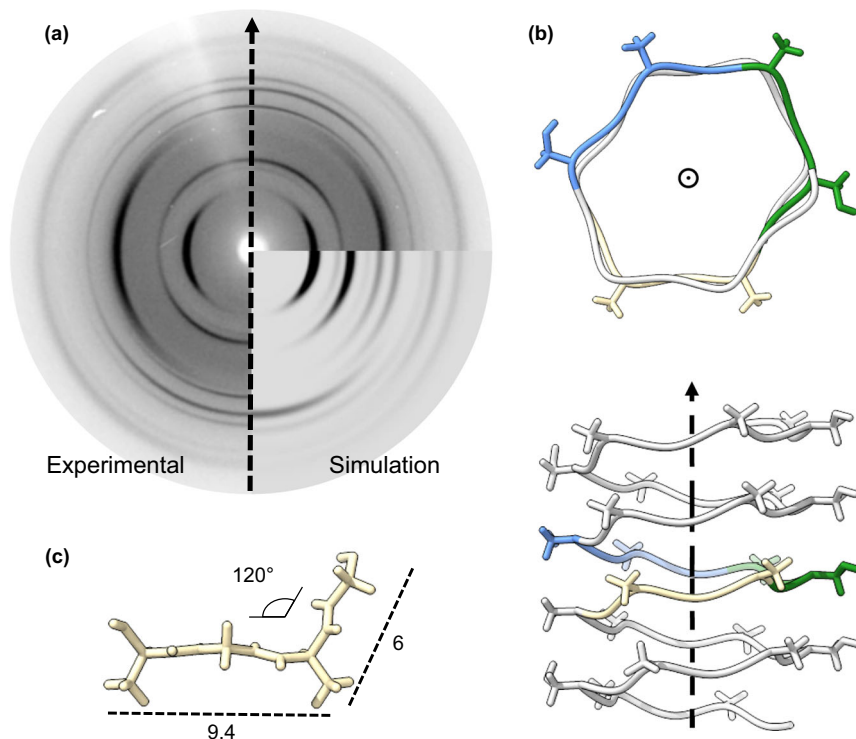
conformational changes in this mesogen that result in concomitant supramolecular assembly or reorganization consistent with the observed optical textures inside the silk gland. The abrupt loss of the cellular optical texture derives from disruption of the preceding cholesteric phase, and is a key intermediate to facilitate the minimum energy transition to the ordered nematic phase, as previously proposed^{34,36}. To investigate this hypothesis, we employed a comprehensive approach combining both computational simulations and experimental methods, encompassing various length scales ranging from the molecular to macroscopic. Our methodology involved the utilization of a fibroin solution achieved by dissolving silk fibres, after a proprietary degumming method, in concentrated LiBr solutions, resulting in a material referred to as native-like silk fibroin (NLSF). This NLSF exhibits numerous characteristics consistent with those of the native silk fibroin (NSF), including electrophoretic mobility and rheological properties. It is important to differentiate NLSF from standard regenerated silk fibroin (RSF)³⁷, as the latter fails to accurately replicate the molecular integrity and rheological properties of the native system.

Results and discussion

Molecular model creation and validation

Dilute solutions of NLSF were cast and left to dry slowly prior to being characterised by fibre X-ray diffraction (fXRD) experiments. When the incidence of the X-ray beam was perpendicular to the film, we obtained a pattern similar to those obtained by powder X-ray diffraction (pXRD), yet, as we rotated the film, sharper reflections were observed and eventually a highly oriented pattern emerged (Fig. 1a and Fig. S1), consistent with previous Silk-I data^{30,38}, see Table S2 for reflections. This data coupled with the observation of liquid crystalline phases suggest the presence of rod-like molecules, which would be preferentially oriented parallel to the surface of the film³⁹. After running AlphaFold2 (AF2) simulation of the repetitive domains of FibH^{40–42}, we found that these repetitive domains are predicted to adopt a novel β -solenoid conformation (Fig. S2). Briefly, the fold results from a supersecondary folding of strands, stabilised by inter-strand hydrogen bonds running along the solenoid axis, as depicted in Fig. 1b; the basic strand motif might explain the proposal of packing of almost extended strands by others for this polymorph^{38,43–45}. Although in this case

Fig. 1 | Structural model for fibroin heavy chain. **a** Experimental and simulated Silk-I diffraction pattern obtained from drop-cast films of NLSF and derived unit cell showing the proposed fibre axis. **b** Proposed unit cell comprised of AGXG motifs. **c** β -solenoid molecular model obtained for fibroin heavy chain, observed from the top and side views (top and bottom, respectively).



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 spectroscopy⁴⁶. It is possible that other G-rich proteins, such as fibroin heavy chain-like, spidroins, and others with hexapeptide repeated motifs of GX fold in similar structures⁴⁷. Structural evidence of different spidroins folding into flexible worm-like structures was recently found by SAXS⁴⁸, pointing at the possibility that β -solenoids might be widespread among silk proteins. Comparable conformation and interpretations were found for the G-rich snow flea antifreeze protein, which was crystallized using racemates^{49,50}. Similar β -helical configurations have also been suggested before for FibH, albeit mostly via computational approaches^{12,31}.

Although the predictions showed low confidence, and relatively high predicted Local Distance Difference Test (pLDDT, average ~ 50), all generated models maintained the solenoidal topology (Fig. S3) and were consistent with the topologies we observed by TEM (Fig. S4). Based on these results, we propose a trigonal unit cell containing a single curved strand, as depicted in Fig. 1c; initial unit cell parameters are a : 9.4, b : 3.4 and c : 6 Å, and $\alpha = \beta$: $\sim 90^\circ$ and γ : 120° , although further refinement is required. We note that the proposed unit cell is similar to that proposed for the PolyGly-II polymorph⁵¹, known to resemble Silk-I²⁸ and gives a simulated diffraction pattern that recapitulated the reflections obtained experimentally (Fig. 1a), allowing us to index most of the observed reflections as seen in Table S3. Nevertheless, observed discrepancies can be attributed to experimental constraints in aligning the sample, inherent heterogeneities in a real sample, and the software's inability to account for helical symmetry. This unit cell is

shown in context of the proposed solenoidal model in Fig. 1b. Given the markedly repetitive characteristics of FibH, featuring extensive low-complexity domains, characterized by a hexapeptide sequence (GAGAGX), where X may assume any of the residues A, S, Y, or V in descending order of abundance, twelve solenoidal domains are expected to form. The projected architecture of the protein consists of 12 solenoid bodies, interconnected by short linker regions that appear structurally unorganized, as depicted in Fig. S5. Significantly, within each repetitive domain, there are more regular and seemingly rigid subdomains (as shown in Fig. S3) that might be the ones giving rise to the deconvolved solid-state NMR Silk I* structures, with the remainder of the dynamic and unordered subdomains contributing to the overall Silk I spectrum^{52,53}.

Furthermore, both 1D proton (Fig. S6) and 2D ^1H - ^1H TOCSY (Fig. S7), ^1H - ^{13}C HSQC, ^1H - ^{15}N HSQC (Fig. S8), ^1H - ^{15}N TROSY (Fig. 2a) and ^1H - ^1H NOESY (Fig. 2a, b) were acquired. Assigned chemical shifts (Table S4) were very close to those reported elsewhere^{22,54}, as well as predicted shifts from our solenoid models (Table S5). Notably, we did not observe changes in the chemical shifts at pH 8 or pH 6, intended to replicate physiological pH change within the silk gland (Fig. S9). The assignment of the amide protons and nitrogen chemical shifts is depicted in Fig. 2a for simplified motifs within FibH. In addition, we note that when predicting the torsion angles for reduced motifs in FibH, two populations for G residues are possible within the generic GX motif²². One of these (ϕ , $\phi = 77^\circ, 10^\circ$) was used to derive the type-II β -turn model for Silk-I*. However, the second population of torsion angles (ca. ϕ , $\phi = -80^\circ, 180^\circ$) was not used to constrain the modelling and matches closely those found in the solenoidal model predicted by AF2 as

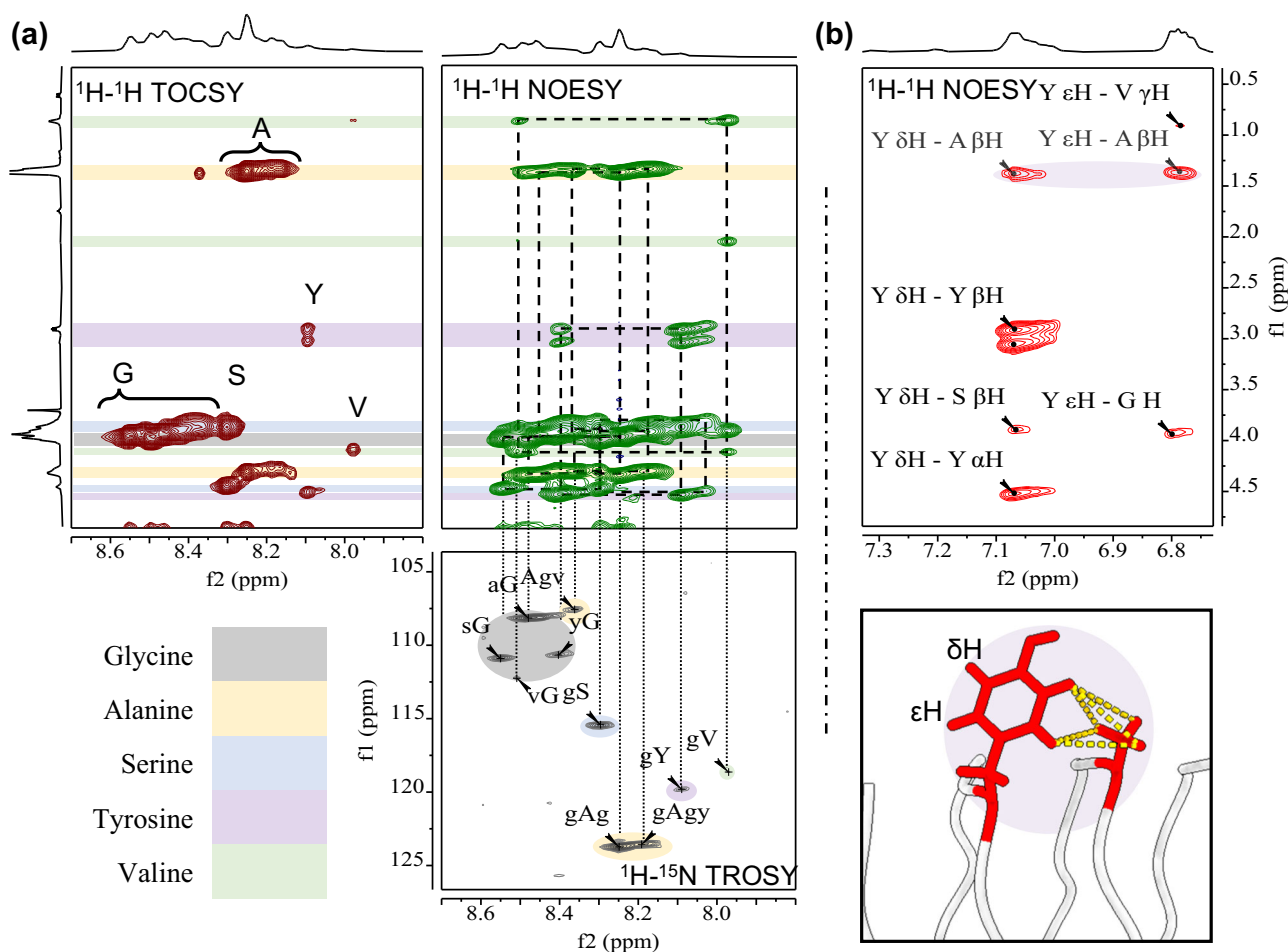
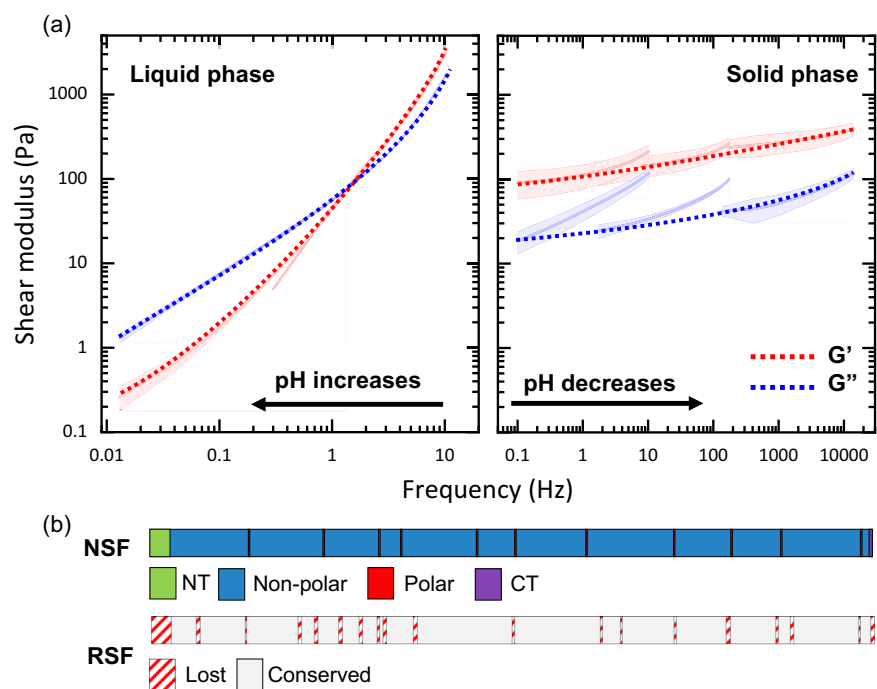


Fig. 2 | NMR analysis of NLSF. a Amide shift assignment via ^1H - ^1H TOCSY, Amide region ^1H - ^1H NOESY analysis, and ^{15}N chemical shift assignment of simplified motifs via ^1H , ^{15}N TROSY. **b** Analysis of the Tyrosine (Y) residue NOE signals

showing evidence of proximity to Alanine-CH, and cartoon representation of motif found within the AlphaFold2 model showing positions like the measured via NOESY experiment.

Fig. 3 | The effect of pH on rheological properties of NLSF, and the lost domains during regeneration. **a** Master curves created using oscillatory shear data of the liquid and solid-like samples of NLSF; guide lines were added to aid in trend visualisation. **b** Cartoon 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. Curves showed averages (continuous lines) and standard deviation (shadowed area around average), technical replicates $N = 5$.



well as previously proposed angles from density functional theory (DFT) calculations on NMR values^{13,55}. 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 Fig. S10)⁵⁶, thus reducing the overall energetic cost of the Silk-I to Silk-II transformation. Further to these observations, NOESY experiments, shown in Fig. 2b evidences Y residues close to (~3.3 to 3.4 Å) the methyl sidechain of A, (Fig. S11), suggesting a possible methyl- π interaction⁵⁷, which 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 Fig. 2b, bottom). Similarly, we observed other NOE intensities, at the amide proton from S residues, which although showing the strongest correlation with G, it also showed correlation with A- β H. Such patterns can be readily explained by our model, where the strongest correlation corresponds to abundant motifs gS (ca. 2.2 Å), and the weaker correlation with A- β H corresponding to across strand gS--A distance (2.7 Å), both the estimated distances and ratios are in good agreement with the experimental data (more in Supplementary Note: NOESY).

The role of NTD, pH effect on rheological properties

It has largely been recognised that in native spinning both flow stress and pH changes are responsible for the transformation of the soluble fibroin into the solid fibre^{7,58–60}. 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^{61,62}. However, direct evidence of the structural changes induced by pH and their effects on rheological properties has proved elusive.

Here we used NLSF samples obtained at concentrations of 60 to 80 mg mL⁻¹, lower than those found inside the gland (190–300 mg mL⁻¹)⁶², buffered at 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 section⁶³. We observed that the material goes through a reversible sol-to-gel 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 studies⁶². 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^{64,65}. Figure 3a shows two master curves demonstrating the effect of pH on the obtained material; master curves were created from the data shown in Fig. S12 and the applied shifts are presented in Table S6; the linear viscoelastic limit (LVE) was determined beforehand, and plots are shown in Fig. S13. 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 a reciprocal relationship 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^{6,62}. 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 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 (Fig. S14)⁵³. 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 LiBr⁶⁶, in congruence with the dominant liquid character and the decreased stability of the interaction that gives rise to the slowest relaxation mode⁶⁷).

The proximity of the critical pH observed here with that reported for NTD (below pH 7)⁵¹, 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, Figs. S15–17. Consequently,

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^{68,69}. 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 RSF at a similar pH of 7^{70,71}. Similarly, reversible sol-gel transitions have been observed only for NSF when changing pH⁵⁸, 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)⁷², well below the physiological pH of the gland^{68,70,73}.

Many authors argue that the molecular weight (MW) reduction in RSF is solely responsible for these observations^{74–78}. However, we considered this reduction insufficient to explain the major differences in literature found viscosities⁷⁹, 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)^{62,80}. Such a large difference, ~4 orders of magnitude, is not substantiated by the MW difference, which on average only falls by approximately half after 30 min of boiling during degumming⁷⁸, as depicted in Fig. S18. This reduction in MW would be translated in a proportional reduction of ~0.088 (calculation in Supplementary Note: polymer viscosity), close to one order of magnitude drop in viscosity (3 orders of magnitude away from the difference observed experimentally). Notably, when comparing NLSF and RSF solutions in LiBr (fully denatured) at similar concentrations a drop of this magnitude is observed (Fig. S19).

Thus, we hypothesised that coupled with the reduction in MW, RSF is also losing its NTD, which as noted previously, is highly implicated in the improved rheological properties of NLSF and NSF. A molecular model of the NTD with the first repetitive domain is shown in Fig. 4a. To prove this hypothesis, we conducted proteomics analysis on the peptide fragments (0.1–6 kDa) obtained directly during the dialysis of the liquid solutions of both RSF and NLSF. Interestingly in this procedure whilst 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 which 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% (Fig. S20a–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 RSF⁷⁸. Of the 35 detected peptides from FibH (Fig. 3b), 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 (Fig. S20d, e). 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 NTD homology and presence across all *Ditrysia*^{64,81,82}, and the absence of the latter two proteins in *Saturniids*^{19,83,84}, 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 (Fig. S21a), i.e., hydrolysis during degumming in Na₂CO₃ is likely limited by accessibility. Furthermore, besides G, the most abundant residues 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 N, despite its low abundance with only 20 in FibH, was found predominantly at the C-terminal of the fragments, with S and T more abundant at the N-terminal positions (Fig. S21b). Overall, during degumming chain scission occurs mainly at S residues, as suggested by previous studies⁷⁶, 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. Residues R, H, and K are only found in the terminal domains, and E and D both in the spacers and terminals (Fig. S21).

Structure and supramolecular ordering

Despite the important 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 TEM of NLSF samples 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 (Fig. 4b, c and Fig. S22a, b, for details). 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 their small size and 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 within FibH might be changing (such as NTD and the flexible linkers). This would allow 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^{85,86}. In contrast, RSF is known to be insensitive in this pH range, only showing featureless globular morphologies^{72,87}, consistent with the hypothesis that the protein is losing vital structurally functional domains.

Beyond the morphological change, we observed the presence of supramolecular structures resembling bottle-brush structures (Fig. 4f and Fig. S23) similar to those observed previously from *Samia ricini* (*Saturniid*) fibroin⁸⁸, and recently in NSf². 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 as illustrated in Fig. 4h.

The structure of the NTD solved by X-ray crystallography, a dimer (n2) where each molecule donates two β -hairpins. This structure has a saddle shape, with highly hydrophobic patches on the surfaces, proposed to form tetrameric units (dimer of dimers); herein we refer to this first interface as n2/n2. The remaining β -sheet surfaces are covered by a short α -helices linked to the rest of the structure by a flexible spacer (Fig. 4d). Given that this helix can flip out of the hydrophobic space⁶⁴, we propose that NTD tetrameric units could stack, through what we called tetramer-tetramer interfaces (n4/n4), as proposed in Fig. 4g, leaving the rest of the multidomain protein body protruding laterally. These structures impart great ordering by creating supramolecular brush-like fibres (Fig. 4e, f, h) which might be responsible for the cholesteric textures observed previously³⁶. Hence, we conducted docking simulations from NTD crystal units at pH 8, 7 and 6 to verify that these can undergo pH-dependent oligomerisation (Fig. S24a). Our results suggest that NTD likely forms higher-order oligomers, with the tetrameric interactions being strongly favoured at pH below 7 (Fig. S24b–e). At this pH, the protonation of residues Asp44 and Asp89 promote the stability of the interaction; both strictly conserved throughout all FibH in *Ditrysia* and belonging to acidic clusters^{82,89}. Moreover, it was noted that all the interfaces (dimer/dimer and tetramer/tetramer), even if slightly different in estimated ΔG values, were in a similar order of magnitude. Current work is undergoing to obtain high resolution experimental data of this transition. 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 are possible and equally favourable parallelly or antiparallelly driven by the exposed A

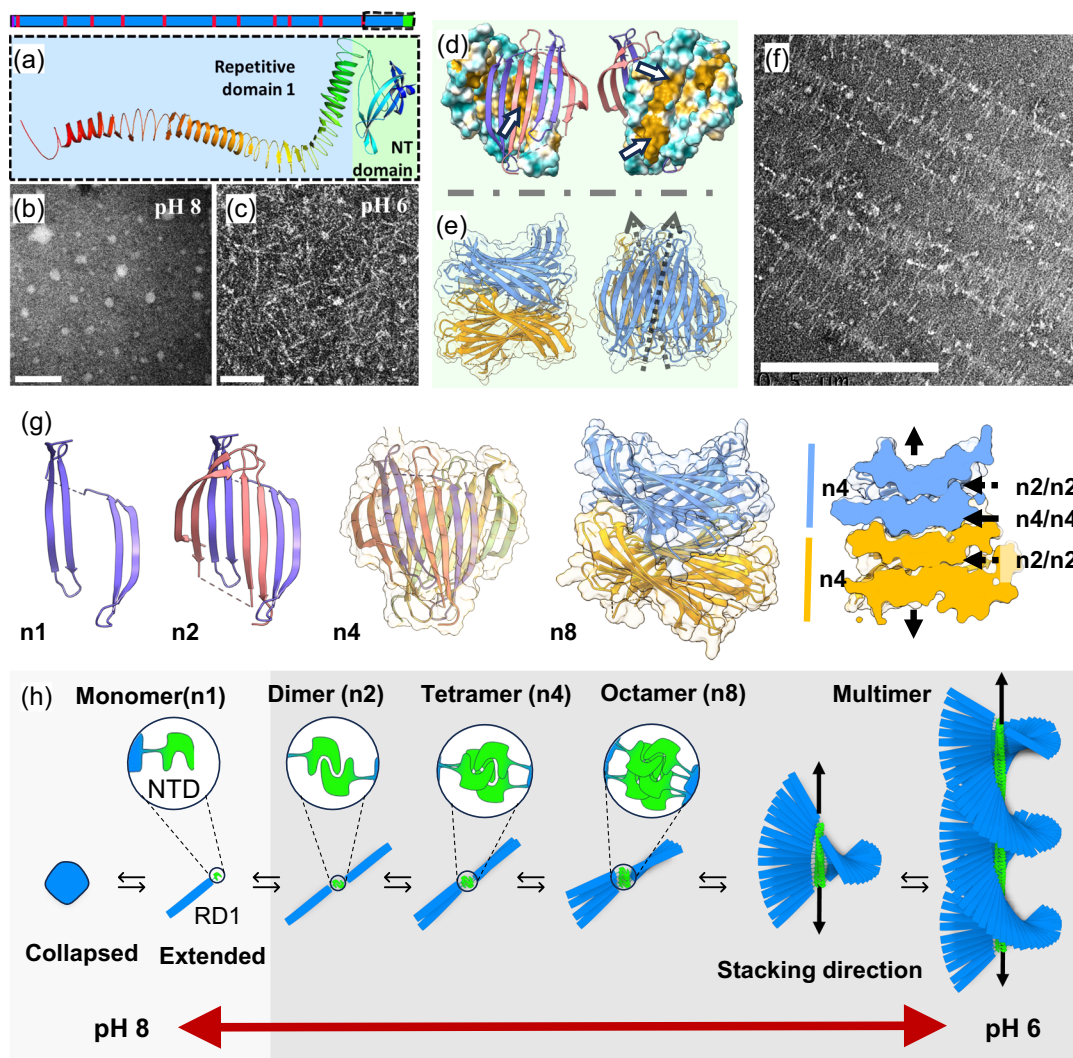


Fig. 4 | Evidence of a reversible pH-driven assembly in NLSF. **a** Cartoon of Fibroin heavy Chain with 12 repetitive blocks (blue), 11 flexible linkers (red), NTD (green) and CTD (purple), under this a molecular cartoon of one of the folded models corresponding to NTD and first repetitive domain in 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 tetramer⁵⁴ coloured by hydrophobicity (cyan hydrophytic and wheat hydrophobic) with top dimer represented as a cartoon (left) and as seen from the back (right) with white arrows pointing at hydrophobic patches involved in dimer-dimer (left) and tetramer-tetramer (right). **e** Proposed NTD tetramer stacking shows 2 tetrameric units stacked from the side (left) and the top (right) with arrows indicating the twist of the stacking. **f** Negative stained TEM of NLSF at pH 6 showing supramolecular

brush-like structures formed via NTD stacking. **g** Molecular model of the transition, left to right, from monomer (n1), dimer (n2), tetramer (n4) and octamer (n8) coloured by chain (n1-4), and by tetrameric unit (n8) and a cross-section of n8 indicating the dimer-dimer (n2/n2) and tetramer-tetramer (n4/n4) interfaces by discontinuous arrows and the stacking direction by continuous arrows on the far right; n4 and n8 models also show a ghostly representation of their surface.

h Proposed pH driven self-assembly of the protein, from globular unimer, extended conformation, dimer (each molecule donates two β -hairpins that interlock), tetramers (as shown in D, formed by dimer units stacking with a central symmetry plane), octamer (as shown in E, formed by the stacking of tetrameric units), and multimeric bottlebrush-like fibril showing proposed cholesteric order (left to right). Scale bars 100 nm for b and c, and 500 nm for f.

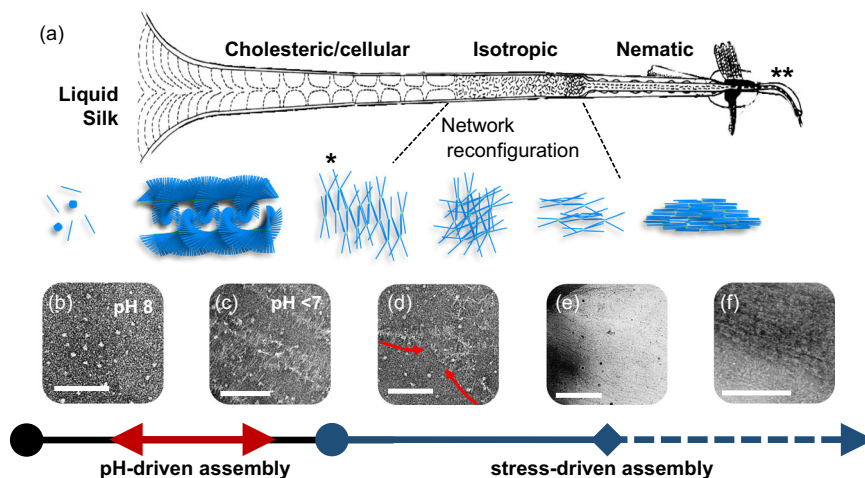
residues on the surface (Fig. S25), providing a foci for network formation (Fig. S26).

These structures likely fulfil an early role in the assembly pathway, given the conservation of the NTD domain across all species. However, no evidence of these bottle-brush structures can be found within the fibre (Figs. S27 and S28) suggesting that these may exist transiently. NTD-driven oligomerisation facilitates 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 forms a network. Thus, in these first steps, the system is characterised by two main interactions, NTD-NTD and solenoid-solenoid (the absence of CTD and FibL in *Saturniids* indicates that these are not essential for fibre formation, as has been suggested recently)⁹⁰. As the silk duct diameter is reduced, a critical

stress is reached, and NTD interactions are disrupted, leaving a network dominated entirely by lateral solenoid interactions (Fig. S29). The order imparted by the NTD oligomers is lost, and so the network can undergo a reorganisation, with the solenoid axis aligning with the flow direction and giving rise to the recently observed fractal network, both in NSF and high-quality RSf⁸⁵. Figure 5 summarises the proposed fibre formation model, with illustrations and experimental microscopy data showing snapshots of the observed structures and correlation with the optical textures.

The described transition explains the transformation of cholesteric-isotropic-nematic observed before³⁴. Moreover, the last step 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. Similar denaturation of α -helix fold of fibroin in 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP) under stretching has been observed recently,

Fig. 5 | Summary of the proposed assembly pathway for fibroin into the silk fibre. **a** Cartoon with the schematic representation of the proposed self-assembly pathway for fibroin. In the first instance, fibroin at pH greater than 7 exists as a globular unit **(b)**. At pH lower than 7 extend and form supra-molecular brush-like fibrillar structures that would align with the direction of the flow and originate a cholesteric texture **(c)**. At critical stress, the supra-molecular structures break **(d)**, and the formed solenoid network aligns under flow to generate the nematic order **(e, f)**. Scale bars 100 nm in **b, d, f** and **g**, 500 nm in **e**. Illustration with optical textures was reprinted (adapted) with permission from Asakura et al. Some Observations on the Structure and Function of the Spinning apparatus in the Silkworm *Bombyx mori*, 2007 Biomacromolecules, ACS. Copyright 2007 American Chemical Society³⁸.



with a α -to- β transformation prompted by stretching⁹¹. These hypotheses are supported by observing two maxima in yield stress in stress-controlled experiments on gel-like material at pH 7 (Fig. S30). Similar stress maxima were recently observed when pulling fibres from other native-like fibroin solutions^{92,93}. 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, similar to the critical amount of work observed for fibrillation⁷⁷. The described process leads to the extension of the backbone and β -sheet formation with “nano-fishnet” architecture proposed before^{92,94,95}. 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), in agreement with force reeling work the observation that the rate of work, and not the amount of work in itself are essential to promote the transition to solidification^{60,96–98}.

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 meaningful advances in understanding the in-solution structure of FibH and uncovered a pH-induced reversible sol-gel transition driven by NTD interactions, which we subsequently correlated with morphological and structural alterations of the protein at the molecular and supra-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 credit 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 supra-molecular assembly. As the pH drops below 7, NTD oligomerises forming large bottle-brush supra-molecular 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 nano-fishnet molecular architecture with β -sheet crystallites as nodes^{94,99}. 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. Our mass spectrometry analysis of proteolysis can therefore provide a powerful method to assess feedstock quality.

Although the β -solenoid structure might be common among silk proteins, we believe it to be incidental, as other fibrillar structures are compatible with the proposed assembly mechanism. NTD being the only essential feature in pH-driven assembly. In silk-spinning *Samia ricini*, the repetitive domains are rich in polyA motifs folding in α -helices^{100,101}, and also form the bottle-brush structures⁸⁸. Therefore, our model provides a general description for silk-fibre formation independent of specific sequences. Beyond insect silk, our work indicates that other low complexity proteins, especially those rich in GX motifs, such as the class III of G-rich proteins and other non-characterised fibroin heavy chain-like proteins widely found across all kingdoms of life. We also note the uncanny resemblance of FibH with known toxic GA repeat proteins associated with disease (e.g., amyotrophic lateral sclerosis, ALS), which in the monomeric species might have a similar fold, and their aggregation pathway might be similar as the proposed here; it has already been proven that (AG)₁₅ peptides show similar conformation to FibH in the solid state^{52,54,102,103}.

In a wider context, our work opens the opportunity for further interrogating the structure of simple repeat proteins, associated both with health and in disease development. Many of these are believed to be disordered, but superstructures of polyglycine-II or polyproline-II helices might underpin their function. Within the silk community, our work also paves the way for the generation of materials based on native-like silk fibroins. Overall, through our findings, we provide a unifying model that accounts for the observations on silk-fibre formation made through decades of research on the fibroin assembly process, from a liquid state to the remarkable structural material that constitutes silk fibre and might serve as inspiration for the design of bioinspired materials.

Materials and methods

Native-like silk fibroin solution (NL-SF)

Fibroin solutions of 8.5 M LiBr and 20 wt% of protein were kindly provided by Orthox LTD in 60 mL syringes and stored at 4 °C until used. Briefly, silk fibres were initially treated as per patent number US 8,128,984B2 using mild degumming conditions to minimize fibroin hydrolysis, after washing with MilliQ water, fibres were gently dried and dissolved in 8.5 M LiBr at 60 °C over 4 h. A stock solution at 2 wt% was made for studies requiring diluted solutions by diluting the provided solution with 8.3 M LiBr. The 20 wt% solutions look clear, free of precipitates, with a distinctive pale-yellow tint and low-shear viscosity of 60 ± 3 Pa.s. The stock solutions were then dialysed against MilliQ water using 12–14 kDa molecular weight cut-off (MWCO) dialysis membranes, with regular medium- to fresh MilliQ water changes. LiBr removal was followed by conductivity measurements of the dialysate and contrasted against a conductivity calibration curve, often requiring 3–4 days of dialysis to remove 99.9% of LiBr salt. After dialysis of the 20 wt% solution, a clear gel is obtained. In contrast, the 2 wt% yields a clear liquid solution with viscosity like pure water. Concentrations were

determined gravimetrically by cutting out small fragments or aliquoting about 500 μL of the solution and lyophilising them overnight in a Labconco Freezone 2.5 L Freeze-drier. The high concentration solution produces a gel with concentrations in the range of 60–80 mg/mL, whereas the 2 wt% solution produces a solution in the range of 5–7 mg/mL.

Standard Regenerated silk fibroin solution (RSF)

A standard regeneration protocol was followed³⁷. Briefly, about 10 g of raw silk, kindly provided by Orthox LTD, were boiled for 30 min in 4 L of aqueous 0.02 M Na_2CO_3 , the supernatant was then discarded, and the fibres were rinsed with copious MilliQ water at room temperature. The wet degummed fibres were dried in a convection oven at 30 °C overnight. The dry degummed fibres were weighted to verify degumming efficiency. Finally, degummed fibres were dissolved by adding specified volumes of 9.3 M LiBr to obtain a 20 wt% silk fibroin solution and left in an oven at 60 °C for 4 h, or until total dissolution was observed (whichever occurred first). The solution was then centrifuged at 5000 units of relative centrifugal force (RCF) for 5 min at room temperature to remove small undissolved particles. After these steps, the solution looks clear, free of precipitates with similar pale-yellow tint and with low-shear viscosity of 2 ± 0.5 Pa.s. The solutions were then dialysed against MilliQ water following an identical procedure as before. After dialysis, the 20 wt% solution produces a clear solution with a viscosity not different to pure water. Concentration was determined gravimetrically by aliquoting 500 μL of solution and lyophilising it overnight. The 20 wt% LiBr solution produces an aqueous solution with 60–80 mg/mL. Given the liquid nature of this sample, dilutions were prepared directly from the concentrated stock.

Sodium Dodecyl sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

For SDS-PAGE measurements, Tris-acetate 8–12% gels were used from Sigma. Samples of 0.1 mg/mL of NL-SF and RSF were used by mixing with stock solutions of running buffer and Coomassie blue. A high molecular weight ladder, HiMark Pre-stained Protein ladder from Invitrogen was used. Gels were run as recommended, using 200 V and 1 Amp for 40 min. Gel images were then analysed using FIJI by measuring grey-scale values and plotting coaligned with the standard ladder.

Rheology

The rheological characterisation was done in a Malvern Kinexus Pro rheometer with a Peltier lower plate for temperature control and with either a 20 mm parallel plate (PP20) upper geometry for gel samples, or with a truncated conical geometry with 4 degrees angle and 40 mm in diameter (CP4/40) for liquid samples. Samples were equilibrated for 24 h at room temperature in x50–100 volume of 100 mM Tris buffer at indicated pH values, titrated using stock solutions of 2 M HCl or 5 M NaOH. Although the buffering efficacy of Tris at pH 6 is very low, this buffer is preferred as other common buffers such as phosphates reduces the storage time considerably, promoting early conformational changes to the NL-SF concentrated solutions/gels.

Briefly, 500 mL of buffer solution were prepared at the desired pH. About 5–10 mL of aqueous SF solution were placed in them, contained by similar dialysis tubing membrane from previous steps. Samples were then left under slow stirring (ca. 150 rpm) at room temperature for 24 h to equilibrate. Following this, samples were taken out from the buffer solution and either carefully poured into clean 50 mL conical end centrifuge falcon tubes for liquid samples or taken to standard plastic Petri dishes (100 mm diameter) for gel samples. These last samples were taken out of the dialysis tube by carefully cutting open the membrane with a clean new scalpel. Samples were used immediately afterwards.

The following procedures were adapted from different works in the literature that work with native silk fibroin solutions, with minor modifications. In every case, samples were treated with the utmost care to prevent early conformational changes prompted by poor handling. Gel phase samples were carefully collected by excising small fragments from the bulk

using a wet scalpel and disposing of any sample that would show evidence of aggregation (opaque fibrillar aggregates). Enough gel sample was used to fill the entire geometry, carefully positioned in the centre of the lower plate, with the upper plate lowered slowly with a maximum allowed normal force of 1 N to a 1 mm gap. Water saturated tissue was left around the geometry, making sure it was not in contact with it, and the entire geometry was enclosed by a plastic cover to prevent dehydration during experiments. For liquid samples, a 5 mL pipette with tips with cut ends were used to prevent applying high shears when handling, with a standard quick loading setting from the software and similar procedures to prevent dehydration.

For the determination of the Linear Viscoelastic Limit (LVL), both elastic (G') and viscous (G'') moduli were measured in amplitude sweeps done in oscillatory mode at a constant frequency of 1 Hz, in the strain range between 0.1 and 100%, with 10 points per decade. Measurements at each point were taken until convergence was detected by the Kinexus software. Liquid samples underwent a pre-homogenisation step, as proposed elsewhere^{60,100}, by conducting fully rotational low-shear rate (1 s^{-1}) deformation for 100 s. At least 3 repeats were done for every sample.

After determination of LVL, G' and G'' were measured under frequency sweeps on fresh samples at constant strain of 0.01 (1%, within LVL of all samples) between 10 and 0.1 Hz. Similarly, 10 points per decade were measured, with measurements being taken until data convergence was observed by the software. Liquid samples underwent homogenisation, as mentioned before, by applying fully rotational deformation at 1 s^{-1} for 100 s before conducting the experiments. At least five repeats were done for every sample.

Only liquid samples were used for viscometry. Viscosity against shear rate measurements was obtained by a shear rate table in the range between 0.1 and 100 s^{-1} , with 10 points per decade. Measurements were taken until convergence was obtained at each shear rate point. Shear viscosity, as well as normal force values, were recorded.

Fibre X-ray diffraction

These experiments were done using Rigaku rotating anode (Cu $K\alpha$) with Saturn CCD detector with exposure times of 30–60 s and specimen to detector distance of 50 or 100 mm. Reflections were measured using CLEARER software.

For diffraction obtained from degummed silk fibres, about 1 cm long fibre bundles were mounted on glass capillary tubes using epoxy-based glue and then mounted directly onto a goniometer head for alignment. Alignment was ensured by monitoring a microscope camera focused on the beam path, and alignment was adjusted for the nominal 0 and 90° orientations. Spectra were acquired at 50 and 100 mm of distance at 0 and 90° orientations, for 30 and 60 s, respectively.

The silk-I film was fabricated by slow drying about 100 μL of NLSF solution at 6 mg/mL at room temperature, drop-casted onto a clean petri dish surface. Once dried, the film was carefully removed using forceps, glued to a capillary glass tube, and observed as the fibres before. Here, the film was orientated so the beam would strike normal to the surface at 0°, parallel to the film surface at 90° and later a range of acquisitions were done by varying orientation by 5° between 0° and 90°.

Liquid chromatography-mass spectrometry (LC-MS/MS) and proteomics analysis

Samples were analysed following modified Shotgun proteomics protocols. For the analysis of the hydrolysis patterns, concentrated solutions in LiBr of both samples were dialysed against MilliQ water while being contained within a dialysis tubing membrane with 6 kDa MWCO, which in turn was contained by a 100 Da dialysis tubing membrane. Care was taken not to cross-contaminate the different spaces, and the space between membranes was filled with 15 mL of MilliQ water. After 99.99% of LiBr was removed by dialysis, determined by the conductivity of an 8 h dialysate from fresh, the samples collected from the volume between the 6 kDa and 100 Da membranes were transferred to sterile 50 mL conical centrifuge falcon tubes, from which samples were directly submitted to be analysed. Obtained

sample from RSF contained about 2.0 ± 0.5 mg/mL of peptides, but no material was recovered from NL-SF (<0.1 mg/mL).

Samples were injected in a Dionex RS3000 High pressure Liquid Chromatographer (HPLC), coupled with a Orbitrap Elite in Electrospray ionisation mode (ESI) to analyse the fragmentation of the submitted peptides by ESI-MS/MS.

The collected data was then analysed using PeptideShaker and contrasted against the sequence of FibH, FibL, and P25 from the UniProt database (accession P05790, P21828, and P04148, respectively).

Nuclear magnetic resonance (NMR)

For NMR studies, only samples at pH 8 and 6 of NL-SF were studied. Briefly, stock 2 wt% solution was dialysed against MilliQ water until $>99.9\%$ of LiBr was removed to obtain a solution of about 1 mg/mL of protein. These solutions were then passed through a NAP45 non-interacting column, following manufacturer instructions, pre-equilibrated with 10 mM Tris in deuterium oxide, titrated beforehand to the desired pH with aqueous 2 M HCl or 5 M NaOH. The collected buffered samples were then analysed using a Bruker Avance III HD 700 with a 1.7 mm cryo-enhanced probe. ^1H -NMR were collected, as well as proton-proton Total Correlation Spectroscopy (^1H - ^1H TOCSY) for residue assignment, Nuclear Overhauser Effect Spectroscopy (NOESY) for distance determination and natural abundance carbon-proton Heteronuclear Single Quantum Coherence (^1H - ^{13}C HSQC) for secondary structure determination of A and G residues.

Fold simulations and model generation

Fold simulations were carried over at two different servers to verify results congruency. However, in general, the AlphaFold2 prediction methodology was used. The services used were those published on Google Colab books by DeepMind and ColabFold teams^{41,42}. For the simulations, Colab Pro, was used to ensure higher assigned memories, better processors and longer runtimes.

Fibroin heavy chain fold simulation. Given the detected modularity of the multidomain protein and the size limitation on the simulations through ColabFold (1000–1400 residues), only the NTD and the first repetitive domain were used initially for a total of 650 amino acids. The multisequence alignment (MSA) was run using Jackhammer and mmSeqs2, with a minimum pair coverage with a query of 50% and minimum sequence identity of 20%. Jackhammer offered the best coverage overall, with 7456 Sequences found in Uniref90, 9062 Sequences found in SmallBFD, and 482 Sequences found in Mgnify, or 17000 sequences in total; most of these covered the repetitive domain. Next, AlphaFold was run, and five models were generated after running the trunk of the AlphaFold neural network 8 times with different random choices across the MSA. Finally, the generated five models were relaxed using Amber-relax forcefield, and the models were downloaded, visualized and post-processed using Pymol. On the other hand, mmSeqs2 found about 45 sequences, primarily for NTD, and the repetitive domain was modelled as a random coil.

Later, the same sequence was also run on the DeepMind colab book, and very similar models were obtained. Following these, the rest of the domains were modelled in fragments below 1000 residues that contained entire domains (either repetitive domains alone, in combination with one linker, or as two repetitive domains with a single linker). Models obtained through this method were used to generate the full FibH structure.

Docking simulations

NTD docking. In the first instance, the solved N-terminal tetramer model (PDB: 3UA0) was relaxed using Rosetta. Then, using the relaxed model, an octameric (tetramer-tetramer) stacked unit was generated with manually adjustments, and the model was further relaxed as an octamer.

Docking simulations under pH control were done by first determining the protonation state of charged residues on the relaxed octamer model.

Next, protonation states were predicted by PDB2QR server (<https://server.poissonboltzmann.org/pdb2pqr>) at pH 6, 7 and 8. These generated models were used as references for root-mean-square deviation (RMSD) analysis during docking experiments. Finally, local docking was run in Rosetta on each of the three models, generating 2000 independent simulations for each. Estimated free energy of interaction measured in Rosetta energy units (REU), $\Delta\Delta G$, was then plotted against RMSD (a measure of the deviation from reference models).

Solenoid lateral dockings. Models obtained from the first repetitive domain was used here. It was noted that the first repetitive domain could be divided into 3 sub domains linked by less regular solenoidal structures that are thought to allow for flexibility within the repetitive domain. Furthermore, the first and second subdomains were unique to the first repetitive domain, whereas the third subdomain was more represented across all repetitive domains. Hints at this were already observed after self-similarity analysis. Consequently, the model was separated into three sliced models containing the subdomains, each relaxed in Rosetta previous to global docking simulations, where only models with $\Delta\Delta G$ lower than -30 REU were used for further analysis.

Far UV circular dichroism (CD)

CD measurements were performed using low concentrations of protein (ca. 0.05 mg/mL) in 10 mM sodium phosphate buffer at the desired pH (6–9) at a temperature of 25 °C. Buffer pH value was adjusted with aqueous dilutions of phosphoric acid or NaOH. Samples from NL-SF were obtained by dialysing stock 2 wt% solution in 8.3 M LiBr and equilibrated in x100 volume of buffer over 24 h. After equilibration, samples were centrifuged at 6000 RCF for 10 min at room temperature to remove any precipitate and later diluted with fresh buffer until obtaining High-tension (HT) values under 700 for the whole studied range. Spectra were acquired from 240 to 190 nm in a 1-mm path length quartz cuvette using a JASCO J-820 spectropolarimeter. The wavelength step for the CD measurements was 1 nm, and the scan rate was 100 nm/min. The shown spectra were background corrected and averaged over five scans. For temperature analysis, similar procedures were followed, the sample was heated in 5 °C steps from 25 to 95 °C and equilibrated for 30 s at each temperature.

Dynamic light scattering (DLS)

DLS measurements were performed using 1 mg/mL protein solutions. After dialysing the 2 wt% in 8.3 M LiBr solution, samples are collected with approximate concentrations of 1 mg/mL, equilibrated in buffer at desired pH (10 mM sodium Phosphate) over 24 h and used directly after centrifugation at 6000 RCF for 10 min. Disposable PS large volume cuvettes were used in a Malvern Instruments ZetaSizer nano ZS where data was acquired using software that included macros for size analysis and temperature trends. Software provided materials parameter values for protein sample, and water for media was used. For temperature trends, samples were equilibrated for 60 s at each temperature, 3 scans were averaged, and the temperature ranges from 25 to 75 °C with 1 °C increment.

Transmission electron microscopy (TEM)

Jeol JEM 1200 with a tungsten filament and acceleration voltage of 120 kV and a Jeol JEM 2100 with a LaB6 electron gun at 200 kV. Both were used for brightfield and negatively stained images.

Tecnai T20 with a LaB6 filament and acceleration voltage of 200 kV for negatively stained samples, and a Tecnai F20 with a field emission gun (FEG) with an acceleration voltage of 200 kV for cryo-TEM. For the latter, a Vitrobot Leica EM GP with humidity set to 100%, blotting time 1 s, was used after adding 5 μL of the sample before plunging into liquid ethane, and grids transferred to a cryo-holder (Gatan, Inc.).

The higher resolution cryo-TEM were taken using a Talos Arctica equipped with a 200 kV X-FEG, Ceta 16 M CCD detector, Gatan K2 DED, and Gatan GIF Quantum LS energy filter. Samples were vitrified in this case using a FEI Vitrobot.

Statistics and reproducibility

We did not observe batch-to-batch variability when using NLSF, and thus, we consider all our replicates as technical replicates. All data was collected in, at least, triplicates (or otherwise stated in the legend) and results are reported as the average with their corresponding standard deviation when applicable.

Data availability

All data are available in the main text or the supplementary information. The source data behind the graphs in the paper can be found in Supplementary Data 1. Additionally, raw data files for NMR, MS, fibre XRD and AlphaFold (.cif) are available at the University of Bristol data repository, data.bris, at <https://doi.org/10.5523/bris.2rm8nb1bualvs2sphf60f5wfy7>.

Received: 17 January 2024; Accepted: 19 June 2024;

Published online: 29 June 2024

References

- Ma, Y. et al. A single-cell transcriptomic atlas characterizes the silk-producing organ in the silkworm. *Nat. Commun.* **13**, 1–16 (2022).
- Song, K. et al. Decoding silkworm spinning programmed by pH and metal ions. *Sci. Bull.* <https://doi.org/10.1016/J.SCI.2023.12.050> (2023).
- Zhou, P. et al. Effects of pH and calcium ions on the conformational transitions in silk fibroin using 2D Raman correlation spectroscopy and ¹³C solid-state NMR †. *Biochemistry* **43**, 11302–11311 (2004).
- Wang, X. et al. In vivo effects of metal ions on conformation and mechanical performance of silkworm silks. *Biochim. Biophys. Acta Gen. Subj.* **1861**, 567–576 (2017).
- Liu, Q. et al. Disruption of the metal ion environment by EDTA for silk formation affects the mechanical properties of silkworm silk. *Int. J. Mol. Sci.* **20**, 3026 (2019).
- Holland, C. et al. Silk and synthetic polymers: reconciling 100 degrees of separation. *Adv. Mater.* **24**, 105–109 (2012).
- Laity, P. R. & Holland, C. The rheology behind stress-induced solidification in native silk feedstocks. *Int. J. Mol. Sci.* **17**, 1812 (2016).
- 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).
- 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).
- Hu, X., Kaplan, D. & Cebe, P. Dynamic protein and water relationships during β-sheet formation. *Macromolecules* **41**, 3939–3948 (2008).
- Asakura, T. Structure of silk I (Bombyx mori Silk Fibroin before Spinning) -Type II β-Turn, Not α-Helix-. *Molecules* **26**, 3706 (2021).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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. Sericol.* **86**, 105–112 (2017).
- 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).
- 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).
- Yonemura, N., Mita, K., Tamura, T. & Sehna, F. Conservation of silk genes in Trichoptera and Lepidoptera. *J. Mol. Evol.* **68**, 641–653 (2009).
- 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).
- Takahashi, Y., Gehoh, M. & Yuzuriha, K. Structure refinement and diffuse streak scattering of silk (Bombyx mori). *Int. J. Biol. Macromol.* **24**, 127–138 (1999).
- Marsh, R. E., Corey, R. B. & Pauling, L. An investigation of the structure of silk fibroin. *Biochim. Biophys. Acta* **16**, 1–34 (1955).
- Chatterley, A. S. et al. Broadband multidimensional spectroscopy identifies the amide II vibrations in silkworm films. *Molecules* **27**, 6275 (2022).
- Giubertoni, G. et al. In situ identification of secondary structures in unpurified bombyx mori silk fibrils using polarized two-dimensional infrared spectroscopy. *Biomacromolecules* **23**, 5340–5349 (2022).
- Kratky, O., Schauenstein, E. & Sekora, A. An unstable lattice in silk fibroin. *Nature* **165**, 319–320 (1950).
- Lotz, B. Rippled Sheets: The Early Polyglycine Days and Recent Developments in Nylons. *ChemBioChem* e202100658 <https://doi.org/10.1002/CBIC.202100658> (2022).
- Lotz, B. & Keith, H. D. Crystal structure of poly(l-Ala-Gly)III: a model for silk I. *J. Mol. Biol.* **61**, 201–215 (1971).
- Okuyama, K. et al. Analysis of Silk I structure by X-ray and electron diffraction methods. *J. Sericult. Sci. Jpn.* **57**, 23–30 (1988).
- 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).
- Akai, H. Ultrastructure of fibroin in the silk gland of larval Bombyx mori. *Exp. Cell Res.* **69**, 219–223 (1971).
- Walker, A. A., Holland, C. & Sutherland, T. D. More than one way to spin a crystallite: multiple trajectories through liquid crystallinity to solid silk. *Proc. Biol. Sci.* **282**, 20150259 (2015).
- 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).
- Rey, A. D. Liquid crystal models of biological materials and processes. *Soft Matter* **6**, 3402–3429 (2010).
- 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).
- Rockwood, D. N., Preda, R. C., Wang, X., Lovett, M. L. & Kaplan, D. L. Materials fabrication from Bombyx mori silk fibroin. *Nat. Protoc.* **6**, 1612–31 (2011).
- He, S.-J., Valluzzi, R. & Gido, S. P. Silk I structure in Bombyx mori silk foams. *Int. J. Biol. Macromol.* **24**, 187–195 (1999).

39. 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).
40. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* 1–4 <https://doi.org/10.1038/s41592-022-01488-1> (2022).
41. Mirdita, M. et al. ColabFold - Making protein folding accessible to all. *bioRxiv* 2021.08.15.456425 <https://doi.org/10.1101/2021.08.15.456425> (2021).
42. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
43. Konishi, T. & Kurokawa, M. The structure of silk fibroin- α . *Sen'i Gakkaishi* **24**, 550–554 (1968).
44. Muñoz-Guerra, S., Fita, I., Aymamí, J. & Puiggali, J. Electron microscopy and structural modeling of the 3-fold helical form of Poly(glycyl- α -alanine). *Macromolecules* **21**, 3464–3468 (1988).
45. Fossey, S. A., Némethy, G., Gibson, K. D. & Scheraga, H. A. Conformational energy studies of β -sheets of model silk fibroin peptides. I. Sheets of poly(Ala-Gly) chains. *Biopolymers* **31**, 1529–1541 (1991).
46. 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).
47. Kajava, A. V. Tandem repeats in proteins: from sequence to structure. *J. Struct. Biol.* **179**, 279–288 (2012).
48. Greving, I. et al. Structural diversity of native major ampullate, minor ampullate, cylindrical, and flagelliform silk proteins in solution. *Biomacromolecules* **21**, 3387–3393 (2020).
49. Pentelute, B. L. et al. X-ray structure of snow flea antifreeze protein determined by racemic crystallization of synthetic protein enantiomers. *J. Am. Chem. Soc.* **130**, 9695–9701 (2008).
50. Graham, L. A. & Davies, P. L. Biochemistry: Glycine-rich antifreeze proteins from snow fleas. *Science* **310**, 461 (2005).
51. Crick, F. H. C. & Rich, A. Structure of polyglycine II. *Nature* **176**, 780–781 (1955).
52. Asakura, T., Ogawa, T., Naito, A. & Williamson, M. P. Chain-folded lamellar structure and dynamics of the crystalline fraction of Bombyx mori silk fibroin and of (Ala-Gly-Ser-Gly-Ala-Gly) $_n$ model peptides. *Int. J. Biol. Macromol.* **164**, 3974–3983 (2020).
53. Asakura, T. & Naito, A. Structure of silk I (Bombyx mori silk fibroin before spinning) in the dry and hydrated states studied using ^{13}C solid-state NMR spectroscopy. *Int. J. Biol. Macromol.* **216**, 282–290 (2022).
54. 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 ^{13}C solid-state NMR and X-ray diffraction methods. *Macromolecules* **38**, 7397–7403 (2005).
55. 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).
56. Ho, B. K. & Brasseur, R. The Ramachandran plots of glycine and proline. *BMC Struct. Biol.* **5**, 1–11 (2005).
57. Platzer, G. et al. PI by NMR: probing CH- π Interactions in protein-ligand complexes by NMR spectroscopy. *Angew. Chem. Int. Ed.* **59**, 14861–14868 (2020).
58. 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).
59. Ceniz, J. L. et al. Mechanical behaviour and formation process of silkworm silk gut. *Soft Matter* **11**, 8981–8991 (2015).
60. Koepfel, A. et al. Spinning beta silks requires both pH activation and extensional stress. *Adv. Funct. Mater.* **31**, 2103295 (2021).
61. Boulet-Audet, M., Holland, C., Gheysens, T. & Vollrath, F. Dry-Spun silk produces native-like fibroin solutions. *Biomacromolecules* **17**, 3198–3204 (2016).
62. Laity, P. R., Gilks, S. E. & Holland, C. Rheological behaviour of native silk feedstocks. *Polymer* **67**, 28–39 (2015).
63. Domigan, L. J. et al. Carbonic anhydrase generates a pH gradient in Bombyx mori silk glands. *Insect Biochem. Mol. Biol.* **65**, 100–106 (2015).
64. 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).
65. Mu, X., Fitzpatrick, V. & Kaplan, D. L. From silk spinning to 3D printing: polymer manufacturing using directed hierarchical molecular assembly. *Adv. Healthcare Mater.* **9**, 1901552 (Wiley-VCH Verlag, 2020).
66. Koepfel, A., Laity, P. R. & Holland, C. The influence of metal ions on native silk rheology. *Acta Biomater.* **117**, 204–212 (2020).
67. Koepfel, A. & Holland, C. Progress and trends in artificial silk spinning: a systematic review. *ACS Biomater. Sci. Eng.* **3**, 226–237 (2017).
68. Koebley, S. R. et al. Silk reconstitution disrupts fibroin self-assembly. *Biomacromolecules* **16**, 2796–2804 (2015).
69. Vollrath, F., Porter, D. & Holland, C. There are many more lessons still to be learned from spider silks. *Soft Matter* **7**, 9595–9600 (2011).
70. 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).
71. Sparkes, J. & Holland, C. The rheological properties of native sericin. *Acta Biomater.* **69**, 234–242 (2018).
72. Chen, P. et al. pH-Triggered transition of silk fibroin from spherical micelles to nanofibrils in water. *Macromol. Res.* **16**, 539–543 (2008).
73. Holland, C., Urbach, J. S. & Blair, D. L. Direct visualization of shear dependent silk fibrillogenesis. *Soft Matter* **8**, 2590–2594 (2012).
74. Wang, Q., Chen, Q., Yang, Y. & Shao, Z. Effect of various dissolution systems on the molecular weight of regenerated silk fibroin. *Biomacromolecules* **14**, 285–289 (2013).
75. Aoki, M., Masuda, Y., Ishikawa, K. & Tamada, Y. Fractionation of regenerated silk fibroin and characterization of the fractions. *Molecules* **26**, 6317 (2021).
76. Partlow, B. P. et al. Silk fibroin degradation related to rheological and mechanical properties. *Macromol. Biosci.* **16**, 666–675 (2016).
77. Salapaka, S. & Salapaka, M. Scanning probe microscopy. *IEEE Control Syst. Mag.* **28**, 65–83 (2008).
78. Wray, L. S. et al. Effect of processing on silk-based biomaterials: reproducibility and biocompatibility. *J. Biomed. Mater. Res. B Appl. Biomater.* **99B**, 89–101 (2011).
79. Holland, C., Terry, A. E., Porter, D. & Vollrath, F. Natural and unnatural silks. *Polymer* **48**, 3388–3392 (2007).
80. Frydrych, M., Greenhalgh, A. & Vollrath, F. Artificial spinning of natural silk threads. *Sci. Rep.* **9**, 1–10 (2019).
81. Zhou, C. Z. et al. Silk fibroin: structural implications of a remarkable amino acid sequence. *Proteins* **44**, 119–122 (2001).
82. 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).
83. Collin, M. A., Mita, K., Sehna, F. & Hayashi, C. Y. Molecular evolution of lepidopteran silk proteins: insights from the ghost moth, *hepialus californicus*. *J. Mol. Evol.* **70**, 519–529 (2010).
84. 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).
85. 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* <https://doi.org/10.1021/ACS.NANO>.

- 3C00105/ASSET/IMAGES/LARGE/NN3C00105_0010.JPEG (2023).
86. Eliaz, D. et al. Micro and nano-scale compartments guide the structural transition of silk protein monomers into silk fibers. *Nat. Commun.* **13**, 1–12 (2022).
 87. Jia, B., Jia, L. & Zhu, J. Effect of pH on the conformational transition of silk fibroin in aqueous solution monitored by Thioflavin-T fluorescence. *J. Fluoresc.* **32**, 389–395 (2022).
 88. 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).
 89. Urovec, M. Z. & Sehnal, F. Unique molecular architecture of silk fibroin in the Waxmoth, *Galleria mellonella**. *J. Biol. Chem.* **277**, 22639–22647 (2002).
 90. Mi, J. et al. High-strength and ultra-tough whole spider silk fibers spun from transgenic silkworms. *Matter* **6**, 3661–3683 (2023).
 91. Yoshioka, T., Tashiro, K. & Ohta, N. Molecular orientation enhancement of silk by the hot-stretching-induced transition from α -Helix-HFIP complex to β -sheet. *Biomacromolecules* **17**, 1437–1448 (2016).
 92. Wan, Q. et al. Mesoscale structure development reveals when a silkworm silk is spun. *Nat. Commun.* **12**, 1–8 (2021).
 93. Sparkes, J. & Holland, C. The energy requirements for flow-induced solidification of silk. *Macromol. Biosci.* **19**, 1800229 (2019).
 94. Liu, R. et al. Nano-fishnet[™] structure making silk fibers tougher. *Adv. Funct. Mater.* **26**, 5534–5541 (2016).
 95. Wan, Q. et al. Carbon nanotubes facilitate silk hierarchical assembly by dry drawing. *Small Struct.* 2300435 <https://doi.org/10.1002/SSTR.202300435> (2024).
 96. Mortimer, B., Guan, J., Holland, C., Porter, D. & Vollrath, F. Linking naturally and unnaturally spun silks through the forced reeling of *Bombyx mori*. *Acta Biomater.* **11**, 247–255 (2015).
 97. Mortimer, B., Holland, C. & Vollrath, F. Forced reeling of *bombyx mori* silk: separating behavior and processing conditions. *Biomacromolecules* **14**, 3653–3659 (2013).
 98. Holland, C., O’Neil, K., Vollrath, F. & Dicko, C. Distinct structural and optical regimes in natural silk spinning. *Biopolymers* **97**, 368–373 (2012).
 99. Kim, Y. et al. Nano-fishnet formation of silk controlled by Arginine density. *Acta Biomater.* **128**, 201–208 (2021).
 100. 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).
 101. 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).
 102. Asakura, T. & Williamson, M. P. A review on the structure of *Bombyx mori* silk fibroin fiber studied using solid-state NMR: an antipolar lamella with an 8-residue repeat. *Int. J. Biol. Macromol.* **245**, 125537 (2023).
 103. Asakura, T. et al. Determination of accurate ¹H positions of (Ala-Gly)_n as a sequential peptide model of *bombyx mori* silk fibroin before spinning (Silk I). *Macromolecules* **46**, 8046–8050 (2013).

Acknowledgements

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, J. Mantell from the Wolfson Bioimaging Facility, U. Borucu from the GW4 Cryo-EM Facility 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 700 MHz spectrometer; 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 authors would also like to express their gratitude for and acknowledge the funder contributions: EPSRC National Productivity Investment Fund grant EP/R51245/XF (R.O.M.T.), EPSRC Doctoral Prize Fellowship at the University of Bristol grant EP/W524414/1 (R.O.M.T.), Wellcome Trust grants 086906/Z/08/Z, and 100917/Z/13/Z (N.S. and R.W.), The EIC Accelerator grant 947454 (N.S. and R.W.), the NIHR i4i Invention for Innovation award II-LB-0417-20005 (N.S. and R.W.), EPSRC early career fellowship grant EP/S017542/1 (F.P.), EPSRC grants EP/K035746/1 and EP/M028216/1 (TEM) and EP/X015416/1 (C.H.), Wellcome Trust grants 202904/Z/16/Z and 206181/Z/17/Z (TEM), BBSRC grant BB/R000484/1 (TEM), BrisSynBio, a BBSRC/EPSRC Synthetic Biology Research Centre, grant BB/L01386X/1 (NMR), BBSRC Alert 20, grant BB/V019163/1 (NMR) and China Scholarship Council (Y.L.). 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

Conceptualization: R.O.M.T., S.A.D., L.S., C.H., N.S., and R.W.; Methodology: R.O.M.T., S.A.D., C.H., and L.S.; Investigation: R.O.M.T., L.S., and Y.L.; Visualization: R.O.M.T. and Y.L.; Funding acquisition: S.A.D., N.S., R.W., and R.O.M.T.; Project administration: R.O.M.T. and S.A.D.; Supervision: S.A.D., N.S., and R.W.; Resources: N.S., R.W., Y.L., F.P., and R.O.M.T. Writing – original draft: R.O.M.T. and S.A.D. Writing – review & editing: R.O.M.T., Y.L., F.P., N.S., R.W., L.S., C.H., and S.A.D.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-024-06474-1>.

Correspondence and requests for materials should be addressed to Rafael O. Moreno-Tortolero or Sean A. Davis.

Peer review information *Communications Biology* thanks Ulyana Shimanovich, Tetsuo Asakura and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Joanna Timmins and Laura Rodríguez Pérez. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024