Complete sequences of the velvet worm slime proteins reveal that slime formation is enabled by disulfide bonds and intrinsically disordered regions

Yang Lu^{1*}, Bhargy Sharma^{1*}, Wei Long Soon¹, Xiangyan Shi², Tianyun Zhao³, Yan Ting Lim³, Radoslaw M. Sobota³, Shawn Hoon⁴, Giovanni Pilloni⁵, Adam Usadi⁶, Konstantin Pervushin⁷ and Ali Miserez^{1,7,*}

¹ Centre for Sustainable Materials (SusMat), School of Materials Science and Engineering, Nanyang Technological University (NTU), 639798 Singapore.

² Department of Biology, Shenzhen MSU-BIT University, No. 1 International University Park Road, Longgang District, Shenzhen, Guangdong Province, China, 518172.

³ Functional Proteomics Laboratory, Institute for Molecular and Cell Biology (IMCB), Agency for Science, Technology, and Research (A*Star), 138673 Proteos, Singapore.

⁴ Molecular Engineering Lab, IMCB, A*Star, 138673 Proteos, Singapore.

⁵ ExxonMobil Research and Engineering, New Jersey, 08801, United States.

⁶ ExxonMobil Asia Pacific Pte Ltd, 098633, Singapore.

⁷ School of Biological Science, NTU, 637551 Singapore.

* Author for correspondence: ali.miserez@ntu.edu.sg.

Y.L and B.S contributed equally to the manuscript.

Abstract

The slime of velvet worms (Onychophora) is a strong and fully biodegradable protein material, which upon ejection undergoes a fast liquid-to-solid transition to ensnare prey. However, the molecular mechanisms of slime self-assembly are still not well understood, notably because the primary structures of slime proteins are yet unknown. Combining transcriptomic and proteomic studies, we have obtained the complete primary sequences of slime proteins and identified key features for slime self-assembly. The high molecular weight slime proteins contain Cys residues at the N- and C-termini that mediate the formation of multi-protein complexes via disulfide bonding. Low complexity domains in the N-termini were also identified and their propensity for liquid-liquid phase separation established, which may play a central role for slime biofabrication. Using solid-state nuclear magnetic resonance, rigid and flexible domains of the slime proteins were mapped to specific peptide domains. The complete sequencing of major slime proteins is an important step towards sustainable fabrication of polymers inspired by the velvet worm slime.

Introduction

Many living organisms secrete structural fibrous materials in their immediate external environment to secure locomotion¹, attach themselves to solid substrates², or capture prey³. These abilities provide valuable lessons for biomimetic and sustainable biopolymer processing⁴. A classic example is spider silk, whose bio-fabrication of mechanically-tough fibers has been investigated for several decades⁵. An organism that has more recently garnered interest as a model system for biopolymer fabrication is the velvet worm (Onychophora).

Onychophora are carnivorous invertebrates that are placed in their own phylum and have existed for more than 500 million years. The preying mechanism of velvet worms is unique, consisting in rapid secretion of a highly adhesive slime out of oral papillae to capture small insects (**Fig. 1**). The secreted slime has many remarkable characteristics: (i) it exhibits high extensibility and tensile strength (ultimate tensile strength of 101.9 ± 20.1 MPa⁶); (ii) it rapidly phase-separates from a concentrated dope solution into the adhesive slime as soon as it is ejected out of the slime gland; (iii) it is water soluble and (iv) it can reversibly form fibers after being re-dissolved. The slime is mainly composed of highly hydrated proteins (90% water in the wet slime)⁷, with a small fraction of lipids. Separation of slime protein components by electrophoresis fractionates them into high, middle and low molecular weight (MW) proteins, with high MW proteins being the most dominant⁸. Using Sanger sequencing of expressed sequence tag library of total RNA extracted from the slime gland, Haritos *et al.*⁸ obtained partial sequences of the high MW protein from *Euperipatoides rowelli*, which we hereby call ER_P1. They found ER_P1 to be a proline (Pro)-rich protein with a highly disordered structure and potential glycosylation sites⁷. Subsequently, Baer *et al.*⁶ revealed through atomic force microscopy and cryo-transmission electron microscopy observations that prior to secretion, the slime precursor consists of nanoglobular building blocks made of proteins and lipids.

Figure 1. Fibrous slime of Onychophora (velvet worm). **a.** The *Eoperipatus* sp. characterized in this study. **b.** Slime ejection from the velvet worm when threatened. **c.** A close look at the slime showing droplets decorating a few threads.

Fiber formation and aqueous recycling of the slime are key physiochemical characteristics that have been investigated in recent years, but the molecular mechanisms behind fiber self-assembly are still uncertain. The most plausible explanations for intermolecular interactions governing fiber formation from the concentrated solution are lipid-protein interactions⁸, charge-charge interactions⁶, and disulfide bonding⁹. Baer *et al.*¹⁰ revealed that fiber formation is pH- and ionic strength-dependent. With additional prediction of phosphorylation sites on the high MW protein sequences, they proposed that the non-specific non-covalent interaction is the major driving force connecting proteins and lipids. Furthermore, crystalline β -sheet structures comprising more than 20 strands were detected in the solid slime using wide-angle X-ray diffraction¹¹. β -sheet is a relatively common structural motif in adhesive materials¹². In the velvet worm slime, it was suggested that elongated β -sheets are formed during protein unfolding triggered by shear forces, leading to rapid fiber formation and curing. The primary structure domains nucleating β -sheets are still unknown, although alternating patches with positive and negative charged residues were suggested as a possible source. The presence of divalent ions in native slime may also enhance the formation of β -sheet through phosphorylation/ Ca^{2+} interactions similar to caddisfly larvae silk¹³. Thus, a reversible fiber assembly model combining both electrostatic interactions and shear-induced β -sheet elongation has been proposed¹⁴. However, experimental evidence is still lacking to fully explain fiber formation. Disulfide bonding is also a common linkage in fibrous proteins¹⁵ and

was suggested to be related to the sticky and elastic properties of the slime¹⁶. However, the limited number of detected cysteine (Cys) in the partial sequence and absence of a reducing agent to reverse aggregation challenged this theory⁸.

In order to elucidate the mechanism of fiber formation, it is critical to obtain the full-length sequence of slime proteins, which has thus far eluded researchers due to the very high MW of the main slime proteins. Indeed, the aggregation of fibers is often governed by the end termini of their constitutive proteins. For example, silk fiber formation is regulated by pH-, ionic strength-, and temperature-dependent aggregation of the C- and N-terminus of silk fibroins^{17,18}. To achieve this, we combined RNA-sequencing of the slime gland of an *Eoperipatus* velvet worm with high-throughput proteomic studies¹⁹ and successfully obtained the full-length sequences of ES_P1 and ES_P2, which are the most abundant proteins of the *Eoperipatus* sp. Further, with proteomic analyses we identified post-translational modifications (PTMs) of the slime proteins including phosphorylation of Ser, Thr and Tyr and hydroxylation of Pro. Although phosphorylation is detected, it was not identified in the high MW proteins. Instead, we find that large complexes of ES_P1, ES_P2, and smaller MW proteins are stabilized by disulfide bond. Finally, we find that the N-termini of ES_P1 and ES_P2 are highly enriched in Gly and Ser, with sequence features reminiscent of the low complexity (LC) domain of the fused in sarcoma (FUS) protein that is well-known to exhibit liquid-liquid phase separation (LLPS)²⁰. We then demonstrate with recombinant protein construct that this Gly- and Ser-rich domain also exhibits phase separation *in vitro*, suggesting that LLPS of slime protein is a central mechanism facilitating the solution-to-fiber transition during slime fabrication. Finally, we conducted solid state nuclear magnetic resonance (ssNMR) studies of ¹³C labelled slime and identified molecular entities of the slime that are located either in the flexible or the rigid domains of the slime proteins. Combined together, our results unveil new molecular insights into the velvet worm slime and its self-assembly mechanisms that provide bioinspired lessons for future synthesis of green biopolymers.

Results and Discussion

Phylogenetic identification

The animals, found locally in the secondary forest in Singapore, contain 21 pairs of legs with a body size ranging from 15 mm at 1 week up to 50 mm for mature animals. 12S rRNA

sequences of the current specimen were compared with all known 12S rRNA sequences of Onychophora downloaded from NCBI and a phylogenetic tree on 12S rRNA was constructed with max likelihood method (**Supplementary Fig. 1**). The specimens are placed near to genus *Eoperipatus* sp. previously found in Thailand (JX568982.1).

Full-length slime protein sequences by combined RNA-seq and proteomics

Total RNA was extracted directly from the isolated slime gland, subjected to RNAseq using Illumina Hiseq platform, and the raw sequences were assembled *de novo* using the Trinity software suite to build the transcriptome²¹. By searching the transcriptome library with Transdecoder²², 9,772 complete coding sequences were predicted (the assembled transcriptome and raw RNAseq data is submitted to NCBI under Bioproject PRJNA806368). Notably, we identified one transcript with very high similarity to the slime ER_P1 from the *Euperipatoides rowelli*⁸ (56% similarity and 99% alignment with E-value of 0) having a translated MW of 116 kDa, hereafter named ES_P1. The full-length sequence of the gene encoding ES_P1 was verified by PCR using cDNA as template, which led to 37 single nucleotide polymorphisms (**Supplementary Table 1**). Another transcript translating for a large MW protein (188 kDa) that shared homology with ES_P1 was also detected in the transcriptome and called ES_P2. Both proteins were characterized by a high content of Pro (17%) and Lys (10%), a limited number of Cys residues (**Fig. 2 and Supplementary Table 2**), as well as a Gly- and Ser-rich domain in the N terminal sharing intriguing homology with the intrinsically disordered region of the FUS protein (**Supplementary Fig. 2**)²³.

To further corroborate that the identified transcripts translated into slime proteins, liquid chromatography with tandem mass spectrometry (LC-MS/MS) was conducted on native slime separated by gel electrophoresis. Slime proteins separated into high, middle and low MW regions by SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE) (**Fig. 2a**). The bands from different MW regions were carefully cut off, in-gel digested, subjected to LC-MS/MS and searched against the transcriptome library of the slime gland. ES_P1 was identified as the major protein in the bands from the high MW region or reduced slime bands (RS_band 1 and RS_band_2), with 71 and 22 peptide fragments matching the translated ES_P1 transcript after trypsin and AspN digestion, respectively (**Fig. 2d**). Furthermore, 13 and 7 peptide fragments from RS_band_3 in the high MW region also matched ES_P2 after trypsin and AspN digestion, respectively.

Bands cut from the middle and low MW regions were also subjected to LC-MS/MS and probed against the slime gland transcriptome. Identified peptides matched additional transcripts with translated MWs of 71 kDa and 12 kDa, respectively. Peptides of both proteins were recovered in high abundance from the bands in their respective MW regions. The middle MW protein ES_P3 matched the middle MW proline-rich proteins from *E. rowelli* (44.6%, with 83% alignment and E-value of $4e^{-141}$), with a high Pro content (8.6%) in ES_P3. Similar to ES_P1 and ES_P2, a Gly- and Ser-rich domain was found at the N-terminal. The low MW protein ES_P4 did not align to any Onychophoran slime protein from the literature and its Pro content was low. In contrast, it contained a relative high amount of Cys residues (13 Cys /117 aa).

A significant feature from the LC-MS/MS data was the detection of 39 and 6 hydroxyproline (HyPro) residues (detected as oxidized Pro) in ES_P1 and ES_P2, respectively, which indicates that at least 19.6% of Pro in ES_P1 and ES_P2 are hydroxylated into HyPro. To verify the presence of HyPro in the native slime, we also conducted amino acid analysis (AAA) of hydrolyzed slime, with the HyPro retention time obtained from an external standard (**Supplementary Fig. 3**). From the overall AAA composition of the slime, we identified 4.3 ± 0.1 % of Pro and 0.38 ± 0.02 HyPro residues. Since the AAA composition measured here is for the total slime, the significantly lower content of HyPro compared with LC-MS/MS results from the individual ES_Ps suggests that HyPro residues are mostly located within the high MW proteins, a result that parallels earlier results obtained by AAA of isolated slime proteins⁷.

Figure 2. SDS-PAGE of slime proteins and full-length protein sequences of high MW slime proteins, ES_P1 and ES_P2. **a.** Long-range SDS-PAGE of the whole slime with Coomassie blue staining on native (N), dephosphorylated (DP), reduced and thermally-denatured (TD) and reduced (R) slime. **b.** ProQ stain for phosphorylated protein on N, DP, and TD slime. **c.** Alcian blue stain/silver for glycoproteins for N, deglycosylated and reduced (DG-R), and deglycosylated and non-reduced (DG) slime. **d.** Primary structure of ES_P1 and ES_P2 with peptide coverage obtained by tandem MS on fragments recovered after digestion with trypsin (green) and AspN (cyan). HyPro was also identified by LC-MS/MS for both slime proteins. **e.** Primary structures of ES_P1 and ES_P2 can be divided into 3 main domains: long predicted disordered domains (grey), Gly- and Ser-rich domains at the N-termini (purple), and repeat domains (red) located along the primary structure. The location of all Cys residues is also highlighted. **f.** Peptide sequences of repeat motifs.

Primary structure features ES_P1 and ES_P2

Prior to this study, only partial sequences of velvet worm slime proteins from *E. rowelli* had been obtained⁸. Thus, ES_P1 and ES_P2 are the first velvet slime proteins with complete and verified end-to-end primary structures. Based on homology search and structural predictions, we identify four significant features in ES_P1 and ES_P2, as highlighted in **Fig. 2f**. First, both proteins appear to be largely disordered based on predictions for intrinsically disordered proteins as discussed in more details later. Second, they contain repeat motifs about 30 amino acid long in ES_P1 and *ca.* 20 amino acid in ES_P2. Notable characteristics within the ES_P1 repeats are the presence of basic di-peptides (RR, RK, and KK motifs), a high abundance of Pro appearing as di-peptides (PP) or flanked with non-polar amino acids (IP, IIP, PI, and IP motifs), whereas anionic and polar residues appear as single residues. In ES_P2, basic di-peptides (RR, KR, KK) are also present, while Pro is usually found in the hydrophobic tripeptide PGP and FGP. In contrast to ES_P1, there are almost no anionic residues in the repeat motifs of ES_P2. Third, the N-termini of both proteins contains LC domains highly enriched in Gly and Ser (**Fig. 2f**) residues with a pattern (Gx(0,1)SSGGsXGS-xGGsX(2)SGGxYGxSxGGsX(4)GSxGx(2)GxSx(5)G) (where x can be any amino acid) predicted with bioinformatic tools²⁴. These domains share intriguing homology with LC domains of the FUS RNA-binding protein²³, which is well-established to exhibit LLPS. Fourth, both proteins contain a small number of Cys residues (6 ES_P1 and 3 in ES_P2) that are mostly located near their extremities, namely at positions 16, 2073, and 2080 in ES_P1, and positions 1713 and 1720 in ES_P2. Since the end termini sequences of slime proteins were unknown prior to this work, these Cys residues were not detected in previous studies.

Large MWs complex in the velvet worm slime

In gluten elastomeric proteins²⁵, Cys residues located at both the C- and the N-terminus have been shown to link glutenin subunits through disulfide bonds in a tail-to-head arrangement, which stabilize and provide elasticity to the structure. A similar disulfide bridge is also known to link the heavy and light chains of *Bombyx mori* silk fibroins²⁶. Based on this resemblance, we explored whether disulfide bonds could also assemble slime protein subunits into larger complexes. To this end, we enhanced electrophoretic separation of the slime proteins using long-range SDS-PAGE gels. The proteins from the native slime located in

the high MW region were further separated into 4 distinct bands (band 1, 2, 3 and 4 in **Supplementary Fig. 4**). After treating the slime with DTT to reduce potential disulfide bonds, these bands shifted to the lower MW region (**Supplementary Fig. 4a**, see also **Fig. 2a**) and separated into additional distinct bands, called reduced slime band 1 (RS band 1), 2 (RS band 2, and 3 (RS band 3), confirming that proteins detected in the high MW region of SDS-PAGE gels are in fact complexes linked by disulfide bonds. To identify the components of each disulfide-linked complex in the high MW region, each band of the native slime from the long range gel was cut-off, subjected to DTT reduction, and stacked onto a second SDS-PAGE gel for further separation (**Supplementary Fig. 4b**). Proteomic analysis by LC/MS-MS of bands from this second gel showed that each band consisted of complexes of multiple proteins (**Supplementary Table 3**). Bands 1, 2, and 3 were dominated by ES_P1, whereas band 4 mostly contained ES_P2. Importantly, two additional low MW proteins, called ES_P5 and ES_P6, were detected in these complexes when the peptide fragments were probed against the translated transcriptome (**Supplementary Table 3**). These proteins were distinct from ES_P3 and ES_P4 (no significant match) identified from the middle and low MW regions of the SDS-PAGE gels of the non-reduced slime shown in **Fig. 2a**. ES_P5, with a MW of 74 kDa, was detected in all bands in high abundance. There were 16 Cys, 9.5% Ser/ 5.3% Thr evenly distributed along the primary structure of ES_P5, indicating possible sites for disulfide bond and phosphorylation, respectively. ES_P6 was only detected in band 2 and was rich in charged residues Glu (13.4%) and Lys (14.4%), often in the form of DD or KK dipeptides. Only four Cys residues were found in ES_P6: three at the N-terminus region and one at C-terminal, similar to ES_P1 and ES_P2. The relative content of Ser and Thr was similar to that of ES_P5. However, no phosphorylation modifications were detected in either ES_P5 or ES_P6.

Both phosphorylation and glycosylation staining indicated that the high MW complexes were phosphorylated and glycosylated (**Fig. 2b and 2c**). Phosphorylated residues were previously suggested to mediate nanoglobules formation⁶ or to enhance β -sheet crystalline stacking¹¹, and a similar function has been identified in aquatic caddisworm silk, whereby phosphoserines of heavy fibroin interact with Ca^{2+} ions to β -rich structures and fibers²⁷ via electrostatic interactions. However, phosphorylation sites in the velvet worm slime were only predicted in the high MW slime proteins and not experimentally confirmed⁶. Our Pro-Q stain after disulfide bond reduction (**Fig. 2b**) clearly indicate that phosphorylation was present mostly in the lower MW proteins of the large complexes, particularly in band 2. However,

phosphorylation sites were not detected in either ES_P1 nor ES_P2 but only in peptides from a smaller MW protein (ES_P7) present in bands 1 to 3 at low abundance, at position Thr 133, Tyr 135 and Ser 136. Since phosphorylation was only detected in lower MW slime proteins, its role may be to mediate linkage of the different slime proteins in the larger complex. While our data indicate that the large MW proteins ES_P1 and P2 are not phosphorylated, we cannot rule out that additional phosphorylated sites on the smaller MW slime proteins could be detected by LC MS/MS using different enzymatic treatments.

In comparison, after disulfide bond reduction glycosylation was mainly found in the complex dominated by ES_P1 (**Fig. 2c**). To identify the carbohydrate moieties linked to the slime proteins, we conducted lectin binding assays²⁸. FITC-labelled lectins showed binding to ES slime using a dot blot assay, demonstrating the presence of the specific glycans, including β -D-galactosyl(1-3)-D-N-acetyl-D-galactosamine, L-fucose, terminal α -D-galactosyl, α -D-mannose, and N-acetyl- β -D-glucosaminyl sugars (**Supplementary Fig. 5**). FITC-fluorescence based binding assay measurements indicated that α -D-mannose, and N-acetyl- β -D-glucosaminyl sugars are the most abundant glycans (**Supplementary Fig. 6**).

Figure 3. Structural predictions and LLPS of ES_P1. **a.** Secondary structure domains predicted by AlphaFold indicated as straight vertical lines along the sequence (green: α -helices; red: β -sheets). **b.** Prediction of intrinsically disordered regions within ES_P1 using various bioinformatics tools. **c.** Predicted structure of ES_P1 based on AlphaFold, with regions of secondary conformation mapped within the structure (green: α -helices, red: β -sheets). **d.** Microdroplets of recombinantly expressed N-terminus region of ES_P1 (ES_P1₃₁₋₈₃, located within the purple blocks in Fig. 2e) observed by light microscopy (top), and GFP-encapsulated microdroplets of ES_P1₃₁₋₈₃ observed by fluorescent microscopy (bottom).

Structural predictions and characterization of slime proteins

We used the AlphaFold tool²⁹ for protein structure prediction, which projected that eleven and thirteen β -sheet rich domains existed in ES_P1 and ES_P2 respectively (**Fig. 3a and Supplementary Fig. 7a**), with high predicted local distance difference test (LDDT) scores, indicating correctly predicted local domains based on interatomic distances²⁹. We also subjected both ES_P1 and ES_P2 to a suite of bioinformatic tools used to identify intrinsically disordered regions (IDRs), which all predicted the proteins to be largely disordered (**Fig. 3b and Supplementary Fig. 7b**). Although a high degree of disorder was confirmed, it is interesting to note that AlphaFold prediction indicates the ability of both proteins to acquire localized short secondary structures (**Figure 3c and Supplementary Fig. 7c**), corroborating

crystalline β -sheets identified in the native slime by wide-angle X-ray diffraction Raman spectroscopy¹¹.

Yet, the N-terminal regions of both proteins containing the LC domains enriched in Gly and Ser were predicted to lack any ordered structure, as evidenced by the low LDDT scores. Based on the striking resemblance of the N-termini domains with the LC domains of the FUS protein that exhibits LLPS, we hypothesized that these domains may play a similar role in ES_P1 and ES_P2, namely they may trigger the formation of concentrated droplets through LLPS. To verify this hypothesis, a construct of ES_P1 from position G31 to Y83 (referred as ES_P1₃₁₋₈₃) was recombinantly expressed in *E. coli* and purified. When ES_P1₃₁₋₈₃ construct solubilized in citrate-phosphate buffer (at concentration above 50 μ M) was pipetted within buffer solutions at pH 4.5 to 8, microdroplets 0.4 to 5 μ m in diameter spontaneously formed (**Fig. 3d**), confirming its ability to exhibit LLPS. These droplets preferably formed at 37 °C but were less visible at room temperature (**Supplementary Fig. 8**). To enhance visualization, GFP was added to the buffer solution and subsequently recruited within the droplets during phase separation as verified by fluorescence microscopy. Overall, these data suggest that the LC domain at the N-terminus in ES_P1 may promote LLPS of the slime proteins as a way to concentrate them prior to slime ejection while at the same time preventing premature aggregation. This mechanism is reminiscent of spider silk formation, although in the latter case silk spidroin precursors are concentrated through fully structured domains³⁰ as opposed to IDRs in the velvet worm slime.

Nuclear magnetic resonance (NMR) of the slime

To correlate structural and molecular features of the slime, we conducted ssNMR measurements on ¹³C-enriched slime from velvet worms. Amino acid signals with greater than 5% abundance in either ES_P1 or ES_P2 proteins were assigned to the 1D ¹³C cross polarization with magic angle spinning (CP-MAS) and direct polarization with magic angle spinning (DP-MAS) ssNMR spectra of ¹³C-enriched slime (**Fig. 4a and Supplementary Table 4**). To achieve a higher peak resolution in CP-MAS spectra as well as the acquisition of insensitive nuclei enhanced by polarization transfer (INEPT) spectra, the dried slime was wetted with a few μ l of water. Peaks between 100-160 ppm can be assigned to aromatic carbons of Tyr and Phe, and C ζ of Arg, all of which consist of more than 3 mol. % in both ES_P1 and ES_P2 (**Fig. 4b**). Carboxylic peaks were not observed in the INEPT spectrum due to the absence of directly

bonded protons for polarization transfer. Glu C_γ and Asp C_β peaks were more prominent in the 1D INEPT but suppressed in CP-MAS, indicating their presence in flexible regions of the proteins as well as their involvement in hydrogen bonding in the slime, which could potentially stabilize the nanoglobules in solution. Due to the overall abundance of Gly in the slime (27% predicted by AAA, **Supplementary Table 2**), the ¹³C spectrum of ¹³C-Gly labelled slime was used to identify the Gly C_α peak, and this peak was deconvoluted in the DP-MAS spectrum to identify the relative abundance of Gly in difference secondary conformations in the slime³¹. The resonance at 44.4 ppm could be assigned to Gly C_α in random coil given that the signal was significantly enhanced in the INEPT spectrum, whereas the 45.3 ppm resonance was assigned to Gly C_α in β-sheets³² (**Fig. 4c**). Based on the areas under the deconvoluted peaks, 16% of Gly residues were estimated to be within β-sheets while 84% of Gly were found in random conformations, allowing the latter to flexibly interact with different residues forming both inter- and intra-molecular bonds in the slime. In comparison, 5 out of 187 Gly (2.7%) and 41 out of 166 Gly residues (24.7%) in ES_P1 and ES_P2, respectively, were predicted by AlphaFold to lie within secondary structures in these two proteins (**Fig. 4d**). Next, amino acid residues within the flexible regions of the slime were assigned using 2D INEPT (**Fig. 4e**). Peaks located at 57.4, 71.8, 80.3, and 108.6 ppm in the 2D INEPT spectrum were assigned to glycan carbons C6, C2, C4 and C1, respectively, strongly indicating the presence of 6-carbon monomeric sugar moieties on the slime proteins³³ (**Supplementary Fig. 9**). These results corroborate the presence of galactose/glucose/mannose sugars bound to proteins detected using the lectin-binding assay. Additionally, since the Ser peaks were very prominent in the 1D and 2D INEPT spectra compared to the CP-MAS spectra, we posit that glycosylation likely occurred on these Ser residues. The unassigned peak at 3.4 ppm could arise from an aliphatic group of the lipid moiety³⁴. In the 2D DARR spectrum collected for 500 ms, the observed 45.3 ppm-54.0 ppm cross-peaks were assigned to Gly C_α-Lys C_α inter-residue correlations. Given that 28 pairs of consecutive GK residues exist in ES_P1 and 12 GK pairs in ES_P2, these results suggest that those pairs are located in the rigid structure of the solid slime and therefore involved in its structural arrangement (**Supplementary Fig. 10**).

Figure 4. NMR characterization of the slime. **a.** 1D ^{13}C ssNMR DP-MAS (black), CP-MAS (green) and INEPT (blue) spectra overlaid for ES slime. Residues of interest are labelled in red. **b.** Peaks beyond 100 ppm overlaid in 1D ^{13}C spectra. **c.** Deconvolution of the Gly peak in DP-MAS spectrum. **d.** Mapping of Gly residues (red) on the predicted ES_P1 structure. **e.** 2D ^1H - ^{13}C INEPT of hydrated ES slime.

Fiber formation after thermo-mechanical treatments

From structural predictions/characterization and SDS-PAGE of reduced samples combined with LC/MS-MS, it appears that β -sheet domains and disulfide bonds are important structural and biochemical characteristics of the slime complex. To gain further insights into the role of these features on slime formation, we subjected the re-dissolved slime to thermal denaturation (to disrupt β -sheets) and/or treated it with DTT (to reduce disulfide bonds) and then attempted to draw fibers from the treated solutions. In native conditions, long fibers could readily be drawn (**Supplementary Movie 1**) as expected. When the re-dissolved slime was heat-treated to 70°C, fiber formation was inhibited (**Supplementary Movie 2**), although some weak fibers could sometimes be obtained. In contrast, incubating the slime with a high concentration of DTT (to ensure complete reduction of the disulfide bonds as verified by SDS-PAGE, see **Supplementary Fig. 11**) did not inhibit fiber formation (**Supplementary Movie 3**). Finally, the combination of both heat-treatment and DTT was the most efficient at inhibiting fiber formation, with no fibers observed in all cases (**Supplementary Movie 4**). These results indicate that thermal denaturation of the slime proteins is most efficient at preventing subsequent fiber formation.

Updated fiber formation model

Fiber formation has recently been proposed to occur by shear-induced β -sheet unfolding followed by stabilization mediated by electrostatic interactions of phosphorylated residues¹¹. An earlier suggestion was that intermolecular disulfide bonds help in stabilizing the slime upon ejection¹⁶. Based on our findings, we propose an updated fiber formation model in Onychophora, as illustrated in **Fig. 5**. Our results clearly support that native slime proteins do not exist as monomeric units but as multi-protein complexes linking high MW with low MW proteins by disulfide bonds through Cys residues located at the termini of either ES_P1 or ES_P2 (**Fig. 2a** and **Supplementary Fig. 11**). However, since reducing inter-molecular disulfide bonds in the resolubilized slime does not inhibit fiber drawing (**Supplementary Fig. 11** and **Supplementary Movie 3**), we conclude that that this latter mechanism is not critical for fiber formation. Multi-protein complexes linked by disulfide bridge have previously been identified in *Bombyx mori* silk fibroins consisting of heavy and light protein chains, with the linkage located at the C-terminus of the heavy protein chain³⁵. Disulfide bridges may also occur in the

slime proteins given that Cys residues in ES_P1 and ES_P2 are similarly placed at both termini. Interestingly, in silk fibroins, the light chain component does not directly contribute to protein structure or fiber formation but is crucial for proper cellular secretion of the heavy chain³⁶. β -sheets in the heavy chains, on the other hand, are well-established to provide mechanical stability to silk fibers³⁷. Disulfide-linked complexes in Onychophora slime are discovered here for the first time, and we suggest that the low MW slime proteins may also assist in cellular secretion by preventing early aggregation, but this remains to be validated.

Figure 5. Proposed fiber formation model in Onychophora slime (SMWSP: small molecular weight slime proteins). Slime proteins are mostly concentrated within nanodroplets as disulfide-bonded complexes that also contain lipids. Protein complex and lipid phases may also be found between the droplets but at lower concentration. Upon slime ejection, β -sheets domains in ES_P1 and ES_P2 are aggregation hot spots that mediate shear-induced fiber formation and lipids migrate towards the outside of the fiber to form a hydrophobic coating (note that in the cartoon, β -sheets domains are not to scale).

Baer *et al.*¹¹ suggested that β -sheets present in the initial nanodroplets undergo shear-induced unfolding, such that the content of β -sheets in the drawn fibers decreases compared to the nanodroplets. Our fiber formation experiments from the re-solubilized slime following

heat treatment suggests a different mechanism. Indeed, while bioinformatic predictions clearly indicate that ES_P1 and ES_P2 are mostly in random coil configuration in the soluble state, they also confirm the presence of a few β -sheet domains that were experimentally identified in the re-dissolved slime by WAXS and FTIR¹¹. The latter are thus the only structural elements that can be thermally denatured and since heat-treatment of the re-solubilized slime inhibited fiber formation, these domains appear to be critical for fiber formation. In this picture, the short predicted β -sheets domains of ES_P1 and ES_P2 (**Fig. 3b**) may act as nucleation sites for shear-induced intermolecular aggregation and fiber stabilization, concomitantly occurring with the fusion of nanoglobules into the nascent fiber.

PTMs previously detected in the slime include phosphorylation⁸ and proline hydroxylation⁷, but it was previously unknown which slime proteins are modified. According to LC MS/MS results, the large MW ES_P1 and ES_P2 are confirmed to contain HyPro whereas phosphorylation is restricted to smaller MW proteins. HyPro is the major PTM reported in spidroin proteins³⁸, and has been linked to increase protein stability³⁹ and mechanical properties of fibrous proteins by enhancing protein-protein interactions. It is thus tempting to suggest a similar stabilization role in slime proteins. With regard to phosphorylation, its location within a low MW slime protein is intriguing and may suggest a role in linking different proteins in the complexes, similar to the more robust disulfide bonds. We note that according to the 2D DARR NMR spectrum (**Supplementary Fig. 10**), the rigid region of the slime contains a high number (28) of the positively-charged dipeptide **GK**, most likely from ES_P1, possibly indicating electrostatic pairing with the negatively-charged phosphorylation side-chains.

In addition to proteins, the slime also contains lipids and carbohydrates. While the presence of proteins and lipids within the nanoglobules has been observed with stimulated emission depletion microscopy⁶, further evidence seems necessary to confirm their spatial distribution, especially for the proteins located outside the nanoglobules. At this stage, we assume that proteins and lipids both present inside and outside the nanoglobules. During slime ejection, we suggest that lipids move outwards and quickly coat the fibers (**Fig. 5**) possibly to enhance their hydrophobicity and prevent early disaggregation. Glycosylation is another modification whose role in slime proteins remains elusive⁷. According to our INEPT NMR spectra (**Supplementary Fig. 9**), glycosylation is located on the flexible regions of the slime, most likely on Ser residues (**Fig. 4e**), thus making it unlikely that carbohydrate side-

chains play a structural role. Given the sticky nature of the slime, an adhesive functionality similar to that of the sericin in silk fibroins⁴⁰ is more probable. However, unlike sericins that are distinct proteins from silk fibroins, glycoprotein staining on SDS-PAGE gels (**Fig. 2c**) suggests that glycosylation is post-translationally incorporated to the structural proteins ES_P1 and ES_P2.

Finally, an important finding of our study is the identification of LC sequence domains in the N-termini of both ES_P1 and ES_P2, with the construct ES_P1₃₁₋₈₃ demonstrated to exhibit LLPS (**Fig. 3c**). Hence, our data suggest that the velvet worm slime may be added to the growing list of extracellular biological materials, including mussel fibers⁴¹, spider silk⁴², or squid beak⁴³, whose biofabrication is mediated by intermediate phases formed by LLPS. Exploiting this mechanism, the velvet worm may be able to stockpile protein complexes in a concentrated state within the slime gland while at the same preventing their aggregation prior to fast ejection.

While there remain outstanding questions pertaining to the reversible liquid-to-solid transition of the viscous slime into strong adhesive fibers, the complete molecular characterization of all main slime proteins is an important step in this direction and paves the way towards sustainable fabrication of fully recyclable (bio)polymeric materials.

Methods and materials

Specimens and slime collection.

Specimens of Onychophora were collected in the local secondary forest in Singapore (Permit No NP/RP19-037 obtained from the National Parks Board, Singapore) near the Island's coast and maintained in the plastic boxes with perforated lids. The specimens were kept at cool temperature (20-26 °C) and sphagnum moss were used to fill 2-3 cm layer to maintain sufficient moisture. The worms were fed every week with crickets. To enrich the slime with ¹³C for NMR studies, each cricket was injected with ca. 100 µl of 5 M isotope labelled D-Glucose (U-¹³C₆, 99% from Cambridge Isotope Laboratories) prior to feeding. The slime collection was performed every week, by stimulating the specimens with a brush and directing the ejection of slime to a pipette tip. To maximum the collection, slime was allowed to dry on the tips and stored at -20 degree until further usage. To obtain a liquid slime, the

dried off slime was scratched off the tip as pieces under a dissection microscope. Slime samples are prepared as dried pieces for NMR or redissolved in minimum amount of water for other experiments.

RNA extraction, sequencing and analysis

Total RNAs were extracted from isolated slime glands by Trizol (Thermo Fisher, United State) followed by the manufacture's protocol. In brief, 1mL Trizol solution was added to the isolated slime gland (<50 mg). The sample was mixed by vortex for 5 minutes and further sonicated on ice with microtips, 40% power for 2-3 seconds. After 5 minutes of incubation at room temperature, 0.2 mL chloroform was added into the mixture and the mixture was subjected to centrifuge at 12,000 *g*, 4 °C for 15 minutes. The top aqueous phase was then transferred to a new tube with equal volume of 70% ethanol. Once gently mixed with the pipette, the solution was transferred to RNeasy mini kit column (Qiagen, Germany) and centrifuged at 8,000 *g* for 30 seconds to capture the RNA on the membrane. After two times wash with 700 μ L RW1 solution and 600 μ L RPE solution provided by the kit respectively, the spin column was transferred to a new collection column and total RNA was eluted by 30 μ L RNase-free water with 2 minutes centrifuge at 8,000 *g*. The extracted RNA was stored at -80 °C until further usage in RNA sequencing.

RNA seq libraries were then prepared using Illumina compatible NEXTflex™ Rapid Directional RNA-Seq kit according to manufacturer's protocol. Sequencing were then performed with using 2x150 bp read length on the HiSeq 2000. The raw fastq reads were checked with FastQC⁴⁴ and trimmed by Trimmomatic⁴⁵. The paired end reads were *de novo* assembled by Trinity²¹ with default settings. Protein sequences were then predicted with TransDecoder²².

Verification of full-length sequencing

Re-sequencing primer sets (as shown in SM) were designed in NCBI with the whole RNA seq result as database with default setting. Primers with Tm close to 65 degree with PlantiumTaq were selected for PCR. PCRs are performed according to the manufacture with adaptations. Each 50 μ L reaction consisted of 250 μ M dNTPs, 1x phusion plus buffer (with 1.7 mM MgCl₂, Thermo Fisher, United State), 400 nM/L of each primer, 0.05 U Phusion plus polymerase

(Thermo Fisher, United State) and 20 ng cDNA. PCR products were send for Sanger sequencing (1st base, Singapore) and results were analyzed by Unipro UGENE⁴⁶.

Protein sample preparation & SDS-PAGE staining

Protein concentration in collected slime sample were detected in Qubit (Thermo Fisher, United State) with protein broad range assay kits. ~100 µg were loaded into 4-15% mini gel (Biorad, England). Slime was treated as follows: (i) native slime was directly collected from the worm without further treatment; (ii) dephosphorylated (DP) slime was dephosphorylated with lambda protein phosphatase (New England biolabs, United State) according to manufacturer's protocol; reduced (R) and reduced with heat (RH) was carried out with 5 mM (1 mg/mL) DTT without/with heat to 70 degree for 10 minutes respectively; deglycosylated (DG) slime was treated with protein deglycosylation mix II (New England biolabs, United State) according to manufacturer's protocol with prior denaturing. Electrophoresis was carried out at 150 V constant voltage for 60 minutes until the tracking dye reached the bottom of acrylamide gel. The gel was then incubated in sensitization buffer (30% v/v ethanol, 10% acetic acid and 10% methanol) for 1 hour. After washed with RO water, the gel was subjected to different staining chemical accordingly. Coomassie blue staining were performed according to Amini et al. ⁴⁷ to visualize all protein contents. Pro-Q diamond phosphoprotein gel stain was performed according to the manufacture. Glycoproteins were identified by the Alcian blue/silver staining.

PAGE on slime samples was conducted using Bio-Rad Protean II electrophoresis system by running on hand-cast 6% Bis-tris large format gels prepared at pH 6.4 with MOPS running buffer at pH 7.7 (50 mM MOPS, 50 mM Tris, 0.03% EDTA, 0.1% SDS). 15 µl of 1 mg/ml slime added to 4X NuPAGE™ LDS Sample Buffer was added per well. Electrophoresis was carried out at 150 V constant voltage initially and the current was maintained under 100 A throughout to prevent excessive heating of the PAGE system until the dye-front reached the bottom of gel. The gels were incubated in sensitization buffer containing 30% v/v ethanol, 10% v/v acetic acid and 10% v/v methanol to fix the gel and remove the excess running buffer dye, followed by quick rinse with water. Based on the sample treatment, the gels were stained using ProQ diamond stain based on manufacturer's protocol (Invitrogen) to detect phosphoproteins, Alcian blue stain for glycoprotein detection, or Blue-silver Coomassie stain, and imaged

accordingly. Ovalbumin was used as positive control for phosphorylated protein stain, and fetuin was used as control for glycosylated protein stain.

Liquid Chromatography Tandem Mass Spectrometry (LC MS/MS)

The LC MS/MS experiments was performed according to Amini et al.⁴⁷ with modifications. Briefly, protein bands were excised from the Coomassie-blue stained gel, and subjected to in-gel digestion with 1.25 µg trypsin (PROMEGA, sequential grade) or 1 µg Asp-N (NEB, England) for overnight digestion respectively. Digested peptides were extracted, dried and desalted with Oasis HLB 1cc 30 mg columns (Waters, WAT094225) for LC MS/MS analysis using the Easy-nLC system coupled with a Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Scientific) and separated on a 50 cm x 75 µm Easy-Spray column. Peptides were separated over a 70 min gradient, using mobile phase A (0.1% formic acid in water) and mobile phase B (0.1% formic acid in 95% acetonitrile), and eluted at a constant flow rate of 300 nl/min using 2%- 27% acetonitrile over 45 min, ramped to 50% over 15 min, then to 90% over 5 min and held for 5 min. Acquisition parameters: data dependent acquisition (DDA) with survey scan of 60,000 resolution, AGC target of 4e5, and maximum injection time (IT) of 100 ms; MS/MS collision induced dissociation in ion trap, AGC target of 1.5e4, and maximum IT of 50 ms; collision energy 35%, isolation window 1.2 m/z.

The further separated bands on long range gel were excised from zinc imidazole stain and reduced with 10mM DTT at 70 degree for 10 minutes. Individual bands were then stacked on a 2nd long range gel and electrophoresis was carried out at 150 V constant voltage for another 40 minutes. After Coomassie blue staining, the protein bands from 2nd long range gel were excised and in-gel digestion was performed with 1 µg Asp-N in 100 mM TEAB. Digested peptides were extracted and desalted for LC MS/MS analysis as above, but with a Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific). Separation parameters were as above, and eluted at a constant flow rate of 300 nl/min using 2%- 27% acetonitrile over 45 min, ramped to 55% over 15 min, then to 95% over 5 min and held for 5 min. Acquisition parameters were as above, with isolation window of 1.6 m/z instead.

Peak lists were generated in Proteome Discoverer 2.4 (Thermo Scientific) using Mascot 2.6.1 (Matrix Science) and concatenated forward/decoy protein sequences obtained from RNAseq.

Search parameters: MS precursor mass tolerance 10 ppm, MS/MS fragment mass tolerance 0.8 Da, 3 missed cleavages; static modifications: Carbamidomethyl (C); variable modifications: Acetyl (Protein N-term), Deamidated (NQ), Nitro (Y), Oxidation (M), Oxidation (P), Phospho (ST), Phospho (Y). False discovery rate estimation with 2 levels: Strict = FDR 1%, Medium = FDR 5%. Precursor mass peak (MS1) intensities were quantified by label-free quantification (LFQ) using the Minora feature detector. The peptides presenting in only one of the replicate were removed prior to quantification analysis. Total abundance of unique peptides from the same master protein was then filtered by threshold of 10^7 , and sorted to identify the major proteins in each sample/band.

Data availability

The assembled transcriptome and raw RNAseq data is submitted to NCBI under Bioproject PRJNA806368). The raw spectra and search data were uploaded to the Jpost repository with the following accession numbers: JPST001465 (jPOST) and PXD031722 (ProteomeXchange).

Prediction tools for disordered regions and structure

Full sequences of ES_P1 and ES_P2 were submitted to web servers to detect disordered regions from primary sequences- Espritz⁴⁸, DisEMBL⁴⁹, FoldIndex⁵⁰, PONDR⁵¹, Globplot⁵², IUPred⁵³, Anchor⁵⁴. Secondary structures for ES_P1 and ES_P2 were predicted using AlphaFold's google colab notebook by predictions of each protein as two halves (<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>). Regions with pLDDT score below 50 were specified to be disordered. The PDB structure files were combined using ab initio domain assembly (AIDA) server (<https://aida.godziklab.org/>).

Solid state Nuclear Magnetic Resonance (ssNMR)

Dried slime collected over 3 months was loaded into Bruker 3.2 mm regular wall ZrO₂ magic angle spinning (MAS) rotor. NMR data were collected on 800 MHz Bruker Advance III instrument equipped with a 3.2 mm H/C/N EFree MAS solid-state probe. One dimensional (1D) ¹H–¹³C cross-polarization (CP), ¹³C direct-polarization (DP), 1D and 2D ¹³C Insensitive Nuclei Enhanced Polarization Transfer (INEPT) and 2D ¹³C–¹³C Dipolar-assisted Rotational Resonance (DARR) experiments were performed with the MAS spinning frequency set at

15.151 kHz and the variable temperature set at -12 °C, which gave an actual sample temperature of 11°C based on the external calibration. Chemical shifts were referenced using the DSS scale with adamantane as a secondary standard for ¹³C (downfield signal at 40.48 ppm) and were calculated indirectly for ¹H. The ¹H-¹³C CP spectrum was collected with a contact time of 1200 μs. 83 kHz SPINAL-64 1H decoupling was implemented during data acquisition. 83 kHz SPINAL-64 1H decoupling was implemented during data acquisition. The recycle delays were 1.5 s and 30 s in the 1D CP and DP MAS experiments, respectively, and acquisition time was 19.1 ms in both the experiments. Additional parameters of the 2D ¹³C-¹³C DARR experiment included 1.5 s recycle delay, 45453 Hz sweep width in both dimensions, 14.1 and 0.9 ms acquisition times in the direct and indirect dimensions, respectively. Data were processed using Topspin 3.6.

NMR peak assignment

1D ¹³C peaks of Ile (Cβ, γ, δ at 36.1, 17.4, 14.1 ppm), Val (Cβ, γ at 32.2, 21.4 ppm), Thr (Cβ, γ at 72.6, 21.4 ppm), Leu (Cβ, δ at 42.1, 25.2 ppm), Lys (Cβ, γ, ε at 32.2, 25.2, 42.1 ppm), Glu (Cγ, δ at 36.1, 183.8 ppm), Pro (Cβ, γ, δ at 32.2, 27.2, 50.8 ppm), Ser (Cβ at 63.9 ppm), Asp (Cβ, γ at 38.2, 180.2 ppm) and Gly (Cα) were assigned based on the average chemical shifts⁵⁵. Cδ, Cε, Cζ of Phe, and Cγ, Cδ of Tyr are assigned to peak cluster at 129.8-132.4 ppm, Phe Cγ at 139.2 ppm, Tyr Cε at 118.2 ppm, and Cζ of Tyr/ Arg at 159.5 ppm.

Lectin binding assay

Lectin-binding was tested using dot blot assay and fluorescence measurement using ELISA on FITC-labelled lectins²⁸. Bovine Serum Albumin was used as negative control and Fetuin (WGA-binding glycoprotein) was used as positive control. For dot blot assay, nitrocellulose membrane strips were wetted with 1x Tris-buffered saline (50 mM Tris, 0.5 M NaCl, pH 7.4) with 0.05% Tween® 20 detergent (TBST) for 1 h with constant shaking. Membranes were taken out of the buffer, and 5 μl each of 0.1 mg/ml BSA, slime and fetuin were added onto the nitrocellulose membrane and the drops for left to dry for 30 min. The membranes were incubated in blocking buffer (5% w/v skim milk powder in TBST) for 2 h with continuous shaking. The membranes were washed 2x for 10 min each with TBST buffer to remove any remaining blocking buffer. The membranes were incubated in 1:200 dilution of 1 mg/ml FITC-labelled lectins in TBST buffer, and kept for continuous shaking in the dark for 2 h. The

membranes were washed 2x with TBST buffer, and imaged using ChemiDoc system using Alexa fluor 488 dye settings.

For fluorescent measurements of binding assay, 50 µl triplicates each of varying concentrations of slime were added per well of transparent Nunclon™ Delta Surface 96-well plate (Thermo scientific), and left to incubate overnight at 4°C to ensure protein binding. Each well was washed twice with 100 µl 1X Phosphate-Buffered Saline (1x PBST: 137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 0.1% w/v Tween-20) to remove unbound slime. 1 mg/ml each of FITC-labelled lectins were dissolved in PBST in 1:200 dilution to make working concentrations. 100 µl of FITC-labelled lectins were added to each well, and the plate was incubated for 2 hours in dark. Lectin solution was removed and each well was washed 1x with 100 µl PBST. 100 µl PBST was added to each well and fluorescence values were measured (ex: 490 nm; em: 525 nm). The values were plotted using Originlab.

Recombinant purification of slime proteins

Plasmid for ES_P1₃₁₋₈₃ protein was synthesized (Bio-basic Asia) and cloned into kanamycin-resistant pSUMO-LIC vector with cleavable N-terminal His-SUMO tag and expressed in chloramphenicol-resistant BL21(DE3) Rosetta T1R *E.coli* cells. Glycerol stocks were stored at -80 °C, and were used to inoculate fresh autoclaved Luria Bertani growth medium for overnight growth at 37 °C at 220rpm. These starter cultures were used to grow *E.coli* at large scale in Terrific Broth media at 37 °C until OD_{600nm} reached 1, and induced with induced with 0.5 mM isopropyl-beta-D-1-thiogalactopyranoside (IPTG) at 18°C for 18-20 h. Cells pellets were harvested by centrifugation at 6000 rpm and resuspended in lysis buffer containing 50 mM Tris-HCl, pH 7.8, 0.2 % Triton X-100, with 1 mM phenylmethylsulfonyl fluoride (PMSF) added immediately before lysis using sonicator followed by lyophilizer. Lysed sample is centrifuged at 40000 xg and the protein present in the supernatant was bound to Ni-NTA column (Cytiva) and eluted in 50 mM Tris-HCl buffer at pH 8 containing 500 mM Imidazole, 100 mM NaCl, and 1 mM TCEP. Imiadazole was removed using HiTrap column (Cytiva) before protease cleavage of the N-terminal SUMO tag. The ES_P1₍₃₁₋₈₃₎ was collected by reverse-IMAC binding to Ni-NTA whereby the protein eluted in the flowthrough and the SUMO tag remained bound to the resin. The ES_P1₍₃₁₋₈₃₎ protein stock for LLPS conditions was stored at 4 °C in CHAPS buffer, pH 11.

Amino Acid Analysis

Triplicates of slime solutions in water (100 mg) were hydrolyzed in 500 mL of 6 M HCl solution containing 0.5% phenol in vacuo at 110 °C for 24 hrs. Solvents were removed by centrifugal vacuum and the hydrolysates were washed twice with water. Dried samples were kept at -20 °C prior to analysis. For cysteine determination, triplicates of slime solutions in water were oxidized with 500 mL of performic acid (9:1 formic acid: hydrogen peroxide) in ice for 18 hrs. 100 mL of hydrobromic acid was added to the solution, after which solvents were removed by speed vacuum. Samples were then hydrolyzed in 500 mL of 6 M HCl solution with 0.5% phenol under vacuum at 110 °C for 24 hrs. Solvents were removed by speed vacuum, after which hydrolysates were washed twice with water and kept at -20 °C prior to analysis. For analysis, hydrolyzed samples (0.5 mg/ml) were dissolved in a pH 2.2 citrate buffer containing 0.1 mM Nitrotyrosine as internal standard. For performic acid-treated samples, 0.1 mM Norleucine was used as the internal standard. Composition analysis was performed with an amino acid analyzer S 433 (Sykam, Germany) using a ninhydrin buffer system.

Acknowledgments

This research is funded by ExxonMobil through the Singapore Energy Research Center (SgEC). YTL and RMA thank the support of A*STAR Core funding and the Singapore National Research Foundation under its NRF-SIS “SingMass” scheme (RMS).

Conflict of interest

G.P and A.U are employees of ExxonMobil who funded the project.

References

1. Sahni, V., Harris, J., Blackledge, T. A. & Dhinojwala, A. Cobweb-weaving spiders produce different attachment discs for locomotion and prey capture. *Nat Commun* 3, 1106 (2012).
2. Waite, J. H. Mussel adhesion – essential footwork. *J Exp Biol* 220, 517–530 (2017).
3. Wolff, J. O., Grawe, I., Wirth, M., Karstedt, A. & Gorb, S. N. Spider’s super-glue: thread anchors are composite adhesives with synergistic hierarchical organization. *Soft Matter* 11, 2394–2403 (2015).
4. Sharma, B., Malik, P. & Jain, P. Biopolymer reinforced nanocomposites: A comprehensive review. *Mater Today Commun* 16, 353–363 (2018).

- 706 5. Hayashi, C. Y., Shipley, N. H. & Lewis, R. V. Hypotheses that correlate the sequence,
707 structure, and mechanical properties of spider silk proteins. *Int J Biol Macromol* 24, 271–275
708 (1999).
- 709 6. Baer, A. *et al.* Mechanoresponsive lipid-protein nanoglobules facilitate reversible fibre
710 formation in velvet worm slime. *Nature Communications* 1–7 (2017) doi:10.1038/s41467-
711 017-01142-x.
- 712 7. Benkendorff, K., Beardmore, K., Gooley, A. A., Packer, N. H. & Tait, N. N. Characterisation
713 of the slime gland secretion from the peripatus, *Euperipatoides kanangrensis* (Onychophora:
714 Peripatopsidae). *Comp Biochem Physiology Part B Biochem Mol Biology* 124, 457–465 (1999).
- 715 8. Haritos, V. S. *et al.* Harnessing disorder: onychophorans use highly unstructured proteins,
716 not silks, for prey capture. *Proceedings of the Royal Society B: Biological Sciences* 277, 3255–
717 3263 (2010).
- 718 9. Röper. Analytical investigations on the defensive secretions from *Peripatopsis moseleyi*
719 (Onychophora). *Zeitschrift für Naturforschung* 57–60 (1977).
- 720 10. Baer, A., Hänsch, S., Mayer, G., Harrington, M. J. & Schmidt, S. Reversible Supramolecular
721 Assembly of Velvet Worm Adhesive Fibers via Electrostatic Interactions of Charged
722 Phosphoproteins. *Biomacromolecules* 19, 4034–4043 (2018).
- 723 11. Baer, A. *et al.* Shear-Induced β -Crystallite Unfolding in Condensed Phase Nanodroplets
724 Promotes Fiber Formation in a Biological Adhesive. *ACS nano* 13, 4992–5001 (2019).
- 725 12. Johnston, E. R., Miyagi, Y., Chuah, J.-A., Numata, K. & Serban, M. A. Interplay between Silk
726 Fibroin's Structure and Its Adhesive Properties. *Acs Biomater Sci Eng* 4, 2815–2824 (2018).
- 727 13. Addison, J. B. *et al.* Reversible Assembly of β -Sheet Nanocrystals within Caddisfly Silk.
728 *Biomacromolecules* 15, 1269–1275 (2014).
- 729 14. Baer, A., Schmidt, S., Mayer, G. & Harrington, M. J. Fibers on the Fly: Multiscale
730 Mechanisms of Fiber Formation in the Capture Slime of Velvet Worms. *Integr Comp Biol* 59,
731 1690–1699 (2019).
- 732 15. Matsuhira, T. & Osaki, S. Molecular weight of *Nephila clavata* spider silk. *Polym J* 47, 456–
733 459 (2015).
- 734 16. Graham, L. D., Glattauer, V., Li, D., Tyler, M. J. & Ramshaw, J. A. M. The adhesive skin
735 exudate of *Notaden bennetti* frogs (Anura: Limnodynastidae) has similarities to the prey
736 capture glue of *Euperipatoides* sp. velvet worms (Onychophora: Peripatopsidae). *Comp*
737 *Biochem Physiology Part B Biochem Mol Biology* 165, 250–259 (2013).
- 738 17. Hagn, F. *et al.* A conserved spider silk domain acts as a molecular switch that controls fibre
739 assembly. *Nature* 465, 239–242 (2010).
- 740 18. Askarieh, G. *et al.* Self-assembly of spider silk proteins is controlled by a pH-sensitive relay.
741 *Nature* 465, 236–238 (2010).
- 742 19. Guerette, P. A. *et al.* Accelerating the design of biomimetic materials by integrating RNA-
743 seq with proteomics and materials science. *Nature Biotechnology* 31, 908–915 (2013).
- 744 20. Li, P. *et al.* Phase Transitions in the Assembly of Multi-Valent Signaling Proteins. *Nature*
745 483, 336–340 (2012).

- 746 21. Grabherr, M. G. *et al.* Trinity: reconstructing a full-length transcriptome without a genome
747 from RNA-Seq data. *Nat Biotechnol* 29, 644–652 (2011).
- 748 22. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the
749 Trinity platform for reference generation and analysis. *Nat Protoc* 8, 1494–1512 (2013).
- 750 23. Murray, D. T. *et al.* Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and
751 Phase Separation of Low-Complexity Domains. *Cell* 171, 615–627.e16 (2017).
- 752 24. Jonassen, I., Collins, J. F. & Higgins, D. G. Finding flexible patterns in unaligned protein
753 sequences. *Protein Sci* 4, 1587–1595 (1995).
- 754 25. Shewry, P. R., Halford, N. G., Belton, P. S. & Tatham, A. S. The structure and properties of
755 gluten: an elastic protein from wheat grain. *Philosophical Transactions Royal Soc Lond Ser B*
756 *Biological Sci* 357, 133–142 (2002).
- 757 26. Inoue, S. *et al.* Silk Fibroin of Bombyx mori Is Secreted, Assembling a High Molecular Mass
758 Elementary Unit Consisting of H-chain, L-chain, and P25, with a 6:6:1 Molar Ratio*. *J Biol Chem*
759 275, 40517–40528 (2000).
- 760 27. Ashton, N. N. & Stewart, R. J. Aquatic caddisworm silk is solidified by environmental metal
761 ions during the natural fiber-spinning process. *Faseb J* 33, 572–583 (2019).
- 762 28. Uchiyama, N. *et al.* Development of a Lectin Microarray Based on an Evanescent-Field
763 Fluorescence Principle. *Methods Enzymol* 415, 341–351 (2006).
- 764 29. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 596,
765 583–589 (2021).
- 766 30. Rising, A. & Johansson, J. Toward spinning artificial spider silk. *Nat Chem Biol* 11, 309–315
767 (2015).
- 768 31. Holland, G. P., Creager, M. S., Jenkins, J. E., Lewis, R. V. & Yarger, J. L. Determining
769 Secondary Structure in Spider Dragline Silk by Carbon–Carbon Correlation Solid-State NMR
770 Spectroscopy. *J Am Chem Soc* 130, 9871–9877 (2008).
- 771 32. Spera, S. & Bax, A. Empirical correlation between protein backbone conformation and
772 C.alpha. and C.beta. ¹³C nuclear magnetic resonance chemical shifts. *J Am Chem Soc* 113,
773 5490–5492 (1991).
- 774 33. Arnold, A. A. *et al.* Identification of lipid and saccharide constituents of whole microalgal
775 cells by ¹³C solid-state NMR. *Biochimica Et Biophysica Acta Bba - Biomembr* 1848, 369–377
776 (2015).
- 777 34. Knothe, G. & Nelsen, T. C. Evaluation of the ¹³C NMR signals of saturated carbons in some
778 long-chain compounds. *J Chem Soc Perkin Transactions 2* 2019–2026 (1998)
779 doi:10.1039/a801617h.
- 780 35. Zhou, C. *et al.* Silk fibroin: Structural implications of a remarkable amino acid sequence.
781 *Proteins Struct Funct Bioinform* 44, 119–122 (2001).
- 782 36. Mori, K. *et al.* Production of a chimeric fibroin light-chain polypeptide in a fibroin
783 secretion-deficient naked pupa mutant of the silkworm Bombyx mori. *J Mol Biol* 251, 217–28
784 (1995).

785 37. Keten, S., Xu, Z., Ihle, B. & Buehler, M. J. Nanoconfinement controls stiffness, strength and
786 mechanical toughness of β -sheet crystals in silk. *Nat Mater* 9, 359–367 (2010).

787 38. Santos-Pinto, J. R. A. dos, Arcuri, H. A., Esteves, F. G., Palma, M. S. & Lubec, G. Spider silk
788 proteome provides insight into the structural characterization of *Nephila clavipes* flagelliform
789 spidroin. *Sci Rep-uk* 8, 14674 (2018).

790 39. Jr., W. G. K. PROLINE HYDROXYLATION AND GENE EXPRESSION. *Annu Rev Biochem* 74,
791 115–128 (2005).

792 40. Kunz, R. I., Brancalho, R. M. C., Ribeiro, L. de F. C. & Natali, M. R. M. Silkworm Sericin:
793 Properties and Biomedical Applications. *Biomed Res Int* 2016, 1–19 (2016).

794 41. Priemel, T. *et al.* Microfluidic-like fabrication of metal ion–cured bioadhesives by mussels.
795 *Science* 374, 206–211 (2021).

796 42. Malay, A. D. *et al.* Spider silk self-assembly via modular liquid-liquid phase separation and
797 nanofibrillation. *Sci Adv* 6, eabb6030 (2020).

798 43. Tan, Y. *et al.* Infiltration of chitin by protein coacervates defines the squid beak mechanical
799 gradient. *Nat Chem Biol* 11, 488–495 (2015).

800 44. Andrews & S. FastQC: A Quality Control Tool for High Throughput Sequence Data.
801 <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2010).

802 45. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
803 sequence data. *Bioinformatics* 30, 2114–2120 (2014).

804 46. Okonechnikov, K., Golosova, O., Fursov, M. & team, U. Unipro UGENE: a unified
805 bioinformatics toolkit. *Bioinformatics* 28, 1166–1167 (2012).

806 47. Amini, S. *et al.* A diecast mineralization process forms the tough mantis shrimp dactyl club.
807 *Proceedings of the National Academy of Sciences* 116, 8685–8692 (2019).

808 48. Walsh, I., Martin, A. J. M., Domenico, T. D. & Tosatto, S. C. E. ESpritz: accurate and fast
809 prediction of protein disorder. *Bioinformatics* 28, 503–509 (2012).

810 49. Linding, R. *et al.* Protein Disorder Prediction Implications for Structural Proteomics.
811 *Structure* 11, 1453–1459 (2003).

812 50. Prilusky, J. *et al.* FoldIndex: a simple tool to predict whether a given protein sequence is
813 intrinsically unfolded. *Bioinformatics* 21, 3435–3438 (2005).

814 51. Romero, P. *et al.* Sequence complexity of disordered protein. *Proteins Struct Funct*
815 *Bioinform* 42, 38–48 (2001).

816 52. Linding, R., Russell, R. B., Neduva, V. & Gibson, T. J. GlobPlot: exploring protein sequences
817 for globularity and disorder. *Nucleic Acids Res* 31, 3701–3708 (2003).

818 53. Erds, G., Pajkos, M. & Dosztnyi, Z. IUPred3: prediction of protein disorder enhanced
819 with unambiguous experimental annotation and visualization of evolutionary conservation.
820 *Nucleic Acids Res* 49, gkab408- (2021).

821 54. Dosztnyi, Z., Mszros, B. & Simon, I. ANCHOR: web server for predicting protein binding
822 regions in disordered proteins. *Bioinformatics* 25, 2745–2746 (2009).

823 55. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. & Sykes, B. D. ^1H , ^{13}C and ^{15}N random
824 coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor
825 effects. *J Biomol Nmr* 5, 67–81 (1995).

826