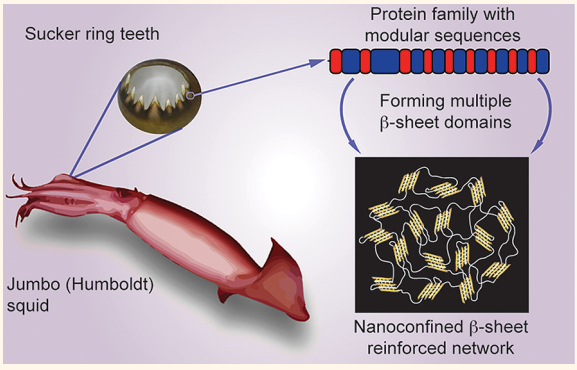
Nanoconfined β-Sheets Mechanically Reinforce the Supra-Biomolecular Network of Robust Squid Sucker Ring Teeth

Paul A. Guerette,[†,‡,⊗] Shawn Hoon,[§,⊥,⊗] Dawei Ding,[†] Shahrouz Amini,[†] Admir Masic,[∥] Vydianathan Ravi,[#] Byrappa Venkatesh,[#] James C. Weaver,[▽] and Ali Miserez[†,⊥,*]

[†]School of Materials Science and Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, [‡]Energy Research Institute at Nanyang Technological University (ERI@N), 50 Nanyang Drive, Singapore 637553, [§]Molecular Engineering Lab, Biomedical Sciences Institutes, A*STAR, 61 Biopolis Drive, Proteos, Singapore 138673, [⊥]School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, Singapore 637551, [∥]Department of Biomaterials, Max-Planck Institute of Colloids and Interfaces, Research Campus Golm, 14424 Potsdam, Germany, [#]Institute of Molecular and Cell Biology, A*STAR, 61 Biopolis Drive, Proteos, Singapore 138673, and [▽]Wyss Institute for Biologically Inspired Engineering, Harvard University, 60 Oxford Street, Cambridge, Massachusetts 02138, United States. [⊗]These two authors contributed equally to this manuscript.

ABSTRACT The predatory efficiency of squid and cuttlefish (superorder Decapodiformes) is enhanced by robust Sucker Ring Teeth (SRT) that perform grappling functions during prey capture. Here, we show that SRT are composed entirely of related structural "suckerin" proteins whose modular designs enable the formation of nanoconfined β-sheet-reinforced polymer networks. Thirty-seven previously undiscovered suckerins were identified from transcriptomes assembled from three distantly related decapodiform cephalopods. Similarity in modular sequence design and exon−intron architecture suggests that suckerins are encoded by a multigene family. Phylogenetic analysis supports this view, revealing that suckerin genes originated in a common ancestor ~350 MYa and indicating that nanoconfined β-sheet reinforcement is an ancient strategy to create robust bulk biomaterials. X-ray diffraction, nanomechanical, and micro-Raman spectroscopy measurements confirm that the modular design of the suckerins facilitates the formation of β-sheets of precise nanoscale dimensions and enables their assembly into structurally robust supramolecular networks stabilized by cooperative hydrogen bonding. The suckerin gene family has likely played a key role in the evolutionary success of decapodiform cephalopods and provides a large molecular toolbox for biomimetic materials engineering.



KEYWORDS: biopolymer · nanoconfinement · β-sheets · biomimetic · modular proteins · silks · squid · suckerin

The evolutionary success of squid and cuttlefish is intimately linked to their predatory prowess, which is greatly enabled by the use of sucker ring teeth (SRT) that line their arms and tentacles and perform essential grappling functions during prey capture. SRT display a regular nanotubular architecture and exhibit an elastic modulus of 6−8 GPa in the dry state and 2−4 GPa under hydrated conditions.[1] Natural and synthetic bulk polymers are usually mechanically stabilized by either (i) chain entanglements as observed in classical amorphous thermoplastic polymers; (ii) dense covalent interchain cross-linking as found in thermoset resins, insect exoskeletons,[2] and squid beaks;[3] or (iii) the addition of a much stiffer reinforcement phase such as fine mineral particles in bone and other biominerals[4] or in engineered nanocomposites.[5] A less-common mechanism discovered in marine worm jaws and arthropod predatory appendages involves transition metal-coordination bonding.[6,7] In contrast, SRT are entirely proteinaceous, contain silk-like β-sheet structures, and are devoid of minerals, metals, and covalent cross-links. The SRT therefore represents a distinct and intriguing model system for the bio-inspired design of robust structural polymers.[8]

* Address correspondence to ali.miserez@ntu.edu.sg.

Received for review April 18, 2014 and accepted June 9, 2014.

Published online June 09, 2014
10.1021/nn502149u

Elucidation of the molecular design of the SRT at multiple length scales is key in order to reveal fundamental insights into the molecular origins of these high-performance materials. From a biological perspective, obtaining the complete sequence of SRT constitutive proteins across species that diverged hundreds of millions years ago has the potential to reveal the molecular basis for the role of SRT in the evolution and diversification of squid and cuttlefish. From an engineering perspective, a comparative analysis of the range of molecular designs may reveal commonalities and differences that impact material performance and would therefore inform our ability to tune the properties of "suckerin"-based synthetic materials.

In this study, we generated transcriptomes from the sucker tissue of three distantly related cephalopods and discovered that SRT in each species are assembled entirely from highly modular suckerin proteins containing peptide building blocks that are reminiscent of those found in silk proteins. Phylogenetic analysis suggests that the suckerins are encoded by an ancient gene family that arose through gene duplication and diverged into five distinct architectures that are conserved across the three species. Synchrotron Wide-Angle X-ray Scattering (WAXS) showed that SRT contain isotropically oriented β-sheet nanocrystals of precise dimensions that reinforce an amorphous network. The β-sheet dimensions are shown to be directly dictated by the suckerin primary amino acid sequence, including the closely conserved lengths of silk-like β-sheet forming modules and the precise location of proline residues that constrain β-sheet size. Finally, we conducted concomitant nanomechanical and micro-Raman spectroscopy measurements, which provided direct evidence that the nanoconfined β-sheets are responsible for the SRT's impressive mechanical properties. This study thus provides a novel and comprehensive molecular toolbox for the design of precisely tuned, robust, and readily processable biopolymers that are stabilized by hydrogen bond interactions and nanoconfined β-sheets.

RESULTS

Identification and Characterization of the Suckerin Gene/Protein Family. We previously identified and sequenced a major constituent of the SRT, a protein named suckerin-39 from the Humboldt squid *Dosidicus gigas* (*D. gigas*) that exhibits an extreme modular architecture.[8] In the current study, we investigated the protein composition and molecular design of SRT proteins from three distantly related decapodiform cephalopods, namely, *D. gigas* (Order Oegeopsida), the bigfin reef squid *Sepioteuthis lessoniana* (*S. lessoniana*, Order Myopsida), and the golden cuttlefish *Sepia esculenta* (*S. esculenta*, Order Sepiida) (Figure 1A). SDS-PAGE and 2D isoelectric-focusing show that, for all three species examined, the SRT (Figure 1B) are assembled from a mixture of proteins whose molecular weights range from ~5 to 60 kDa and whose isoelectric points (p*I*'s) fall in a relatively narrow window between pH 7 and 10 (Figure 1C and Figure S1). These data also revealed that *D. gigas* SRT contain a larger repertoire of proteins than those from *S. lessoniana* and *S. esculenta*.

RNA-seq transcripts generated from the tentacle sucker tissue of each of the three species were searched using the amino acid composition profile and primary amino acid sequences of *D. gigas* suckerin-39. This approach, combined with RACE-PCR, resulted in the identification and complete sequencing of an additional 37 unique genes encoding highly modular proteins with molecular weights ranging from ~5 to 57 kDa and p*I*'s in the 7−10 range. Some of the suckerin transcripts were among the most highly expressed in the sucker tissue while others were expressed at much lower levels (Table S1). An exhaustive search of the NCBI and Uniprot protein databases did not yield statistically significant hits (E-value cutoff of 10) to known proteins, supporting the view that the suckerins represent a unique class of structural proteins. The suckerins typically have a ~17−23 amino acid signal peptide and the full length proteins exhibit similarity in amino acid composition, with a heavy bias toward Gly, Tyr, and His, and to a lesser extent Leu, Ala, Thr, Ser, and Val (Figure S2). Similarity at the primary amino acid sequence level is also evident, with ~20−90% identity between all 38 suckerins. The primary sequences and modular designs of the suckerins are shown in Figures 1D and S3−S5. Most suckerins contain sets of commonly occurring small peptide modules, including GGY and GGLY (Figures S6 and S7). GGY peptides are present in other structural proteins including silk proteins,[9−11] shell matrix proteins,[12,13] crocodile skin β-keratins[14] and insect cuticle proteins,[15] suggesting convergent evolutionary origins of this peptide motif. Medium sized modules are also evident in the suckerins, where we use [M1] to designate ~3−15 amino acid long Ala-, Val-, Thr-, Ser-, and His-rich modules and [M2] to designate the ensemble of repetitive and nonrepetitive Gly-rich sequences. [M1] and [M2] frequently occur in tandem and are often flanked by Pro residues. The combined Pro[M1]Pro[M2] unit (designated as a large module [L]) ranges in length from ~15 to 68 amino acids and is reiterated 3−13 times in the different proteins. Similarities and distinctions in the large-scale modular architectures between suckerins are also evident (Figure 2). Strikingly, some suckerins exhibit extreme conservation of their large-scale modular design, while others do not. It is also notable that different suckerins exhibit length differences in the Gly-rich [M2] regions, a feature that may have a direct impact on SRT assembly and mechanics.

Similarity in amino acid composition, primary sequence, and modular design suggests that the suckerins are encoded by members of a multigene family.

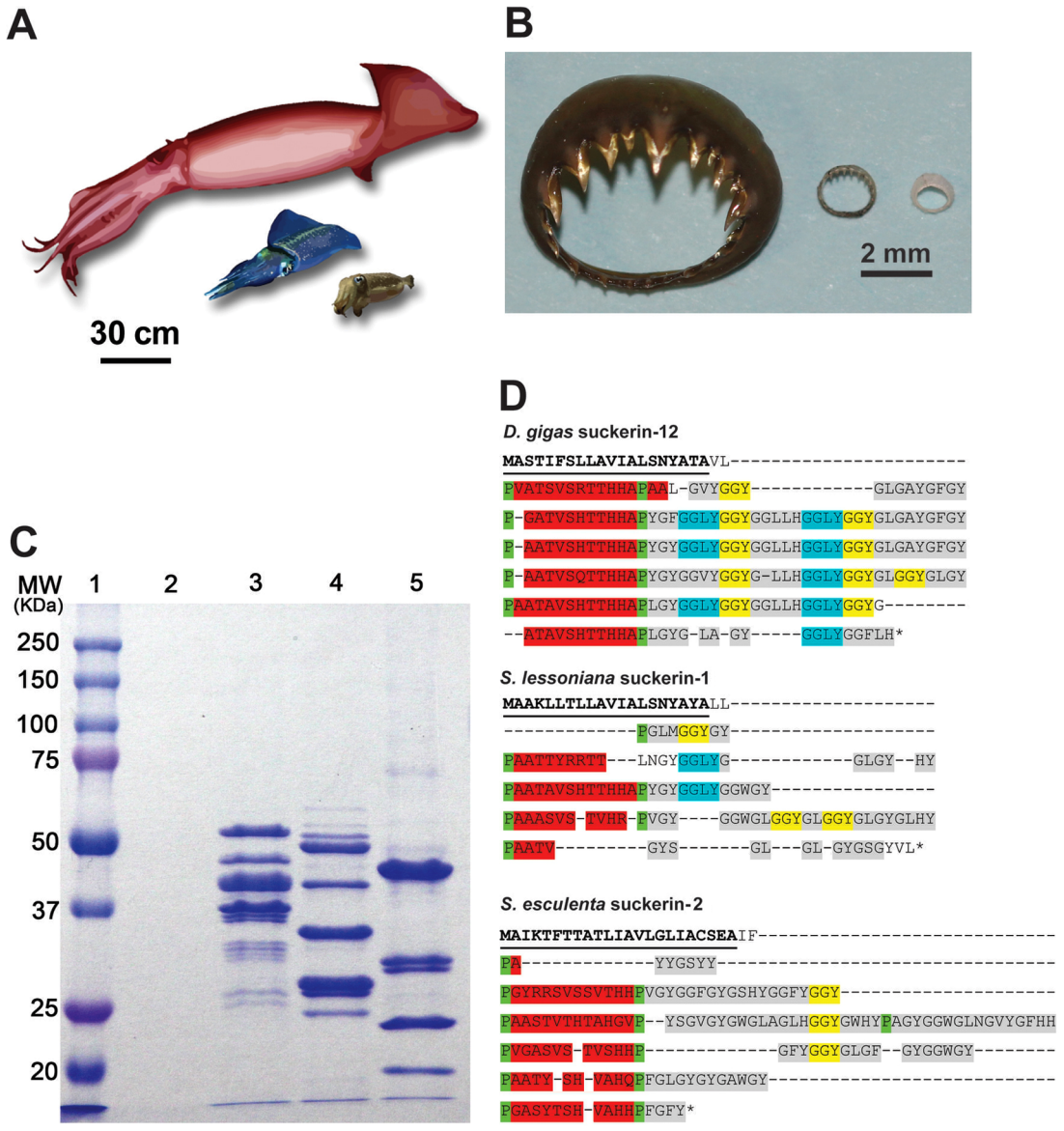


Figure 1. Morphology and composition of SRT from three distantly related cephalopods. (A) Scalar relationships of *D. gigas* (top), *S. lessoniana* (middle), and *S. esculenta* (bottom) and their respective SRT (B). (C) SDS-PAGE of SRT proteins from *D. gigas* (lane 3), *S. lessoniana* (lane 4), and *S. esculenta* (lane 5). (D) Primary amino acid sequence and modular sequence alignments for representative suckerin proteins from all three species. [M1] are highlighted in red and common tri- and tetra-peptide modules in [M2] are highlighted in yellow and blue, respectively. Nonrepetitive Gly-rich sequences in [M2] are shown in gray. Note the regular placement of proline residues (green).

To determine their evolutionary relationships, we generated phylogenetic trees[16] of all suckerins using Maximum Parsimony (MP), Maximum Likelihood (ML), and Neighbor Joining (NJ) methods. On the basis of bootstrap support values and congruence between the three phylogenetic methods, we were able to classify the genes into at least six distinct phylogenetic clades (Figures 3 and S8). All six clades are supported by the three (NJ, ML, and MP) phylogenetic methods. This implies that the present day suckerin genes arose from at least six ancestral suckerin genes through gene duplication. Five of the six clades contain proteins from all three species studied. Because these species shared a common ancestor ~354 MYa,[17] our analysis suggests that the suckerin gene family arose during or before the Devonian period.

Across the 38 suckerin proteins, variations in size and the presence or absence of small and large modules indicate that genetic divergence has involved both gene duplication and segmental expansion/deletion. This is most evident in *D. gigas*, which exhibits an approximately 3-fold expansion in its suckerin repertoire and a concomitant increase in modularity and the number of tandem repeats within individual genes and proteins. To clarify the relationship between the suckerin genes, we used genomic DNA analysis to determined the exon−intron structure of three representative *D. gigas* suckerin genes predicted to belong

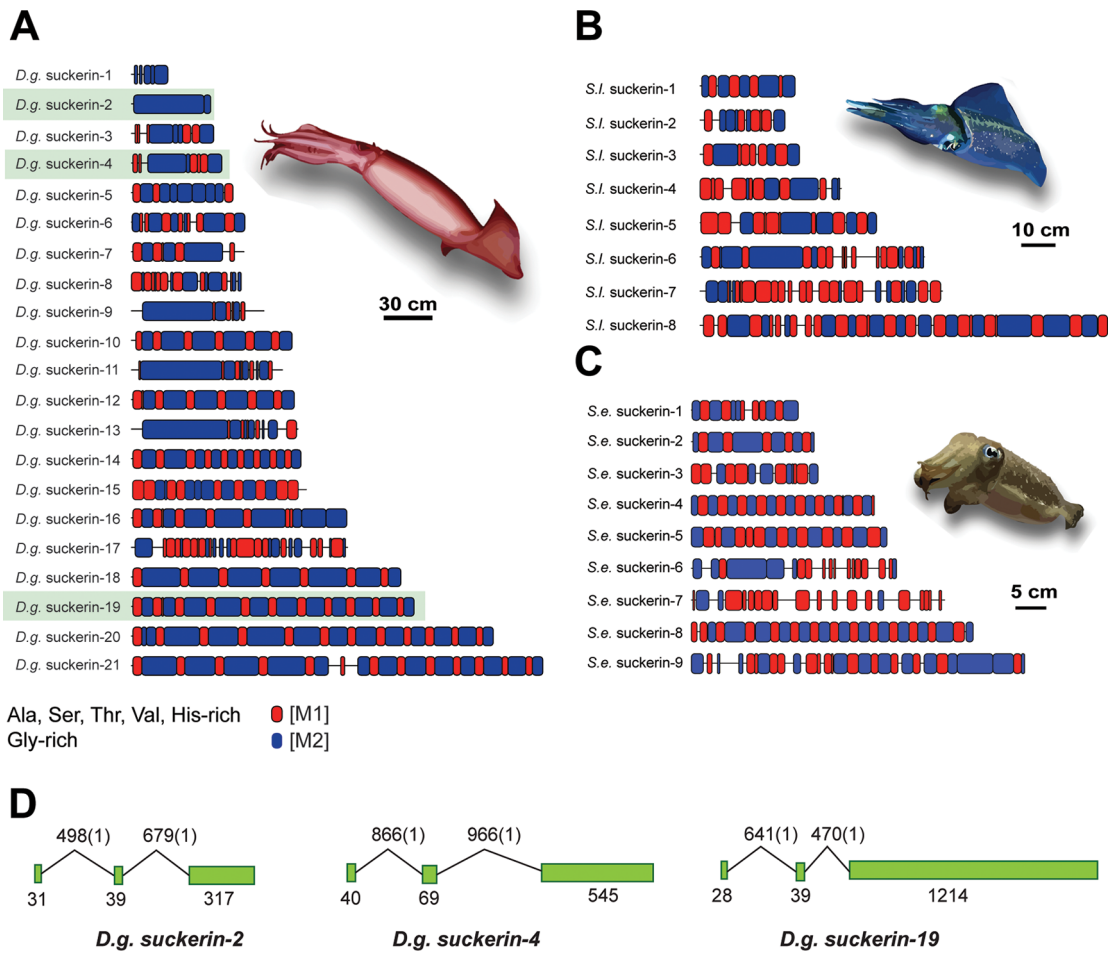


Figure 2. Large-scale modular architecture of suckerin proteins. (A) *D. gigas*, (suckerins for which genomic organization was determined are highlighted in green), (B) *S. lessoniana*, and (C) *S. esculenta*. (D) Exon−intron structure of representative *D. gigas* suckerin genes. Exon lengths are indicated under the green exon boxes. Intron lengths are indicated above and intron phases in parentheses.

to different clades (Figures 2D and S9). Despite markedly different modular architectures, all three genes had a strikingly similar exon−intron organization with all introns present in phase 1 suggesting they were all derived from a common ancestral gene (Figure 2D). The coding sequences of all three genes are divided among three exons and separated by relatively short introns. While the signal peptide is split between the first two exons, the bulk of the coding sequence is located in the third exon where molecular divergence has primarily occurred after gene duplication. Mechanisms by which this organization arises may include slippage of DNA polymerase, nonreciprocal homologous crossing-over and/or gene conversion, as is observed in genes encoding other highly modular proteins.[18,19] Like spider silk proteins, suckerins are encoded by GC-rich sequences that are known recombination hot-spots.[18,20,21] The occurrence of these regions, combined with a high degree of sequential homologous modular sequences that can facilitate unequal crossing over, may have influenced the modular design and divergence of the suckerin gene family. Understanding the molecular mechanisms that underlie the evolution of modular proteins remains a major challenge.[19] Elucidation of these mechanisms in the suckerin gene family should provide unique insights into the biomechanical and adaptive roles of the suckerin proteins.

Identification and Characterization of Nanoconfined β-Sheets in the SRT. The structure of native SRT was then investigated with the goal of linking the primary amino acid sequence designs of the suckerins with SRT structure, hierarchical design, and mechanical function. Previous polarized micro-Raman spectroscopy studies on the *D. gigas* SRT revealed randomly oriented β-sheets that stabilize a silk-like protein polymer network.[8] We characterized the nanoscale organization of the SRT in greater detail by conducting high-energy Wide Angle X-ray Scattering (WAXS), which revealed a circular scattering pattern with equal intensity at all azimuths, denoting a random orientation of crystalline domains (Figure 4A). Integrating across all azimuthal angles (Figure 4B) revealed reflection positions consistent with silkworm[22] and spider silks,[23] including the most

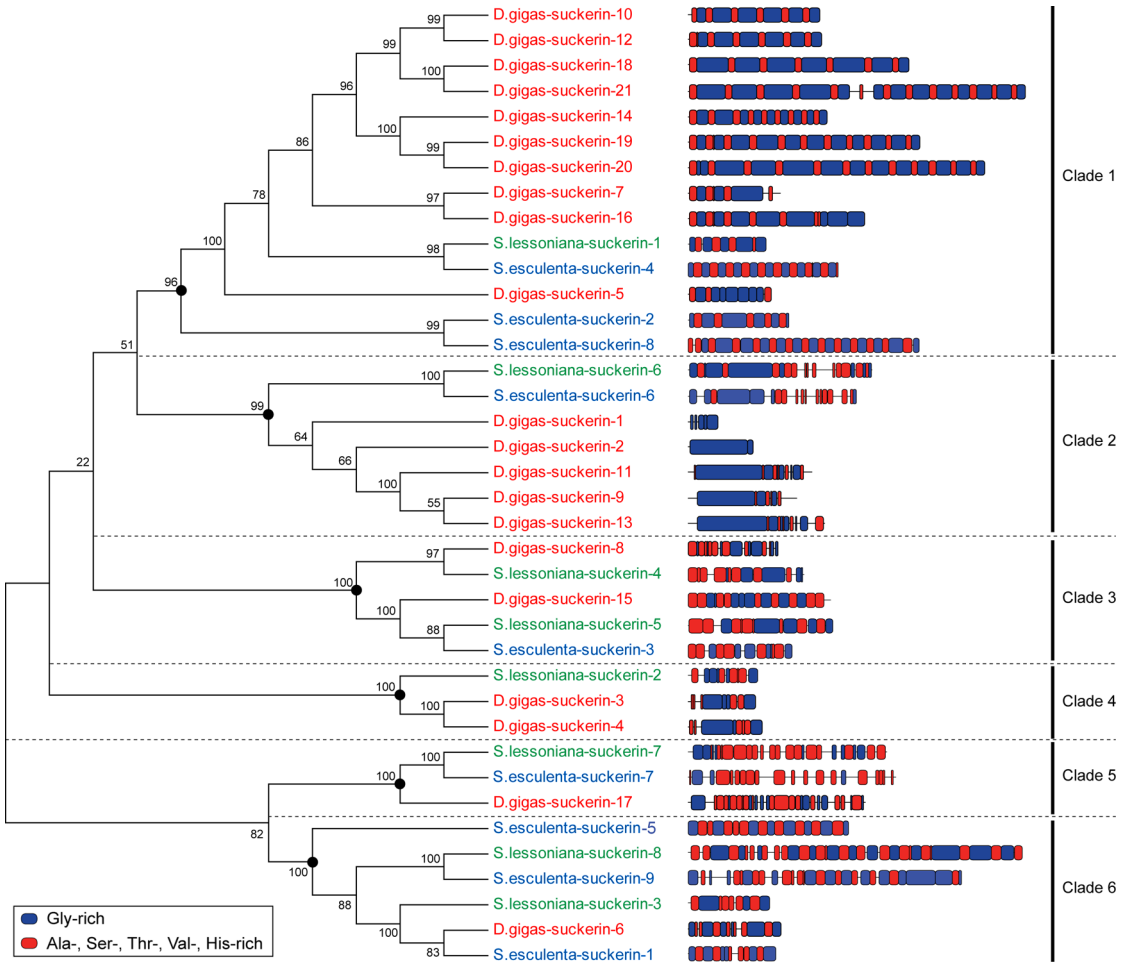


Figure 3. Phylogenetic relationships of all known suckerin proteins. (Left) Neighbor Joining tree of all known suckerin proteins from *D. gigas* (red), *S. lessoniana* (green), and *S. esculenta* (blue). Values at nodes denote bootstrap percentages. (Right) Large-scale modular architecture of suckerin proteins from Figure 2. Black circles denote clades that are supported by all three phylogenetic methods (NJ, MP and ML, see Materials and Methods and Figure S8).

intense peaks attributed to the combination of [120] and [200] reflections, the [002] reflection (along the chain direction), and the [100] reflection at lower *Q*-values. Notably, the overall 2D pattern is highly reminiscent of microfibers assembled from a fibroin solution,[24] which also exhibit a powder-type diffraction pattern. Using peak broadening analysis with the classic Scherrer equation on the deconvoluted pattern, we estimated β-sheet dimensions of 2.4−2.6 nm in the H-bond direction and 3−3.5 nm along the peptide backbone direction (Figure 4B, inset), corresponding to 5 strands in width and ∼8−10 amino acids in length, respectively. Underlying the crystalline reflections, an amorphous halo was also detected, suggesting an overall semicrystalline structure made of amorphous domains strengthened by randomly oriented nanoconfined β-sheets (Figure 4C). β-sheets of strikingly similar dimensions have previously been shown by computational methods to confer enhanced strength and toughness *via* the cooperative involvement of hydrogen bonds.[25]

We next sought to determine how the primary suckerin amino acid sequence enables β-sheet formation while simultaneously placing limits on their overall dimensions. Sequences in [M1] bear similarity to β-sheet forming poly-Ala sequences of spider dragline silks and the Val-Thr motifs that form β-sheets in spider viscid silk.[26] However, the [M1] motifs are less repetitive than spidroin or silkworm fibroin sequences, suggesting that irregularity in side chain chemistry may limit intersheet crystal packing. The suckerin proteins exhibit rigorous positional conservation of Pro (Figures 4D and S3−S5), a known β-sheet disruptor,[27,28] that likely limits β-sheet size and could thus influence the SRT's mechanical properties. We determined the mean [M1] inter-Pro distances in the suckerins for each species to be ∼12−13 residues (Figures 4F and S10). Considering that residues directly adjacent to Pro are likely involved in disordered structure or β-turns, it is reasonable to assume that ∼10−11 residues within [M1] can form β-strands ∼3.1−3.5 nm in length. This is in very good agreement with the 3−3.5 nm long β-strands calculated from WAXS data and suggests that local stretches of ∼10−11 amino acids within [M1] adopt β-sheet conformations where Pro constrains the dimensions

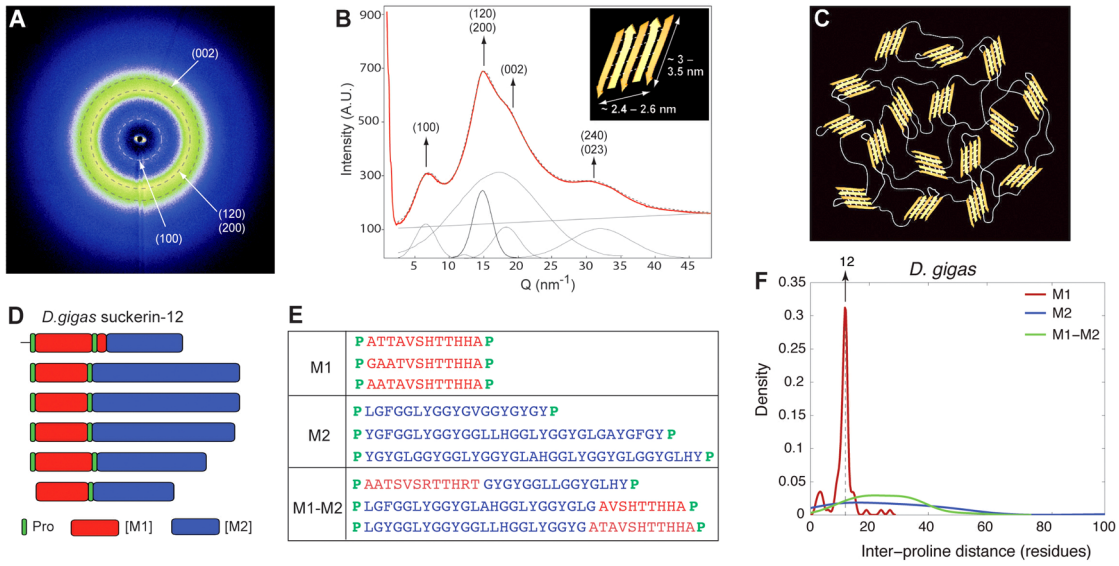


Figure 4. Synchrotron X-ray diffraction of SRT. (A) WAXS pattern of SRT (tip region) and (B) integrated spectrum (all azimuthal angles), indicating reflection positions consistent with β-sheets (inset: estimation of β-sheet size using the Scherrer analysis). Deconvolution indicates a broad amorphous halo. (C) Semicrystalline structure of SRT, consisting of randomly oriented β-sheets within an amorphous matrix. (D) Modular architecture of representative suckerin showing the regular placement of Pro flanking the [M1] modules. (E) Representative inter-proline spacing of [M1], [M2], and [M1−M2] sequences. (F) Distribution of the number of residues between consecutive Pro for all *D. gigas* suckerins.

of the β-sheets. While Gly residues maintain conformational flexibility and often reduce the propensity for β-sheet formation,[27,28] it cannot be fully ruled out that the Gly-rich suckerin domains may also participate in β-sheet formation.

Contribution of Hydrogen Bonds and Nanoconfined β-Sheets to SRT Mechanics. To further dissect the relationship between suckerin structure and SRT mechanics, we conducted complementary nanoindentation and micro-Raman spectroscopy measurements, using conditions designed to target and disrupt hydrophobic interactions and hydrogen bonds. In the dry, hydrated, and ethanol treated samples, the elastic modulus (E) did not vary significantly over time (Figure 5A). While hydrophobic residues are present in [M1] and [M2], ethanol treated samples exhibited similar properties to dry SRT samples suggesting only nominal contributions of hydrophobic interactions. On the other hand, urea treatments, which disrupt hydrogen bonds, resulted in clear decreases in E, with decay plateaus correlating with urea concentration. At the highest concentration of urea, the final modulus was ~20 MPa, which is dramatically lower than values obtained under both dry and hydrated conditions. These data provide the first direct evidence of the key contribution of H-bonding to SRT mechanics and function. In addition, Raman spectroscopy measurements were conducted in parallel to evaluate changes in secondary structure with urea treatment (Figure 5B). Incubation of SRT in 0.2 M urea lead to a distinct intensity decrease of the specific amide III Raman shifts 1236 and 1339 cm^{-1}, which indicates a loss of β-sheet structure.[29] We thus infer that there exists a direct correlation between decrease in modulus and the disruption of β-sheets as illustrated in Figure 5C. Taken together, these data indicate that hydrogen bonds localized to nanoconfined β-sheets play a key role in stabilizing and strengthening SRT.

CONCLUSIONS

The SRT system reveals unique insights relevant to both materials engineering and the biology of cephalopods. From a materials science perspective, SRT represent a new class of intriguing load-bearing biological polymers with a distinctive structural reinforcement strategy. They are composed solely of proteins and do not contain secondary, mechanically distinct phases such as minerals or polysaccharides. Interchain covalent cross-links are also absent, and as a result SRT can be thermally processed like classical thermoplastic polymers in a fully reversible fashion,[8] while exhibiting bulk structural properties that rival those of strong engineering thermoset polymers. The present study elucidates in greater detail the key molecular mechanisms that contribute to this intriguing combination of physicochemical characteristics. Our results demonstrate that SRT are assembled from a family of highly modular suckerin proteins. These natural block copolymers are arranged into large supramolecular networks consisting of an amorphous structure stabilized by nanoconfined and randomly oriented β-sheets, resulting in the impressive functional mechanics of the SRT. The dimensions of the nanoscale β-sheets were found to be precisely governed by the regular placement of Pro residues flanking Ala-, Val-, Thr-, Ser-, and His-rich [M1] domains that are reminiscent of β-sheet forming sequences in silks. The primary amino acid sequences of

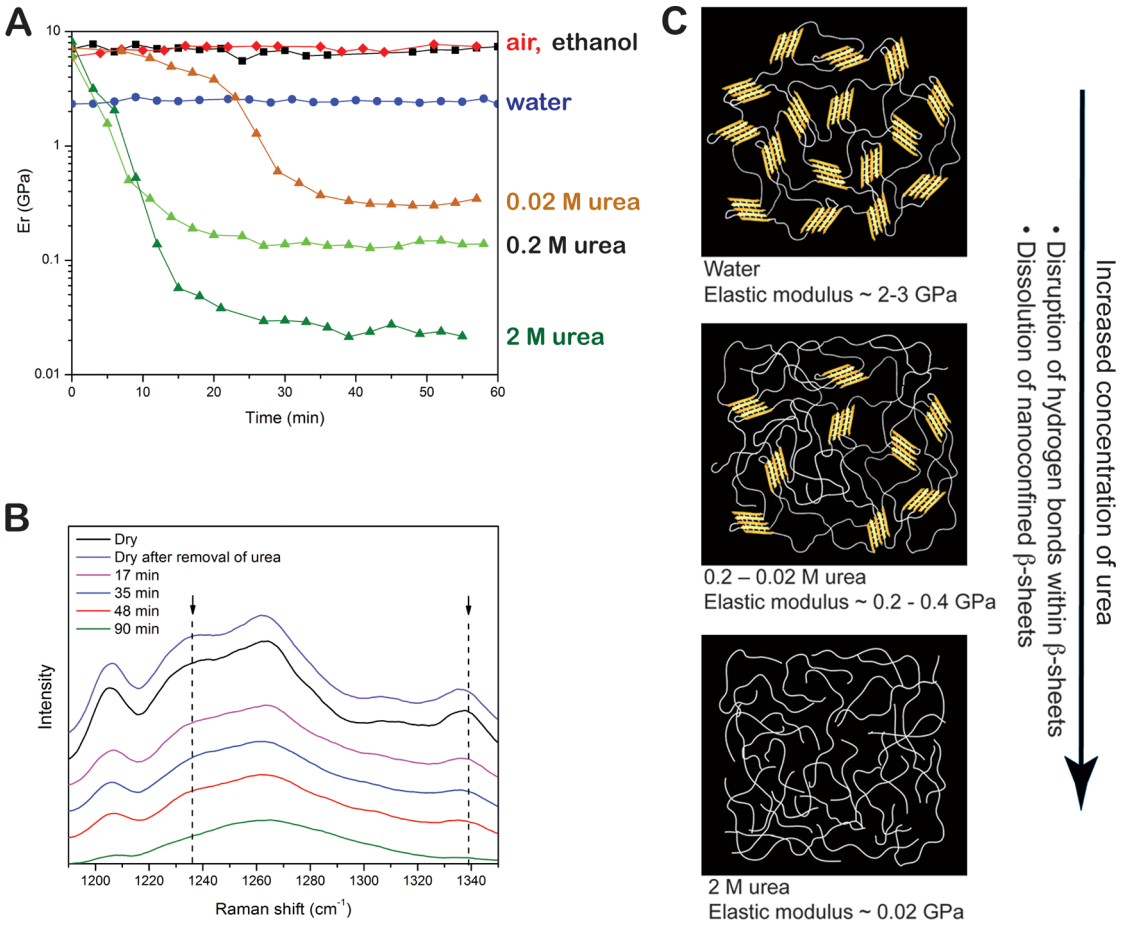


Figure 5. Mechanical properties and secondary structure of *D. gigas* SRT tip. (A) Elastic modulus *versus* time measured by nanoindentation in air, water, 100% ethanol, and various concentrations of urea. (B) Micro-Raman spectroscopy of *D.gigas* SRT incubated in 0.2 M urea. Dotted lines correspond to β-sheet structure and its disruption with exposure to urea (Note the recovery of β-sheet structure upon removal of urea and drying). (C) Schematic of hydrogen bond disruption accompanied by β-sheet dissolution upon contact of SRT surface with urea.

the suckerins bear distinct similarities with silk proteins, which have gained significant attention in recent years in fields ranging from tissue engineering and drug delivery to photonics.[30] Favorable properties of the suckerins include ease of recombinant protein production, robust mechanical properties, mechanical tunability, biocompatibility, and environmentally friendly processability, thus demonstrating that suckerins have the potential to rival the properties of their silk-based counterparts.[8] The identification of the suckerin gene and protein family now greatly expands this molecular toolbox. For example, the mechanical properties of polymers are strongly influenced by molecular weight, cross-link density, and intervening amorphous chain length.[31] The diversity of sequences across all three cephalopod species suggests an additional key facet of these unique protein polymers. Specifically, the suckerin proteins exhibit variability both in the relative proportion and spacing of β-sheet nanocrystal forming domains (Figure 2) and their expression levels (Table S1). These differences thus offer the potential to create polymer networks that are reinforced with distinct volume fractions of nanoconfined β-sheets. These distinctions in modular architectures between suckerins therefore provide design lessons for producing tailored polymer networks with precisely tuned bulk mechanical properties or the creation of mechanically graded biopolymers. Furthermore, the suckerins offer unique peptide designs that may be used in applications involving peptide self-assembly, biosensing, and molecular switching.

From a biological perspective, our comparative analysis across multiple species shows that suckerins are encoded by an ancient gene family that dates back to the Devonian period. Variation in suckerin modularity may have provided squid and cuttlefish a degree of flexibility and adaptability in the network designs and functional mechanics of their SRT. We suggest that the suckerins represent important evolutionary markers that will likely provide key information regarding the evolutionary history of SRT and cephalopods in a manner similar to how our view of spider evolution has been shaped by in-depth studies of modular spider silk proteins.[10,32−34] Of particular interest will be to elucidate the basis for the expansion of the suckerin gene family in the mesopelagic *D. gigas* compared

to epipelagic *S. lessoniana* and *S. esculenta*, which may relate SRT mechanics with prey type and predatory behavior. The fossil record, molecular systematics, and comparative embryological studies continue to reveal a rich and intriguing evolutionary history of the cephalopods[35] and our results now indicate that suckerin-based nanoconfined β-sheets have likely played a key role in the foraging success, evolution, and diversification of squid and cuttlefish over the past 350 million years.

MATERIALS AND METHODS

Research Specimens and Tissue Samples. Humboldt squid (*D. gigas*) were caught off the east coast of the Baja Peninsula La Paz, Mexico, and *S. lessoniana* and *S. esculenta* were acquired live from commercial sources in Singapore. In each case, the animals were euthanized, and materials and tissues were harvested immediately. Tissues were stored in RNA-later (Qiagen).

2D Isoelectric Focusing. SRT from the tentacles of *D. gigas*, *S. lessoniana*, and *S. esculenta* were pulverized separately in liquid nitrogen and resuspended in rehydration buffer with the appropriate ampholytes (BioRad) at 2 mg/mL concentration. A total of 125 μL of each sample was incubated overnight with pH 7−10 BioRad ReadyStrips in preparation for isoelectric focusing. Samples were focused with a 0−4000 V ramp using the Protean IEF cell (BioRad). The strips were then subjected to reduction and alkylation steps according to manufacturers' instructions. The second dimension gels (BioRad Mini-PROTEAN TGX) were run with a low initial voltage of 10 V for 20 min followed by 200 V for 30 min. Gels were stained with Sypro Ruby stain according to manufacturers' instructions.

RNA-seq Transcriptome Library Preparation. Total RNA was extracted with a Qiagen RNeasy mini kit. An amount of 2−10 μg total RNA per tissue was used in the construction of each RNA-seq library. Poly-A mRNA was enriched with oligo dT beads (Invitrogen) and used for constructing strand-specific paired-end libraries according to manufacturers' instructions (ScriptSeq mRNA-seq library kit v1, Epicenter, Illumina). Phusion PCR polymerase (Thermo Scientific) was used for the final library amplification (12 cycles). PCR cleanup was performed with the MinElute PCR purification kit (Qiagen). Library quantification and quality assessment was performed on an Agilent 2100 Bioanalyzer.

Library Sequencing. Each library was diluted to 8 pM, and clusters were generated on paired-end-read flow cells on an Illumina cBot. For each library, 2 × 76 bp paired-end-reads were collected on a Illumina GA IIx. The *D. gigas* transcriptome was collected from one lane, while the *S. lessoniana* and *S. esculenta* transcriptomes were collected from 2 lanes.

Transcriptome Assembly. Raw reads were converted to fastq format using the Illumina's Offline Base Caller (v1.6). *De novo* transcript assembly was performed with the Trinity software suite[36] using standard parameters on a computational cluster. The final Butterfly output predicted transcript files were used for subsequent analysis.

Protein Annotation. Open reading frames were predicted using custom Perl scripts and the Transdecoder utility provided in the Trinity software suite. Predicted transcripts were quantified by RSEM software.[37] Predicted protein sequences were searched using USEARCH[38] against NR (GenBank) and Pfam protein databases to identify homologous sequences. Signal peptide predictions were obtained with SignalP 4.0.[39]

Protein Module Alignments. The most prominent modular repeats were initially detected visually and by using the MEME SUITE.[40] ClustalW[41] combined with manual sequence alignment was subsequently used to generate large and small scale alignments.

Sequence Completion and Verification. RACE-PCR was used to determine the full length coding sequence of target transcripts and to verify the RNA-seq sequences. Five micrograms of total RNA was subjected to RACE-PCR using Invitrogen's Generacer Kit. PCR primers were designed based on Trinity assembled transcripts, and KOD extreme Taq-polymerase (Merck Milipore) was used for amplification. Products were subcloned into pCR2.1 by Topo TA cloning (Invitrogen) and subjected to Sanger sequencing. Sequence alignments were performed with ClustalW. All sequences have been deposited in NCBI.

Genomic Locus Amplification. *D. gigas* beak buccal mass tissue was stored in 70% ethanol at −80 °C. The tissue was pulverized in liquid nitrogen and quickly incubated in 500 μL of CTAB extraction buffer (3% cetyltrimethylammonium bromide, 1.4 M NaCl, 100 mM Tris (pH 8.0), 3% polyvinylpyrrolidone, 0.2% β-mercaptoethanol) at 65 °C for 1 h. An equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) was added, and the sample was mixed by vortexing and centrifuged for 10 min at 16 000 rpm. The aqueous phase was extracted, and another round of phenol−chloroform extraction was performed. Genomic DNA was then precipitated by adding 3 vol of 95% ethanol and 1/10 vol of 3 M sodium acetate and chilled at −20 °C for 1 h. The precipitated DNA was pelleted at 16 000 rpm for 20 min and then washed once with 70% ethanol. The air-dried DNA pellet was resuspended in TE buffer and stored at −20 °C until further use.

Primers were designed based on the full-length transcript sequence to amplify the genomic locus of *D. gigas-19* (*D. gigas-19-Fwd*, TGAAGGAGTAGAAAGTAGTCTCCA; *D. gigas-19-Rev*, TCTTTGTTCACTGGGATGTTCG); *D. gigas-2* (*D. gigas-2- Fwd*, CGTTTCCTGATCAGTAAAGATGGC; *D. gigas-2-Rev*, TATCCAAGATAACCACCATATCCTCCA); and *D. gigas-4* (*D. gigas-4-Fwd*, ATGGCATCTACCAAATTGATTTTTGTTGTTTTAC; *D. gigas-4-Rev*, TTAATGCAGTCCGAGATATCCTCCATAAA). PCR was performed with KOD Xtreme Hot Start DNA Polymerase (Merck) using the following conditions: 94 °C 3 min; 35 cycles of 98 °C 10 s, 55 °C 30 s, 68 °C 4−8 min. Final extension of 68 °C for 10 min.

Phylogenetic Analysis. Full-length suckerin protein sequences were aligned using ClustalW as implemented in the BioEdit sequence alignment editor.[42] Manual inspection of the alignments were performed and adjusted accordingly. We used three different methods of phylogenetic estimation: Neighbor Joining (NJ), Maximum Parsimony (MP), and Maximum Likelihood (ML). Since homology searches against the NCBI nonredundant database did not pick up any sequences related to suckerins, we generated trees without specifying an out-group sequence. The NJ tree was generated using MEGA6[43] using the p-distance method, pairwise-deletion of gaps and 1000 bootstrap replicates for node support. The MP tree was also generated using MEGA6 with 1000 bootstrap replicates for node support. The ML tree was generated using PhyML-3.1[44] with the WAG + I + G substitution model as deduced by ModelGenerator[45] and 100 bootstrap replicates for node support. On the basis of the bootstrap support values and congruence between the three phylogenetic methods, the 38 suckerin sequences could be assigned to at least six distinct phylogenetic clades. All six clades are reproduced by the three phylogenetic methods (NJ, MP, ML). This suggests that the common ancestor of *D. gigas*, *S. lessoniana* and *S. esculenta* possessed at least six suckerin genes which subsequently expanded through gene duplications in the three lineages with a greater expansion in the Humboldt squid lineage. The protein modular architectures of the suckerin sequences are shown juxtaposed with the NJ tree (Figure 3) to clarify the evolutionary history of this protein family.

X-ray Diffraction. WAXS measurements applying synchrotron radiation were performed at the μspot beamline, BESSY II, at the Helmholtz Zentrum Berlin (Berlin, Germany). X-ray patterns were recorded with a 2D CCD detector (MarMosaic 225, Rayonix Inc., Evanston, IL) with a pixel size of 73 μm and an array of 3072 × 3072 pixels. For the acquisition of the 2D pattern, the energy of the X-ray beam (100 μm in diameter) of 17 keV and a

sample-to-detector distance of 200 mm were calibrated using a quartz standard. Patterns were corrected for empty beam background and variations in incident beam intensity using the Fit2D software,[46] and the integrated 1D intensity profiles were fitted using OPUS software to estimate the peak positions and fwhm.

Calculation of Inter-Proline Distances. The number of residues between each pair of consecutive proline residues was calculated and then classified according to whether the residues overlapped with [M1], [M2], or both [M1−M2] modules.

Nanoindentation. *D. gigas* SRT were embedded in glass ionomer cement (Riva Luting, SDI) and flat tip cross sections were obtained by polishing with diamond paste down to a particle size of 0.25 μm. The embedded cross-sectioned samples were then secured to the base of a custom-built fluid cell machined from a plastic Petri dish, which was designed to enable the simultaneous immersion of the sample surface and probing in the longitudinal direction by Nanoindentation using a TriboScan 950 (Hysitron, MN). An elongated cube-corner tip was employed for all measurements. Lines of 20 indents were performed in the core region of the SRT cross-section where the mechanical properties were previously shown to exhibit very little variability (namely the indentation grid was set so as to prevent probing the higher modulus exterior region of the SRT which contained a lower fraction of nanotubules), with an indent spacing of 15 μm. The loading cycle conditions were the following: peak loads were 200 μN (urea immersed samples) and 1000 μN (all other treatments); loading/unloading rates were 50 μN/s; the time at peak load was 2 s. To maintain the sample under immersed conditions, the fluid cell was carefully refilled at regular intervals with the solutions.

Raman Spectroscopy. Polished SRT cross-sections were probed with a confocal Raman microscope (LabRAM HR800, Horiba Jobin Yvon) equipped with a diode-pumped 785 nm near-infrared laser (PilotPC 500, Sacher Lasertechnik) and a 50× long working distance objective (Olympus, numerical aperture of 0.5). Labspec 5 (Horiba Jobin Yvon) software was used for measurement. Protein Amide I spectra (with 60 s integration time) were obtained at different time intervals from 0 to 90 min after immersing the sample in 0.2 M urea and the 1190−1350 cm^{-1} spectral range was integrated.

Conflict of Interest: The authors declare the following competing financial interest(s): Three authors (Dr. Paul A. Guerette, Dr. Shawn Hoon, and Prof. Ali Miserez) have filed a patent describing the suckerin proteins and peptides and uses thereof.

Acknowledgment. We are grateful to Dr. Sydney Brenner and Dr. Peter Fratzl for insightful discussions and suggestions. We also thank Kiat Whye Kong and Vamshidhar Gangu for technical assistance. This research was funded by a Singapore Ministry of Education Tier 2 Grant (MOE 2011-T2-2-044) (A. Miserez), the Singapore National Research Foundation (NRF) through a NRF Fellowship (A. Miserez), and the Biomedical Research Council (BMRC) of the Agency for Science, Technology, and Research (A*STAR) of Singapore (S. Hoon). We also thank the anonymous reviewer for their valuable comments.

Supporting Information Available: A document containing the following figures and table: 2D isoelectric focusing gels of SRTs for the three species; complete amino acid composition of all suckerins; primary acid sequences of all suckerins; relative frequencies of short peptides in M1 and M2 modules; MP and ML phylogenetic trees of suckerins; genomic DNA sequences of three representative suckerins from *D. gigas*; distribution of amino acids between consecutive Pro residues in suckerins of *S. lessoniana* and *S. esculenta*; gene expression levels for all three species in the sucker tissues. This material is available free of charge *via* the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

1. Miserez, A.; Weaver, J. C.; Pedersen, P. B.; Schneberk, T.; Hanlon, R. T.; Kisailus, D.; Birkedal, H. Microstructural and Biochemical Characterization of the Nano-porous Sucker Rings from *Dosidicus gigas*. *Adv. Mater.* **2009**, *21*, 401–406.
2. Andersen, S. A. Insect Cuticular Sclerotization: A Review. *Insect. Biochem. Mol. Biol.* **2010**, *40*, 166–178.
3. Miserez, A.; Schneberk, T.; Sun, C.; Zok, F. W.; Waite, J. H. The Transition from Stiff to Compliant Materials in Squid Beak. *Science* **2008**, *319*, 1816–1819.
4. Wang, R.; Gupta, H. S. Deformation and Fracture Mechanisms of Bone and Nacre. *Annu. Rev. Mater. Res.* **2011**, *2011*, 41–73.
5. Studart, A. R. Towards High-Performance Bioinspired Composites. *Adv. Mater.* **2012**, *24*, 5024–5044.
6. Broomell, C. C.; Waite, J. H.; Zok, F. W. Role of Transition Metals in Sclerotization of Biological Tissue. *Acta. Biomater.* **2008**, *4*, 2045–2051.
7. Politi, Y.; Priewasser, M.; Pippel, E.; Zaslansky, P.; Hartmann, J.; Siegel, S.; Li, C. H.; Barth, F. G.; Fratzl, P. A Spider's Fang: How to Design an Injection Needle Using Chitin-Based Composite Material. *Adv. Funct. Mater.* **2012**, *22*, 2519–2528.
8. Guerette, P. A.; Hoon, S.; Seow, Y.; Raida, M.; Masic, A.; Tay, G. Z.; Amini, S.; Demirel, M. C.; Abdon, F.; Wong, F. T.; Miserez, A. Accelerating the Design of Biomimetic Materials by Integrating RNA-Seq with Proteomics and Materials Science. *Nat. Biotechnol.* **2013**, *31*, 908–915.
9. Guerette, P. A.; Ginzinger, D.; Weber, B.; Gosline, J. M. Silk Properties Determined by Gland-Specific Expression of a Spider Fibroin Gene Family. *Science* **1996**, *272*, 112–115.
10. Gatesy, J.; Hayashi, C. Y.; Motriuk, D.; Woods, J.; Lewis, R. Extreme Diversity, Conservation, and Convergence of Spider Silk Fibroin Sequences. *Science* **2001**, *291*, 2603–2605.
11. Zhou, C. Z.; Confalonieri, F.; Medina, N.; Zivanovic, Y.; Esnault, C.; Yang, T.; Jacquet, M.; Janin, J.; Duguet, M.; Perasso, R.; *et al.* Fine Organization of *Bombyx mori* Fibroin Heavy Chain Gene. *Nucleic Acids Res.* **2000**, *28*, 2413–9.
12. Yano, M.; Nagai, K.; Morimoto, K.; Miyamoto, K. Shematrin: A Family of Glycine-Rich Structural Proteins in the Shell of the Pearl Oyster Pinctada fucata. *Comp. Biochem. Physiol., Part B: Biochem. Mol. Biol.* **2006**, *144*, 254–262.
13. Suzuki, M.; Murayama, E.; Inoue, H.; Ozaki, N.; Tohse, H.; Kogure, T.; Nagasawa, H. Characterization of Prismalin-14, a Novel Matrix Protein from the Prismatic Layer of the Japanese Pearl Oyster (*Pinctada fucata*). *Biochem. J.* **2004**, *382*, 205–213.
14. Dalla Valle, L.; Nardi, A.; Gelmi, C.; Toni, M.; Emera, D.; Alibardi, L. β-keratins of the Crocodilian Epidermis: Composition, Structure, and Phylogenetic Relationships. *J. Exp. Zool., Part B* **2009**, *312*, 42–57.
15. Zhong, Y. S.; Mita, K.; Shimada, T.; Kawasaki, H. Glycine-Rich Protein Genes, which Encode a Major Component of the Cuticle, Have Different Developmental Profiles from other Cuticle Protein Genes in *Bombyx mori*. *Insect Biochem. Mol. Biol.* **2006**, *36*, 99–110.
16. Thornton, J. W.; DeSalle, R. Gene Family Evolution and Homology: Genomics Meets Phylogenetics. *Annu. Rev. Genomics. Hum. Genet.* **2000**, *1*, 41–73.
17. Strugnell, J.; Jackson, J.; Drummond, A. J.; Cooper, A. Divergence Time Estimates for Major Cephalopod Groups: Evidence from Multiple Genes. *Cladistics* **2006**, *22*, 89–96.
18. Hayashi, C. Y.; Lewis, R. V. Molecular Architecture and Evolution of a Modular Spider Silk Protein Gene. *Science* **2000**, *287*, 1477–1479.
19. Moore, A. D.; Bjorklund, A. K.; Ekman, D.; Bornberg-Bauer, E.; Elofsson, A. Arrangements in the Modular Evolution of Proteins. *Trends. Biochem. Sci.* **2008**, *33*, 444–451.
20. Beckwitt, R.; Arcidiacono, S.; Stote, R. Evolution of Repetitive Proteins: Spider Silks from *Nephila clavipes* (Tetragnathidae) and *Araneus bicentenarius* (Araneidae). *Insect Biochem. Mol. Biol.* **1998**, *28*, 121–130.
21. Mita, K.; Ichimura, S.; James, T. C. Highly Repetitive Structure and its Organization of the Silk Fibroin Gene. *J. Mol. Evol.* **1994**, *38*, 583–592.
22. Martel, A.; Burghammer, M.; Davies, R. J.; Riekel, C. Thermal Behavior of *Bombyx mori* Silk: Evolution of Crystalline Parameters, Molecular Structure, and Mechanical Properties. *Biomacromolecules* **2007**, *8*, 3548–3556.

23. Ulrich, S.; Glisovic, A.; Salditt, T.; Zippelius, A. Diffraction from the β-Sheet Crystallites in Spider Silk. *Eur. Phys. J. E: Soft Matter Biol. Phys.* **2008**, *27*, 229–242.
24. Martel, A.; Burghammer, M.; Davies, R. J.; Di Cola, E.; Vendrely, C.; Riekel, C. Silk Fiber Assembly Studied by Synchrotron Radiation SAXS/WAXS and Raman spectroscopy. *J. Am. Chem. Soc.* **2008**, *130*, 17070–17074.
25. Keten, S.; Xu, Z.; Ihle, B.; Buehler, M. J. Nanoconfinement Controls Stiffness, Strength and Mechanical Toughness of β-Sheet Crystals in Silk. *Nat. Mater.* **2010**, *9*, 359–367.
26. Pézolet, M.; Lefèvre, T. Unexpected β-sheets and Molecular Orientation of Flagelliform Spider Silk as Revealed by Raman Spectromicroscopy. *Soft Matter* **2012**, *8*, 6350–6357.
27. Rauscher, S.; Baud, S.; Miao, M.; Keeley, F. W.; Pomes, R. Proline and Glycine Control Protein Self-Organization into Elastomeric or Amyloid Fibrils. *Structure* **2006**, *14*, 1667–1676.
28. Monsellier, E.; Chiti, F. Prevention of Amyloid-Like Aggregation as a Driving Force of Protein Evolution. *EMBO Rep* **2007**, *8*, 737–42.
29. Lefevre, T.; Paquet-Mercier, F.; Rioux-Dube, J. F.; Pezolet, M. Review Structure of Silk by Raman Spectromicroscopy: From the Spinning Glands to the Fibers. *Biopolymers* **2012**, *97*, 322–336.
30. Omenetto, F. G.; Kaplan, D. New Opportunities for an Ancient Material. *Science* **2010**, *329*, 528–531.
31. Rubinstein, M.; Colby, R. H. *Polymer Physics*; Oxford University Press: New York, 2003.
32. Ayoub, N. A.; Garb, J. E.; Tinghitella, R. M.; Collin, M. A.; Hayashi, C. Y. Blueprint for a High-Performance Biomaterial: Full-Length Spider Dragline Silk Genes. *PLoS One* **2007**, *2*, e514.
33. Chaw, R. C.; Zhao, Y.; Wei, J.; Ayoub, N. A.; Allen, R.; Atrushi, K.; Hayashi, C. Y. Intragenic Homogenization and Multiple Copies of Prey-Wrapping Silk Genes in Argiope Garden Spiders. *BMC Evo. Biol.* **2014**, *14*, 31.
34. Garb, J. E.; DiMauro, T.; Lewis, R. V.; Hayashi, C. Y. Expansion and Intragenic Homogenization of Spider Silk Genes Since the Triassic: Evidence from Mygalomorphae (Tarantulas and Their Kin) Spidroins. *Mol. Biol. Evol.* **2007**, *24*, 2454–2464.
35. Kroger, B.; Vinther, J.; Fuchs, D. Cephalopod Origin and Evolution: A Congruent Picture Emerging from Fossils, Development and Molecules: Extant Cephalopods are Younger than Previously Realised and Were under Major Selection to Become Agile, Shell-less Predators. *BioEssays* **2011**, *602−613*, 8.
36. Grabherr, M. G.; Haas, B. J.; Yassour, M.; Levin, J. Z.; Thompson, D. A.; Amit, I.; Adiconis, X.; L, F.; Raychowdhury, R.; Zeng, Q.; Chen, Z.; *et al.* Full-Length Transcriptome Assembly from RNA-Seq Data Without a Reference Genome. *Nat. Biotechnol.* **2011**, *29*, 644–652.
37. Li, B.; Dewey, C. N. RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome. *BMC Bioinf.* **2011**, *12*, 323.
38. Edgar, R. C. Search and Custering Orders of Magnitude Faster than BLAST. *Bioinformatics* **2010**, *26*, 2460–2461.
39. Nordahl Petersen, T.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 4.0: Discriminating Signal Peptides from Transmembrane Regions. *Nat. Methods* **2011**, *8*, 785–786.
40. Bailey, T. L.; Boden, M.; Buske, F. A.; Frith, M.; Grant, C. E.; Clementi, L.; Ren, J.; Li, W. W.; Noble, W. S. MEME SUITE: Tools for Motif Discovery and Searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208.
41. Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948.
42. Hall, T. A. BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95–98.
43. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725–2729.
44. Guindon, S.; Dufayard, J.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321.
45. Keane, T. M.; Creevey, C. J.; Pentony, M. M.; Naughton, T. J.; Mclnerney, J. O. Assessment of Methods for Amino Acid Matrix Selection and their Use on Empirical Data Shows that *ad hoc* Assumptions for Choice of Matrix Are Not Justified. *BMC Evol. Biol.* **2006**, *6*, No. 29.
46. Hammersley, A. P. FIT2D: An Introduction and Overview. ESRF Internal Report, ESRF97HA02T; Grenoble: France, **1997**.