Molecular tandem repeat strategy for elucidating mechanical properties of high-strength proteins

Huihun Jung^{a,b,1}, Abdon Pena-Francesch^{a,b,1}, Alham Saadat^{c,d}, Aswathy Sebastian^{c,d,e}, Dong Hwan Kim^f, Reginald F. Hamilton^{a,b}, Istvan Albert^{c,d,e}, Benjamin D. Allen^{c,d,2}, and Melik C. Demirel^{a,b,d,2}

^aMaterials Research Institute, Pennsylvania State University, University Park, PA 16802; ^bDepartment of Engineering Science and Mechanics, Pennsylvania State University, University Park, PA 16802; ^cDepartment of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA 16802; ^dThe Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA 16802; ^eBioinformatics Consulting Center, Pennsylvania State University, University Park, PA 16802; and ^fDepartment of Biology, Pennsylvania State University, University Park, PA 16802

Edited by Stephen L. Mayo, California Institute of Technology, Pasadena, CA, and approved May 2, 2016 (received for review November 24, 2015)

Many globular and structural proteins have repetitions in their sequences or structures. However, a clear relationship between these repeats and their contribution to the mechanical properties remains elusive. We propose a new approach for the design and production of synthetic polypeptides that comprise one or more tandem copies of a single unit with distinct amorphous and ordered regions. Our designed sequences are based on a structural protein produced in squid suction cups that has a segmented copolymer structure with amorphous and crystalline domains. We produced segmented polypeptides with varying repeat number, while keeping the lengths and compositions of the amorphous and crystalline regions fixed. We showed that mechanical properties of these synthetic proteins could be tuned by modulating their molecular weights. Specifically, the toughness and extensibility of synthetic polypeptides increase as a function of the number of tandem repeats. This result suggests that the repetitions in native squid proteins could have a genetic advantage for increased toughness and flexibility.

tandem repeat | high strength | protein | thermoplastic | squid ring teeth

Proteins are heteropolymers that provide a variety of building blocks for designing biological materials (1). Proteins have several advantages as natural materials: (i) their chain length, sequence, and stereochemistry can be easily controlled, (ii) the molecular structure of proteins is well-defined (e.g., secondary, tertiary, and quaternary structures), (iii) they provide a variety of functional chemistries for conjugation to other biomolecules or polymers, and (iv) they can be designed to exhibit a variety of physical properties (2). Proteins are diverse but often display substantial similarity in sequence and 3D structure. Duplication of structural units is a natural evolutionary strategy for increasing the complexity of both globular and fibrous/structural proteins (3). For example, collagen has polyproline- and glycine-rich helices, whereas silk and elastin have β -spiral [GPGXX], linker [GP(S,Y,G)], and 3_{10} -helix [GGX] repeats. These repetitions are advantageous because of the intrinsic promotion of stability through the periodic recurrence of favorable interactions (4–7).

A new family of repetitive structural proteins was recently identified in the tentacles of several squid species (8, 9). Squid have teeth-like structures inside their suckers that allow the animals to grip tightly on a diverse array of objects (10). Using the tools of molecular biology and proteomics, it has been shown that these squid ring teeth (SRT) proteins have segmented semicrystalline morphology with repetitive amorphous and crystalline domains. SRT-based materials were shown to have high elastic modulus: 4–8 GPa in air and 2–4 GPa underwater below the glass transition temperature (11). However, a clear relationship between the molecular structure and the mechanical properties of this material remains elusive. This problem is complex, because SRT proteins are polydispersed in chain length, and the crystalline and amorphous segments within each SRT protein also vary in length and amino acid sequence (12).

To investigate the genetic basis of material properties in natural and artificial SRT sequences, we have developed a new approach

for the design and production of structural proteins that comprise one or more tandem repeats (TRs) of a single unit with distinct amorphous and crystalline regions. In general, our design strategy uses three parameters to modulate the properties of the protein: (i) the composition of the crystalline/ordered or amorphous regions, (ii) the length ($L = L_a + L_c$) and fraction ($f = L_a/L_c$) of the amorphous (L_a) and crystalline regions (L_c), and (iii) the repeat number n : the number of tandem copies of the amorphous plus crystalline unit.

This approach requires the efficient construction of DNA sequences that encode artificial TR proteins. Popular methods for the synthesis of TR genes rely on recursive in vitro ligation of DNA fragments or controlled doubling by iterative cloning (13). Recursive ligation allows many repeats to be assembled in a single step, but the product size is difficult to control. Iterative cloning allows TR sequences of any size to be produced in a controlled fashion but is extremely laborious, requiring several months to produce larger products (14). Neither method is amenable to pooled processing of repeat unit libraries: if multiple sequences are present in a single reaction, they will be ligated together randomly rather than each separately, giving rise to heterogeneous TR products.

To enable the work that we report here and more expansive future studies, we developed an alternative TR DNA assembly method to (i) produce TR sequences of various lengths in a single reaction, (ii) offer better control over the resulting lengths,

Significance

Squid have teeth-like structural [squid ring teeth (SRT)] proteins inside their suckers, which have segmented semicrystalline morphology with repetitive amorphous and crystalline domains. These proteins have high elastic modulus and toughness. However, a clear relationship between molecular structure and mechanical properties of this material remains elusive. To investigate the genetic basis of material properties in SRT sequences, we developed a new approach for the design and production of structural proteins. We show that the toughness and flexibility of these synthetic SRT mimics increase as a function of molecular weight, whereas the elastic modulus and yield strength remain unchanged. These results suggest that artificial proteins produced by our approach can help to illuminate the genetic basis of protein material behavior in SRT.

Author contributions: B.D.A. and M.C.D. designed research; H.J., A.P.-F., A. Saadat, D.H.K., I.A., B.D.A., and M.C.D. performed research; A. Sebastian, R.F.H., I.A., B.D.A., and M.C.D. analyzed data; and B.D.A. and M.C.D. wrote the paper.

Conflict of interest statement: The authors have a pending patent application.

This article is a PNAS Direct Submission.

Data deposition: The sequence reported in this paper has been deposited in the National Center for Biotechnology Information BioProject database, www.ncbi.nlm.nih.gov/bioproject/ (accession no. PRJNA320263).

¹H.J. and A.P.-F. contributed equally to this work.

²To whom correspondence may be addressed. Email: bda3@psu.edu or mdemirel@engr.psu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1521645113/-DCSupplemental.

and (iii) allow pooled processing of unit sequence libraries. In this approach, long TR products from a short sequence unit are produced by rolling circle amplification (RCA). The RCA reaction is tuned to incorporate noncanonical nucleotides at random positions. These nucleotides block digestion by key restriction endonucleases; the resulting partial digestion products can be separated by size and cloned into an expression vector for protein production. This method, which we call “protected digestion of rolling-circle amplicons” (PD-RCA), can be used to prepare a library of TR sequences with a controlled distribution of lengths in a single cloning step.

To validate our approach to mapping sequence–structure–property relationships in segmented structural proteins, we applied PD-RCA and recombinant expression in *Escherichia coli* to produce a panel of artificial SRT-based proteins that vary only in the repeat number but not in the lengths or compositions of their crystalline and amorphous regions. We show that the toughness and flexibility of these synthetic SRT mimics increase as a function of molecular weight, whereas the elastic modulus and yield strength remain unchanged. These results suggest that artificial proteins produced by PD-RCA can help to illuminate the genetic basis of protein material behavior and that SRT proteins provide a promising platform for the design of previously unidentified materials with custom properties.

Results and Discussion

SRT is a protein complex that is composed of polypeptides with repetitive amino acid sequences similar to a semicrystalline segmented copolymer (12). The unique architecture of SRT is the key to the creation of high-strength materials using the TR strategy. Fig. 1 shows SRT’s compositional variations in different squids. We studied four selected species around the world that are commonly found in the fishing areas shown in Fig. 1A. The protein gel electrophoresis results (Fig. 1B) show the molecular weight distribution of the SRT proteins from different squids. A combination of RNA sequencing (15) and protein MS (16) was performed to identify several sequences of the SRT complex for these four species. mRNA extracted from the suction cups of the squid epithelium tissues was sequenced to identify the transcripts that matched the protein sequences observed in the SRT complex. High-throughput sequencing produced paired end reads with read lengths of at least 250 bp, which were used to assemble a preliminary transcriptome. The sequence data has been deposited in

the National Center for Biotechnology Information BioProject database (PRJNA320263). Peptide sequences from the whole-SRT protein complex were sequenced using MS to provide N-terminal biased partial protein sequences that were matched against the putative transcripts. Details of the iterative bioinformatics approach can be found in our earlier publication (9).

The crystal-forming polypeptide sequence and the amorphous-structured polypeptide sequence (Fig. 1C) are derived from SRT proteins from any of the following species: *Loligo vulgaris*, *Loligo pealei*, *Todarodes pacificus*, and *Euprymna scolopes* (Fig. 1D). These polypeptides are studied with jalview, a sequence analysis tool for protein alignment (*SI Appendix, Fig. S1*). The sequence analysis of SRT protein shows a repetitive crystalline/amorphous architecture (AVSHT-rich/GLY-rich) that can form antiparallel β -sheets with turns. Because of the presence of two histidine and two alanine amino acids at opposite ends of each crystalline segment (next to each proline amino acid that divides the sequence), we suggested that the antiparallel arrangement of β -sheets is more favorable than parallel β -sheets (12). This alignment is an excellent strategy for the stability of the β -sheets, because parallel β -sheets would position neighboring histidine side chains next to each other, resulting in a less stable asymmetric β -sheet stacking because of the large volume of the aromatic ring in histidine side chains and the smaller volume of methyl group in alanine. However, antiparallel β -sheets alternate the position of the histidine and alanine groups in neighboring chains, resulting in a more compact and ordered structure. Amorphous domains of SRT also show sequence repetition (*SI Appendix, Fig. S2*). However, this repetition is not surprising, because the amorphous domain of the structural proteins typically comprises TRs of structural units, such as [GP(S,Y,G)] that provide mechanical flexibility between crystallites.

Native SRT proteins already show considerable diversity (variable AVSHT-rich) in their crystal-forming sequences (9). Our designed sequences are based on the crystal-forming polypeptide sequence of PAAASVSTVHHP and the amorphous polypeptide sequence of YGYGGLYGGLYGGLGY (Fig. 2A). This unit is one of several possible consensus sequences derived by inspection of the alignments from all four squid species (*SI Appendix, Figs. S1 and S2*). We used this unit to construct three TR sequences that differ only by their repeat numbers and hence, their total lengths. These sequences, with repeat numbers of 4, 7, and 11, are named

Fig. 1. (A) Fishery information for four common squid species and (B) corresponding protein gels and optical images of SRT are shown. The individual molecular weight (MW) distribution is nonuniform as seen from protein gels, but the repeats in protein sequences are similar (*SI Appendix*). (C) Repetitions in protein sequences can be visualized by segmented (nonhomogenous) copolymer architecture that has crystalline (green) and amorphous (red) regions as shown in the schematic. (D) Taxonomic classification of squid species reveals the separation between these species.

Fig. 3. (A) SDS/PAGE showing the sizes of the synthetic proteins with $n = 4$, $n = 7$, and $n = 11$. (B) Cartoon representation of the segmented polymer architecture of assembled polypeptides containing ordered β -sheet crystals and amorphous Gly-rich regions. Amorphous and crystalline are colored in green and red, respectively. The (C) FTIR and (D) XRD spectra for all three samples are shown. α , α -helix; β , β -sheet; MW, molecular weight; rc, random coil; sc, side chain; t, turn.

structure elements is determined by calculating the ratio of the fitted bands area to the total deconvoluted amide I band area (excluding the side chains band) (sc in *SI Appendix*, Fig. S5). The secondary structure composition of synthetic polypeptides is summarized in *SI Appendix*, Table S3. The differences in secondary structure quantification might arise from analyzing the raw data vs. the deconvoluted spectra of amide I band (26).

Representative XRD spectra for three synthetic proteins are shown in Fig. 3D and *SI Appendix*, Fig. S6. The diffraction spectra for all three synthetic proteins are very similar. The crystallite size (i.e., $\sim 3 \times 2 \text{ nm}$) is estimated from XRD according to the Scherrer equation (27). The Miller indices are assigned consistently with the native SRT from a related species (*Dosidicus gigas*) (28). The major crystalline peaks can be observed at $2\theta = 9.50^\circ$, 19.15° , and 24.85° corresponding to lattice distances $d_{100} = 9.31 \text{ \AA}$, $d_{200} = 4.63 \text{ \AA}$, and $d_{002} = 3.58 \text{ \AA}$, respectively (Fig. 3D and *SI Appendix*, Fig. S8). Additionally, a weak diffraction peak is observed at $2\theta = 36.73^\circ$ with lattice distance $d_{240} = 2.44 \text{ \AA}$ accompanied with a broad peak. The intense peak at $2\theta = 19.15^\circ$ is attributed to the combination of (120) and (200) reflections, and the peak at $2\theta = 36.73^\circ$ is attributed to the combination of (240) and (023) reflections. These lattice distances correspond to the hydrogen bond distance between two β -sheet chains, the distance between alternating β -sheet chains (i.e., unit cell dimension in the hydrogen bond direction fitting two β -sheet chains), and the chain length of a single amino acid in an antiparallel β -sheet structure (with a two-residue repeat distance of $7.0 \text{ \AA}$), respectively (29). According to the XRD results, β -sheet crystals can accommodate ~ 11 residues along the backbone direction and ~ 4 strands along the hydrogen bonding direction, which agree well with the initial sequence design (i.e., 10-aa length between proline residues in crystalline segments). The β -sheet crystal structure is fitted into an orthorhombic unit cell referencing to other known β -sheet crystals, such as silk (30). Although $(0k0)$ diffraction peaks cannot be resolved in the current diffraction pattern, the unit cell dimension b (amino acid side chain direction) is calculated from the d_{120} , d_{240} , and d_{023} spacing values. The unit cell parameters obtained by the diffraction data are $a = 9.31 \text{ \AA}$ (H bond direction), $b = 11.06 \text{ \AA}$ (amino acid side chain direction), and $c = 7.16 \text{ \AA}$ (chain backbone direction). The resulting crystal structure for synthetic polypeptides has a similar symmetry to the crystal structure of *Nephila clavipes* spider silk, which is classified into the Warwicker system group 3b and has an orthorhombic unit cell (31). We should mention that predicting the dimension in stacking direction is very complex. The crystalline segments of synthetic polypeptides are rich in Ala, Thr, Val, Ser, and His amino acids, which increase the complexity in the intersheet stacking (especially when incorporating large side groups, such as His). It is known that different amino acids in the crystalline chains can lead to varying intersheet spacing distances (known as nonperiodic lattice crystals) because of the effect of the different side groups (32). For example, silk β -sheet crystals from different species, such as *N. clavipes* spider or *B. mori* silkworm, have conserved sequences (i.e., polyalanine or alternating Gly-Ala) with repeating units (33). However, because of the alternating order of Gly and Ala amino acids, one side of the silk chain is populated by methyl groups, whereas hydrogen side groups populate the other. This order results in an alternating stacking of the β -sheets, where the methyl faces have a greater intersheet separation ($5.7 \text{ \AA}$) than the glycyl faces ($3.5 \text{ \AA}$) (29). Thus, the more diverse β -sheet sequences of native SRT proteins and the SRT mimics that we report here may give rise to even more complex stacking assemblies, including nonperiodic lattice crystals. We also calculated the crystallinity percentage of the synthetic polypeptides by fitting the crystalline and amorphous peaks in the Lorentz-corrected wide-angle X-ray scattering (WAXS) intensity data (*SI Appendix*, Fig. S6) (34). The crystallinity index is calculated as the ratio of the deconvoluted crystalline area to the total area. The crystallinity index of these proteins is between 43% and 45% as listed

in *SI Appendix, Table S4*. This crystallinity is slightly higher than the FTIR results because of increased noise inherent to WAXS analysis.

We studied the mechanical response of all three synthetic polypeptides using DMA (Fig. 4A). The initiation and progression of deformation are shown in the digital image correlation (DIC) snapshots in Fig. 4B for syn-n4, syn-n7, and syn-n11 samples. Syn-n4 is brittle and shows linear elastic behavior at low strains and then fracture. In contrast, both syn-n7 and syn-n11 can be deformed to larger strains compared with syn-n4, and they exhibit irreversible plastic deformation. Craze lines are shown with arrows in Fig. 4A, *Inset*. The drawability of the syn-n11 is significantly larger than for the other two samples. Young modulus (~ 0.7 – 0.8 GPa) for the synthetic polypeptides can be estimated from the linear region of the stress–strain curve in Fig. 4A and *SI Appendix, Fig. S7*. Compared with elastic modulus of recombinant 18-kDa SRT protein from *L. vulgaris* (~ 1 – 2 GPa), this value is slightly lower. The lower modulus could be because of ambient water in the sample ($\sim 5\%$) or trace amounts of 1,1,1,3,3,3-hexafluoro-2-propanol retained from casting ($<1\%$). We also point out that elastic modulus of synthetic polypeptides or recombinant proteins are typically lower compared with native proteins (e.g., ~ 4 – 6 GPa for SRT protein from *L. vulgaris* or ~ 8 – 10 GPa for silk protein from *B. mori*) because of intermolecular interactions of multiple protein sequences in native complexes. Although the elastic modulus and the yield strength for three samples are similar (i.e., ~ 14 MPa for syn-n4 and syn-n7 and a slightly higher value of 18 MPa for syn-n11), their toughness (i.e., 0.14, 0.46, and 2.37 MJ/m³, respectively) and extensibility (i.e., 2%, 4.5%, and 15%, respectively) increase as a function of polypeptide molecular weight (*SI Appendix, Table S5*). Fig. 4B shows strain contour maps, which were measured using the DIC analysis technique (*SI Appendix, Fig. S9*), for each sample to scrutinize the material response in a pointwise manner over three sample surfaces (35). The contours for syn-n7 (column f in Fig. 4B) and syn-n11 (column j in Fig. 4B), which follow the elastic σ – ϵ response, exhibit localized regions of concentrated strains (nearly 9.5%) that exceed the corresponding average strains in the σ – ϵ curve. Contour maps for syn-n4 do not show similar strain concentrations accompanying the lowest extensibility. Thus, the concentrations are likely forming near initial microcracks. The concentrated regions in the maps grow across the sample surface with increasing deformation, and the magnitudes exceed 20%, which are considerably higher than average strains. For the syn-n11, the concentrated regions show that residual strains and deformation on the fracture surface are the most diffuse of the three synthetic polypeptides. The results suggest that the diffuse nature of stress concentration for the higher repeat numbers/longer lengths can facilitate toughening.

Several models have been developed for understanding the mechanism of fracture in polymers (36). However, prediction of maximum fracture is still an active research area because of difficulties modeling the nucleation of microcracks in polymers. Following the structure–property relationship (37) for the yield strength of the synthetic proteins as 17.5 MPa, which agrees well with the experimental data of 14–18 MPa observed in Fig. 4A and *SI Appendix, Fig. S7*. The amorphous region of the synthetic protein has a loose network of chains that are tied together through secondary interactions (e.g., hydrogen bonds and van der Waals interactions). Therefore, we propose that the amorphous chains and reordering of β -sheets should dominate the fracture mechanism and that the secondary bonds are broken on tensile deformation. A deconvoluted FTIR spectrum shows that the crystallinity content of deformed syn-n11 samples does not change (*SI Appendix, Table S6*), whereas individual β -sheet peaks vary (i.e., reorganization of crystalline domains), the turn content increases, and the α -helix content decreases (Fig. 4C). This result agrees well with the observed macroscopic tensile

Fig. 4. Mechanical testing of syn-n4, syn-n7, and syn-n11 samples. (A) Stress–strain curves show that toughness and extensibility of synthetic polypeptides increase as a function of protein molecular weight. (*Inset*) Fractured samples show brittle fracture for syn-n4, whereas syn-n7 and syn-n11 show ductile fracture (craze lines marked with arrows). (B) DIC shows full-field strain measurement for all three samples at the locations marked with point labels (labels a–t) in the stress–strain graph. Syn-n4 sample shows homogeneous strain along the gauge length, whereas syn-n7 and syn-n11 samples show local strain concentration during yielding. (C) FTIR analysis of pristine and drawn syn-n11 samples. α , α -helix; β , β -sheet; rc, random coil; sc, side chain; t, turn.

behavior of an initial linear elastic regime followed by a large plateau regime, at which the secondary bonds break.

Conclusion

We designed and characterized a new polypeptide sequence based on the native amino acid content of semicrystalline SRT proteins and then generated TRs of this sequence with a range of chain lengths using our PD-RCA approach. We show that toughness and extensibility of the synthetic polypeptides increase as a function of their molecular weights, whereas the elastic modulus and the yield strength remain unchanged. This result suggests that the repetitions in native SRT could have a genetic advantage for increased toughness and flexibility. Similar to their natural and recombinant counterparts, synthetic SRT mimics such as those described here can be processed to form any of a variety of 3D shapes, including but not necessarily limited to ribbons, lithographic patterns, and nanoscale objects, such as nanotube arrays. The ability to easily manufacture protein-based materials with

tunable self-healing properties (38) will find applications in a broad array of useful applications, including textiles, cosmetics, and medicine.

Materials and Methods

Construction of a TR Template. A 111-bp gene fragment (Fig. 2A) encoding an 18-aa amorphous region and an 11-aa crystalline region was synthesized by Genewiz, cloned into plasmid pCR-Blunt by standard methods, and verified by Sanger sequencing. The insert contains five restriction sites to enable the PD-RCA process described below: two Scal sites to allow the insert to be removed from its vector by digestion, a BbvCI site to allow a phi29-polymerase priming site to be generated by the nicking enzyme nt.BbvCI and an Acc65I site and an Apal site, which can each be blocked through the incorporation of 5mC in place of cytosine. A circular, nicked version of the insert sequence was prepared as a template for RCA as follows. The plasmid was digested with Scal-HF, and the resulting 105-bp fragment was isolated on a 1% agarose-Tris-acetate-EDTA (TAE) gel and purified with an Omega Bio-Tek E.Z.N.A Gel Extraction Kit. The purified 105-bp fragment was then circularized with T4 ligase at room temperature followed by 10 min at 65 °C to inactivate the ligase; 1 μ L heat-inactivated ligation reaction was then nicked using nt.BbvCI to create a priming site for RCA. The nicking enzyme reaction was heat-inactivated for 20 min at 80 °C.

RCA. The 1.5 μ L of the heat-inactivated nicking reaction was used as the template in a 10- μ L RCA reaction with 1 $\times$ New England Biolabs (NEB) phi29 polymerase buffer, 1 μ g BSA, 1 mM dATP, 1 mM dGTP, 1 mM dTTP, 0.5 mM dCTP, 0.5 mM 5-methyl-dCTP, and 2.5 U NEB phi29 polymerase. The reaction was incubated at 30 °C for 24 h and then heat-inactivated for 10 min at 65 °C.

Sizing and Cloning of TR Products. The heat-inactivated RCA reaction was sequentially digested with Apal and Acc65I, yielding TRs of various sizes because the random protection of their recognition sites by 5mC (Fig. 2B). TR fragments between 500 and 1,500 bp were isolated from a 1% agarose-TAE gel and purified with an Omega Bio-Tek E.Z.N.A Gel Extraction Kit. The purified fragments were cloned through the Acc65I and Apal sites into the ORF of an expression vector prepared by site-directed mutagenesis of pET14b. Colony PCR was used to screen for clones with inserts of the desired sizes; diagnostic digestion and Sanger sequencing confirmed the lengths and compositions of the clones after plasmid isolation.

Protein Expression of TR-Syn. A single colony was inoculated and grown overnight in 5 mL LB with ampicillin (100 μ g/mL). The overnight culture was scaled up to 2 L (i.e., four by 500 mL LB media) and grown on a shaker at 210 rpm and 37 °C for 5 h. When the cultures reached OD₆₀₀ of 0.7–0.9, isopropyl β -D-1-thiogalactopyranoside was added to the final concentration of 1 mM, and shaking was continued at 37 °C for 4 h. Then, the cells were pelleted at 21,612 $\times$ g for 15 min and stored at –80 °C. After thawing, cell pellets were resuspended in 300 mL lysis buffer (50 mM Tris, pH 7.4, 200 mM NaCl, 1 mM PMSF, and 2 mM EDTA) and lysed using a high-pressure homogenizer. The lysate was pelleted at 29,416 $\times$ g for 1 h at 4 °C. The lysed pellet was washed twice with 100 mL urea extraction buffer [100 mM Tris, pH 7.4, 5 mM EDTA, 2 M urea, 2% (vol/vol) Triton X-100] and then washed with 100 mL washing buffer (100 mM Tris, pH 7.4, 5 mM EDTA). Protein collection in the washing step (urea extraction and final wash) was performed by centrifugation at 3,752 $\times$ g for 15 min. The resulting recombinant protein pellet was dried with a lyophilizer (FreeZone 6 Plus; Labconco) for 12 h. The final yield of expressed protein was ~15 mg/1 L bacterial culture.

Sample Preparation and Characterization. Syn-n4, syn-n7, or syn-n11 protein was dissolved in 1,1,1,3,3,3-hexafluoro-2-propanol to a concentration of 50 mg/mL in a sonication bath for 1 h. The solution was then cast into polydimethylsiloxane dog bone-shaped molds to produce the desired geometry for mechanical testing, and solvent was evaporated at room temperature under a fume hood overnight. Resulting films were ~55 μ m in thickness (*SI Appendix*, Fig. S8). All three samples were characterized by XRD, FTIR, DMA, and DIC (details in *SI Appendix*).

ACKNOWLEDGMENTS. The authors thank Dr. Tim Miyashiro (Pennsylvania State University) for providing the bobtail squid samples and Dr. Tugba Ozdemir for helping with RNA extraction from squid suction cup tissues. The authors acknowledge technical support (Dr. Tatiana Laremore and Dr. Craig Praul) from the Genomics and Proteomic Facilities of the Huck Institutes of the Life Sciences at the Pennsylvania State University. H.J., A.P.-F., D.H.K., and M.C.D. were supported partially by Office of Naval Research Grant N000141310595, Army Research Office Grant W911NF-16-1-0019, Materials Research Institute Humanitarian Funding, and the Pennsylvania State University internal funds. A. Saadat, A. Sebastian, I.A., and B.D.A. were supported by the Huck Institutes of the Life Sciences and the Department of Biochemistry and Molecular Biology.

- Kaplan D, McGrath K (2012) *Protein-Based Materials* (Birkhäuser, Boston).
- Langer R, Tirrell DA (2004) Designing materials for biology and medicine. *Nature* 428(6982):487–492.
- McLachlan AD (1972) Repeating sequences and gene duplication in proteins. *J Mol Biol* 64(2):417–437.
- Cetinkaya M, Xiao S, Markert B, Stacklies W, Gräter F (2011) Silk fiber mechanics from multiscale force distribution analysis. *Biophys J* 100(5):1298–1305.
- Lin S, et al. (2015) Predictive modelling-based design and experiments for synthesis and spinning of bioinspired silk fibres. *Nat Commun* 6:6892.
- Nova A, Ketten S, Pugno NM, Redaelli A, Buehler MJ (2010) Molecular and nano-structural mechanisms of deformation, strength and toughness of spider silk fibrils. *Nano Lett* 10(7):2626–2634.
- Söding J, Lupas AN (2003) More than the sum of their parts: On the evolution of proteins from peptides. *BioEssays* 25(9):837–846.
- Guerette PA, et al. (2013) Accelerating the design of biomimetic materials by integrating RNA-seq with proteomics and materials science. *Nat Biotechnol* 31(10):908–915.
- Pena-Francesch A, et al. (2014) Materials fabrication from native and recombinant thermoplastic squid proteins. *Adv Funct Mater* 24(47):7401–7409.
- Nixon M, Dilly P (1977) Sucker surfaces and prey capture. *Symp Zool Soc Lond* 38:447–511.
- Pena-Francesch A, et al. (2014) Pressure sensitive adhesion of an elastomeric protein complex extracted from squid ring teeth. *Adv Funct Mater* 24(39):6227–6233.
- Demirel MC, Cetinkaya M, Pena-Francesch A, Jung H (2015) Recent advances in nanoscale bioinspired materials. *Macromol Biosci* 15(3):300–311.
- Tokareva O, Michalczyk-Lacerda VA, Rech EL, Kaplan DL (2013) Recombinant DNA production of spider silk proteins. *Microb Biotechnol* 6(6):651–663.
- Teulé F, et al. (2009) A protocol for the production of recombinant spider silk-like proteins for artificial fiber spinning. *Nat Protoc* 4(3):341–355.
- Haas BJ, et al. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8(8):1494–1512.
- Pevtsov S, Fedulova I, Mirzaei H, Buck C, Zhang X (2006) Performance evaluation of existing de novo sequencing algorithms. *J Proteome Res* 5(11):3018–3028.
- Wong K-K, Markillie LM, Saffer JD (1997) A novel method for producing partial restriction digestion of DNA fragments by PCR with 5-methyl-CTP. *Nucleic Acids Res* 25(20):4169–4171.
- Goormaghtigh E, Cabiaux V, Ruyschaert J-M (1994) *Determination of Soluble and Membrane Protein Structure by Fourier Transform Infrared Spectroscopy. Physicochemical Methods in the Study of Biomembranes* (Springer, Berlin), pp 329–362.
- Hu X, Kaplan D, Cebe P (2006) Determining beta-sheet crystallinity in fibrous proteins by thermal analysis and infrared spectroscopy. *Macromolecules* 39(18):6161–6170.
- Chen X, Knight DP, Shao Z, Vollrath F (2002) Conformation transition in silk protein films monitored by time-resolved Fourier transform infrared spectroscopy: Effect of potassium ions on Nephila spidroin films. *Biochemistry* 41(50):14944–14950.
- Nilsson MR (2004) Techniques to study amyloid fibril formation in vitro. *Methods* 34(1):151–160.
- Mouro C, Jung C, Bondon A, Simonneaux G (1997) Comparative Fourier transform infrared studies of the secondary structure and the CO heme ligand environment in cytochrome P-450cam and cytochrome P-420cam. *Biochemistry* 36(26):8125–8134.
- Jackson M, Mantsch HH (1991) Protein secondary structure from FT-IR spectroscopy: Correlation with dihedral angles from three-dimensional Ramachandran plots. *Can J Chem* 69(11):1639–1642.
- Taddei P, Monti P (2005) Vibrational infrared conformational studies of model peptides representing the semicrystalline domains of Bombyx mori silk fibroin. *Biopolymers* 78(5):249–258.
- Teramoto H, Miyazawa M (2005) Molecular orientation behavior of silk sericin film as revealed by ATR infrared spectroscopy. *Biomacromolecules* 6(4):2049–2057.
- Lórenz-Fonfría VA, Padrós E (2004) Curve-fitting of Fourier manipulated spectra comprising apodization, smoothing, derivation and deconvolution. *Spectrochim Acta A Mol Biomol Spectrosc* 60(12):2703–2710.
- Scherrer P (1918) Bestimmung der Grösse und der inneren Struktur von Kolloidteilchen mittels Röntgenstrahlen. *Nachr Akad Wiss Gott Math Physik Kl* 1918:98–100.
- Guerette PA, et al. (2014) Nanoconfined β -sheets mechanically reinforce the supra-molecular network of robust squid Sucker Ring Teeth. *ACS Nano* 8(7):7170–7179.
- Marsh RE, Corey RB, Pauling L (1955) An investigation of the structure of silk fibroin. *Biochim Biophys Acta* 16(1):1–34.
- Warwicker J (1954) The crystal structure of silk fibroin. *Acta Crystallogr* 7(8-9):565–573.
- Warwicker JO (1960) Comparative studies of fibroins. II. The crystal structures of various fibroins. *J Mol Biol* 2(6):350–362.
- Thiel BL, Guess KB, Viney C (1997) Non-periodic lattice crystals in the hierarchical microstructure of spider (major ampullate) silk. *Biopolymers* 41(7):703–719.
- Lotz B, Colonna Cesari F (1979) The chemical structure and the crystalline structures of Bombyx mori silk fibroin. *Biochimie* 61(2):205–214.
- Glatter O, Kratky O (1982) *Small Angle X-Ray Scattering* (Academic, London).
- Landi A, Hamilton RF (2015) The impact of martensite deformation on shape memory effect recovery strain evolution. *Mater Mater Trans A* 46(8):3481–3489.
- Kausch HH (2012) *Polymer Fracture* (Springer, Berlin).
- Seitz J (1993) The estimation of mechanical properties of polymers from molecular structure. *J Appl Polym Sci* 49(8):1331–1351.
- Sariola V, et al. (2015) Segmented molecular design of self-healing proteinaceous materials. *Sci Rep* 5:13482.