Title

Nanolattice-Forming Hybrid Collagens in Protective Shark Egg Cases

Journal

Biomacromolecules, 23(7)

ISSN

1525-7797

Authors

Goh, Rubayn

Yoshida, Eric

Schaible, Eric

et al.

Publication Date

2022-07-11

DOI

10.1021/acs.biomac.2c00341

pubs.acs.org/Biomac

Article

Nanolattice-Forming Hybrid Collagens in Protective Shark Egg Cases

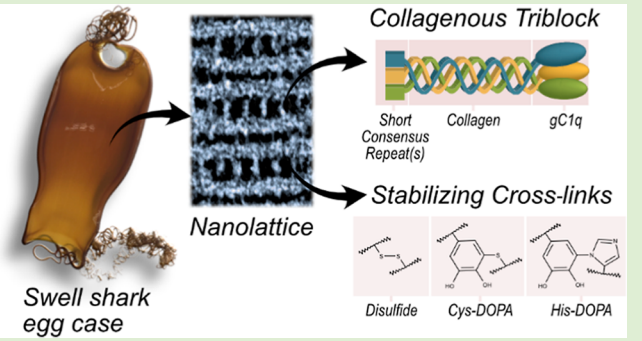
Rubayn Goh,* Eric Yoshida, Eric Schaible, Rachel Behrens, Christophe A. Monnier, Bradley Killingsworth, Kiat Whye Kong, Shu Hui Hiew, Ali Miserez, Shawn Hoon, and J. Herbert Waite*

Cite This: *Biomacromolecules* 2022, 23, 2878–2890

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Nanoscopic structural control with long-range ordering remains a profound challenge in nanomaterial fabrication. The nanoarchitectured egg cases of elasmobranchs rely on a hierarchically ordered latticework for their protective function—serving as an exemplary system for nanoscale self-assembly. Although the proteinaceous precursors are known to undergo intermediate liquid crystalline phase transitions before being structurally arrested in the final nanolattice architecture, their sequences have so far remained unknown. By leveraging RNA-seq and proteomic techniques, we identified a cohort of nanolattice-forming proteins comprising a collagenous midblock flanked by domains typically associated with innate immunity and network-forming collagens. Structurally homologous proteins were found in the genomes of other egg-case-producing cartilaginous fishes, suggesting a conserved molecular self-assembly strategy. The identity and stabilizing role of cross-links were subsequently elucidated using mass spectrometry and *in situ* small-angle X-ray scattering. Our findings provide a new design approach for protein-based liquid crystalline elastomers and the self-assembly of nanolattices.

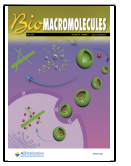


INTRODUCTION

The survival of all living organisms depends on the presence and functional properties of passive protective structures. Passive structures such as skin provide effective barriers against episodic environmental threats such as infectious agents, corrosives, foulants, and predators and are multifunctional with critical roles in maintaining heat and water balance.[1] In fertilized eggs, eggshells[2] and egg cases[3,4] ensure insulation while maintaining metabolic and respiratory waste exchange. Shark and skate egg cases, with their exquisite textures and shapes, have long been objects of fascination as their sobriquet "mermaid's purses" attests. Recent studies of the swell shark egg case have shown that this fascination is not undeserved—despite being composed of ∼62% water, the egg case features excellent mechanical properties, selective permeability, and an intricate hierarchical architecture ranging from centimeters down to the nanometer length scale.[3] Swell sharks are bottom dwellers and reside among rocky crevices. Unsurprisingly, the egg cases of swell sharks are often found deposited on rocky reef structures in the benthic zone with their tendrils often wrapped around rocks and kelp to prevent them from being swept away by current and surges. Despite this, dislodged wave-swept egg cases can occasionally be found along the harsher, turbulent intertidal zones. The adaptions of the egg cases allow them to resist and survive the buffeting waves and abrasive environment. The architecture of each egg case is composed of nanoribbons assembled in Bouligand-like twists resembling synthetic fiber-reinforced composites. However, a distinguishing factor, the nanolattice—resembling mechanical metamaterials—that constitutes these nanoribbons. The nanolattice provides lattice-governed deformation mechanisms that include uniform nucleation of nanoscale shear bands, with an orientation-dependent lattice rotation that facilitates shear band formation.[3] The exquisite architecture enables the multifunctional egg case to achieve a finely tuned balance between two opposing attributes that are mandatory for survival, that is, porosity and mechanical toughness (Figure 1). Above all, the aqueous and bottom-up fabrication process of elasmobranch egg cases provides an exemplary model for the self-assembly of long-range, hierarchically ordered nanoarchitectures.

Earlier studies on the formation of elasmobranch egg cases revealed multiple intermediate lyotropic liquid crystalline phases that exist in storage granules of the egg-case-producing nidamental gland. These precursors undergo an extrusion-like

Received: March 18, 2022
Revised: June 2, 2022
Published: June 24, 2022



https://doi.org/10.1021/acs.biomac.2c00341
Biomacromolecules 2022, 23, 2878–2890

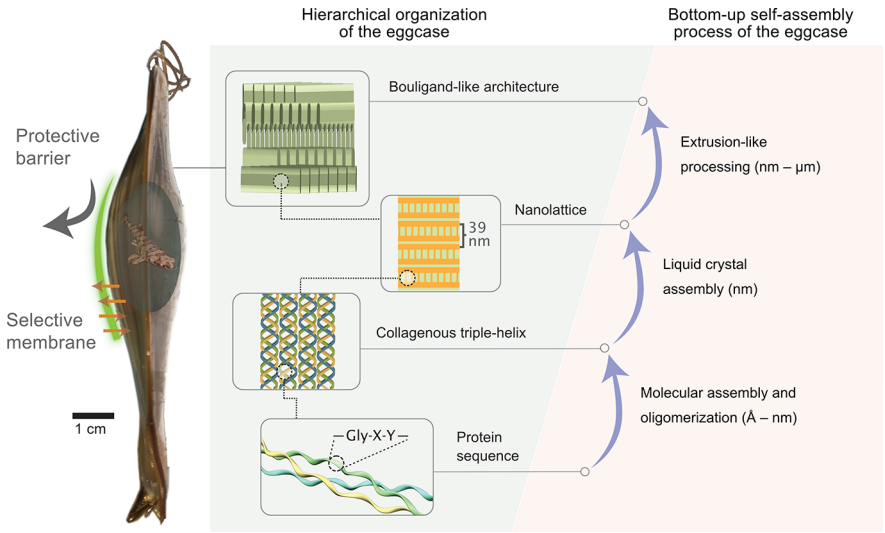


Figure 1. Illustration of the multifunctionality, hierarchical organization, and the enabling multistep self-assembly process of the collagenous egg case. The protein sequence serves as the fundamental basis for the self-assembly of the complex architecture seen in the egg case.

process with intermediate lyotropic phases that include micellar, lamellar, cholesteric, hexagonal, and a phase equivalent to the nanolattice seen in the final egg case.[5] These transitions have been proposed to be influenced by both pH and concentration.[6] The eventual structure is then further stabilized *via* oxidative covalent cross-linking,[7] which results in a highly ordered, anisotropic, and mechanically stable nanoarchitecture that is comparatively analogous to synthetic liquid crystalline elastomers.[8] Being a proteinaceous material, the basis of the architecture is determined by the primary protein sequence (Figure 1). Our aim here was to identify the egg case proteins behind the self-assembly of the nanolattice, as well as to distill the molecular design principles that will provide bioinspired guidelines to develop materials where nanoscopic structural control, strength, and well-defined selective permeability are desirable.

Structural proteins have been notoriously challenging to sequence due to their highly cross-linked nature and ensuing insolubility.[9] In this work, we leveraged RNA-sequencing and proteomic techniques[10] to obtain the complete sequence of the nanolattice-forming proteins and identify cross-links involved in mechanical reinforcement. This newly identified cohort of load-bearing proteins is distinguished by a contiguous modular series of collagens flanked by short consensus repeats (SCRs) and C1q globular (gC1q) domains—two protein domains that are classically associated with the immune system.[11,12] A database search based on our results revealed proteins with similar structural configurations in the genomes of both oviparous and ovoviviparous elasmobranchs and chimaera, suggesting a conserved molecular strategy for the assembly of egg cases from cartilaginous fishes. Using amino acid analysis and liquid chromatography coupled with tandem mass spectrometry (LC−MS/MS) of egg case hydrolysates, along with *in situ* small-angle X-ray scattering (SAXS) during tensile deformation, we identified the chemical cross-links within the egg case and elucidated the structural and mechanical role of chemically reducible cross-links. Our findings advance the understanding of distinctive network-forming collagens[13] and will foster the development of supramolecular self-assembled nanomaterials. The molecular strategy identified in the egg case provides insights into a new paradigm for the self-assembly of mechanical metamaterials,[14] liquid crystalline elastomers,[15] and selective membranes where well-defined pore sizes are sought-after such as ultrafiltration[16] and biomedical applications.[17]

MATERIALS AND METHODS

Egg Cases and Nidamental Gland Collection. Nidamental gland and egg case specimens for protein identification and amino acid analyses were donated by a parasitology class. The gravid female swell shark intended for the parasitology class was freshly collected from the wild and euthanized with an overdose of MS-222 by immersion into sodium-bicarbonate-buffered MS-222 bath (>250 mg/L). The nidamental glands were stored in RNAlater and stored at −80 °C until needed. Egg cases for cross-link identification were obtained from tank-reared swell sharks, sacrificed within 2 months of being laid, and stored at −80 °C until needed. The embryos were carefully removed from the egg case and euthanized with an overdose of MS-222 by immersion into sodium-bicarbonate-buffered MS-222 bath (1 g/L). The procurement of egg cases is in accordance with Institutional Animal Care and Use Committee (IACUC) guidelines and reviewed by UCSB IACUC.

Transcriptome Library Construction. RNA-seq analysis was performed as previously described.[10] The nidamental gland was homogenized at 4 °C and total RNA extracted with a Monarch Total RNA Miniprep kit. Samples were dried in RNAstable and stored at −20 °C. RNA was then resuspended in diethylpyrocarbonate-treated water and poly-A mRNA was enriched with oligo dT beads. Subsequent library preparation was done with a NEXTflex Rapid Directional RNA-Seq Library Prep kit. Library quality was analyzed with an Agilent 2100 bioanalyzer.

The library was then sequenced on an Illumina HiSeq 4000 and 2 × 151 bp paired-end reads were collected. *De novo* transcript assembly was performed on quality-filtered FASTQ sequences with the Trinity software suite on a computational cluster using standard parameters. Putative coding regions from the assembled transcript sequences were extracted with TransDecoder that is part of the Trinity software suite.[18]

Preparation and Pepsin Digestion of Egg Case. Egg cases were dialyzed against Milli-Q water thrice, 10 min each, prior to further processing. Inner and outer cuticles of the egg cases were peeled off, and the mid layer which comprises the bulk of the egg case was homogenized in Milli-Q water at 4 °C. The homogenized egg case was then lyophilized and stored at −80 °C.

Samples were treated in 10 mM dithiothreitol (DTT-treated) and compared against untreated samples (native). Egg cases were

https://doi.org/10.1021/acs.biomac.2c00341
Biomacromolecules 2022, 23, 2878−2890

Table 1. Primer Sequences

Pursein-1a Forward Primer	CATTGCAAGCAGTGGTATCAACATTGCAAGCAGTGGTATCAA
Pursein-1a Reverse Primer 1	CATTCCTCACAATGACCTGACATTCCTCACAATGACCTGA
Pursein-1a Reverse Primer 2	CAAATAATTGGCCCAGGAGCCAAATAATTGGCCCAGGAGC
Pursein-1b Forward Primer	AGAGAGATTCTGAAAGCCTGAGAGAGATTCTGAAAGCCTG
Pursein-1b Reverse Primer	TTATCCTCTCAGCTCTCAACTTATCCTCTCAGCTCTCAAC
Pursein-2a Forward Primer	AAGAGAGATTCTGAAAGCCTAAGAGAGATTCTGAAAGCCT
Pursein-2a Reverse Primer	CACGCTGTGATTAAGGAAATCACGCTGTGATTAAGGAAAT

subsequently pepsinized at a concentration of 1 mg/mL (4 M urea, 5% acetic acid) at a 1:50 pepsin (Promega) to substrate ratio for 48 h at 4 °C. Samples were clarified by centrifugation and the soluble fractions were dialyzed twice against 0.1 M sodium borate (pH = 8.0). Dialyzed samples were centrifuged, and the insoluble fractions were washed once with 0.05 M sodium borate (pH 8.0), resuspended in 0.1% acetic acid, and dialyzed thrice against 0.1% acetic acid. Dialyzed samples were lyophilized and resolubilized prior to SDS−PAGE. All steps were performed at 4 °C.

Samples were mixed 1:1 with 2× Laemmli buffer (Bio-Rad) containing 100 mM DTT and heated at 70 °C for 10 min. Samples were loaded onto a 12% Bis−Tris gel (Bolt) and run at 110 V for 1 h 30 min in MOPS running buffer (Bolt). The gel was rinsed thrice with water, stained with 0.1% Coomassie Brilliant Blue R-250 (Bio-Rad) in 20% methanol, 0.1% acetic acid, and destained with 30% methanol followed by water. Protein bands were excised using a sterile scalpel blade, covered with 0.1% acetic acid, and shipped to Bioproximity for tryptic digestion and LC−MS/MS (Thermo Q-Exactive HF-X Orbitrap mass spectrometer). Acetic acid and water used were LC−MS grade (Optima).

Protein Identification, Verification, and Analysis. LC−MS/MS results were *de novo* assembled and searched against open-reading frames of the transcriptomic library using PEAKS Studio XPro (Bioinformatics Solutions) with post-translational modifications being identified in the process. Results were filtered with a 1% false discovery rate threshold and a minimum of 5% ion intensity for post-translational modifications.

cDNA library for sequence verification by Sanger sequencing was generated from poly-A enriched mRNA. Poly-A mRNA was enriched from total RNA using a NEB Magnetic mRNA isolation kit whereas cDNA library was prepared with NEBNext Single Cell/Low Input cDNA Synthesis and Amplification Module. Gene-specific primers were designed using Primer-BLAST[19] for PCR amplification and Sanger sequencing. PCR amplification was performed using Phusion High-Fidelity DNA Polymerase. Pursein-1a had an incomplete 5′ end, therefore sequence of the template-switching oligo from the NEBNext cDNA synthesis kit was used as the forward primer along with pursein-1a reverse primer 1 for PCR amplification. Sequencing reads for pursein-1a forward primer were low in quality and therefore reverse primer 2 was designed for sequencing of the 5′ end of pursein-1a (Table 1).

***In Silico* Analyses.** Signal peptides were identified using SignalP 5.0.[20] All signal peptides had a probability of >98%. However, the exact cleavage site for pursein-1a was ill-defined with the most probable site having a probability of 49.8%. Multiple sequence alignment of proteins was performed on ClustalW.[21] UniProt accession numbers of proteins used in the alignments are P02745, P02746, P02747, P27658, P25067, Q03692, and P17927.

Homologous proteins were searched using BlastP with pursein sequences against the non-redundant protein database. Sequences identified were then scanned using ScanProsite for contiguous SCR and collagenous and gC1q domains. Signal peptides were identified as before. Signal peptides in the proteins of the cloudy catshark and thorny skate were detected only with start codons further downstream than the one indicated. Accession numbers of the protein sequences are XP_032902287.1, XP_007894850.1, XP_007894575.1, GCC26270.1, XP_020368837.1, XP_020368836.1, and GCB65488.1. These proteins were previously labeled uncharacterized or loosely categorized based on similarities to the recognized proteins such as collagen and C1q.

The sequence of gC1q domains in purseins was identified using ScanProsite[22] and submitted to SwissModel[23] for homology modeling. The crystal structure of the homotrimer of collagen X[24] was used as the template for the homology model. GMQE scores were 0.67, 0.69, and 0.61 and QMEAN Z-scores were −3.27, −2.70, and −2.57 for purseins-1a, 1b, and 2a, respectively. Regions that are best modeled corresponded to the conserved core β-sandwich scaffold.[11] Results were exported and the protein surface was visualized using ChimeraX (version 1.1).[25]

Hydropathy plots of purseins were obtained with Expasy ProtScale using the hydrophobic scale developed by Miyazawa and Jernigan[26] with a 21-residue window size. Sequence of the signal peptides was not included in the hydropathy plots.

Amino Acid Hydrolysis and Analysis. Homogenized and lyophilized egg case was treated with 10 mM DTT for 24 h at 4 °C and treated with 50 mM iodoacetamide in 100 mM ammonium bicarbonate for 40 min in the dark. Samples were centrifuged at 14000×*g*, the supernatant was removed, and the pellet resuspended in Milli-Q water for a total of three times. DTT treatment was omitted to compare the amount of cysteines in the native and the reduced state. Cysteine was detected as carboxymethylcysteine. Samples were then lyophilized and resuspended in 500 μL of 6 M HCl and 5% phenol at ∼2 mg/mL, vacuum sealed, and hydrolyzed at 110 °C for 24 h. The hydrolysates were cleaned up by complete evaporation on a SpeedVac (Savant) followed by resuspension and evaporation, thrice with 200 μL of Milli-Q water and thrice with 200 μL of ethanol. Samples were then resuspended in 0.01 N HCl and loaded onto the automated amino acid analyzer (Hitachi High-Technologies L-8900).

Isolation of DOPA-Containing Cross-Links and Identification with LC−MS/MS. Homogenized and lyophilized egg case was resuspended in 6 M HCl and 5% phenol, vacuum sealed, and hydrolyzed at 110 °C for 2 h. The hydrolysates were cleaned up by complete evaporation on a SpeedVac (Savant) followed by resuspension and evaporation, thrice with 200 μL of Milli-Q water and thrice with 200 μL of ethanol. The hydrolysate was then resuspended in 0.1 M sodium phosphate buffer (pH = 7.4) and centrifuged at 14,000×*g*. The soluble supernatant was passed through a phenylboronate affinity column (Affi-Gel boronate, Bio-Rad) that captures *cis*-diols. After loading the supernatant, the column was washed extensively with 10 column volumes of 0.1 M phosphate buffer (pH = 7.4), 10 volumes of 2.5 mM ammonium bicarbonate, and 10 volumes of Milli-Q water. Bound ligands were then eluted with 10 volumes of 5% acetic acid and lyophilized. The isolated sample was redissolved in 1% acetic acid for UPLC−MS/MS analyses. Mass spectrometry was carried out on a Waters Xevo G2-XS QTof with positive mode electrospray ionization coupled to an ACQUITY UPLC H-Class system. Separations were achieved on a Waters BEH C18 column with a linear gradient of 0 to 60% ACN (0.1% formic acid) over 9 min. For MS/MS analysis, a fixed collision energy of 20 kEV was employed.

***In Situ* SAXS during Tensile Testing.** Egg cases were punched into dogbone-shaped samples with a gauge length of 5 mm and a width of 2 mm. DTT-treated samples were treated with 10 mM DTT in artificial seawater (450 mM NaCl, 10 mM KCl, 9 mM $CaCl_2$, 30 mM $MgCl_2$, and 16 mM $MgSO_4$) for at least 24 h at 4 °C. Samples were then marked with a pair of permanent marker lines for optical strain tracking.

https://doi.org/10.1021/acs.biomac.2c00341
Biomacromolecules 2022, 23, 2878−2890

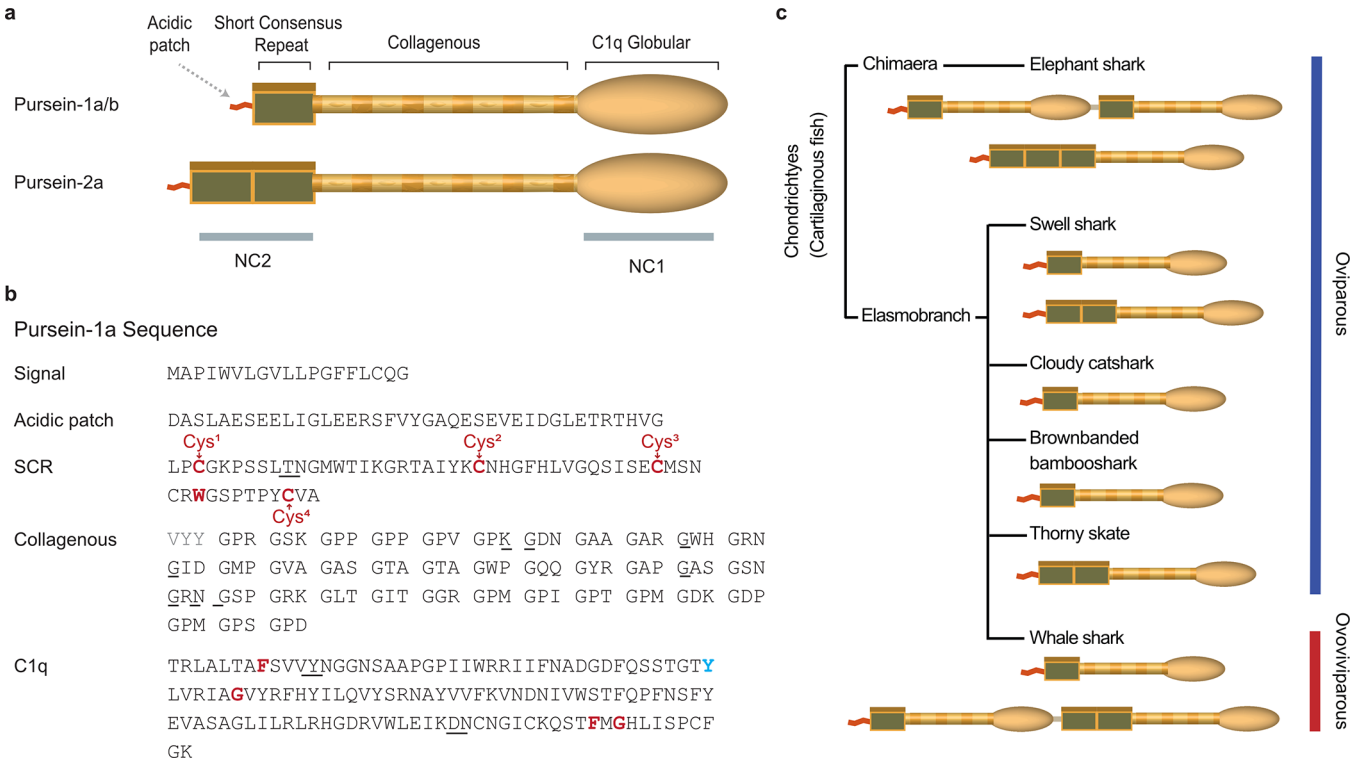


Figure 2. Full-length sequence and structural analysis of purseins and homologous proteins. (a) Illustration of purseins showing structural homology and the contiguity of the domains. (b) Sequence of pursein-1a with consensus sequences of the domains in red. Deviation from the gC1q consensus sequence, F → Y^{272}, is in blue. Intermittent sequences are in gray and Asn−Gly pairs are underlined. (c) Structurally homologous sequences found in the annotated genomes of other oviparous and ovoviviparous cartilaginous fishes.

Three samples of each condition, native and DTT-treated, were uniaxially and symmetrically strained on a Linkam TST350 tensile stage at a rate of 50 μm s^{-1} during *in situ* SAXS measurements at Beamline 7.3.3 of the Advanced Light Source.[27] The force, displacements of the markers (CCD camera), and X-ray scattering were recorded synchronously during *in situ* SAXS during mechanical testing. X-ray scattering was collected on a Dectris Pilatus3 2M detector whereas samples were exposed to 10 keV X-rays for 0.5 s at 2.5 s intervals. Calibration of the beam center and sample-to-detector distance with silver behenate and all SAXS data reduction were done using the Nika Package[28] on Igor Pro 8.

Macroscopic strains were measured by determining the change in centroid-to-centroid distance of the lines (Δl) with respect to the initial distance (l_0), using the equation $\varepsilon = \Delta l/l_0$. Stress and strain values were interpolated at 2% strain intervals and averaged with the calculated standard deviation. As the optical tracking of strain is subjected to inaccuracies due to centroid detection and camera resolution, corresponding stress and peak intensity of duplicate strain values and non-sequential strain values were averaged to allow interpolation of the data. Diffraction intensities plotted in Figure 4b were obtained by line integrations parallel to the stress axis—along the vertical axis of the 2D SAXS with a line width of 3 px. The integrated intensities were then subtracted with an exponential baseline and fitted to a Gaussian function. Comparison of peak intensities were calculated with the heights of the fitted curves in Figure 4b, and the full widths at half maximums were plotted for comparison in Figure 4c. To calculate the average and standard deviations of stress and peak intensities, the corresponding values were interpolated in 2% strain intervals and averaged. Data points with less than three samples were not included in the plots (*i.e.,* due to peak intensities being lower than the detection limits).

Approximation of nanoribbon strain distribution as shown in Figure 4e was processed by masking the q^* peak for each image in the egg case tensile data set and successive fitting of the radial intensity of the $2q^*$ peaks. Fitting was done from 0 to 180°, in 1° increments, with the axis of tension defined at 0°. Nanoribbon strains were calculated by converting q-space to d-space ($2\pi/q$) and applying the equation $\varepsilon_r = \Delta d/d_0$ where Δd is the change in d-spacings and d_0 is the d-spacing of the unstrained material. All images in the data set were similarly processed to yield complete measurements of the evolution of the d-spacings and strain of the nanoribbons, at every angle from 0 to 180°, at every time point in the experiment.

Polarized Optical Microscopy. Egg case was cryosectioned with a cryostat (Leica CM1850) and sections were washed with Milli-Q water to remove residual OCT compound. Cross sections of the egg case were then examined under an optical microscope (Olympus BX51) with cross polarizers (BX-POL) and imaged using a colored camera (Olympus DP73).

Transmission Electron Microscopy. The fixation and analysis of egg cases with transmission electron microscopy (TEM) was performed as previously described.[3] Briefly, egg cases were immersed in 90 mM PIPES, 37.5 mM HEPES, 3 mM $MgCl_2$, and 9% sucrose (pH = 8.0, 1.5× MarPHEM) for 30 min at 4 °C followed by fixation in a solution of 2.5% glutaraldehyde (EMS #16220) and 1.5% formaldehyde (EMS #15714) in 1.5× MarPHEM for 2 h at 4 °C. Samples were washed with Milli-Q water to remove residual fixatives and post-fixed in a solution of 2% osmium tetroxide for 90 min at 4 °C followed by washing thrice in excess Milli-Q water. Subsequently, the samples were dehydrated by solvent exchange with 0, 20, 40, 60, 80, and 100% ethanol in water followed by an exchange with a graded series of 0, 20, 40, 60, 80, and 100% propylene oxide in ethanol. Gradual infiltration of Embed 812 and embedding of samples were then performed in accordance with the manufacturer's instructions. 50 nm ultrathin sections were then prepared with an ultramicrotome (Leica EM UC6), mounted on a copper grid, stained with 0.6% uranyl acetate (Ted Pella #19481) in 30% ethanol at 60 °C and 3% lead citrate (EMS #22410), and imaged in HAADF-STEM mode on a ThermoFisher Talos G2 200× with an accelerating voltage of 200 kV.

Atomic Force Microscopy and Force Spectroscopy. Cryosectioned samples were mounted under and investigated on a MFP-3D

https://doi.org/10.1021/acs.biomac.2c00341
Biomacromolecules 2022, 23, 2878−2890

atomic force microscope (Asylum Research, Santa Barbara, USA) mounted on an inverted light microscope. AppNano FORTA silicon tips were used for all experiments, and their exact spring constant values (~2−3.6 nN/nm) were assessed experimentally by the thermal tune method before each use. Deflection sensitivity was established using glass as an indefinitely stiff reference material. All images were recorded at 0.7 Hz in AC mode.

Force spectroscopy measurements were performed with a loading force of 50 nN and a loading/unloading rate of 250 nm/s was chosen with respect to Buckle's one-tenth law[29] to mitigate any influence of the underlying layers. Force maps with resolutions of 24 × 24 pixels (576 force curves) were then recorded over the previously imaged areas. An Asylum Research MFP-3D Hertz analysis tool was used to determine the stiffness values. The upper 60% of the approach curve, a half-angle of 20°, and a Poisson ratio of 0.33 were used to fit the model. The calculated stiffness values were then spatially plotted to yield a color-coded stiffness map with OriginLabs 8.5. A two-dimensional spline interpolation was performed on these maps to smoothen the visual presentation of the data. Peak values in the binned stiffness distributions were identified *via* the Origin Peak Analysis feature.

RESULTS

C. *ventriosum* Collagenous Nanolattice-Forming Proteins. The existence of triple-helical collagens in elasmobranch egg cases was previously established by X-ray scattering, amino acid composition, and thermal shrinkage studies.[7] However, efforts to characterize these nanolattice-forming proteins from the gland extracts of a related elasmobranch (lesser-spotted dogfish) by Edman sequencing yielded only limited protein sequences.[30] To further investigate the full-length sequence of the collagenous egg case proteins, we established an mRNA transcriptome library isolated from the nidamental gland of a gravid swell shark, an elasmobranch endemic to the coast of central California. The library was assembled by *de novo* transcript assembly of short reads using the Trinity software suite.[18]

Owing to native collagen's resistance to pepsin,[31] we narrowed the search for the collagenous nanolattice-forming proteins by pepsinization of the egg case and sequencing the pepsin-resistant fragments using in-gel tryptic digestion followed by LC−MS/MS. Full-length sequences of collagenous proteins and their post-translational modifications were identified by searching *de novo* assembled partial protein sequences against protein sequences predicted from the open-reading frames of the transcriptome library.

Nanolattice-forming proteins, hereafter called "purseins", were identified and are distinguished by a modular architecture consisting of SCR, collagen, and gC1q domains. We retain the nomenclature of collagens and refer to the non-collagenous C-terminal domain and N-terminal domain(s) as NC1 and NC2 (Figure 2a), respectively. The pursein-1 and pursein-2 variants differed by the number of SCRs. In all cases, an acidic patch with an isoelectric point of ~4 and a signal peptide were identified at the N-terminus of the translated sequences but were not detected by LC−MS/MS. Highly conserved nonhomologous acidic sequences have also been reported in silk fibroin and on the C-terminal side of the collagen domain in byssal thread preCOL precursors.[32,33] With regard to the fibroins, Schaefer *et al.* proposed that acidic domains engage in electrostatic interactions with Ca and K ions in concentration-dependent viscosity adjustments[33] during the self-assembly process. This may or may not be relevant to the N-termini of purseins as well.

Partial coverage of the pursein-1b NC2 domain and NC1 domains of all three purseins suggests that these domains were not proteolytically processed in the final assembly of the egg case (Figure S1), unlike mammalian fibrillar collagens where N- and C-termini of procollagens are enzymatically cleaved during self-assembly.[13] Alignment of the pursein sequences showed similarities in the NC2 and collagenous domains of the partially sequenced dogfish egg case proteins[30] (Figure S2).

Variable hydroxylation of Pro, deamidation of Asn, and oxidation of Met and Trp were also predicted from the *de novo* assembled peptide sequences (Figure S1). At least five incidences of the Asn−Gly dipeptide, which is susceptible to spontaneous deamidation,[34,35] were seen in all purseins (Figures 2b, S1) and correlate well with the locations of the variably deamidated Asn residues. Analysis of the collagenous region of the pursein sequences also indicated the unusual presence of the helix-destabilizing aromatic residues Trp and Tyr.[36] By contrast, Cys were localized to the non-collagenous flanking domains.

Although the tandem configuration of the three modules is unique to purseins, the collagen-gC1q configuration has also been identified in C1q proteins and network-forming collagens, types VIII and X.[37] Multi-sequence alignment of gC1q modules indicated the conservation of key hydrophobic core residues in purseins except for a $F \rightarrow Y^{272}$ substitution in pursein-1a.[37] Alignment of the SCR domains confirmed that the consensus sequence was conserved, which includes four invariant Cys and one Trp and other mostly invariant Pro, Gly, and hydrophobic residues (Figures 2b, S3).[12] The four invariant Cys residues are also known to form specific intradomain disulfide bonds having a Cys^1−Cys^3 and Cys^2−Cys^4 disulfide bond configuration[12] (Figures 2 and S3b).

Conserved Molecular Strategy of Nanolattice-Forming Proteins. Protein sequences of purseins were searched against the non-redundant protein database on NCBI using the BLAST algorithm.[38] Proteins with similar configurations were identified in both oviparous and ovoviviparous elasmobranchs and chimaera (Figure 2c). Extensive scans of other similar proteins originating from species not belonging to either of these classifications did not share the same structural configurations that are unique to purseins. In all cases of pursein-like proteins identified through the search, the proteins were made of varied configurations of the same three modules identified in purseins. These proteins were previously designated as uncharacterized or vaguely classified based on similarities with known proteins such as collagen and C1q. The accession numbers of these proteins are listed in the Supporting Information (In Silico Analyses). Collagenous domains in these proteins were similar in length, ~118 amino acids. Comparatively, the triple-helical domains of collagen types VIII and X are ~456 residues and ~78 residues long for C1q proteins, respectively (Figure S4a).

Signal peptides were identified in all pursein-like proteins except for the proteins identified in the whale shark genome—the largest aquatic non-mammal and the only ovoviviparous elasmobranch under comparison.[39] As a caveat, the predicted sequence in the whale shark has not been backed by transcriptome support, and it is therefore unclear if the lack of signal peptide is due to partial gene prediction of the gene architecture or an indication of a different mode of secretion, for example, by holocrine secretion as in the secretion of hagfish slime, involving the rupture of the cellular plasma membrane.[40]

https://doi.org/10.1021/acs.biomac.2c00341
Biomacromolecules 2022, 23, 2878−2890

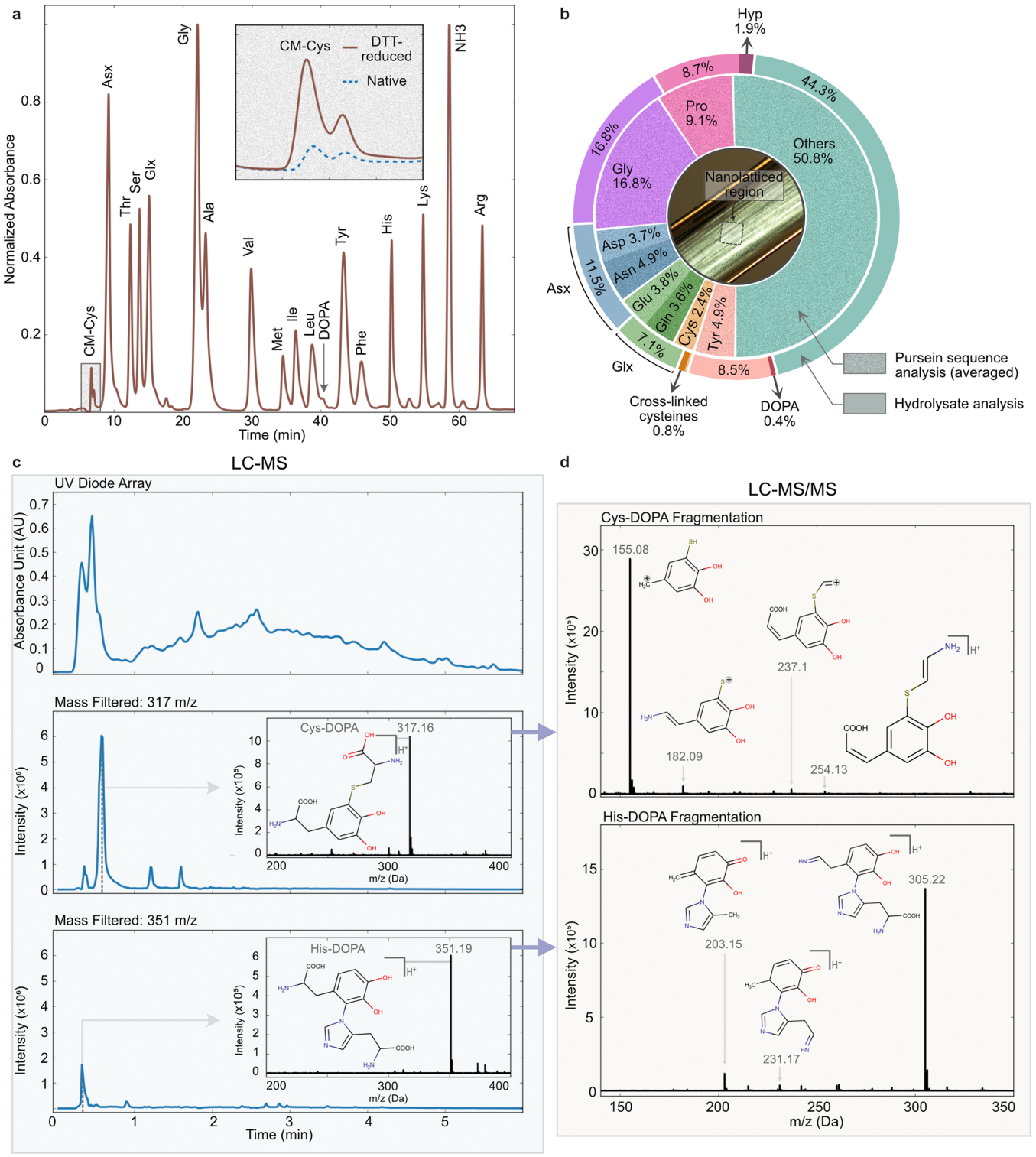


Figure 3. Biochemical analysis of the swell shark egg case. (a) Amino acid analysis of the DTT-reduced egg case hydrolysate. Inset shows the enlarged carboxymethylcysteine (CM-Cys) peak in both DTT-treated and native egg case. Intensities have been normalized to the intensity of Gly. (b) Amino acid composition comparison between averaged pursein sequence and DTT-reduced egg case hydrolysate. Asn/Asp and Glu/Gln were not differentiated during amino acid analysis and are reported as Asx and Glx, respectively. Inset illustrates the birefringence seen in the egg case under polarized optical microscopy and the nanolattice-forming region that was used in the hydrolysate analysis is indicated within the image. (c) LC−MS of *cis*-diol isolates of partially hydrolyzed egg case, that is, catecholic compounds and sugars. Masses corresponding to cysteinyl-DOPA (Cys-DOPA) and histidyl-DOPA (His-DOPA) were identified and verified by their (d) LC−MS/MS fragmentation spectra.

Amino Acid Analysis and Cross-Link Identification. The amino acid compositions of egg case hydrolysates closely resemble the composition profile of the purseins. The presence of ∼17% Gly and ∼9% Pro and *ca.* 2% hydroxyproline (Hyp) is indicative of the collagenous nature of purseins (Figure 3a,b). This agrees with the initial assumption that the egg cases

https://doi.org/10.1021/acs.biomac.2c00341
Biomacromolecules 2022, 23, 2878−2890

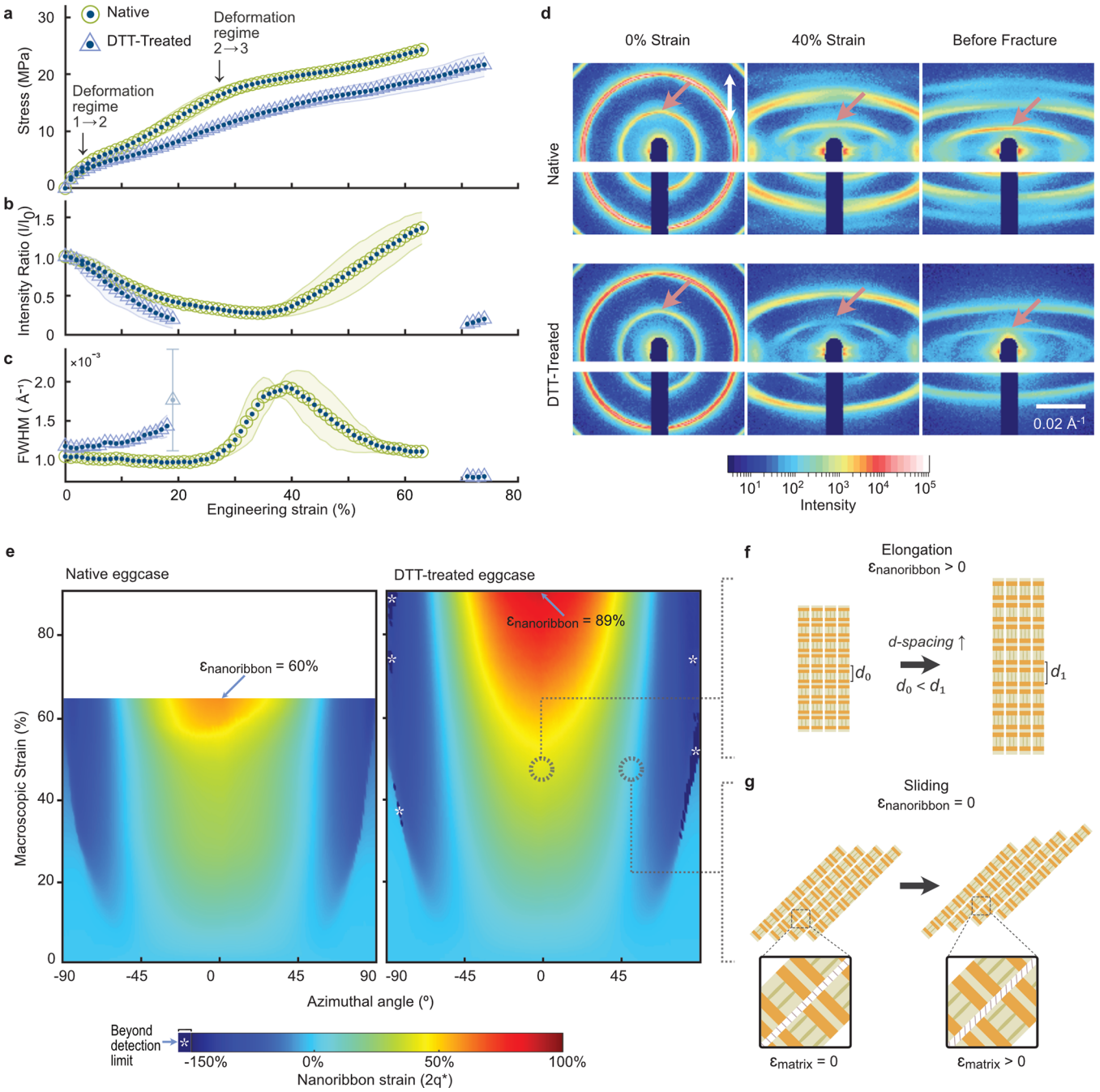


Figure 4. *In situ* SAXS during tensile testing of native and DTT-reduced egg cases. (a) Stress−strain curve of native and DTT-treated egg cases strained along the long axis of the egg case and measured with *in situ* SAXS during tensile deformation. The approximate transitions from deformation regime 1 → 2 and 2 → 3 in the native egg case are indicated by arrows. (b) Intensity (I) of the scattering vector q^* parallel to the stress axis (normalized against the initial intensity I_0) plotted against strain for the corresponding samples and (c) FWHM of q^* parallel to the stress axis with respect to strain. Standard deviations of (a−c) are plotted as shaded regions except for the FWHM of $q^*_{\text{DTT-treated}}$ at 19% strain whereby an error bar is used to acknowledge the uncertainty in that measurement point. (d) 2D SAXS of representative samples at various strains. The white arrow represents the stress axis. The red arrow highlights the decrease and the subsequent increase in q^* intensity along the stress axis at higher strains. (e) Angular distribution of nanoribbon strain for native and DTT-treated egg cases derived by determining $2q^*$ across −90−90°, where 0° is the stress axis. Dark blue regions marked with an asterisk are areas where the signal intensity was too low for quantifying strain. Nanoribbon parallel to the stress axis is indicated with an arrow. (f) Elongation of nanoribbons parallel to the stress axis results in an increase in *d*-spacing. (g) Sliding of nanoribbons ∼50° to the stress axis does not result in a change in the *d*-spacing. Inset illustrates the deformation of the interfacial matrix during sliding.

are primarily collagenous[7] and are predominantly composed of purseins. About 0.4% dihydroxyphenylalanine (DOPA) was also detected. Cys residues were protected during hydrolysis by alkylation with iodoacetamide and were quantified as carboxymethyl-cysteine (CM-Cys), resulting in *ca.* 0.23% of Cys in the native samples. Comparison with the hydrolysate composition of egg case reduced with DTT indicated that *ca.* 0.83% cysteines were present as solvent-accessible cross-links

https://doi.org/10.1021/acs.biomac.2c00341
Biomacromolecules 2022, 23, 2878−2890

that could be reduced by DTT. Cysteinyl cross-links that were not reduced by DTT were left unprotected and therefore do not contribute to the quantification of Cys. Slight discrepancies between the hydrolysate composition and the translated sequence could potentially arise from differences in the ratio of purseins, unaccounted for Trp which is unstable during hydrolysis by hydrochloric acid, as well as from the potential contribution of a matrix protein that has yet to be characterized.

Empirical evidence of a change in egg case color from dark-green to orange−green upon treatment with 10 mM DTT suggests the presence of reducible aromatic cross-links besides disulfides. Given the involvement of DOPA in the egg case tanning process[7] and other sclerotized hard tissues,[41,42] the egg case was hydrolyzed for 2 h and passed through a phenylboronate affinity column to isolate possible *cis*-diols (*e.g.,* catecholic compounds and sugars). Isolated compounds were subsequently characterized with LC−MS and masses corresponding to DOPA, Cys-DOPA, and His-DOPA were detected. The identities of Cys-DOPA and His-DOPA were further confirmed by their LC−MS/MS fragmentation mass spectra (Figures 3c,d, S5, S6).[41,43,44]

Mechanical and Structural Roles of Reducible Cross-Links. Amino acid composition and cross-link identification results suggest that the intermediate lyotropic liquid crystalline phase reported in earlier studies[6,8] is formed by collagenous proteins with flanking domains stabilized by cysteinyl- and DOPA-derived cross-links. These include disulfides, Cys-DOPA, His-DOPA, and possibly dopaquinones. Both disulfide and Cys-DOPA cross-links are reducible by DTT, although Cys-DOPA would likely bleach without cleavage. Therefore, if disulfide cross-links structurally reinforce the egg case, we would expect DTT to render the egg case more compliant and extensible. To further elucidate the mechanical and structural role of reducible cross-links that are localized in the non-collagenous domains, egg cases were subjected to chemical reduction with DTT for 24 h and characterized with *in situ* SAXS during tensile testing.[45]

In earlier investigations, tensile testing of swell shark egg cases elucidated a nonmonotonic stress−strain curve with three deformation regimes that are distinguished by a change in the slope of the stress−strain response (*i.e.,* modulus) of the egg case (Figure 4a). *In situ* SAXS during tensile testing subsequently revealed that these regimes involve deformation of nanoribbons oriented parallel to the stress axis in the first regime followed by cumulative engagement of nanoribbons oriented off-axis and perpendicular to the stress axis in the second and third regimes.[3] Treatment with DTT resulted in an increase in strain to failure from 65.9 ± 0.03 to 82.1 ± 0.09% with a general decrease in the strain hardening observed in the second and third deformation regime.[3] Changes to the initial elastic modulus were minimal and within standard deviation (Figure 4a). The 39 nm periodic *d*-spacing of the nanoribbons arising from the constituting nanolattices gives rise to the observed primary peak, q^* ($q^* = 2\pi/d$), and the corresponding higher-order reflections. Therefore, analyses of the peak intensities, positions, and distribution were used to gain insights into the relative disorder and strain of the nanolattice.[3] Analyzing the primary q^* peaks parallel to the stress axis revealed a higher rate of nanolattice disordering in the DTT-treated samples. In contrast to the native samples, the DTT-treated samples had a steeper decrease in peak intensity as well as the complete disappearance of the peak between 20 and −60% strain in the second and third regimes (Figure 4b,d). This is further supported by a steeper increase in full width at half-maximum (FWHM) of the DTT-treated samples in the first two deformation regimes which correspond with the increased disordering and broadening of the *d*-spacing distribution (Figure 4c). In the native sample, a remarkable strain alignment was seen parallel to the stress axis in the third deformation regime.[3] This phenomenon was less pronounced in the DTT-treated sample (Figure 4b,d) and demonstrated the importance of the cross-links in maintaining nanolattice structural integrity against deformation and in facilitating strain alignment seen in the third regime.

To better elucidate mechanisms involved in the deformation of the egg case, the *d*-spacings derived from q^* across −90−90° at each macroscopic strain were analyzed and the strains of the nanoribbons were plotted against macroscopic strain. The resulting plot illustrates the strain distribution within the egg case while under tensile stress. Observations of the strain distribution indicated a clear delineation between elongation and contraction of nanoribbons at about ±50° of the azimuthal angle, where nanoribbons along the boundary have nearly zero strain (Figure 4e). As the azimuthal angle of 50° closely corresponds to the orientation with the highest amount of shear (45°), shear-dominant deformation mechanisms that do not affect the *d*-spacing such as fibril sliding might be the primary strain mechanism (Figure 4f,g). Because a weakening of the inter-nanoribbon interface and matrix would promote fibril sliding and influence the strain distribution within the egg case, the semblance of nanoribbon strain distribution between the native and DTT-treated egg cases implies that these mechanisms were not reinforced by reducible disulfide cross-links. Analyzing the internal strains of the nanoribbons by comparing the changes in $2q^*$ during macroscopic deformation indicated a higher strain to failure for DTT-treated nanoribbons that were parallel to the stress axis (Figure 4e). The results of the DTT-treated samples agree with our previous assessment of untreated egg case where the nanoribbon strain correlates with the macroscopic strain.[3] This is contrary to observations of deformation on collagen in tendon and fish scales[46,47] and suggests continuous stress distribution across fibers that are parallel to the stress orientation. In this situation, when $\varepsilon_{sample} = \varepsilon_{nanoribbon,}$ minimal strain will be borne by the matrix or dissipated by failure mechanisms such as fibril sliding (Figure 4f).

DISCUSSION

Embryonic survival of oviparous sharks depends on the tough yet permeable encasement that protects the embryos from the environment. This dual functionality can be attributed to the intricate hierarchical structure and the nanolattice architecture of the egg case. Key to the nanolattice self-assembly is the molecular structure of the pursein building blocks that enable a liquid crystalline assembly leading to the permeable nanolattice.[5,6,8] This study provides a deeper understanding of a molecular strategy whereby biofabrication of a multifunctional membrane is mediated by a multistep supramolecular assembly of proteins capable of forming intermediate, highly ordered, liquid crystal phases.[5]

The most noteworthy finding of this study is the discovery of a new class of modular nanolattice-forming proteins comprising SCR, collagen, and gC1q domains. Structurally homologous proteins were identified from the published genomes of other oviparous and ovoviviparous cartilaginous

https://doi.org/10.1021/acs.biomac.2c00341
Biomacromolecules 2022, 23, 2878−2890

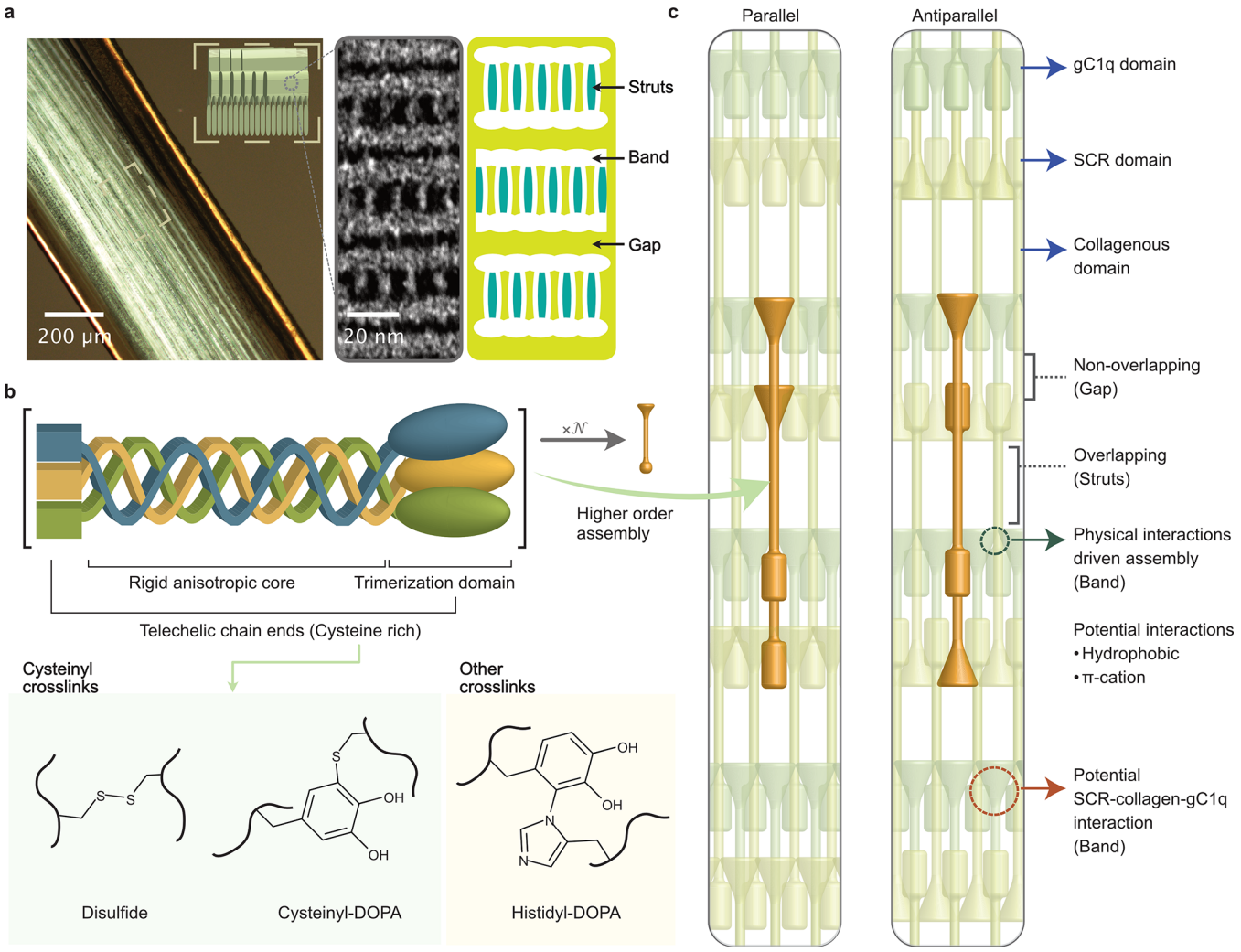


Figure 5. Hierarchical organization of purseins. (a) Birefringence of the egg case indicative of molecular anisotropy. Illustrations highlight the hierarchical assembly and features of the nanolattice architecture seen with TEM.[3] The gap region was proposed by Knupp *et al.* to be composed of a low-density arrangement of struts resulting in obscurity under TEM.[57] (b) Illustration of a pursein protomer, the functional aspect of each domain in the self-assembly of the nanolattice and cross-links identified within the egg case. (c) Adapted nanolattice self-assembly model from Knupp *et al.*[57] including molecular insights from this study. *N* is the number of protomers that form a strut and run parallel to each other, whereby $N_{\text{Knupp's model}} = 4$.

fish, suggesting a conserved strategy among elasmobranchs and chimaeras. Intriguingly, both NC1 and NC2 of purseins are composed of domains that are crucial components of the complement system on which innate immunity depends. In the complement system, complement receptor type 1 (CR1), a protein consisting of 30 tandem SCRs, binds to C1q proteins *via* its collagen stalk and gC1q domain,[48,49] suggesting likely telechelic interactions between the flanking domains of purseins.

Owing to its propensity to trimerize, gC1q has also been demonstrated to be an important trimerization-enabling domain with the ability to form hexamers.[50] Because collagens are prone to misfolding due to the ease of misalignment of the repetitive $(GXY)_n$ motifs, terminal trimerization domains ensure lateral register that is necessary during folding of the collagen triple helix. This improves the stability of the collagen and enables proper refolding.[51] In addition, the exposed strips of hydrophobic aromatic residues identified in the gC1q domains of collagen types VIII and X have been previously proposed to drive higher-order supramolecular assembly leading to non-stochastic hexagonal networks.[24,52] Homology models of the NC1 domains of purseins revealed an intriguing aromatic "crown" on the surface (Figure S4b), suggesting that a similar phenomenon as proposed in type VIII and X collagens might facilitate the supramolecular assembly of the nanolattice architecture in the egg case. Intriguingly, Arg residues were observed in the vicinity of these aromatic residues across all purseins, hinting at the possibility of π-cation contributions to the self-assembly process.[53] Protein hydropathy analysis indicates a consistent pattern of hydrophilic domains flanked by comparatively more hydrophobic domains in purseins—resembling amphiphilic lyotropic mesogens (Figure S7).[8,54] Hence, along with possible π-cation and SCR−collagen−gC1q interactions, hydrophobic interactions may be part of the mechanism that drives the supramolecular formation of a lyotropic lamellar phase bridged by collagenous struts.

The terminal modules of purseins are further reinforced by cysteine-based cross-links, that when chemically reduced result in softening of the egg case and an increased nanolattice

https://doi.org/10.1021/acs.biomac.2c00341
Biomacromolecules 2022, 23, 2878−2890

disordering during deformation. Consistent with this is a decrease in resistance of the DTT-treated terminal modules toward pepsinization (Figure S1). Because pepsin resistance of these domains is due to the occlusion of sites vulnerable to enzymatic cleavage in the native structure, this decreased resistance emphasizes the necessity of cysteinyl cross-links for maintaining the tightly folded terminal secondary structures. These results agree with the observations of the highly conserved Cys within NC2 forming intramolecular cross-links.[12] Furthermore, this also suggests that Cys residues of NC1 form intermolecular or intramolecular cross-linking despite not sharing the same conserved Cys positions as C1q proteins.[11]

To estimate the length of purseins, we assumed a collagenous triple helix with an increase per residue of *ca.* 0.29 nm,[55,56] which would result in a collagen length of *ca.* 34 nm. The gC1q trimer of collagen type X has a globular structure with a diameter of *ca.* 5 nm[24,37] whereas a single SCR module of CR1 has an averaged length *ca.* 2.9 nm.[12] Summing these distances coincides with the length of proteins that could give rise to the 39 nm periodicity in the egg case.[3] Indeed, these dimensions are in rough agreement with TEM observations of the purified, nanolattice-forming, dogfish egg case protein.[7] The mesoscale molecular arrangement of the lower density "gap" region was elucidated by freeze-etched metal-shadowed replicas of the dogfish egg case indicating a regular arrangement of fine filaments spaced laterally apart.[57] Based on unit cell symmetry observations and the continuity of the tetragonal network, a model involving axially staggered molecules with interacting globular domains was proposed for the dogfish egg case.[57] In the proposed nanolattice model, the protomers are assembled into octamers with four protomers axially shifted by ∼13 nm with respect to the other four. Complementing the proposed model with knowledge of the full-length pursein sequence, we propose that these staggered pairs of four protomers can align either in a parallel or antiparallel fashion to form the octamers, with the terminal domains facilitating the self-assembly by serving as both physical and chemical (cysteinyl) cross-linkers. Coupled with TEM observations,[3] the gap region would correspond to the non-overlapping region comprising four protomers, whereas each strut consists of an overlapping region composed of eight protomers (Figure 5).

Evidence of structural,[3] optical (Figure 5a), and mechanical anisotropy (Figure S8) supports the notion that the egg cases are analogous to synthetic liquid crystalline elastomers.[8,15,58] Although the interactions arising from the terminal non-collagenous domains are critical to the final molecular assembly, the intermediate liquid crystalline phases of purseins observed during assembly[5,6] facilitate long-range order of the nanoarchitecture. The mesogenic properties of collagens can be attributed to the rigidity and the high aspect ratio of the triple helix.[59,60] The advantage of a macromolecular mesogen is apparent for a semi-permeable membrane as it determines the dimensions of its constitutive nanolattice as well as its consequent membrane selectivity. The benefits of lyotropic liquid crystalline phases for self-assembly are two-fold: (1) the formation of well-defined, non-stochastic permeable networks and (2) the processability of dense and possibly shear-thinning fluids.[61] These properties are essential for processing dense structural materials from fluidic precursors stored in granules that need to merge into a congruent phase and solidify only at the appropriate time.[62]

CONCLUSIONS

In summary, using a top-down approach we identified a cohort of collagenous nanolattice-forming proteins, purseins, and its sequence was revealed to be of a triblock architecture with a trimeric collagenous midblock flanked by terminal domains with propensities to form supramolecular assemblies. The molecular sequence revealed in this study provides a blueprint of the necessary parameters to enable liquid crystalline assembly of a multifunctional nanolattice with well-defined porosity and selectivity. The ability to recreate nano-architectures with well-defined, low tortuosity aqueous channels, such as in the nanolattices, would be advantageous in applications where selective permeability is desirable, such as in molecular separation, drug delivery, and immunoisolation of transplanted cells.[63−65] Subsequent preservation of the nanoarchitecture in its dehydrated or pyrolyzed state (*i.e.,* pyroprotein) would provide a high specific surface area that is favorable in applications such as gas storage and catalysis.[66] Moreover, nanoarchitectured carbon lattice materials have also been shown to provide mechanical enhancements, including lightweight, strength, stiffness, flaw tolerance, and impact resilience.[14,67,68]

In contrast to industrial manufacturing practices, biological materials are processed under mild aqueous processing conditions and from natural building blocks. Therefore, elasmobranch egg cases not only provide us with molecular and nanostructural design lessons for tough and permeable materials but also offers a valuable lesson in green processing. Although it is unclear if the genomic sequences of the whale shark and elephant shark egg case-forming proteins undergo alternative splicing, it is intriguing to consider the various configurations of the three domains and the possible nanoarchitectures that arise from these proteins. Pursein and its self-assembly strategy offer a timely inspiration for nanoscopic structural control of metamaterials, membranes with well-defined porosities, and liquid crystalline elastomers. On a more fundamental level, purseins provide an ideal comparative system for gC1q-containing network-forming collagen types VIII and X, thereby deepening our understanding of their self-assembly process and identifying the design variables that regulate the hexagonal and tetragonal geometry of non-stochastic collagen networks. Future development in the fundamental understanding for the multiscale self-assembly of purseins could ultimately provide an overarching guideline for molecular design, thus spurring future novel bioinspired molecular designs for nanolattices with advantageous tailorability of lattice parameters and geometry.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.biomac.2c00341.

SDS−PAGE and LC−MS/MS coverage of pepsinized native and DTT-treated egg cases, pursein sequences, multiple sequence alignments and illustrations of pursein domains, surface illustration of purseins, LC−MS/MS of Cys-DOPA and His-DOPA, hydropathy plot of pursein sequences, and AFM images and force map of egg case cross section (PDF)

Comparative *in situ* SAXS during tensile deformation of representative native and DTT-treated egg cases (MP4)

https://doi.org/10.1021/acs.biomac.2c00341
Biomacromolecules 2022, 23, 2878−2890

AUTHOR INFORMATION

Corresponding Authors

Rubayn Goh − *Materials Department, University of California, Santa Barbara, California 93106, United States; Institute of Materials Research and Engineering (IMRE), Agency for Science, Technology and Research (A*STAR), Singapore 136834, Singapore;* orcid.org/0000-0001-8333-8143; Email: rubayn_goh@imre.a-star.edu.sg

J. Herbert Waite − *Department of Molecular, Cellular, and Developmental Biology, University of California, Santa Barbara, California 93106, United States;* orcid.org/0000-0003-4683-7386; Email: hwaite@ucsb.edu

Authors

Eric Yoshida − *Materials Department, University of California, Santa Barbara, California 93106, United States;* orcid.org/0000-0002-2767-7703

Eric Schaible − *Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States*

Rachel Behrens − *Materials Research Laboratory, University of California, Santa Barbara, California 93106, United States*

Christophe A. Monnier − *Marine Science Institute, University of California, Santa Barbara, California 93106, United States*

Bradley Killingsworth − *Department of Molecular, Cellular, and Developmental Biology, University of California, Santa Barbara, California 93106, United States*

Kiat Whye Kong − *Molecular Engineering Laboratory, Institute of Molecular and Cell Biology (IMCB), A*STAR, Singapore 138673, Singapore*

Shu Hui Hiew − *Center for Sustainable Materials (SusMat), School of Materials Science and Engineering, Nanyang Technological University (NTU), Singapore 639798, Singapore; School of Biological Sciences, NTU, Singapore 637551, Singapore*

Ali Miserez − *Center for Sustainable Materials (SusMat), School of Materials Science and Engineering, Nanyang Technological University (NTU), Singapore 639798, Singapore; School of Biological Sciences, NTU, Singapore 637551, Singapore;* orcid.org/0000-0003-0864-8170

Shawn Hoon − *Molecular Engineering Laboratory, Institute of Molecular and Cell Biology (IMCB), A*STAR, Singapore 138673, Singapore*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.biomac.2c00341

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank R. Segalman, S. Danielsen, E. Monroe, and S. Wang for the productive discussion and L.L. Tan, and D. Ignatenko for their technical assistance. We would also like to acknowledge C. Pierre, C. Orsini, and S. Simon from UCSB Marine Operations and The REEF for the collection and generous contribution of the egg cases and D. Morton for facilitating the donation of the nidamental gland from the parasitology class. The research reported here was supported in part by the National Science Foundation (NSF) Materials Research Science and Engineering Center (MRSEC) at UC Santa Barbara (NSF DMR 1720256) through IRG-3. Individual fellowship support of R.G. was provided by the Singapore Agency for Science, Technology and Research (A*STAR). X-ray scattering was performed at the Advanced Light Source, a DOE Office of Science User Facility (DE-AC02-05CH11231; beamline 7.3.3). We acknowledge the use of the shared facilities of the NSF MRSEC (DMR 1720256), the NRI-MCDB Microscopy Facility, and the research facilities within California NanoSystems Institute, supported by UC Santa Barbara, and the UC Office of the President at UC, Santa Barbara. A.M. acknowledges financial support from the Singapore Ministry of Education through an Academic Research Fund Tier 3 grant (# MOE 2019-T3-1-012).

ABBREVIATIONS

SCR, short consensus repeat; gC1q, C1q globular domain; LC−MS/MS, liquid chromatography coupled with tandem mass spectrometry; SAXS, small-angle X-ray scattering; DTT, dithiothreitol; NC, non-collagenous; Hyp, hydroxyproline; DOPA, dihydroxyphenylalanine; CM-Cys, carboxymethyl-cysteine; Cys-DOPA, cysteinyl-DOPA; His-DOPA, histidyl-DOPA; FWHM, full width at half-maximum; CR1, complement receptor type 1

REFERENCES

(1) Madison, K. C. Barrier Function of the Skin: "La Raison d'Être" of the Epidermis. *J. Invest. Dermatol.* **2003**, *121*, 231−241.

(2) Tullett, S. G.; Deeming, D. C. The Relationship between Eggshell Porosity and Oxygen Consumption of the Embryo in the Domestic Fowl. *Comp. Biochem. Physiol.* **1982**, *72*, 529−533.

(3) Goh, R.; Danielsen, S. P. O.; Schaible, E.; McMeeking, R. M.; Waite, J. H. Nanolatticed Architecture Mitigates Damage in Shark Egg Cases. *Nano Lett.* **2021**, *21*, 8080−8085.

(4) Miserez, A.; Wasko, S. S.; Carpenter, C. F.; Waite, J. H. Non-Entropic and Reversible Long-Range Deformation of an Encapsulating Bioelastomer. *Nat. Mater.* **2009**, *8*, 910−916.

(5) Knight, D. P.; Feng, D.; Stewart, M.; King, E. Changes in Macromolecular Organization in Collagen Assemblies during Secretion in the Nidamental Gland and Formation of the Egg Capsule Wall in the Dogfish Scyliorhinus Canicula. *Philos. Trans. R. Soc., B* **1993**, *341*, 419−436.

(6) Feng, D.; Knight, D. P. The Effect of PH on Fibrillogenesis of Collagen in the Egg Capsule of the Dogfish, Scyliorhinus Canicula. *Tissue Cell* **1994**, *26*, 649−659.

(7) Knight, D. P.; Feng, D.; Stewart, M. Structure and Function of the Salachian Egg Case. *Biol. Rev. Cambridge Philos. Soc.* **1996**, *71*, 81−111.

(8) Knight, D. P.; Vollrath, F. Biological Liquid Crystal Elastomers. *Philos. Trans. R. Soc., B* **2002**, *357*, 155−163.

(9) Tan, Y.; Hoon, S.; Guerette, P. A.; Wei, W.; Ghadban, A.; Hao, C.; Miserez, A.; Waite, J. H. Infiltration of Chitin by Protein Coacervates Defines the Squid Beak Mechanical Gradient. *Nat. Chem. Biol.* **2015**, *11*, 488−495.

(10) Guerette, P. A.; Hoon, S.; Seow, Y.; Raida, M.; Masic, A.; Wong, F. T.; Ho, V. H. B.; Kong, K. W.; Demirel, M. C.; Pena-Francesch, A.; Amini, S.; Tay, G. Z.; Ding, D.; Miserez, A. Accelerating the Design of Biomimetic Materials by Integrating RNA-Seq with Proteomics and Materials Science. *Nat. Biotechnol.* **2013**, *31*, 908−915.

(11) Kishore, U.; Ghai, R.; Greenhough, T. J.; Shrive, A. K.; Bonifati, D. M.; Gadjeva, M. G.; Waters, P.; Kojouharova, M. S.; Chakraborty, T.; Agrawal, A. Structural and Functional Anatomy of the Globular Domain of Complement Protein C1q. *Immunol. Lett.* **2004**, *95*, 113−128.

(12) Kirkitadze, M. D.; Barlow, P. N. Structure and Flexibility of the Multiple Domain Proteins That Regulate Complement Activation. *Immunol. Rev.* **2001**, *180*, 146−161.
(13) Shoulders, M. D.; Raines, R. T. Collagen Structure and Stability. *Annu. Rev. Biochem.* **2009**, *78*, 929−958.
(14) Bauer, J.; Meza, L. R.; Schaedler, T. A.; Schwaiger, R.; Zheng, X.; Valdevit, L. Nanolattices: An Emerging Class of Mechanical Metamaterials. *Adv. Mater.* **2017**, *29*, 1701850.
(15) Barclay, G. G.; Ober, C. K. Liquid Crystalline and Rigid-Rod Networks. *Prog. Polym. Sci.* **1993**, *18*, 899−945.
(16) Landsman, M. R.; Sujanani, R.; Brodfuehrer, S. H.; Cooper, C. M.; Darr, A. G.; Davis, R. J.; Kim, K.; Kum, S.; Nalley, L. K.; Nomaan, S. M.; Oden, C. P.; Paspureddi, A.; Reimund, K. K.; Rowles, L. S.; Yeo, S.; Lawler, D. F.; Freeman, B. D.; Katz, L. E. Water Treatment: Are Membranes the Panacea? *Annu. Rev. Chem. Biomol. Eng.* **2020**, *11*, 559−585.
(17) Adiga, S. P.; Jin, C.; Curtiss, L. A.; Monteiro-Riviere, N. A.; Narayan, R. J. Nanoporous Membranes for Medical and Biological Applications: Nanoporous Membranes for Medical and Biological Applications. *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* **2009**, *1*, 568−581.
(18) Grabherr, M. G.; Haas, B. J.; Yassour, M.; Levin, J. Z.; Thompson, D. A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; Chen, Z.; Mauceli, E.; Hacohen, N.; Gnirke, A.; Rhind, N.; di Palma, F.; Birren, B. W.; Nusbaum, C.; Lindblad-Toh, K.; Friedman, N.; Regev, A. Full-Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome. *Nat. Biotechnol.* **2011**, *29*, 644−652.
(19) Ye, J.; Coulouris, G.; Zaretskaya, I.; Cutcutache, I.; Rozen, S.; Madden, T. L. Primer-BLAST: A Tool to Design Target-Specific Primers for Polymerase Chain Reaction. *BMC Bioinf.* **2012**, *13*, 134.
(20) Almagro Armenteros, J. J.; Tsirigos, K. D.; Sønderby, C. K.; Petersen, T. N.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 5.0 Improves Signal Peptide Predictions Using Deep Neural Networks. *Nat. Biotechnol.* **2019**, *37*, 420−423.
(21) Madeira, F.; mi Park, Y.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A. R. N.; Potter, S. C.; Finn, R. D.; Lopez, R. The EMBL-EBI Search and Sequence Analysis Tools APIs in 2019. *Nucleic Acids Res.* **2019**, *47*, W636−W641.
(22) de Castro, E.; Sigrist, C. J. A.; Gattiker, A.; Bulliard, V.; Langendijk-Genevaux, P. S.; Gasteiger, E.; Bairoch, A.; Hulo, N. ScanProsite: Detection of PROSITE Signature Matches and ProRule-Associated Functional and Structural Residues in Proteins. *Nucleic Acids Res.* **2006**, *34*, W362−W365.
(23) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* **2018**, *46*, W296−W303.
(24) Bogin, O.; Kvansakul, M.; Rom, E.; Singer, J.; Yayon, A.; Hohenester, E. Insight into Schmid Metaphyseal Chondrodysplasia from the Crystal Structure of the Collagen X NC1 Domain Trimer. *Structure* **2002**, *10*, 165−173.
(25) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E. UCSF ChimeraX: Structure Visualization for Researchers, Educators, and Developers. *Protein Sci.* **2021**, *30*, 70−82.
(26) Miyazawa, S.; Jernigan, R. L. Estimation of Effective Interresidue Contact Energies from Protein Crystal Structures: Quasi-Chemical Approximation. *Macromolecules* **1985**, *18*, 534−552.
(27) Montanaro, J.; Gruber, D.; Leisch, N. Improved Ultrastructure of Marine Invertebrates Using Non-Toxic Buffers. *PeerJ* **2016**, *4*, No. e1860.
(28) Ilavsky, J. Nika: Software for Two-Dimensional Data Reduction. *J. Appl. Crystallogr.* **2012**, *45*, 324−328.
(29) Westbrook, J. H.; Conrad, H. *The Science of Hardness Testing and its Research Applications*; American Society for Metals: USA, 1973.
(30) Luong, T.-T.; Boutillon, M.-M.; Garrone, R.; Knight, D. P. Characterization of Selachian Egg Case Collagen. *Biochem. Biophys. Res. Commun.* **1998**, *250*, 657−663.
(31) Qin, X.; Waite, J. H. Exotic Collagen Gradients in the Byssus of the Mussel Mytilus Edulis. *J. Exp. Biol.* **1995**, *198*, 633−644.
(32) Waite, J. H.; Qin, X.-X.; Coyne, K. J. The Peculiar Collagens of Mussel Byssus. *Matrix Biol.* **1998**, *17*, 93−106.
(33) Schaefer, C.; Laity, P. R.; Holland, C.; McLeish, T. C. B. Silk Protein Solution: A Natural Example of Sticky Reptation. *Macromolecules* **2020**, *53*, 2669−2676.
(34) Sagert, J.; Waite, J. H. Hyperunstable Matrix Proteins in the Byssus of Mytilus Galloprovincialis. *J. Exp. Biol.* **2009**, *212*, 2224−2236.
(35) Robinson, A. B.; Rudd, C. J. Deamidation of Glutaminyl and Asparaginyl Residues in Peptides and Proteins. *Curr. Top. Cell. Regul.* **1974**, *8*, 247−295.
(36) Persikov, A. V.; Ramshaw, J. A. M.; Kirkpatrick, A.; Brodsky, B. Amino Acid Propensities for the Collagen Triple-Helix. *Biochemistry* **2000**, *39*, 14960−14967.
(37) Kishore, U.; Gaboriaud, C.; Waters, P.; Shrive, A. K.; Greenhough, T. J.; Reid, K. B. M.; Sim, R. B.; Arlaud, G. J. C1q and Tumor Necrosis Factor Superfamily: Modularity and Versatility. *Trends Immunol.* **2004**, *25*, 551−561.
(38) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215*, 403−410.
(39) Joung, S.-J.; Chen, C.-T.; Clark, E.; Uchida, S.; Huang, W. Y. P. The Whale Shark, Rhincodon Typus, Is a Livebearer: 300 Embryos Found in One "Megamamma" Supreme. *Environ. Biol. Fish.* **1996**, *46*, 219−223.
(40) Koch, E. A.; Spitzer, R. H.; Pithawalla, R. B.; Parry, D. A. An Unusual Intermediate Filament Subunit from the Cytoskeletal Biopolymer Released Extracellularly into Seawater by the Primitive Hagfish (Eptatretus Stouti). *J. Cell Sci.* **1994**, *107*, 3133−3144.
(41) Miserez, A.; Schneberk, T.; Sun, C.; Zok, F. W.; Waite, J. H. The Transition from Stiff to Compliant Materials in Squid Beaks. *Science* **2008**, *319*, 1816−1819.
(42) Miserez, A.; Rubin, D.; Waite, J. H. Cross-Linking Chemistry of Squid Beak. *J. Biol. Chem.* **2010**, *285*, 38115−38124.
(43) Zhao, H.; Waite, J. H. Coating Proteins: Structure and Cross-Linking in Fp-1 from the Green Shell Mussel Perna Canaliculus. *Biochemistry* **2005**, *44*, 15915−15923.
(44) Zhao, H.; Sun, C.; Stewart, R. J.; Waite, J. H. Cement Proteins of the Tube-Building Polychaete Phragmatopoma Californica. *J. Biol. Chem.* **2005**, *280*, 42938−42944.
(45) Hexemer, A.; Bras, W.; Glossinger, J.; Schaible, E.; Gann, E.; Kirian, R.; MacDowell, A.; Church, M.; Rude, B.; Padmore, H. A SAXS/WAXS/GISAXS Beamline with Multilayer Monochromator. *J. Phys.: Conf. Ser.* **2010**, *247*, 012007.
(46) Fratzl, P.; Misof, K.; Zizak, I.; Rapp, G.; Amenitsch, H.; Bernstorff, S. Fibrillar Structure and Mechanical Properties of Collagen. *J. Struct. Biol.* **1998**, *122*, 119−122.
(47) Zimmermann, E. A.; Gludovatz, B.; Schaible, E.; Dave, N. K. N.; Yang, W.; Meyers, M. A.; Ritchie, R. O. Mechanical Adaptability of the Bouligand-Type Structure in Natural Dermal Armour. *Nat. Commun.* **2013**, *4*, 2634.
(48) Jacquet, M.; Cioci, G.; Fouet, G.; Bally, I.; Thielens, N. M.; Gaboriaud, C.; Rossi, V. C1q and Mannose-Binding Lectin Interact with CR1 in the Same Region on CCP24-25 Modules. *Front. Immunol.* **2018**, *9*, 453.
(49) Klickstein, L. B.; Barbashov, S. F.; Liu, T.; Jack, R. M.; Nicholson-Weller, A. Complement Receptor Type 1 (CR1, CD35) Is a Receptor for C1q. *Immunity* **1997**, *7*, 345−355.
(50) Zhang, Y.; Chen, Q. The Noncollagenous Domain 1 of Type X Collagen: A Novel Motif for Trimer and Higher Order Multimer Formation without a Triple Helix. *J. Biol. Chem.* **1999**, *274*, 22409−22413.

https://doi.org/10.1021/acs.biomac.2c00341
Biomacromolecules 2022, 23, 2878−2890

(51) Boudko, S. P.; Engel, J.; Bächinger, H. P. The Crucial Role of Trimerization Domains in Collagen Folding. *Int. J. Biochem. Cell Biol.* **2012**, *44*, 21−32.
(52) Kvansakul, M.; Bogin, O.; Hohenester, E.; Yayon, A. Crystal Structure of the Collagen A1(VIII) NC1 Trimer. *Matrix Biol.* **2003**, *22*, 145−152.
(53) Gallivan, J. P.; Dougherty, D. A. Cation-π Interactions in Structural Biology. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9459−9464.
(54) Fischer, P.; Finkelmann, H. Lyotropic Liquid-Crystalline Elastomers. In *Structure, Dynamics and Properties of Disperse Colloidal Systems*; Rehage, H., Peschel, G., Eds.; Steinkopff: Darmstadt, 1998; pp 127−134.
(55) Bhattacharjee, A.; Bansal, M. Collagen Structure: The Madras Triple Helix and the Current Scenario. *IUBMB Life* **2005**, *57*, 161−172.
(56) Harrington, M. J.; Gupta, H. S.; Fratzl, P.; Waite, J. H. Collagen Insulated from Tensile Damage by Domains That Unfold Reversibly: In Situ X-Ray Investigation of Mechanical Yield and Damage Repair in the Mussel Byssus. *J. Struct. Biol.* **2009**, *167*, 47−54.
(57) Knupp, C.; Chew, M.; Squire, J. Collagen Packing in the Dogfish Egg Case Wall. *J. Struct. Biol.* **1998**, *122*, 101−110.
(58) Aßfalg, N.; Finkelmann, H. A Smectic A Liquid Single Crystal Elastomer (LSCE): Phase Behavior and Mechanical Anisotropy. *Macromol. Chem. Phys.* **2001**, *202*, 794−800.
(59) Kirkness, M. W.; Lehmann, K.; Forde, N. R. Mechanics and Structural Stability of the Collagen Triple Helix. *Curr. Opin. Chem. Biol.* **2019**, *53*, 98−105.
(60) Rey, A. D. Liquid crystal models of biological materials and processes. *Soft Matter* **2010**, *6*, 3402−3429.
(61) Bonn, D.; Meunier, J.; Greffier, O.; Al-Kahwaji, A.; Kellay, H. Bistability in Non-Newtonian Flow: Rheology of Lyotropic Liquid Crystals. *Phys. Rev. E* **1998**, *58*, 2115−2118.
(62) Priemel, T.; Palia, G.; Förste, F.; Jehle, F.; Sviben, S.; Mantouvalou, I.; Zaslansky, P.; Bertinetti, L.; Harrington, M. J. Microfluidic-like Fabrication of Metal Ion−Cured Bioadhesives by Mussels. *Science* **2021**, *374*, 206−211.
(63) van Kuringen, H. P. C.; Schenning, A. P. H. J.; Broer, D. J. Liquid Crystal Polymer Membranes. In *Encyclopedia of Membranes*; Drioli, E., Giorno, L., Eds.; Springer: Berlin, Heidelberg, 2015; pp 1−3.
(64) Werber, J. R.; Osuji, C. O.; Elimelech, M. Materials for Next-Generation Desalination and Water Purification Membranes. *Nat. Rev. Mater.* **2016**, *1*, 16018.
(65) Desai, T. A.; West, T.; Cohen, M.; Boiarski, T.; Rampersaud, A. Nanoporous Microsystems for Islet Cell Replacement. *Adv. Drug Delivery Rev.* **2004**, *56*, 1661−1673.
(66) Niu, J.; Shao, R.; Liu, M.; Zan, Y.; Dou, M.; Liu, J.; Zhang, Z.; Huang, Y.; Wang, F. Porous Carbons Derived from Collagen-Enriched Biomass: Tailored Design, Synthesis, and Application in Electrochemical Energy Storage and Conversion. *Adv. Funct. Mater.* **2019**, *29*, 1905095.
(67) Portela, C. M.; Edwards, B. W.; Veysset, D.; Sun, Y.; Nelson, K. A.; Kochmann, D. M.; Greer, J. R. Supersonic Impact Resilience of Nanoarchitected Carbon. *Nat. Mater.* **2021**, *20*, 1491−1497.
(68) Zhang, X.; Vyatskikh, A.; Gao, H.; Greer, J. R.; Li, X. Lightweight, Flaw-Tolerant, and Ultrastrong Nanoarchitected Carbon. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 6665−6672.

https://doi.org/10.1021/acs.biomac.2c00341
Biomacromolecules 2022, 23, 2878−2890