Published in final edited form as:
Biomacromolecules. 2014 January 13; 15(1): 30–42. doi:10.1021/bm401598z.

Structural Proteins from Whelk Egg Capsule with Long Range Elasticity Associated with a Solid-state Phase Transition

S. Scott Wasko[1,#], **Gavin Tay**[2,3,#], **Andreas Schwaighofer**[4], **Christoph Nowak**[3,4], **J. Herbert Waite**[1,*], and **Ali Miserez**[2,3,5,*]

[1]Biomolecular Sciences and Engineering Program, University of California, Santa Barbara, California 93106, USA

[2]School of Materials Science and Engineering, Nanyang Technological University, Singapore 639798

[3]Center for Biomimetic Sensor Science, 50 Nanyang Drive, Singapore 637553

[4]Austrian Institute of Technology GmbH, AIT, Donau-City Str. 1, 1220 Vienna, Austria

[5]School of Biological Sciences, Nanyang Technological University, Singapore 637551

Abstract

The robust, proteinaceous egg capsules of marine prosobranch gastropods (genus *Busycotypus*) exhibit unique biomechanical properties such as high elastic strain recovery and elastic energy dissipation capability. Capsule material possesses long-range extensibility that is fully recoverable and is the result of a secondary structure phase transition from α-helix to extended β-sheet rather than of entropic (rubber) elasticity. We report here the characterization of the precursor proteins that make up this material. Three different proteins have been purified and analyzed, and complete protein sequences deduced from messenger ribonucleic acid (mRNA) transcripts. Circular dichroism (CD) and Fourier transform infrared (FTIR) spectra indicate that the proteins are strongly α-helical in solution and primary sequence analysis suggests that these proteins have a propensity to form coiled-coils. This is in agreement with previous wide-angle *x*-ray scattering (WAXS) and solid-state Raman spectroscopic analysis of mature egg capsules.

[*]Authors for correspondence: waite@lifesci.ucsb.edu, ali.miserez@ntu.edu.sg.
[#]These authors contributed equally to this work

Accession Numbers

The sequence data of Bc-CP 1a, Bc-CP 1b, Bc-CP 2, and Bc-CP 3 have been deposited in the NCBI database with the accession numbers GU205809, GU205810, GU205811, and GU205812 respectively.

SUPPORTING INFORMATION AVAILABLE:

Table S1. N-termini Edman sequencing results of Bc-CPs from individual SDS-PAGE bands of crude extract.
Figure S1. 1% agarose gel of RT-PCR with 3' RACE products.
Figure S2. Complete amino acid sequences of (a) CP-1b and (b) CP-2 as deduced from cDNA.
Figure S3. CD spectra of mixtures of CP-1 and CP-2 in various ratios.
Figure S4. ATR-FTIR spectra of CP-1 and CP-3, showing the preferential formation of β-sheets over α-helices when the proteins are freeze-dried from concentrated, compared to diluted, solutions.
Figure S5. (a) ATR-FTIR spectra of egg case, showing a loss of Amide II at ~1525 cm^{-1} after hydrogen is substituted by deuterium. (b) Curve-fitted Amide I of dehydrated egg case and (c) egg case in D_2O, confirming the presence of β-sheets along with high α-helical content (>50%).
Figure S6. Relative amounts of (a) hydrophobic and (b) charged amino acid residues at various heptad positions in the predicted coiled-coil domains of Bc-CPs, and other known coil-coiled proteins in the CC+ database.
Figure S7. Transmission electron micrographs of purified Bc-CPs.
This material is available free of charge via the Internet at http://pubs.acs.org/.

Keywords

Egg capsule; elasticity; coiled-coil; α-β transition; shape memory; self-healing

1. Introduction

Rubbery proteins such as elastin, resilin, and abductin are well studied from a variety of organisms in which they endow specialized tissues such as aorta, insect wings, and scallop hinge ligaments with low modulus, high extensibility (>100%), high resilience (>90%), and entropy-driven elastic recovery[1–4] Accordingly, the entropy of a rubbery protein chain at rest is much greater than of a chain pulled taut, hence its spontaneous recovery upon unloading. Whereas these structures are mostly disordered, α-helical proteins are also found as building blocks of a diverse variety of extra-organismal load-bearing scaffolds. The keratins found in mammalian hair, wool, nails, horns, claws and hooves[5] are familiar examples, whereas the hagfish slime[6, 7] is a more esoteric one. Unlike the proteinaceous egg capsules of other oviparous species, such as dogfish and spiders which produce collagen and silk-based capsule proteins respectively[8, 9], the whelk egg capsule is predominantly an α-helical biomaterial. This molecular design was also recently reported in the egg case of the praying mantis[10].

Recently, there has been growing evidence that certain proteins or protein assemblies that feature significant long-range order as detected by *x*-ray scattering can undergo a phase transition during uniaxial stretching, from one crystalline structure such as α-helix to another such as β-sheet[11], according to a mechanism initially suggested by Flory[12]. This paradigm differs significantly in that recovery from large extensions is driven mostly by internal energy rather than entropy variations as a function of strain. Aculeate silk was the first fiber shown by wide-angle *x*-ray scattering (WAXS) to undergo an α-to-β transition, but the transformation was irreversible[13]. Wool keratin fiber was then later shown to also undergo an α-to-β transition during extensions of up to 40% but its recovery was slow and incomplete[5, 14–16] In contrast, the egg capsule of the prosobranch gastropod mollusks or whelks of the genus *Busycotypus* (formerly known as *Busycon*) exhibits a completely and spontaneously reversible structural transition during large extensions from α-helical coiled-coils to β-sheets[17] (or extended β-strands according to recent investigations[18]). Other coiled-coil fibrous proteins that are susceptible to an α-to-β molecular transition include myosin, fibrin and intermediate filaments (IFs) from the cell cytoskeleton, and a comparison of their stress-strain behaviors at various length scales has been recently reviewed[11]. The ability of coiled-coil fibers to undergo an α-to-β transition under mechanical stress is also well supported by molecular dynamic simulations carried out on various IFs[19] and fibrinogen[20]

The *Busycotypus* egg case material consists of proteins that encapsulate embryos during their extended development period[17]. Meter-long strings of hundreds of interconnected egg capsules (colloquially known as mermaid's necklaces, Fig. 1) are routinely deposited on the seabed, and rely chiefly on the protein casing to protect the developing embryos from strong hydrodynamic forces associated with wave velocities in excess of 25 m sec^{-1}, abrasion from sand and other particulate debris, predation, and harmful ultraviolet (UV) light[21–23]. Fully processed, or mature, whelk egg capsule is a highly extensible, self-assembled, proteinaceous material which possesses remarkable self-healing or shape-memory properties. The egg capsules approach the extensibility of the biological rubbers such as elastin and resilin, but have an initial stiffness two orders of magnitude higher. Strain energy absorption capacity of the capsule material is more than five times greater while its resilience is two thirds that of elastin[24] (Table 1). Furthermore, even after the material is extended beyond its yield point, upon relaxation it returns to its original size and shape, and

within seconds recovers its high initial stiffness[25] (Fig. 2). This is a rare and highly desirable combination of properties of polymer fibers[26].

Previous studies have detailed the tensile properties[25, 32] and structural conformation[33, 34] of the capsule material, as well as modeled the thermodynamics of its self-healing/shape-memory behavior[10]. However, little is known about the protein chemistry of the egg capsule material. It is therefore of critical importance to characterize and wherever possible, reconcile the biochemistry with the observed tensile properties of the constituent proteins of this material, hereafter referred to as the *Busycotypus canaliculatus* capsule proteins (Bc-CPs).

Capsule proteins undergo at least two separate post-translational processing steps in the female whelk before the mature end product is formed. Firstly, soluble precursor capsule proteins self-assemble into a jelly-like structure which possesses no macroscopic load-bearing capability. Then, this jelly-like structure is covalently cross-linked before the egg capsule is released and abandoned on the ocean floor[32]. As a consequence of the covalent cross-linking or sclerotization, it is extremely difficult to extract full-length, intact protein from the mature capsule. However, the precursor proteins are readily extractable from gravid female nidamental glands to be purified. Previous work has determined that the precursors extracted from the gland tissue are the same as the proteins present in the immature capsules[32]. Using nidamental gland-derived precursors, we determined the protein molecular masses, amino acid compositions, primary amino acid sequences, secondary structures in solution and self-assembly behavior. From these data, as well as data from previous work, plausible structure-function models for these proteins are proposed.

2. Materials and Methods

2.1 Materials

All materials were purchased from Sigma-Aldrich or Fisher Scientific unless otherwise stated.

2.2 Capsule Proteins Purification

Live whelks (*Busycotypus canaliculatus*) were purchased from the Marine Biological Laboratories (Woods Hole, Massachusetts, USA). Females were euthanized by immersion in a cold ethanol bath and were immediately dissected to remove the nidamental gland which was then placed at −80 °C. Crude extract was prepared by treating the gland material with 6 M urea in 5% acetic acid and 50 mM tris-(2-carboxyethyl)phosphine (TCEP) to reduce disulfide bonds. Reverse phase high pressure liquid chromatography (RP-HPLC) was performed using a C8 column (Perkin Elmer Brownlee Aquapore Octyl C8 or Agilent Technologies Zorbax 300SB-C8 with a Varian ProStar or Agilent Technologies 1260 Infinity HPLC system, respectively). Crude extract and purified protein fractions were assayed via sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) using 15% polyacrylamide gels.

2.3 Mass Spectrometry

The mass-to-charge ratio (*m/z*) was determined by matrix-assisted laser desorption/ionization with time of flight (MALDI-ToF) mass spectrometry using a PerSeptive Biosystems Voyager DE mass spectrometer (AB Biosystems, Foster City, California, USA). Measurements were also carried out on an AXIMA Performance mass spectrometer (Shimadzu, Kyoto, Japan). In both cases, protein solutions were dried on a sample plate, over which a saturated solution of sinapic acid in 50% acetonitrile and 0.1% trifluoroacetic acid was subsequently added and dried.

2.4 Amino Acid Composition Analysis

Lyophilized RP-HPLC fractions were hydrolyzed *in vacuo* using 6 N HCl at 110 °C for 24 h. Following hydrolysis, the samples were washed with high purity water and methanol, twice each, and dried. Amino acid compositions were determined using a Beckman Coulter 6300 amino acid analyzer (Brea, California, USA) with ninhydrin-based post column derivatization as described elsewhere[24]. Additional measurements were also acquired using a Sykam S433 amino acid analyzer (Eresing, Germany) with identical sample preparation procedures.

2.5 Edman Sequencing

Lyophilized RP-HPLC fractions were subjected to Edman sequencing reactions using a Porton Instruments PI 2020 as described elsewhere[35]. For CP-2 and CP-3, N-terminal Edman sequencing was also performed by transferring protein from SDS-PAGE gel to Immobilon-P polyvinylidene difluoride (PVDF) membrane (Qiagen, Valencia, California, USA), excising bands and submitting them to the Protein Facility of Iowa State University (Ames, Iowa, USA) for analysis.

2.6 Reverse Transcription Polymerase Chain Reaction (RT-PCR) with Rapid Amplification of cDNA Ends (RACE) and Sequencing

RNA was purified from fresh (non-frozen) nidamental gland tissue immediately following dissection with TRIzol Reagent following the manufacturer's instructions. Total RNA was reverse transcribed to RACE-ready complementary deoxyribonucleic acid (cDNA) using a GeneRacer Kit (Invitrogen, Carlsbad, California, USA), primers were purchased from Integrated DNA Technologies (Coralville, Iowa, USA) and PCR reagents were purchased from Novagen (Darmstadt, Germany). PCR was carried out using a Mastercycler Gradient thermocycler (Eppendorf, Hauppauge, New York, USA). PCR products were purified using 1.1% agarose gels, visualized by ethidium bromide, and bands were excised and purified with a Qiagen Gel Extraction Kit. PCR products were directly sequenced by GENEWIZ (La Jolla, California, USA).

From the partial N-terminal sequence, degenerate primers were designed for RT-PCR. Firstly, 5' RACE-PCR was performed to give the sequence of the 5' untranslated region. At the same time, 3' RACE-PCR was performed to give cDNA sequences of the signal peptide to the start methionine and beyond for each transcript.

Degenerate primers used for the Bc-CPs:

5' RACE

CP1-a: 5'- GAYYTNGAYTTYGAYGTNCCNAARGARGCN −3'

CP1-b: 5'- GCNYTNGAYTTYGAYGTNACNAARGARGCN −3'

CP2/3: 5'- TTYCARGGNTTYAAYTTYGGNAARAARGAY −3'

3' RACE

CP1-a: 3'- CTRRANCTRAARCTRCANGGNTTYCTYCGH −5'

CP1-b: 3'- CGNRANCTRAARCTRCANTGNTTYCTYCGN −5'

CP2/3: 3'- AARGTYCCNAARTTRAARCCNTTYTTYCTR −5'

where Y = C, T; R = A, G; N = A, C, G, T

From the results of the 5' and 3' RACE-PCR, new sets of non-degenerate primers were designed for 3'-RACE to amplify full-length transcripts.

2.7 Circular Dichroism (CD) Spectroscopy

Far UV (190 – 260 nm) CD measurements were carried out using an Olis RSM 1000 spectrophotometer (Olis, Bogart, Georgia, USA) with a 0.5 mm quartz cell at 1 nm resolution. Purified protein solutions were prepared in 50 mM acetate buffer, pH 4.0. Ten spectra were taken at room temperature and the results were averaged. Spectra are presented in mean residue ellipticity with protein molecular weight determined by MALDI-ToF, and protein concentration estimated by quantitative ninhydrin-based amino acid analysis (see above) as described elsewhere[36]. Secondary structure analysis was performed using K2D3 software[37] - each spectrum was scaled linearly until a best fitting curve was found.

2.8 Fourier Transform Infrared (FTIR) Spectroscopy

Infrared absorption measurements were performed using a VERTEX 70 or 70v FTIR spectrometer (Bruker, Karlsruhe, Baden-Württemberg, Germany), respectively equipped with a PIKE MIRacle or Harrick Horizon attenuated total reflection (ATR) measuring unit, each containing a ZnSe ATR crystal (angle of incidence $\theta = 45°$). Purified protein solutions were prepared in 50 mM acetate buffer, pH 4.0. The sample chamber was continuously purged with dry carbon dioxide-free air, and the total reflected IR beam intensity was measured using a liquid nitrogen-cooled photovoltaic mercury cadmium telluride (MCT) detector. Spectra were recorded with a spectral resolution of 4 cm^{-1} in double-sided acquisition mode, and the mirror velocity was set to 80 kHz. At least 1000 scans were taken for each spectrum, which was calculated using a 3-term Blackman-Harris apodization function. Spectra of proteins were obtained by subtracting IR absorption due to acetate buffer. Secondary structure analysis was performed using the OPUS 6.5 software (Bruker, Karlsruhe, Baden-Württemberg, Germany).

IR absorption spectra of solid-state egg capsules were also acquired in the ATR mode under the same conditions.

2.9 Calculation of Protein Dimensions

Based on the results of sequencing and the secondary structure analyses of the FTIR spectra, the average number of amino acid residues per protein chain present in the capsule material was calculated and the relative amounts of α-helices, β-sheets and random coils were determined. Those numbers were used to compute the maximum end-to-end protein length by summing the length of each type of secondary structure, assuming that they are linearly arranged. The translational rise is 1.5 Å per residue for α-helix, 3.5 Å per residue for parallel β-sheet, and 3.3 Å per residue for antiparallel β-sheet. For the length of random coil, statistical analysis of steric restrictions was taken into account and a characteristic ratio $C_\infty$of 9.0 was used as described by Creighton[38]. That resulted in a calculation that was dictated by the equation:

$$\langle r^2 \rangle_\circ^{1/2} = (130n)^{1/2}$$

where the root mean square $\langle r^2 \rangle_\bigcirc^{1/2}$ was taken as the end-to-end length (Å).

3. Results

3.1 Capsule Proteins Purification, Mass Spectrometry and Amino Acid Composition Analysis

Consistent with previous findings, SDS-PAGE analysis of nidamental gland crude extract indicates a family of proteins around 50 kDa with three distinct bands. In order of increasing molecular mass, the bands are named CP-1, CP-2 and CP-3. Using RP-HPLC, each individual protein in this family can be purified from the crude extract (Fig. 3). In order of increasing retention time, the proteins are CP-1, CP-3 and CP-2. MALDI-ToF mass spectrometry indicates the molecular masses of CP-1, CP-2 and CP-3 to be 47.5, 51.6, and 52.1 kDa respectively (Fig. 4). The three proteins have amino acid compositions that are similar to one another as determined by ninhydrin-based analysis. They are also similar to hard α-keratins in terms of richness in Asx, Glx and Leu, but different in two key aspects. Firstly, α-keratins have high Cys, or disulfide, content whereas Bc-CPs have almost none. Secondly, Bc-CPs have a much higher Lys content than α-keratins[39] (Table 2).

3.2 Protein and cDNA Sequencing

N-terminal amino acid sequence of the first ten residues was obtained by automated Edman reaction involving the RP-HPLC purified fractions. CP-1 has a unique N-terminal sequence whereas CP-2 and CP-3 share the first ten amino acids. This sequence identity for CP-2 and CP-3 was verified by repeating the Edman reaction with SDS-PAGE membrane transfers of crude extract (supplementary Table S1).

From the partial N-terminal sequence, degenerate primers were designed for RT-PCR. First 5’ RACE was performed to give the sequence of the 5’ untranslated region. Then, non-degenerate primers were designed for 3’ RACE to amplify full-length transcripts. 3’ RACE of the CP-1 cDNA transcript produced two PCR products, one at ∼1500 bp and another at ∼2400 bp, whereas that of the CP-2/CP-3 cDNA transcript produced only one PCR product at ∼2400 bp (supplementary Fig. S1). These three PCR products were gelexcised and sequenced directly without cloning.

All three transcripts contain an open reading frame, which translates to a full length gene with start (methionine) and termination codon. Despite the difference in the sizes of the PCR products, both CP-1 transcripts translate to proteins of identical length comprising 446 amino acid residues and have 95% sequence identity. Hereafter these variants are referred to as CP-1a and CP-1b. However, unless otherwise stated, when a protein is described as “CP-1”, it is a mixture of CP-1a and CP-1b.

The transcript for CP-2 and CP-3 translates to a single protein comprising 479 amino acid residues (Fig. 5). The fact that this transcript is responsible for the production of two different proteins can be explained by potential early termination in the translation process. Since sequencing was performed directly on gel-purified PCR products without cloning, multiple variants may be present, resulting in different nucleotide sequences at the same position in the transcript. The primary codon at position 476 is AGA (arginine), however, in the raw sequencing chromatogram there is a smaller, but still significant, thymine peak in the first position. This would result in TGA, a termination codon, which reduces the length of the protein by four amino acid residues. These 4 residues have a combined weight of 578.6 Da. This accounts for the observed molecular mass difference between CP-2 and CP-3. The computed molecular masses (excluding signal peptides) of the cDNA–deduced protein sequences are consistent with the actual protein molecular masses observed by MALDI-ToF mass spectrometry (Fig. 4). Amino acid compositions of the purified proteins also agree with the compositions predicted from cDNA–deduced protein sequences (Table 3).

In addition, all three transcripts code for signal peptides, indicating that these proteins are secreted via secretory vesicle pathways. Cleavages between G-21 and D-22 in CP-1a, G-21 and A-22 in CP-1b, as well as G-18 and L-19 in both CP-2 and CP-3 are predicted using SignalP (ExPASy), and match the N-terminal sequences derived by Edman sequencing. Sequence analysis was performed using BLAST (NCBI) which did not produce a match with E-value $<10^{-9}$ for CP-1a and CP-1b, and E-value $<10^{-5}$ for CP-2 and CP-3 in the database of sequences from eukaryotic species, indicating that the Bc-CPs are not homologous to any known eukaryotic protein.

3.3 CD and FTIR Spectroscopy

Both far-UV CD spectroscopy and mid-IR FTIR spectroscopy were used to examine the secondary structures of Bc-CPs in solution. Lyophilized RP-HPLC purified Bc-CPs were re-suspended in 50 mM acetate buffer, pH 4.0 (lyophilized proteins have limited solubility at more neutral pH), and spectra were collected at room temperature.

The CD spectra of both CP-1 and CP-2 (Fig. 6) were characteristic of single coil α-helical structure[41], exhibiting two minima at 208 and 222 nm with a $[\theta]_{222}$: $[\theta]_{208}$ ratio of <1. Secondary structure analysis by K2D3 showed that while both CP-1 and CP-2 are dominated by α-helices (>50%), there are small amounts of β-sheets (∼6%). The existence of β-sheets is confirmed by FTIR spectroscopy, as described below.

The overall FTIR absorption spectra of CP-1 and CP-3 in the ATR mode (Fig. 7c and d) show sharp intense bands centered at ∼1650 and ∼1540 cm^{-1}, corresponding to Amide I and Amide II bands respectively, with a clear presence of methylene peaks between 2800–3000 cm^{-1}. In order to determine the various vibration modes of the peptide backbone and obtain the relative amounts of the different types of secondary structures, the Amide I peak was curve-fitted using the Bruker OPUS 6.5 software. A 80% Gaussian : 20% Lorentzian function was employed to deconvolute the Amide I band into its constituent vibration modes. The positions of the constituent bands are held fixed, while the intensities and bandwidths were allowed to vary. From the area under the fitted curves, the relative fractions of α-helices, β-sheets and random coils are determined. Assignment of Amide I components to the various secondary structures is based on literature[42]. All samples are dominated by a high amount of α-helices, ranging from 70 to 75% for the pure Bc-CPs. The β-sheet content, although lower, remains significant at 15 to 20%, while unordered regions account for about 10% of the secondary structure. FTIR spectroscopy was also performed on an equimolar mixture of CP-1 and CP-3. The mixture exhibits a significant decrease in the relative fraction of α-helices (∼60%), together with an increase in both β-sheets (∼23%) and unordered regions (∼15%), thus resulting in an overall lower α-helix : β-sheet ratio compared to the individual Bc-CPs.

FTIR spectra of pure Bc-CPs that were subjected to freeze-drying from a highly concentrated state show significant differences in the shape of the Amide I band in comparison to Bc-CPs that were freeze-dried from diluted solutions (supplementary Fig. S4). For both proteins, a strong decrease in α-helical content is observed (40 and 20% for CP-1 and CP-3, respectively) and accompanied by an increase in β-sheet content (up to 50% for CP-1). In other words, in going from a state of diluted to concentrated solution, the secondary structure of the Bc-CPs changes from an α-helix to a β-sheet dominated one. As a result, the proteins also became less soluble.

In addition, mid-IR FTIR spectroscopy was used to examine the secondary structure of the *solid-state* native egg capsule in the ATR mode. The overall spectrum (Fig. 7a) similarly shows a clear presence of Amide I and II bands, as well as methylene peaks. Curve-fitting (Fig. 7b) revealed a high α-helical content, but it is significantly lower compared to the

individual CP-1, CP-3 or the equimolar mixture of them. The α-helix: β-sheet ratio is however similar to that of the equimolar mixture. Therefore, in going from the purified proteins to the egg capsule, the relative amount of α-helix decreases while both β-sheet and random coil increase. Additional measurements of the egg case were made in dehydrated state and in D_2O to confirm the presence of β-sheets (supplementary Fig. S5).

4. Discussion

4.1 Egg capsules as α-helical load-bearing materials

Although the whelk egg capsule shares numerous similarities with other α-helical materials, it exhibits unique mechanical and biochemical properties.

Under uniaxial tension, the egg capsule exhibits a relatively high initial Young's modulus (Table 1). The capsule material undergoes pseudo-yielding at 5% engineering strain, in the form of a force plateau up to ∼60% strain, beyond which it re-stiffens up to 170% strain[17, 43] Like intermediate filament (IF) based materials such as keratins and hagfish slime, coordinated rupturing of hydrogen bonds within the α-helices, as well as hydrophobic and ionic interactions between the helices that form the coiled-coil, allows for the elongation of the protein fibers without fracturing the egg capsule as a whole. The α-helical coiled-coils initially unravel into random coils as the capsule material stretches through the post-yield region of the stress-strain curve, and once all the α-helices are sufficiently extended, they form β-sheets or extended β-strands[18] with neighboring protein chains, causing re-stiffening. This phenomenon is responsible for its high strain energy absorption[5, 7, 16]. However, it is its quickly and completely recoverable long-range extension that sets the egg capsule apart from the other materials. Although not quite as extensible as the hagfish slime, the capsule material has the ability to recover its original conformation when the load is removed, whereas the hagfish slime is irreversibly deformed. Hydrated α-keratins recover slowly, dissipating energy, but are not as extensible as the egg capsule. What is most remarkable is the ability of the capsule material to rapidly and fully recover its initial stiffness despite having gone through multiple cycles of tensile strain[17]. Biochemical differences between keratins, hagfish, and *Busycotypus* egg cases are likely to explain the differences in their mechanical properties (Fig. 8).

Interestingly, this overall stress-strain behavior has been observed or predicted at the single-molecule level in other coiled-coil fibrous proteins. In myosin, AFM single-molecule force spectroscopy experiments yielded similar force-extension curves, consisting of an initial linear response, followed by a plateau region, and re-stiffening at large elongations, and the process was reversible[44] More recently, the same behavior was predicted by molecular dynamic simulations of force-induced unfolding of fibrinogen[20], where β-sheets were formed at high extensions, thus suggesting a somehow "universal" mechanism of α-β transition in coiled-coil load-bearing proteins.

4.2 *Busycotypus* egg case chemistry and coiled-coil predictions

Proteins comprise over 90% of the dry weight of the whelk egg capsule[17, 25] and the Bc-CPs characterized in this work represent the most abundant of these[32]. The amino acid composition of the Bc-CPs is highly similar to the capsule material and is predominantly made up of amino acids - such as Ala, Leu, Met, Lys and Glx (Gln or Glu) - that have a high propensity to form α-helix, with relatively few amino acids - such as Val, Ile, Phe, Tyr, Cys and Trp – that have the propensity to form β-sheet. This composition is also very similar to the α-keratin filaments, but with a striking difference in the Cys and Lys contents. Cys is the primary cross-linking residue in the keratins[5] and previous studies have suggested Lys as the potential cross-linker in the egg capsule[45, 46]. As seen from the results of sequencing (Fig. 5

and supplementary Fig. S2), CP-2 has the same sequence as CP-3 and is only four amino acid residues shorter at the C-terminal than CP-3. Since the sequences of CP-2 and CP-3 are more than 99% similar, they are likely to exhibit identical structural properties and are therefore treated as the same protein henceforth, unless otherwise stated.

Although there are striking similarities between the whelk egg capsule material and IF materials in terms of mechanics, amino acid compositions and the ability of the precursor proteins to self-assemble into filaments *in vitro* without accessory molecules, primary sequence analyses indicate that the Bc-CPs are very different from the IF proteins. BLAST (NCBI) analyses produced E-values of $<10^{-9}$ for CP-1a and CP-1b, and $<10^{-5}$ for CP-2 and CP-3, indicating that they are not homologous to every known eukaryotic IF protein in the database. Each of the Bc-CPs contains a signal peptide sequence at the N-terminus. These short sequences, 21 residues long in CP-1a and CP-1b, and 18 residues long in CP-2 and CP-3, indicate that the Bc-CPs are exported from the cell via a controlled vesicle based secretory pathway. This is in direct contrast to the hard α-keratins and the hagfish slime proteins, which remain in the intracellular space due to cornification or enter the extracellular region by apoptosis of the specialized cells in which they are accumulated[47, 48].

As was with IF materials, *x*-ray scattering studies have shown that the α-helical filaments of the whelk egg capsule are in the form of a coiled-coil structure[5] This well-known structure is the result of a heptad repeat of amino acid residues ***a*** through ***g***, in which apolar residues at positions ***a*** and ***d*** form hydrophobic interactions that bring and hold two or more α-helices together, while the flanking charged residues ***e*** and ***g*** provide electrostatic stabilization[49–51]. Thus, we strongly expected that sequence analysis of CP-1 and CP-3 would predict for coiled-coil structures. CP sequences were subjected to coiledcoil analysis using the SCORER 2.0 software[40] which predicts for coiled-coil propensity, while the canonical heptad repeat ***a****bc****d****efg* of coiled-coil domains was assigned based on the MARCOIL software[52] The results are shown in Fig. 5, where the heptad domains are highlighted in yellow and blue. Not surprisingly, large coiled-coil domains are unambiguously predicted for both CP-1 and CP-3. Using a 50% confidence level, both proteins are predicted to contain ∼50% coiled-coils, with domains as large as 25 heptads long. Additional corroboration of coiled-coil predictions was also obtained via the Lupas method[53] (data not shown). Furthermore, the SCORER algorithm predicts that coiled-coil domains are more likely to form trimers, in contrast to the dimers found in hagfish slime and α-keratin materials, and the same prediction was obtained using Berger's method[54]. Based on the strong suggestion of coiled-coil trimers, the capsule proteins may be more closely related to fibrinogen, which is a trimeric coiled-coil protein that also has load-bearing capability, than the α-keratins or hagfish slime filament proteins, but this awaits confirmation by high-resolution structural measurements[55]. A further inspection of heptad assignment revealed the presence of stutters and heptad discontinuities involving deletions of 2 residues. These coiled-coil irregularities may play a critical role in the α-β transition, as discussed later (Section 4.5).

Interestingly, while there is little difference in the hydrophobic and charged residue content (supplementary Fig. S6) in the coiled-coil domains of the Bc-CPs compared to other known coiled-coil proteins in the CC+ database[56] there is a relatively large amount of Gly and polar amino acid residues (Ser, Thr, Asn, Gln). On the other hand, the Bc-CPs contain relatively fewer aromatic residues such as Phe, Trp and Tyr (Fig. 9). These differences are notably evident within the hydrophobic core at heptad positions ***a*** and ***d***. Very likely, with significantly fewer bulky aromatic groups, the coiled-coil Bc-CPs are able to undergo easy and quick uncoiling and recoiling, during loading and unloading respectively, due to less steric hindrance. In addition, the larger presence of polar groups across all heptad positions in the Bc-CPs is likely to yield hydrophilic interactions with water (considering that the

native egg case serves to function in a marine environment), thereby assisting water in its role as a plasticizer that would enable smooth and fast structural transitions both upon loading and unloading by increasing chain flexibility.

4.3 Egg case protein secondary structure by CD and FTIR spectroscopy

In agreement with the helix-favoring amino acid composition and coiled-coil predictions, both CD and FTIR data indicate that the Bc-CPs are strongly α-helical in solution. In order to obtain some insights about post-translational self-assembly and processing, it is instructive to compare the relative secondary structure content obtained by FTIR in the precursor proteins in solution and in the fully processed, mature capsule in a hydrated state. In the solid-state capsule material, the α-helical content is less pronounced, and concomitantly there are relatively more β-sheets and random coils. *In situ* tensile experiments using WAXS[17] and Raman spectroscopy[18] have confirmed the dominance of α-helices in the unstrained capsule material, but β-sheets, on the other hand, were not detected. While Raman and FTIR spectroscopy are complementary techniques, FTIR has a higher sensitivity than Raman, and is more suitable for semi-quantitative analysis[57] In Raman spectroscopy, the signal intensity of a specific band does not vary linearly with the relative abundance of the secondary structure. It is thus very likely that a structure present in significant amount can be buried in the background, especially when its vibration mode has a weak polarizability. Such a signal would show up only when the structure becomes dominant in the sample. In the previous measurements, the extended β-structure (which is also clearly detected by X-ray) was detected at high strain but its intensity was low in comparison to other Raman bands. Hence, it is likely that in these previous investigations, the initial β-sheet content was too low to be detected. On the other hand, the current data obtained by ATR-FTIR spectroscopy indicates a significant presence of β-sheets in the native material.

In the capsule material, the α-helix : β-sheet ratio is lower than the individual CP-1 or CP-3 but is similar to the equimolar mixture of them. In addition, the α-helix : random coil and β-sheet : random coil ratios are lower than the individual CP-1 or CP-3 but are similar to the equimolar mixture of them. These ratios suggest that assembly and cross-linking of the precursor proteins into the mature egg capsule is associated with β-sheets and random coils formation at the expense of α-helices. Since the relative content of α-helices and β-sheets in the equimolar mixture of CP-1 and CP-3 is also significantly closer to the capsule material than the individual CP-1 or CP-3 is, these data suggest that during capsule processing, CP-1 and CP-3 associate to form localized regions of inter-chain β-sheets. However, it is unknown if they would associate and form dimeric or trimeric coiled-coils. Furthermore, it is also unknown if the coiled-coils of the egg capsule are homomeric or heteromeric. The propensity of the Bc-CPs for β-sheet formation is supported by FTIR measurements of individual CP-1 and CP-3, which were freeze-dried from concentrated solutions in 5% acetic acid (supplementary Fig. S4). A clear increase in β-sheet content was noticed with higher concentrations, and was accompanied with a lower protein solubility. However, the Bc-CPs very likely misfolded into non-native amyloid-like structures, as is commonly seen with many other proteins at high concentrations [58]

The SDS-PAGE of immature capsules and crude nidamental gland extract indicated a trace presence of proteins other than the Bc-CPs. Therefore, although TEM showed that the purified Bc-CPs are capable of self-assembling into fibers (supplementary Fig. S7), an equally plausible scenario that could contribute to the increase in the β-sheets and random coils content during the natural assembly process is an incorporation of additional proteins which may serve to bundle the coiled-coil filaments into bigger fibers that make up the

mature egg capsule. The presence of such proteins in the mature capsule is further suggested by a significantly high Gly content which cannot be solely attributed to the Bc-CPs.

CD experiments were carried out at pH 4 in order to keep CP-1 and CP-2 in solution. At this low pH, CD spectra of individual Bc-CPs suggest that they preferentially adopt single-coil rather than coiled-coil conformations, as seen from the $[\theta]_{222} : [\theta]_{208}$ ratios which are <1[59, 60] (Fig. 6). Given that both CP sequences unambiguously predict for coiled-coil structures and the clear coiled-coil signature of mature capsules by WAXS, it is reasonable to assume that single-coil helices assemble into coiled-coils upon secretion into the higher pH extracellular environment (sea water, pH ∼8). CD spectra at seawater pH or even neutral pH could not be obtained since both proteins would not solubilize.

A critical question we sought to answer involves understanding the molecular mechanisms by which the α-helix to β-sheet transition (and its associated unique mechanical properties) is initiated in the mature capsule. We propose two possible mechanisms by which the phase transition can be initiated by means of nucleation, namely (i) by elongation of pre-existing β-sheets and (ii) by local unwinding of the coiled-coil at regions of instabilities due to stutters or where the heptad is incomplete (Fig. 10). In the first case, while the β-sheet content does not vary significantly between the solution states and the solid egg capsule, its presence in the unstrained capsule material is unlikely to be fortuitous, and may have a significant role in the strain-induced α-helix to β-sheet phase transition. Thermodynamically, small domains of β-sheets can serve as nucleation sites for the strain-induced α-helical coiled-coil to β-sheet transformation detected by WAXS. β-sheets are extended crystalline structures and are inherently stiffer than other secondary structures, as is well-established in silks[61] and other IFs that undergo α-helix to β-sheet transition[62]. Hence, when a series of stiff β-sheets and compliant coiled-coils, is pulled axially, a stress concentration occurs at the coiled-coil/β-sheet interface. This creates an ideal driving force for the coiled-coils or random coils at the interfaces to transform into new β-sheets that add to the amount of pre-existing β-sheets. The presence of pre-existing interchain β-sheets in a coiled-coil of CP-1 and CP-2 would then mean that the new β-sheets are formed between the chains of CP-1 and CP-2 as the coiled-coil extends and unfolds. In other words, much like impurities are well-known to help initiate liquid-to-solid phase transitions by lowering the activation energy barrier, initial β-sheets crystals would be able to locally initiate the large-scale strain-induced phase transition by lowering the energy barrier for the transition. The second mechanism postulates that the transition starts occurring at coiled-coil instabilities. Stutters are found in all the egg case proteins and were also identified in the capsular proteins of a related species, *Pugilina cochilidum*[63]. Critically, stutters have been found by comprehensive molecular dynamic simulations to be the preferred site for the initiation of the α-β transition in the IF protein vimentin[64, 65] and their presence is thus viewed as a necessary condition to induce the transition. Both mechanisms could possibly be concurrently active, but further experimental work is necessary before more definitive conclusions can be drawn.

4.4 Proposed micro-mechanical model of α-β transition

Based on an equimolar amount of CP-1 and CP-3, the maximum end-to-end distance of the average capsule protein in the solid-state unstrained egg capsule is 93.3 or 91.0 nm, depending on whether the pre-existing β-sheets are parallel or antiparallel respectively. When the capsule material is stretched, α-helices unravel and form β-sheets. If these newly formed β-sheets are parallel, the end-to-end lengths increase to 153.4 and 151.1 nm, corresponding to an increase of 64.4 and 66.0% respectively. If the new β-sheets are antiparallel, the lengths would increase to 147.3 and 145.0 nm, which correspond to increments of 57.9 and 59.3% respectively. All these values are close to the 60% post-yield

engineering strain that was observed with the bulk egg capsule material. From the amino acid composition analyses, the average Lys content of CP-1 and CP-3 is ∼3% higher than the bulk material. This difference has been thought to be due the formation of Lys-Lys cross-links when the immature egg capsule is processed under the ventral pedal gland (VPG) of the snail. Lys-Lys cross-links are resistant to acid hydrolysis, and so would account for the lower Lys content in the mature capsule as measured by amino acid analysis. As the distribution of Lys is fairly constant throughout both the primary sequences of CP-1 and CP-3 (Fig. 11), the α-helices, β-sheets and random coils in the mature egg capsule are respectively 1.5, 0.6 and 0.9% cross-linked on average. It has been hypothesized that disulfide linkages prevent unraveling of the 1B and 2B helical domains[16] in the hard α-keratin materials, but compared to the high cross-link density of up to 15% in the hard α-keratins, the low cross-link density in the α-helical domains of the egg capsule is unlikely to have a huge effect on the unraveling and hence extensibility of the helices. Indeed, the extensibilities of the α-keratin materials and the egg capsule material are vastly different in terms of the rate and extent of recovery, which could very well be due to the stark differences in the nature and densities of the cross-links.

4.5 Egg case cross-linking and open questions

During the egg laying process, Bc-CPs first assemble into filaments in the immature egg capsule before they are cross-linked under the VPG[32]. Similar to purified IF proteins in solution, the egg capsule filament assembly can occur *in vitro* without the aid of additional accessory proteins or molecules. Based on the similar dimensions of the fibers that are created *in vitro*, observed here by TEM, and those that occur naturally in the egg capsule, observed previously by SEM[17], the purified Bc-CPs form the filaments that the mature egg capsule material is composed of. While the secondary structure analyses by FTIR suggest that the egg capsule filament is composed of a self-assembling mixture of CP-1 and CP-2, separate solutions of CP-1 and CP-2 are equally capable of forming filaments. Further studies are thus required to determine if filaments of CP-1 or CP-2 are present in the mature egg capsule. On the other hand, IF proteins such as α-keratins are known to form heterodimeric coiled-coils *in vivo* but not *in vitro* unless both proteins are present in the solution[66] Therefore, further studies are also required to confirm the presence of filaments containing both CP-1 and CP-2 in the egg capsule. Furthermore, as IF proteins are known to form coiled-coil tetramers, and sometimes octamers, in solution before assembling into filaments[66, 67], a more thorough investigation into the self-assembly of the Bc-CPs will be necessary to ascertain whether they form coiled-coil multimers in solution prior to filament assembly.

After assembly, cross-linking occurs as the final processing step before the egg capsules mature and are deposited on the ocean floor. It is only after cross-linking that the egg capsules exhibit unique and robust mechanical properties. Previous work on both the egg capsule material[45] and the VPG where this final processing step occurs[46] suggested that the cross-links are Lys-derived. As seen in the amino acid composition analyses, all the Bc-CPs have high quantities of Lys which are quite fairly distributed throughout the length of the primary sequences. Based on the heptad repeat assignment by SCORER, Lys residues can be found in positions both within and outside the core of the coiled-coil. This translates to a possibility for inter-chain cross-links to form both between and within coiled-coils. In CP-2 and 3 specifically, there is a high proportion of Lys residues in the C-terminal regions. CP-2 and CP-3 have 53 and 54 Lys residues out of a total of 475 and 479 amino acid residues respectively, but 20 and 21 of those Lys residues are located between position 413 and the C-terminus (Fig. 11). These regions of the primary sequences seem to be responsible for end-to-end cross-linking of the Bc-CPs, but until a chemical structure of the cross-link is determined, nothing definitive can be said about Lys-based cross-links.

There is a variety of unanswered questions regarding the whelk egg capsule post-translation assembly and processing. The mature capsule material is over 90% protein by dry weight, but the composition of the remaining mass is unknown. Furthermore, SDS-PAGE of the immature capsules and crude nidamental gland extract indicate that traces of additional proteins, other than the Bc-CPs, are present in this material. They are in far lower quantities, but their structure and function are yet to be investigated. They are potentially matrix proteins that bundles the egg capsule filament proteins into higher hierarchical fibrous structures. α-keratin materials have globular, high-sulfur content matrix proteins which contribute significantly to the mechanical properties[5, 16] whereas hagfish slime does not[28]

At the moment, precise molecular-scale knowledge of the steps leading to the extended β-sheet conformation and their time-dependent mechanisms remain unknown. Elucidation of the crystal structures of the Bc-CPs would provide key information to further understand the solid-state phase transformation.

5. Conclusions

In this study, the precursors of the primary structural proteins of the whelk egg capsules were purified and characterized. Their amino acid compositions are very similar to IF proteins with the exception of Lys and Cys, which are significantly higher in the egg capsule and the IFs, respectively. CD and FTIR data indicate that, much like the IF proteins, the Bc-CPs are strongly α-helical in solution and are capable of self-assembling into filaments without the assistance of accessory molecules.

Using reverse transcription techniques, the primary amino acid sequences of 4 Bc-CPs were deduced from their cDNA. The primary sequences are unique as BLAST did not produce any significant match to known proteins. Analyses of these sequences indicated that the proteins are likely to form coiled-coils, which is consistent with previous *x*-ray scattering studies. In comparison to coiled-coil proteins from an established database, Bc-CPs appear to have a higher amount of Gly and smaller amounts of aromatic residues in their hydrophobic cores. Furthermore, the terminal regions of the sequences are densely populated by Lys and Arg, both of which are suspected to play a role in the covalent cross-linking of the egg capsule fibers.

This work opens further interesting questions, especially regarding self-assembly and the molecular mechanisms of the α-β transition, which is an emerging theme for a variety of bioelastomeric materials[5] Firstly, the prediction of a trimeric coiled-coil in the egg case capsule that is reminiscent of fibrin needs to be tested further. Secondly, there are uncertainties whether the coiled-coils are homo-oligomeric or hetero-oligomeric. Finally, the exact nature of the initiation of α-β transition, which is critical in driving the reversible transformation, remains as an intriguing question. ATR-FTIR data indicates a small, yet significant β-sheet content in both the purified proteins and in the mature capsule. These β-sheets “nuclei” could act as local sites where new β-sheet formation is initiated. Alternatively, the transition could be initiated at coiled-coil instabilities such stutters.

Elucidation of these mechanisms is currently underway in our laboratory using recombinant and site-directed mutagenesis of the precursor proteins.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Dr. Krystyna Brzezinska for assistance with the CD spectrometry, Dr. Paul Guerette for many useful discussions, Rebekah Ashley (Molecular Cellular and Developmental Biology, UC Santa Barbara) for assistance with cDNA sequencing and Prof. Robert Shadwick (University of British Columbia, Vancouver, Canada) for insightful advice throughout the project. We thank Tan Kang Hong Delford and Loke Jun Jie for assistance with protein purification.

This work was supported by a University of California Biotechnology Research and Education Program (UC BREP) Graduate Research and Education in Adaptive Biotechnology (GREAT) grant (#2007–02), an NIH grant #R01DE018468 (JHW), and the Singapore National Research Foundation (NRF) through a NRF Fellowship (A.M.). The work made use of MRL Central Facilities supported by the MRSEC program of the NSF under Award # DMR05-20415. C.N. and A.S acknowledge the partial support the Austrian Federal Ministry for Transport, Innovation and Technology (GZ BMVIT-612.166/0001-III/I1/2010) and the ZIT (Center of Innovation and Technology of Vienna).

References

1. Gosline JM, Lillie MA, Carrington E, Guerette P, Ortlepp C, Savage K. Elastic Proteins: Biological Roles and Mechanical Properties. Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences. 2002; 357:121–132.
2. Rauscher, S.; Pomès, R. Chapter 10: Structural Disorder and Protein Elasticity. In: Fuxreiter, M.; Tompa, P., editors. Fuzziness: Structural Disorder in Protein Complexes. Landes Bioscience and Springer Science+Business Media; 2012. p. 159-183.
3. Tatham AS, Shewry PR. Comparative Structures and Properties of Elastic Proteins. Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences. 2002; 357:229–234.
4. Urry DW, Hugel T, Seitz M, Gaub HE, Sheiba L, Dea J, Xu J, Parker T. Elastin: A Representative Ideal Protein Elastomer. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences. 2002; 357(1418):169–184.
5. Feughelman, M. Mechanical Properties and Structure of Alpha-Keratin Fibers: Wool, Human Hair and Related Fibres. University of New South Wales Press; 1997.
6. Koch EA, Spitzer RH, Pithawalla RB, Parry DA. An Unusual Intermediate Filament Subunit from the Cytoskeletal Biopolymer Released Extracellularly into Seawater by the Primitive Hagfish (Eptatretus stouti). J Cell Sci. 1994; 107(11):3133–3144. [PubMed: 7535307]
7. Fudge DS, Gardner KH, Forsyth VT, Riekel C, Gosline JM. The mechanical properties of hydrated intermediate filaments: Insights from hagfish slime threads. Biophysical Journal. 2003; 85(3):2015–2027. [PubMed: 12944314]
8. Gathercole LJ, Atkins EDT, Goldbeckwood EG, Barnard K. Molecular Bending and Networks in a Basement Membrane-like Collagen - Packing in Dogfish Egg Capsule Collagen. Int J Biol Macromol. 1993; 15(2):81–88. [PubMed: 8485107]
9. Hu XY, Kohler K, Falick AM, Moore AMF, Jones PR, Sparkman OD. Vierra, C, Egg Case Protein-1 - A New Class of Silk Proteins with Fibroin-like Properties from the Spider *Latrodectus hesperus*. Journal of Biological Chemistry. 2005; 280(22):21220–21230. [PubMed: 15797873]
10. Walker AA, Weisman S, Kameda T, Sutherland TD. Natural Templates for Coiled-Coil Biomaterials from Praying Mantis Egg Cases. Biomacromolecules. 2012; 13(12):4264–4272. [PubMed: 23137042]
11. Miserez A, Guerette PA. Phase transition-induced elasticity of α-helical bioelastomeric fibres and networks. Chemical Society Reviews. 2013; 42(5):1973–1995. [PubMed: 23229440]
12. Flory PJ. Theory of Elastic Mechanisms in Fibrous Proteins. Journal of the American Chemical Society. 1956; 78(20):5222–5235.
13. Rudall K. Silk and other cocoon proteins. Comparative biochemistry. 1962; 4:397–433.
14. Wortman FJ, Zahn H. The Stress/Strain Curve of α-Keratin Fibers and the Structure of the Intermediate Filament. Textile Research Journal. 1994; 64:737–743.
15. Bendit EG. Alpha-beta Transformation in Keratin. Nature. 1957; 179(4558):535–535.
16. Hearle JW. A Critical Review of the Structural Mechanics of Wool and Hair Fibres. Int J Biol Macromol. 2000; 27(2):123–138. [PubMed: 10771062]

17. Miserez A, Wasko S, Carpenter C, Waite JH. Non-Entropic and Reversible Long-Range Deformation of an Encapsulating Bioelastomer. Nat Mater. 2009; 8:910–916. [PubMed: 19838185]
18. Harrington MJ, Wasko S, Masic A, Fisher FD, Gupta HS, Fratzl P. Pseudoelastic Behaviour of a Natural Material is Achieved via Reversible Changes in Protein Backbone Conformation. Journal of the Royal Society Interface. 2012; 9(76):2911–2922.
19. Qin Z, Kreplak L, Buehler MJ. Nanomechanical Properties of Vimentin Intermediate Filaments Dimers. Nanotechnology. 2009; 20:425101. [PubMed: 19779230]
20. Zhmurov A, Kononova O, Litvinov RI, Dima RI, Barsegov V, Weisel JW. Mechanical Transition from α-Helical Coiled Coils to P-Sheets in Fibrin(ogen). Journal of the American Chemical Society. 2012; 134(50):20396–20402. [PubMed: 22953986]
21. Denny, MW. Biology and the Mechanics of the Wave-Swept Environment. Princeton University Press; 1988.
22. Magalhaes H. An Ecological Study of Snails of the Genus Busycon at Beaufort, North Carolina. Ecological Monographs. 1948; 18(3):377–409.
23. Rawlings TA. Adaptations to Physical Stresses in the Intertidal Zone: The Egg Capsules of Neogastropod Molluscs. American Zoologist. 1999; 39(2):230–243.
24. Aaron BB, Gosline JM. Elastin as a Random-network Elastomer - A Mechanical and Optical Analysis of Single Elastin Fibers. Biopolymers. 1981; 20(6):1247–1260.
25. Rapoport HS, Shadwick RE. Mechanical Characterization of an Unusual Elastic Biomaterial from the Egg Capsules of Marine Snails (*Busycon spp*). Biomacromolecules. 2002; 3:42–50. [PubMed: 11866554]
26. Lendlein A, Kelch S. Shape-memory Polymers. Angewandte Chemie (International ed. in English). 2002; 41(12):2035–2057. [PubMed: 19746597]
27. Weis-Fogh T. Molecular Interpretation of the Elasticity of Resilin, a Rubber-like Protein. Journal of Molecular Biology. 1961; 3:648–667.
28. Fudge DS, Gosline JM. Molecular Design of the Alpha-Keratin Composite: Insights from a Matrix-Free Model, Hagfish Slime Threads. Proceedings. Biological sciences / The Royal Society. 2004; 271(1536):291–299. [PubMed: 15058441]
29. Pollock CM, Shadwick RE. Relationship between Body Mass and Biomechanical Properties of Limb Tendons in Adult Mammals. The American journal of physiology. 1994; 266(3 Pt 2):R1016–R1021. [PubMed: 8160850]
30. Kastelic J, Baer E. Deformation in Tendon Collagen. Symposia of the Society for Experimental Biology. 1980; 34:397–435. [PubMed: 7256561]
31. Denny M. Physical-properties of Spiders Silk and their Role in Design of Orb-Webs. Journal of Experimental Biology. 1976; 65(2):483–506.
32. Rapoport HS, Shadwick RE. Reversibly Labile, Sclerotization-induced Elastic Properties in a Keratin Analog from Marine Snails: Whelk Egg Capsule Biopolymer (WECH). Journal of Experimental Biology. 2007; 210(1):12–26. [PubMed: 17170144]
33. Flower NE, Geddes AJ, Rudall KM. Ultrastructure of Fibrous Protein from Egg Capsules of Whelk *Buccinum undatum*. Journal of Ultrastructure Research. 1969; 26(3–4):262–&. [PubMed: 5813350]
34. Goldsmith LA, Hanigan H-M, Thorpe JM, Lindberg KA. Nidamental Gland Precursor of the Egg Capsule Protein of the Gastropod Mollusc *Busycon carica*. Comparative Biochemistry and Physiology Part B: Comparative Biochemistry. 1978; 59(2):133–138.
35. Zhao H, Waite JH. Coating proteins: Structure and cross-linking in fp-1 from the green shell mussel Perna canaliculus. Biochemistry. 2005; 44(48):15915–15923. [PubMed: 16313194]
36. Greenfield NJ. Using Circular Dichroism Spectra to Estimate Protein Secondary Structure. Nature Protocols. 2006; 1(6):2876–2890.
37. Louis-Jeune C, Andrade-Navarro MA, Perez-Iratxeta C. Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. Proteins: Structure, Function, and Bioinformatics. 2012; 80(2):374–381.
38. Creighton, TE. Proteins: Structure and Molecular Properties. Macmillan Higher Education; 2013.

39. Vincent, JFV. Structural Biomaterials. Princeton University Press; 1990.
40. Armstrong CT, Vincent TL, Green PJ, Woolfson DN. SCORER 2.0: An Algorithm for Distinguishing Parallel Dimeric and Trimeric Coiled-Coil Sequences. Bioinformatics (Oxford, England). 2011; 27(14):1908–1914.
41. Zhou NE, Zhu BY, Kay CM, Hodges RS. The 2-Stranded Alpha-helical Coiled-Coil is an Ideal Model for Studying Protein Stability and Subunit Interactions. Biopolymers. 1992; 32(4):419–426. [PubMed: 1623137]
42. Troullier A, Reinstadler D, Dupont Y, Naumann D, Forge V. Transient Non-Native Secondary Structures during the Refolding of Alpha-Lactalbumin Detected by Infrared Spectroscopy. Nature Structural Biology. 2000; 7(1):78–86.
43. Rapoport, HS. Biomechanics, Biochemistry, and Molecular Biology of a Molluscan Scleroprotein Elastomer: Whelk Egg Capsules. San Diego, La Jolla, CA: University of California; 2003.
44. Schwaiger I, Sattler C, Hostetter DR, Rief M. The Myosin Coild-Coil is a Truly Elastic Protein Stucture. Nat Mater. 2002; 1:232–235. [PubMed: 12618784]
45. Price NR, Hunt S. The Occurrence of Reducible Compounds In an Invertebrate Structure Protein of *Buccinum Undatum* (L.). Experientia. 1976; 32(5):557–558.
46. Price NR, Hunt S. An Unusual Type of Secretory Cell in the Ventral Pedal Gland of the Gastropod Mollusc *Buccinum undatum* L. Tissue & cell. 1976; 8(2):217–228. [PubMed: 941131]
47. Koch EA, Spitzer RH, Pithawalla RB, Downing SW. Keratin-like Components of Gland Thread Cells Modulate the Properties of Mucus from Hagfish (*Eptatretus stouti*). Cell and tissue research. 1991; 264(1):79–86. [PubMed: 1711418]
48. Rogers GE. Biology of the Wool Follicle: An Excursion into a Unique Tissue Interaction System Waiting to be Re-discovered. Experimental Dermatology. 2006; 15(12):931–949. [PubMed: 17083360]
49. Woolfson DN. The Design of Coiled-Coil Structures and Assemblies. Advances in Protein Chemistry. 2005; 70:79–112. [PubMed: 15837514]
50. Lupas AN, Gruber M. The Structure of α-Helical Coiled Coils. Advances in Protein Chemistry. 2005; 70:37–78. [PubMed: 15837513]
51. Woolfson DN, Bartlett GJ, Bruning M, Thomson AR. New Currency for Old Rope: From Coiled-Coil Assemblies to α-Helical Barrels. Current Opinion in Structural Biology. 2012; 22:1–10. [PubMed: 22265341]
52. Delorenzi M, Speed T. An HMM Model for Coiled-Coil Domains and a Comparison with PSSM-based Predictions. Bioinformatics (Oxford, England). 2002; 18(4):617–625.
53. Lupas A, Vandyke M, Stock J. Predicting Coiled Coils from Protein Sequences. Science. 1991; 252(5009):1162–1164. [PubMed: 2031185]
54. Wolf E, Kim PS, Berger B. MultiCoil: A Program for Predicting Two- and Three-Stranded Coiled Coils. Protein science : a publication of the Protein Society. 1997; 6(6):1179–1189. [PubMed: 9194178]
55. Bullough PA, Tulloch PA. High-resolution spot-scan electron microscopy of microcrystals of an α-helical coiled-coil protein. Journal of Molecular Biology. 1990; 215(1):161–173. [PubMed: 2398496]
56. Oliver DT, Efrosini M, Derek NW. CC+: a relational database of coiled-coil structures. Nucleic Acids Research. 2009; 37(suppl 1)
57. Moss, D. Royal Society of Chemistry. Cambridge, U.K: 2010. Biomedical Applications of Synchrotron Infrared Microspectroscopy.
58. Dobson CM. Protein Folding and Misfolding. Nature. 2003; 426(6968):884–890. [PubMed: 14685248]
59. Xu C, Liu R, Mehta AK, Guerrero-Ferreira RC, Wright ER, Dunin-Horkawicz S, Morris K, Serpell LC, Zuo X, Wall JS, Conticello VP. Rational Design of Helical Nanotubes from Self-Assembly of Coiled-Coil Lock Washers. Journal of the American Chemical Society. 2013; 135(41):15565–15578. [PubMed: 24028069]
60. Zaccai NR, Chi B, Thomson AR, Boyle AL, Bartlett GJ, Bruning M, Linden N, Sessions RB, Booth PJ, Brady RL, Woolfson DN. A de novo peptide hexamer with a mutable channel. Nat Chem Biol. 2011; 7(12):935–941. [PubMed: 22037471]

61. Keten S, Xu Z, Ihle B, Buehler MJ. Nanoconfinement Controls Stiffness, Strength and Mechanical Toughness of Beta-Sheet Crystals in Silk. Nat Mater. 2010; vol. 9(4):359–367. [PubMed: 20228820]

62. Fudge DS, Hillis S, Levy N, Gosline JM. Hagfish Slime Threads as a Biomimetic Model for High Performance Protein Fibres. Bioinspiration & Biomimetics. 2010; 5 035002 (8pp).

63. Guerette PA, Hoon S, Seow Y, Raida M, Masic A, Wong FT, Ho VHB, Kong KW, Demirel MC, Pena-Francesch A, Amini S, Tay GZ, Ding D, Miserez A. Accelerating the design of biomimetic materials by integrating RNA-seq with proteomics and materials science. Nat Biotech. 2013; 31(10):908–915.

64. Arslan M, Qin Z, Buehler MJ. Coiled-Coil Intermediate Filament Stutter Instability and Molecular Unfolding. Comput Methods Biomech Biomed Engin. 2011; 14(5):483–489. [PubMed: 21516532]

65. Qin Z, Buehler MJ. Molecular Dynamics Simulation of the α-Helix to β-Sheet Transition in Coiled Protein Filaments: Evidence for a Critical Filament Length Scale. Physical Review Letters. 2010; 104(19):198304. [PubMed: 20867006]

66. Kreplak L, Aebi U, Herrmann H. Molecular Mechanisms Underlying the Assembly of Intermediate Filaments. Experimental Cell Research. 2004; 301(1):77–83. [PubMed: 15501448]

67. Coulombe PA, Fuchs E. Elucidating the Early Stages of Keratin Filament Assembly. Journal of Cell Biology. 1990; 111(1):153–169. [PubMed: 1694855]

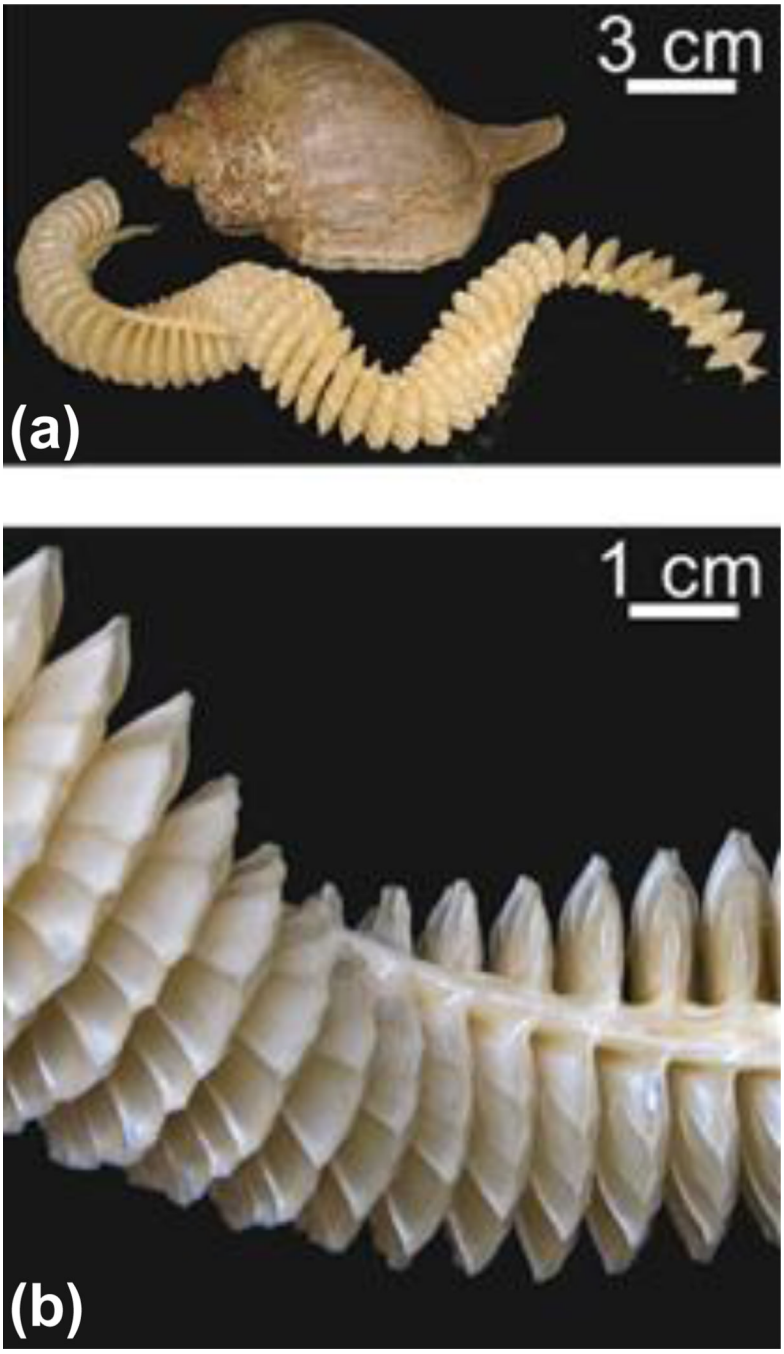

Figure 1.
Photographs of **(a)***Busycotypus canaliculatus* with its string of egg capsules and **(b)** capsules at higher magnification.

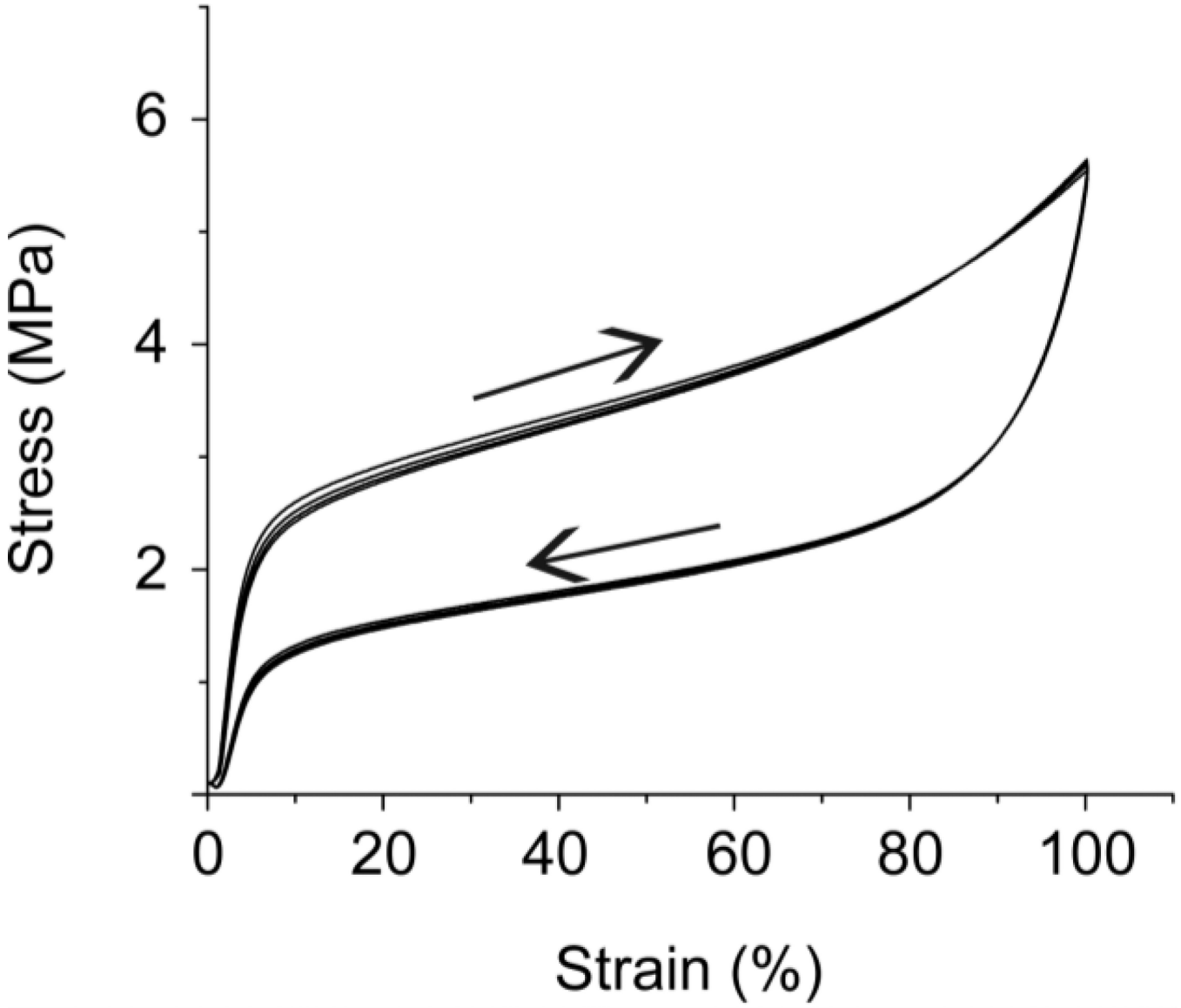

Figure 2.
Typical engineering stress-strain curve of whelk egg capsule material under uniaxial tension. Sample was cyclically strained multiple times at a rate of 5 mm/min with no rest between cycles. Recovery of initial stiffness was complete and rapid.

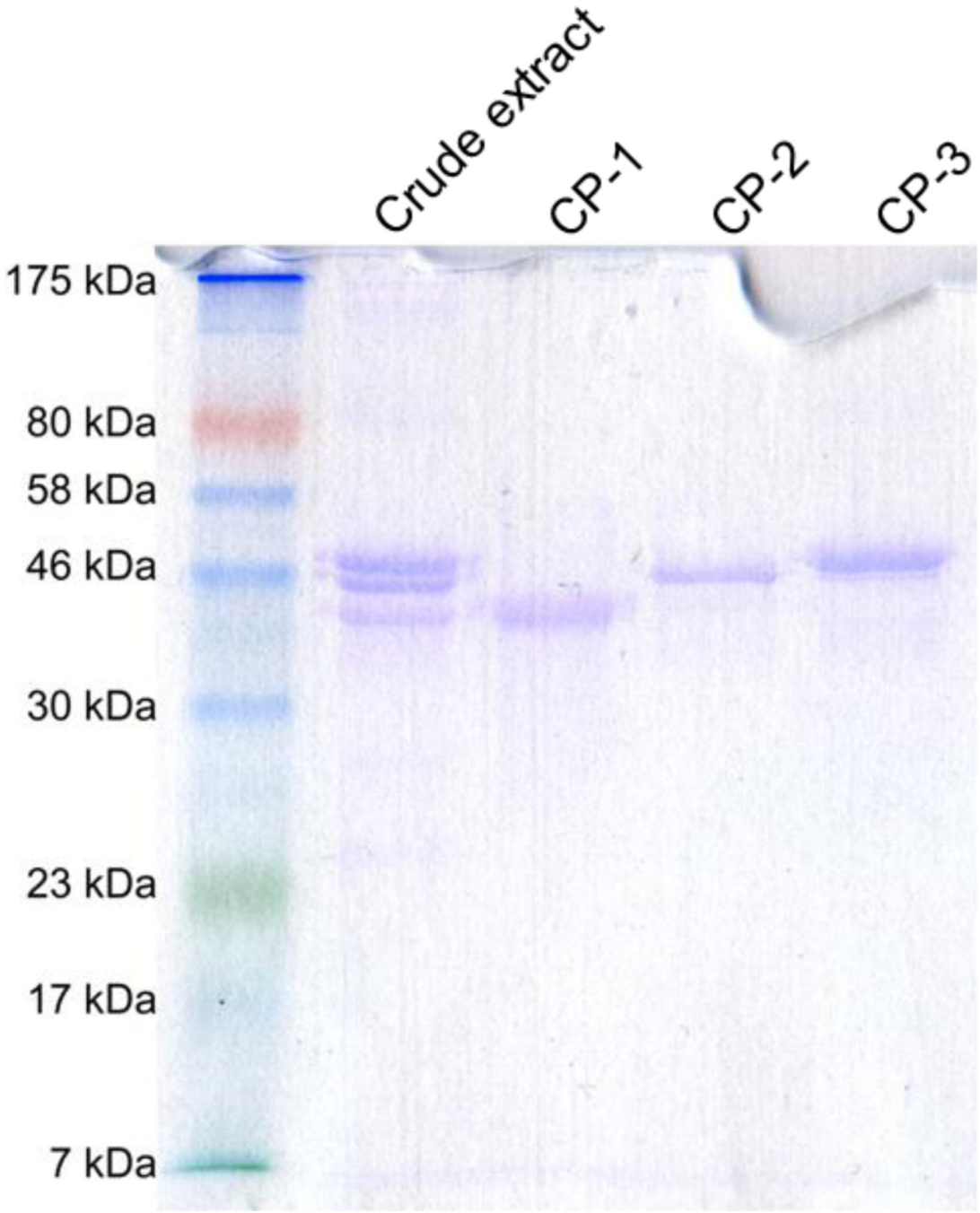

Figure 3.
SDS-PAGE of Bc-CPs purified from nidamental gland crude extract.

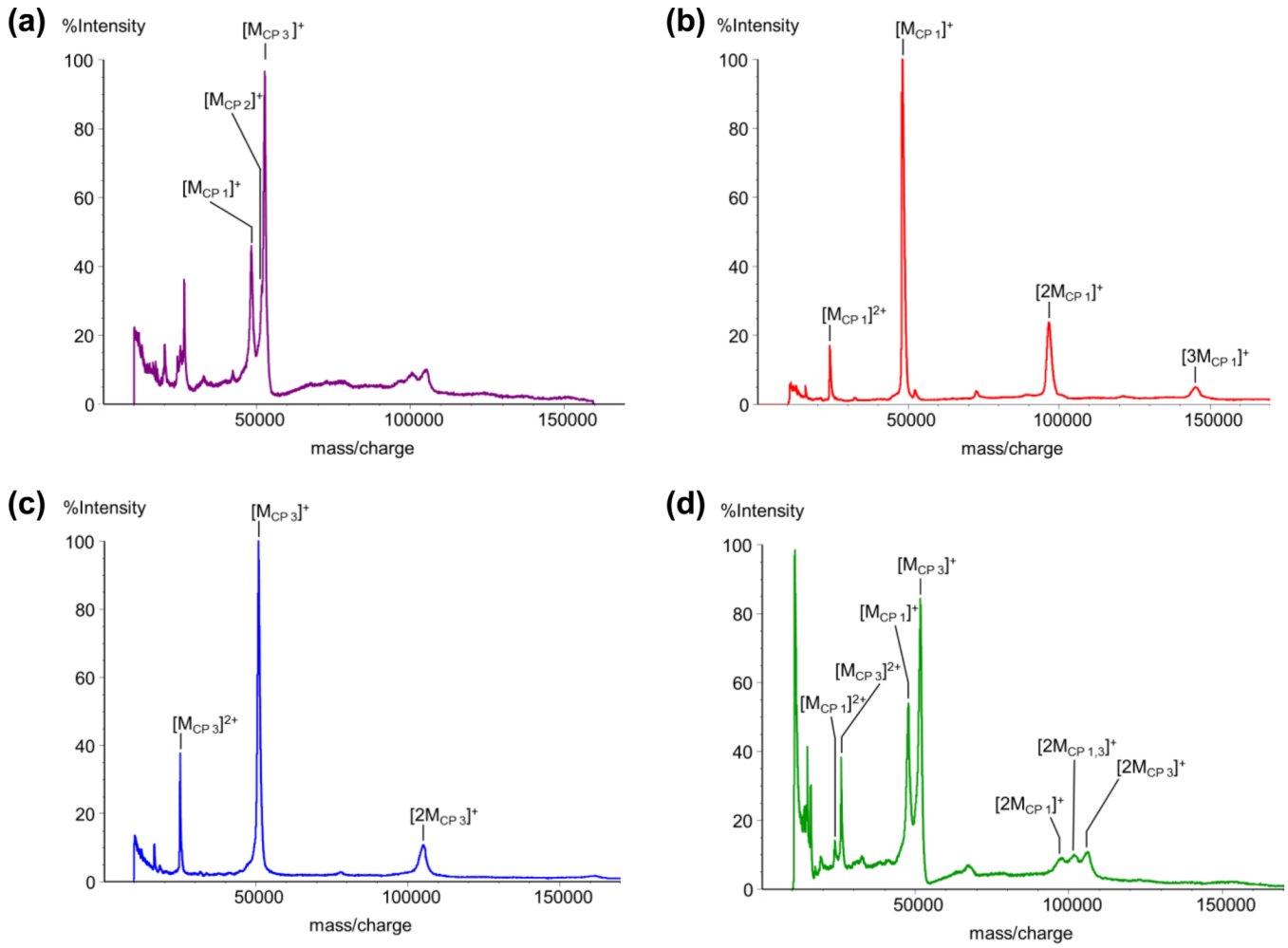

Figure 4.

MALDI-ToF mass spectra of **(a)** nidamental gland crude extract, **(b)** CP-1, **(c)** CP-3 and **(d)** a mixture of CP-1 and CP-3. Observed molecular masses of CP-1, CP-2 and CP-3 are 47.5, 51.6 and 52.1 kDa respectively.

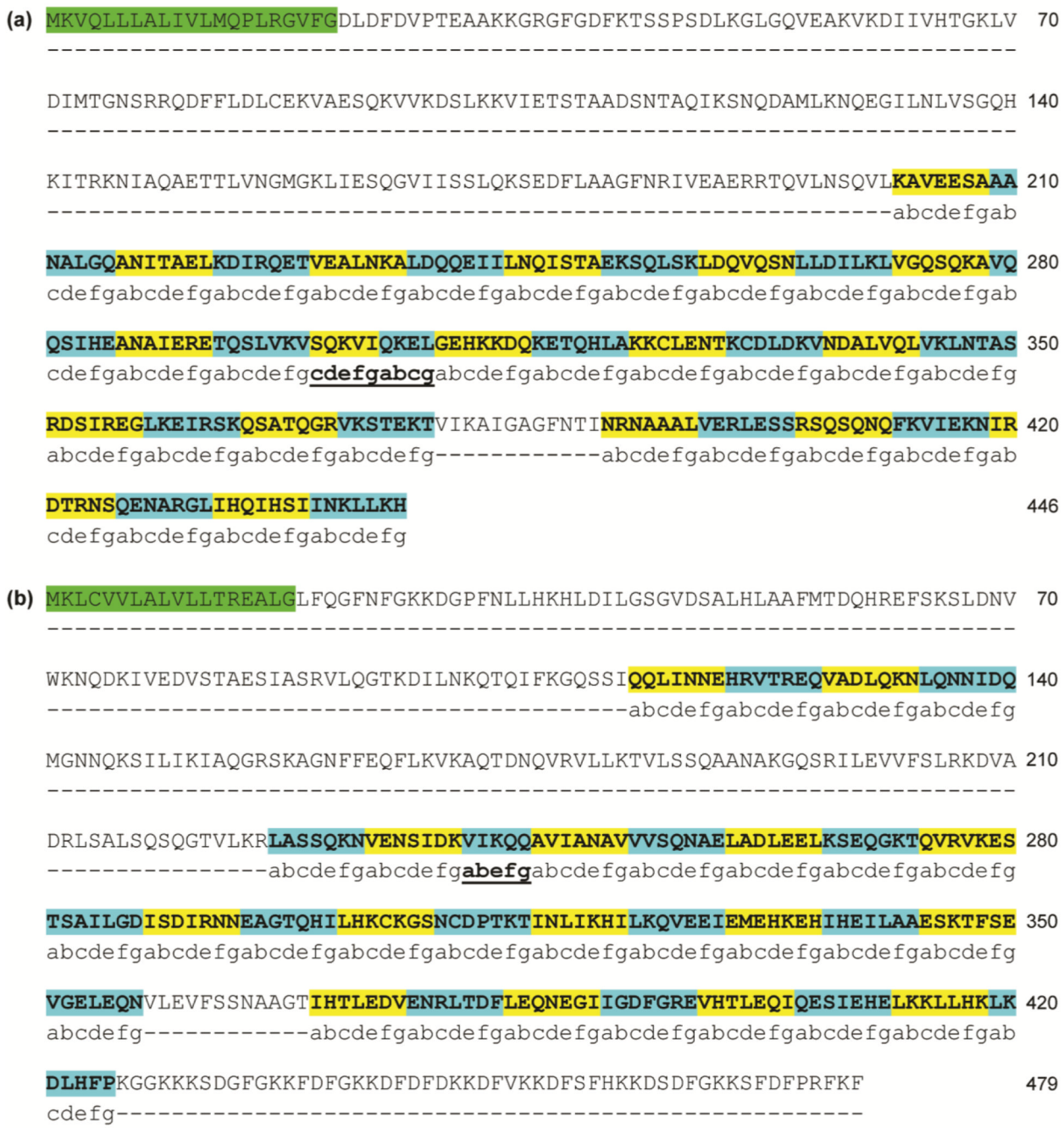

Figure 5.

Complete amino acid sequences of **(a)** CP-1a and **(b)** CP-3 as deduced from cDNA. Highlighted in green are signal peptide sequences, while yellow and blue are predicted coiled-coil heptad repeats ("*abcdefg*") as assigned by the SCORER 2.0 algorithm[40] at 50% confidence level. The heptad repeats account for 57.2 and 47.5% of the CP-1a and CP-3 sequences respectively. Incomplete heptad repeats which are regions of coiled-coil irregularity are underlined (assignment of heptads with greater confidence would require further structural investigations). Complete amino acid sequences of CP-1b and CP-2 are given in supplementary Fig. S2.

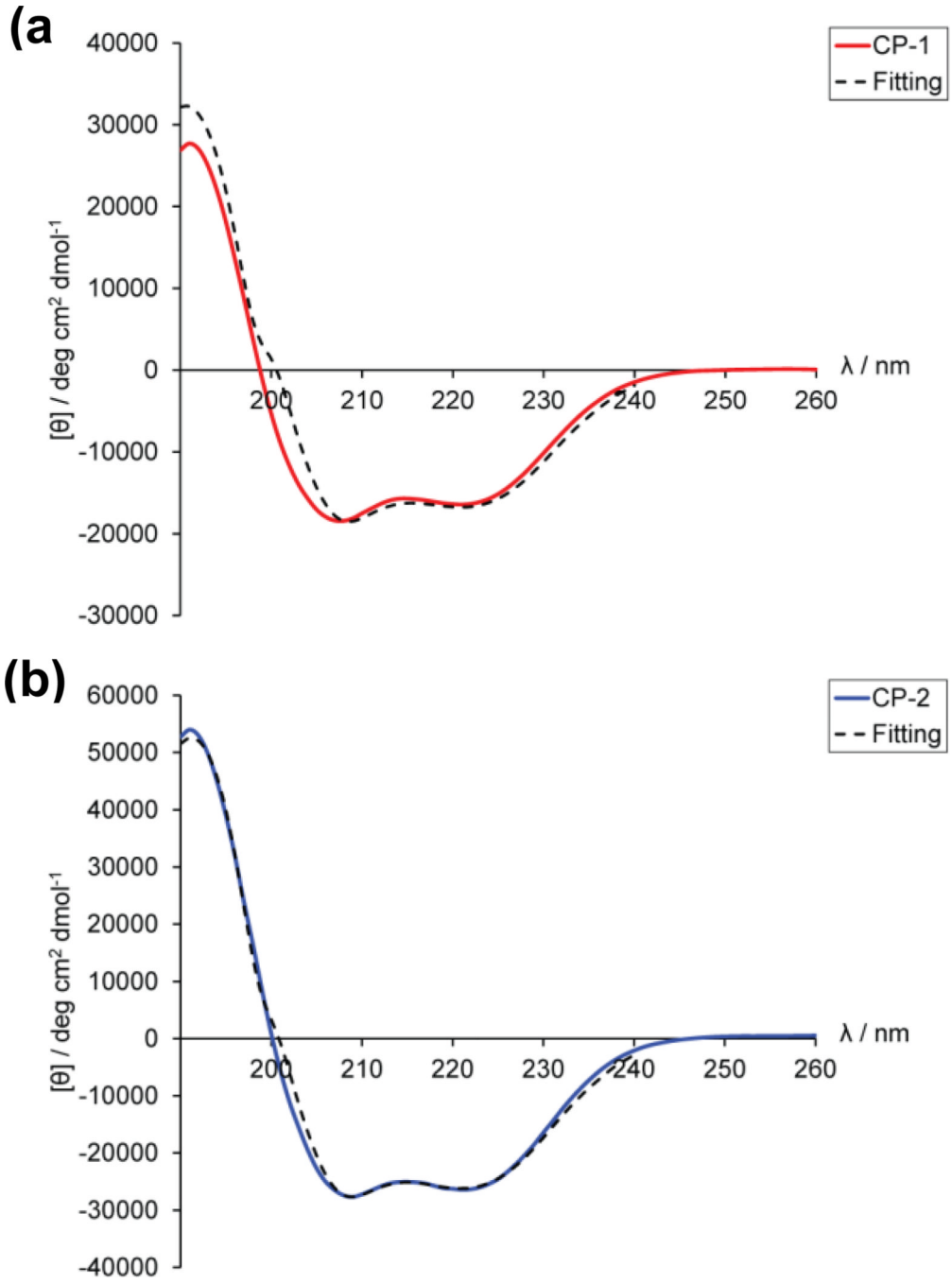

Figure 6.

CD spectra of purified **(a)** CP-1 and **(b)** CP-2 in acetate buffer, pH 4.0. Minima at 208 and 222 nm and a $[\theta]_{222} : [\theta]_{208}$ ratio of <1 are representative of single coil α-helix. Concentrations of CP-1 and CP-2 were ∼0.8 and ∼1.2 mg/mL respectively. K2D3 curve-fitting[37] for both CP-1 and CP-2 revealed secondary structure compositions that are dominated by α-helices (>50%), with small amounts of β-sheets (∼7%). CD spectroscopy was also performed on mixtures of CP-1 and CP-2 in different molar ratios at pH 4.0. This also produced similar α-helical spectra with a $[\theta]_{222} : [\theta]_{208}$ ratio of <1 which indicates single coil α-helix conformation (supplementary Fig. S3).

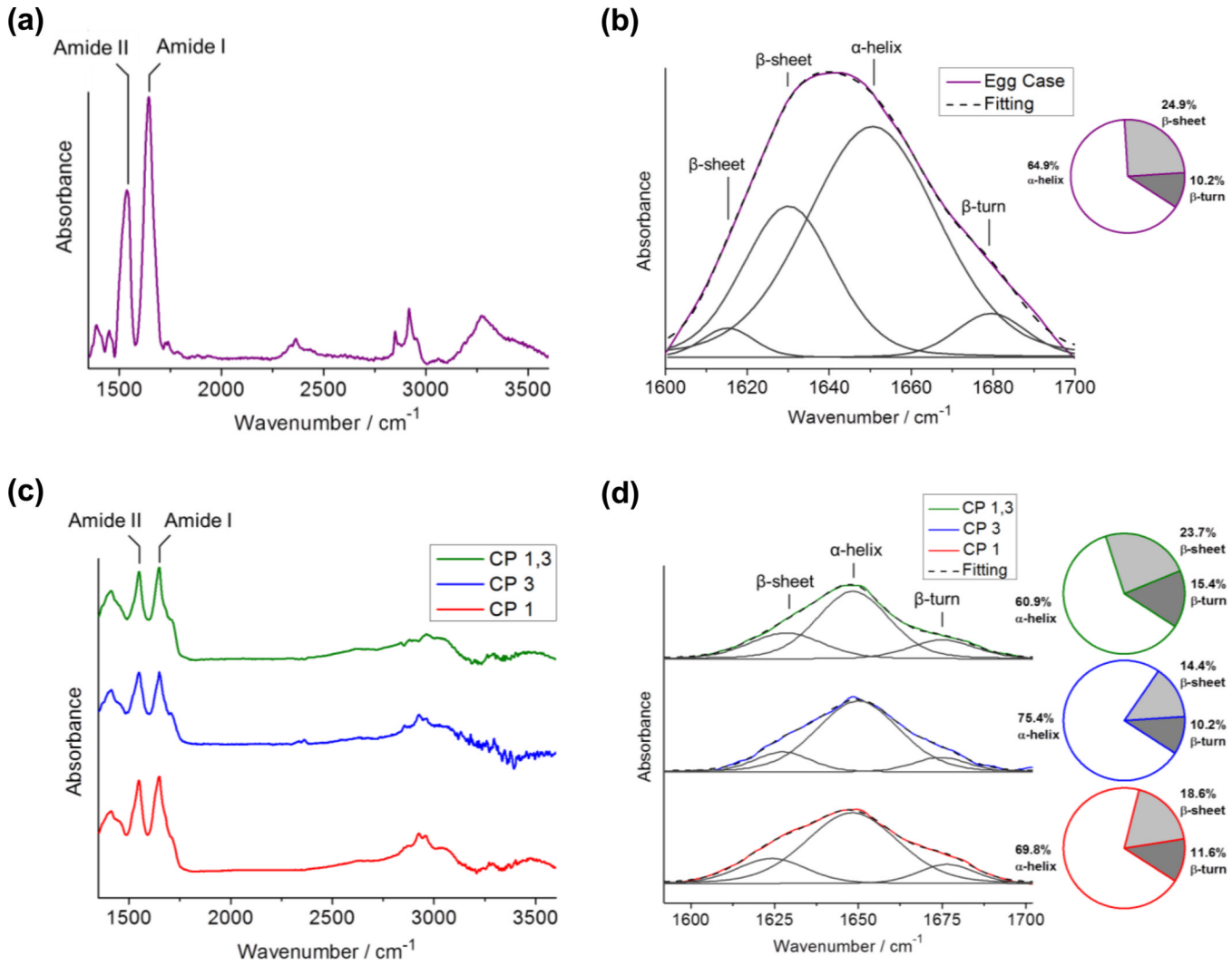

Figure 7.

ATR-FTIR spectra of native (hydrated) egg case and purified egg case proteins. **(a)** Native egg capsule and **(b)** curve-fitted Amide I spectra of the egg case. **(c)** Purified Bc-CPs and **(d)** curve-fitted Amide I of purified Bc-CPs showing secondary structure compositions that are dominated by α-helices (>50%). The α-helix : β-sheet ratio is ∼2.5 in the egg capsule and in the mixture of CP-1 and CP-3, and is significantly lower than that in CP-1 or CP-3 individually.

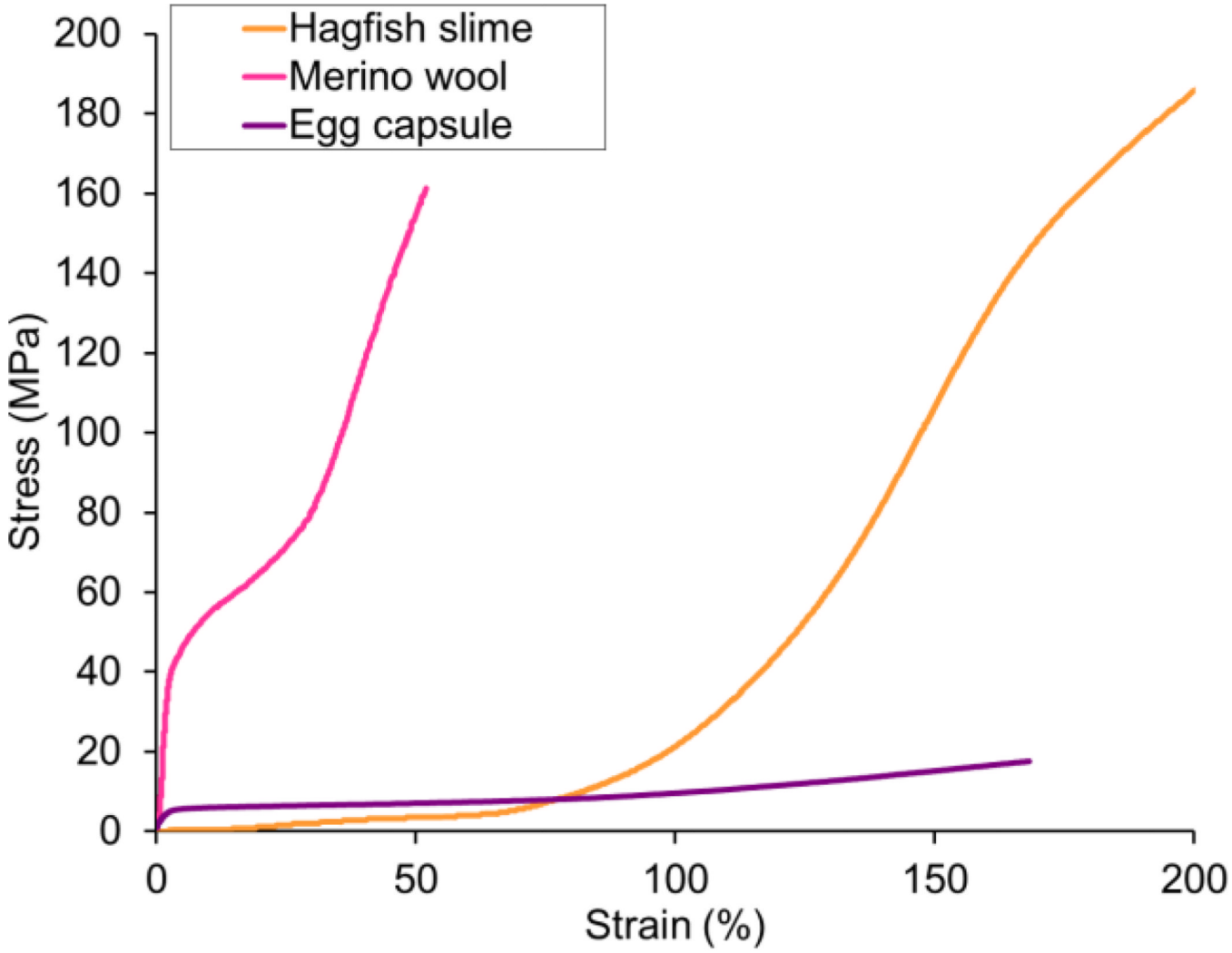

Figure 8.
Engineering stress-strain curves of the whelk egg capsule material, hagfish slime, and merino wool keratin[28]

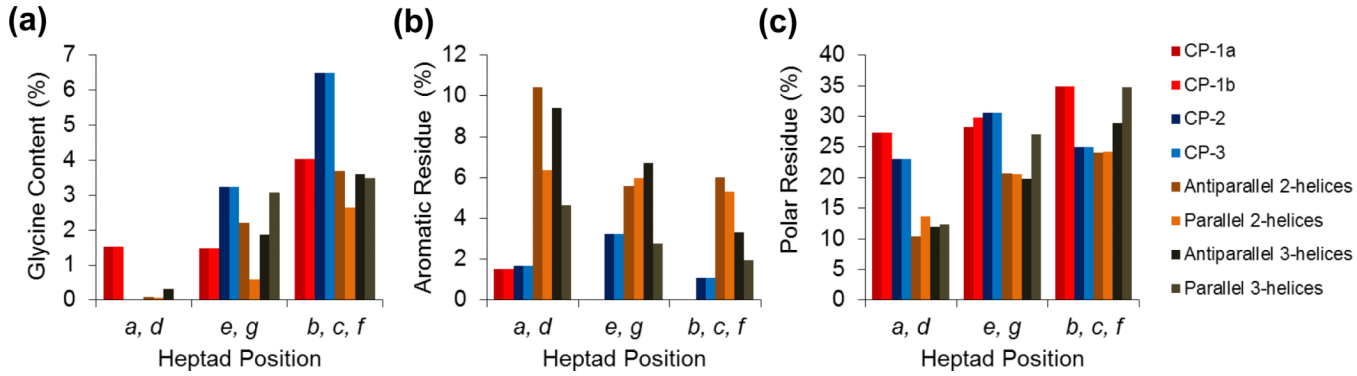

Figure 9.

Relative amounts of **(a)** glycine, **(b)** aromatic and **(c)** polar amino acid residues at various heptad positions in the predicted coiled-coil domains of Bc-CPs, and other known coil-coiled proteins in the CC+ database[56]. In the Bc-CPs, the glycine and polar residue content is notably higher, whereas the aromatic residue content is significantly lower.

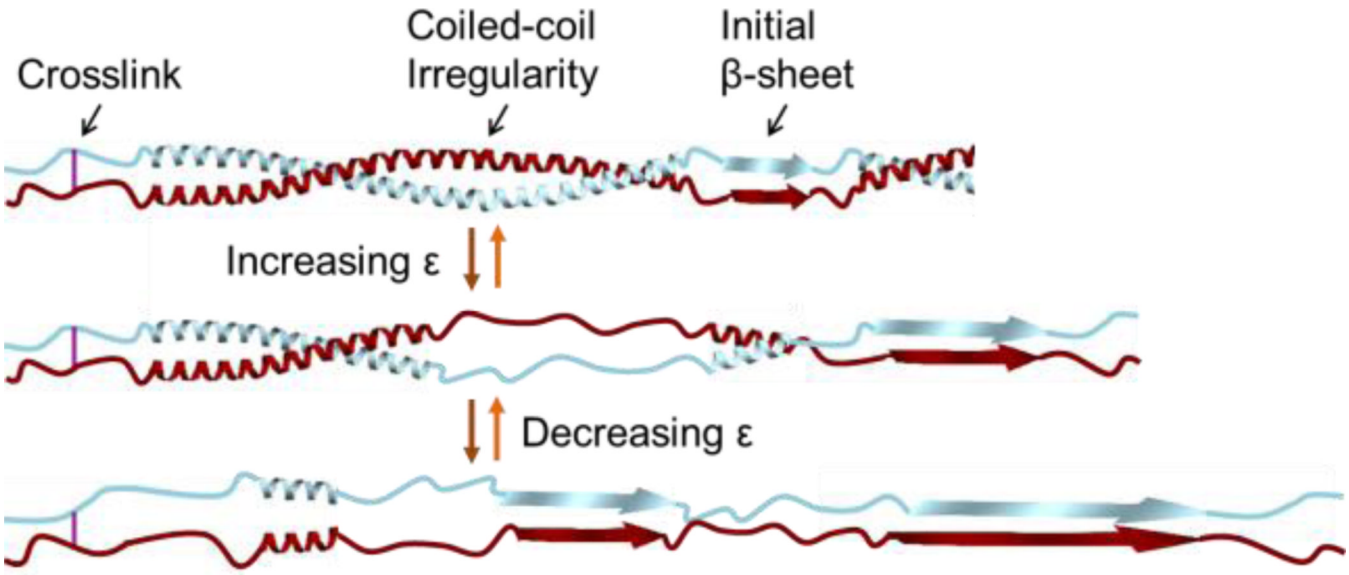

Figure 10.

Proposed model of α-β transition, revised after Ref. [9]. Regions of coiled-coil irregularity, such as stutter or stammer, and the ends of pre-existing intra-chain β-sheets can both serve as stress concentration points where extension and unraveling of coiled-coils occur to form new β-sheets during uniaxial tension. (Coiled-coils are assumed to be dimeric for the purpose of this illustration.)

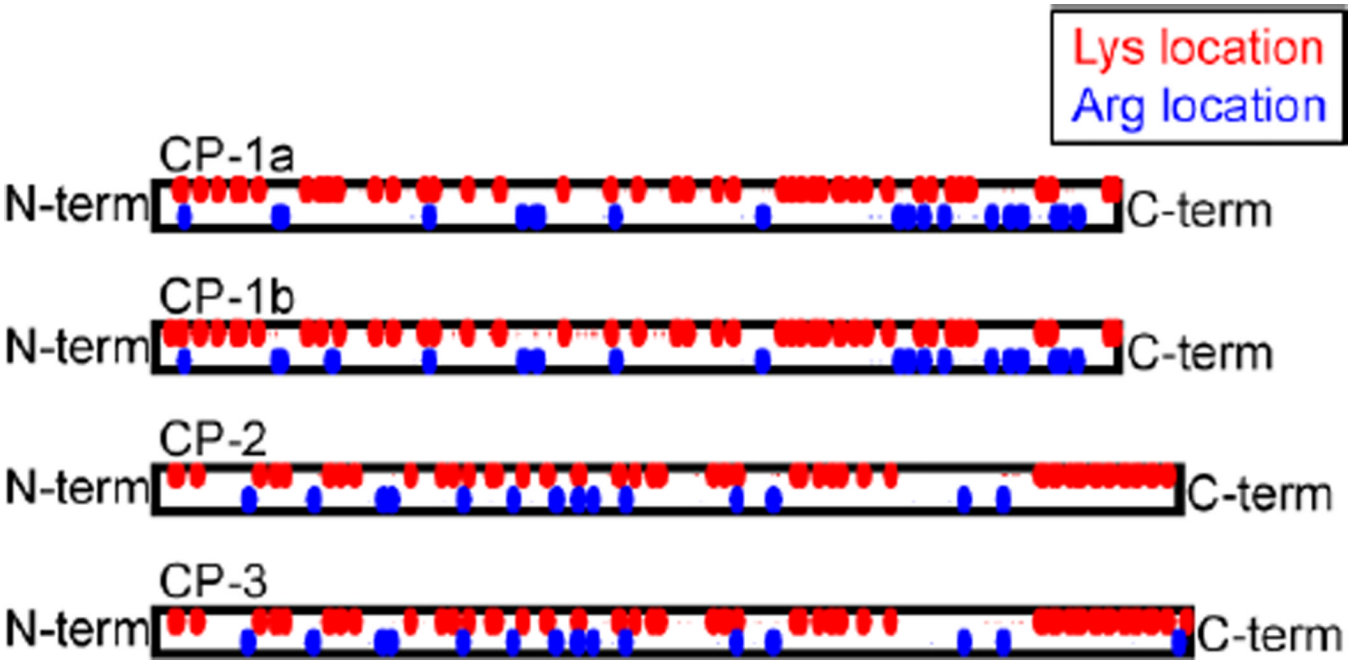

Figure 11.
Distribution of Lys and Arg residues in the primary sequences of the Bc-CPs. There are distinct clusters of not only Lys but also Arg, which preliminary data suggests may contribute to cross-linking.

Table 1

Mechanical properties of the whelk egg capsule, compared with other fibrous protein materials[5, 24, 27–31]. Strain rates for egg capsule, keratin, collagen and silk are all comparable, but are unavailable for resilin and elastin.

Protein Material	E_i (MPa)	Yield Stress (MPa)	Breaking Stress (MPa)	Yield Strain (%)	Breaking Strain (%)	Strain Energy to Failure (MJm^{-3})	Resilience (%)
Whelk Egg Capsule	60	2.5	10.0	5	169	8.5	49
Wool Keratin	3000	30	150	5	52	37.4	55
Resilin	2	NA	4	NA	190	4	92
Elastin	1	NA	2	NA	150	1.6	90
Tendon Collagen	1200	20	39	1.5	12	4.1	90
Dragline Silk	10000	160	1100	4	30	160	35

Table 2

Amino acid composition, in mole percent, of purified Bc-CPs and egg capsule, showing large differences in Cys and Lys content compared to α-keratin[39]. The maximum standard deviation is ~5% of the values shown.

Amino Acid	CP-1	CP-2	CP-3	Bulk Egg Capsule	α-Keratin
Ala	9.0	6.4	6.3	7.5	7.7
Arg	4.7	4.2	4.1	3.9	7.9
Asx	10.1	10.8	10.7	13.5	9.6
Cys	0.0	0.0	0.0	0.0	6.0
Glx	18.6	16.1	16.7	15.6	16.9
Gly	5.1	5.9	6.3	8.0	5.2
His	2.2	3.4	3.7	1.5	0.6
lie	5.9	5.4	6.1	4.9	3.8
Leu	9.5	9.1	8.6	9.5	10.2
Lys	12.0	12.2	12.6	9.4	4.1
Met	0.0	0.0	0.0	0.8	0.6
Phe	2.4	4.1	5.6	4.6	2.0
Pro	0.9	2.9	1.6	1.9	3.3
Ser	7.2	7.9	7.5	8.0	8.1
Thr	5.0	4.7	3.5	5.0	4.8
Tyr	0.3	0.5	0.0	0.4	2.7
Val	7.1	6.4	6.7	5.5	6.4
Total	100	100	100	100	100

Table 3

Amino acid compositions and molecular masses of Bc-CPs. Observed values are obtained with ninhydrin-based analysis and MALDI-ToF mass spectrometry, whereas predicted values are derived from translated cDNA sequences. For the observed values, the maximum standard deviation is ~5% of the values shown.

	CP-1a		**CP-1b**		**CP-2**		**CP-3**	
	Observed	**Predicted**	**Observed**	**Predicted**	**Observed**	**Predicted**	**Observed**	**Predicted**
Ala	9.0	8.5	9.0	8.9	6.4	5.7	6.3	5.6
Arg	4.7	4.5	4.7	4.5	4.2	3.1	4.1	3.3
Asx	10.1	11.5	10.1	11.6	10.8	12.9	10.7	11.8
Cys	0.0	0.7	0.0	0.7	0.0	0.4	0.0	0.4
Glx	18.6	16.2	18.6	16.9	16.1	14.8	16.7	14.8
Gly	5.1	4.9	5.1	4.9	5.9	5.9	6.3	5.9
His	2.2	1.9	2.2	2.1	3.4	3.7	3.7	3.7
Ile	5.9	7.5	5.9	6.8	5.4	6.6	6.1	6.5
Leu	9.5	9.6	9.5	9.9	9.1	9.4	8.6	9.3
Lys	12.0	10.6	12.0	10.4	12.2	11.4	12.6	11.5
Met	0.0	0.7	0.0	0.7	0.0	0.7	0.0	0.7
Phe	2.4	2.1	2.4	2.1	4.1	5.9	5.6	6.3
Pro	0.9	0.5	0.9	0.2	2.9	0.9	1.6	0.9
Ser	7.2	8.2	7.2	7.8	7.9	7.9	7.5	8.8
Thr	5.0	5.4	5.0	5.6	4.7	3.9	3.5	3.9
Trp	-	0.0	-	0.0	-	0.2	-	0.2
Tyr	0.3	0.0	0.3	0.0	0.5	0.0	0.0	0.0
Val	7.1	7.1	7.1	6.8	6.4	6.6	6.7	6.5
Total	100	100	100	100	100	100	100	100
	CP-1a		CP-1b		CP-2		CP-3	
	Observed	Predicted	Observed	Predicted	Observed	Predicted	Observed	Predicted
Mass (Da)	47461	46880	47461	46962	51550	51437	52129	52016