The Journal of Experimental Biology 210, 4307-4318
Published by The Company of Biologists 2007
doi:10.1242/jeb.009753

Holdfast heroics: comparing the molecular and mechanical properties of *Mytilus californianus* byssal threads

Matthew J. Harrington* and J. Herbert Waite

Department of Molecular, Cellular, and Developmental Biology, University of California at Santa Barbara (UCSB), Santa Barbara, CA 93106, USA

*Author for correspondence (e-mail: harringt@lifesci.ucsb.edu)

Accepted 2 October 2007

Summary

The marine mussel *Mytilus californianus* Conrad inhabits the most wave-exposed regions of the rocky intertidal by dint of its extraordinary tenacity. Tenacity is mediated in large part by the byssus, a fibrous holdfast structure. *M. californianus* byssal threads, which are mechanically superior to the byssal threads of other mytilids, are composed almost entirely of a consortium of three modular proteins known as the preCols. In this study, the complete primary sequence of preCols from *M. californianus* was deduced and compared to that of two related species with mechanically inferior byssal threads, *M. edulis* Linnaeus and *M. galloprovincialis* Lamarck in order to explore structure–function relationships.

The preCols from *M. californianus* are more divergent from the other two species than they are from one another. However, the degree of divergence is not uniform among the various domains of the preCols, allowing us to speculate on their mechanical role. For instance, the extra spider silk-like runs of alanine-rich sequence in the flanking domains of *M. californianus* may increase crystalline order, enhancing strength and stiffness. Histidine-rich domains at the termini, in contrast, are highly conserved between species, suggesting a mechanical role common to all three. Mechanical testing of pH-treated and chemically derivatized distal threads strongly suggests that histidine side chains are ligands in reversible, metal-mediated cross-links *in situ*. By combining the mechanical and sequence data, yield and self-healing in the distal region of threads have been modeled to emphasize the intricate interplay of enthalpic and entropic effects during tensile load and recovery.

Key words: mussel, byssus, self-healing, histidine–metal, collagen.

Introduction

The key to survival on rocky wave- and wind-swept seashores for many mussel species is a beard-like array of tethering attachment threads known collectively as the byssus, which is used to secure the mussel to the hard substratum against the lift and drag forces of waves (Yonge, 1962). In addition, the byssus is emerging as an important paradigm for the bio-inspired engineering of water-resistant adhesives (Lee et al., 2006), functionally graded polymers (van Hest and Tirrell, 2001; Waite et al., 2004), and injection molding of liquid crystals (Hassenkam et al., 2004). The utility of such paradigms relies on the depth to which structure–function relationships are understood. A particularly effective strategy for defining structure–function relationships is to investigate homologous structures in closely related species (Hayashi et al., 1999; Brooks et al., 2005). With this report, byssal thread mechanics and the protein chemistry of three species of *Mytilus* are compared. Of these, the byssus of the California mussel *Mytilus californianus* (Conrad), stands out as mechanically superior.

The threads that make up the byssus of *M. californianus* are several centimeters long with a diameter of about 200 μm, function outside the body of the organism, and are formed from soluble protein precursors secreted by the foot of the mussel. Byssal threads are subdivided into four morphologically and mechanically distinct regions: the stem, the plaque, and the proximal and the distal portions of the thread (Fig. 1). The stem attaches the thread to the mussel tissue, and the plaque contains the adhesive that connects the thread to the hard substratum. The thread connecting the stem to the plaque is further divided into two mechanically distinct regions. The proximal end (closest to the organism) is extensible up to 200% of its original length, has a low initial stiffness, and a corrugated appearance. The distal portion of the thread, in contrast, is characterized by a high initial stiffness followed by a yield point at about 15% strain and a noticeable stress softening (Bell and Gosline, 1996). For most engineered polymeric materials, yield is not reversible and leads to permanent deformation rendering the material functionally useless. However, in distal threads, damage due to yield is reversible in a time-dependent, self-healing manner with threads recovering 25% of the lost modulus and strain energy in 10 min following a cycle to 35% strain (Carrington and Gosline, 2004).

Thread formation begins when the mussel foot touches down on a surface it finds suitable for attachment (Waite, 1992). Once

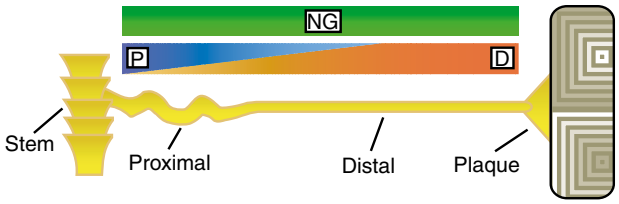

Fig. 1. Schematic of an isolated byssal thread showing important structural features. Morphological and mechanical differences graded from distal to proximal correspond with a gradient in the relative composition of preCol variants D, NG and P.

committed, soluble thread precursors are secreted into the ventral groove of the foot where muscular contractions mold them into functional threads. The whole process takes a total of 2–5 min. The main protein component preCol, also known as byssal collagen, makes up 96% of distal thread and 66% of the proximal (Waite et al., 2002). There are three preCol variants, each of which is divided into several distinct domains with particular motifs resembling known structural proteins.

Each preCol has a central collagen domain, which deviates from fibrillar collagens (Types I, II, III, V and XI) in containing 1–5 interruptions in the canonically repeated Gly-X-Y collagen sequence. These substitutions and deletions are believed to cause bends in the normal rod-like shape of the collagen trimer as exhibited in the C1q heterotrimer (Kilchherr et al., 1985), and several other invertebrate collagens (Sicot et al., 1997; Fowler et al., 2000). It has been experimentally demonstrated that proline or hydroxyproline in the X and Y position produce the most stable triple helix and all other amino acid residues have a tendency to destabilize the helix (Persikov et al., 2005). In the preCol collagen domain the X and Y residues tend to be destabilizing residues. Despite this and the breaks in the repeat structure, the presence of a stable collagen triple helix is suggested in threads by wide angle X-ray scattering of distal threads (Mercer, 1952; Rudall, 1955) as well as by the resistance of the byssal collagens to pepsin treatment (Qin and Waite, 1995).

At either end of the collagen region are flanking domains that differ between the three variants and resemble motifs of known load-bearing proteins. PreCol D, which is dominant in the distal end of the thread and decreases in abundance axially toward the proximal end (Fig. 1), has flanking domains that resemble a motif sequence of spider dragline silk. PreCol P, which exists in a gradient complementary to preCol D, has flanking regions that closely resemble elastin, and preCol NG, which is uniformly present throughout the thread, has Gly-rich flanking domains that resemble plant cell wall proteins (Waite et al., 1998). Since the mechanical properties of dragline silk (as in preCol D) differ substantially from elastin (as in preCol P), the graded distribution of preCols with different flanking regions is suspected to play a major role in the graded mechanical properties along the distal to proximal axis (Waite et al., 2002).

Beyond the flanks, at the N- and C-termini are regions with abundant histidine and 2–4 residues of dihydroxyphenylalanine (DOPA), which is a post-translational modification of tyrosine common to other byssal proteins in the plaque and outer coating of the thread. The histidines are proposed to be reversible ligands for coordination bonds with transition metal ions such as zinc(II) and copper(II), as demonstrated with shorter synthetic peptides (Waite et al., 1998; Schmitt et al., 2000). Such reversibly broken coordination complexes are stronger than non-covalent bonds, but possess only half the strength of a covalent bond (Lee et al., 2006) and are suspected to play an integral role as sacrificial bonds in yield and self-healing in the distal portion of the thread. DiDOPA cross-links have been proposed as covalent bonds linking preCols end to end (Waite et al., 2002).

Many species of mussels produce a byssus, but the component threads of byssi from different species are not mechanically equivalent (Brazee and Carrington, 2006). There are five species in the genus *Mytilus* including *M. galloprovinicialis* Lamarck, *M. edulis* Linnaeus, and *M. californianus* Conrad (Gosling, 1992). *M. californianus* is phylogenetically distant from *M. galloprovincialis* and *M. edulis* and is readily distinguishable from them (Santaclara et al., 2006). Whereas the mechanical properties of the threads of *M. galloprovinicialis* and *M. edulis* are very similar to one another (Lucas et al., 2002), *M. californianus* has much thicker and mechanically superior threads. *M. californianus* distal threads are 2–3 times stiffer and 30% more extensible than distal threads from the other two species (Bell and Gosline, 1996). Distal threads from *M. californianus* also self-heal more quickly than those of *M. edulis* and *M. galloprovincialis* after yield (Carrington and Gosline, 2004). These mechanical differences are probably related to adaptations in *M. californianus* that help it survive in the more exposed, wave-swept parts of the intertidal. Interestingly, these differences appear limited to the distal region since mechanical testing of the proximal region has not revealed significant differences in modulus, strength or extensibility between the three species (Bell and Gosline, 1996).

cDNA sequence has been deduced for all three preCols from *M. galloprovinicialis* and *M. edulis* (Qin et al., 1997; Coyne et al., 1997; Qin and Waite, 1998; Lucas et al., 2002), but not for *M. californianus*. The purpose of this study was twofold: (i) to determine the complete primary sequence of collagenous preCols from *M. californianus* to enable a direct comparison with those from other species, and (ii) to reconcile both divergent and conserved features of preCol biochemistry with the comparative mechanical performance of the distal portion of the byssal thread. The results suggest that the single most influential factor in the tensile superiority of *M. californianus* is the greater abundance of silk-like polyalanine tracts and that the highly conserved histidine-rich domains play an integral role as sacrificial bonds in threads of all three species.

Materials and methods

RACE ready cDNA construction

Whole mussel feet were dissected from *Mytilus californianus* Conrad collected from Goleta Pier (Goleta, CA, USA) and stored in flowing seawater (SW) at 12–15°C. Foot tissue was immediately homogenized under liquid nitrogen, and total RNA was extracted with the RNeasy Plant Mini Kit from Qiagen using the manufacturer's protocol (Qiagen, Valencia, CA, USA). Freshly purified RNA was reverse transcribed to 3′ and 5′ RACE ready cDNA using the SMART RACE cDNA Kit (Clontech, Mountain View, CA, USA).

Table 1. *PCR reactions and primers used to determine the cDNA sequence of M. californianus preCols*

PreCol D	5′RACE/DR-1, DF-2/DR-2, DF-3/DR-3
McDR-1	5′ TCCGCCACCTACAACTGGACC 3′
McDF-2	5′ GCAGCAGCAAACGCAGCAGCAGGAGGAT 3′
McDR-2	5′ CTCCGGCTGTTCTGAGGTCTTCAAT 3′
McDF-3*	5′ GTGGTATGGGTAGACGAG 3′
McDR-3*	5′ AACACTTGCAGATTTTATTGATA 3′
PreCol NG	5′RACE/NGR-1, NGF-2/NGR-2^{dg}, NGF-3/NGR-3, NGf-4/NGR-4
McNGR-1	5′ AAGTGGGAATCCGCCGCCACCAGTTGC 3′
McNGF-2	5′ GCGGCTATCGCCCGCACAGGACTA 3′
McNGR-2^{dg}	5′ YTGWCCWGGDGCWCCWCCDGCATTWCCWGGTTGG 3′
McNGF-3	5′ GCAACTGGTGGCGGCGGATTCCCACTT 3′
McNGR-3	5′ CCACCAGCCTCTCCTGCTGGTCCTTCT 3′
McNGF-4*	5′ AAGGAGAACTTGGACCAGTCG 3′
McNGR-4*	5′ AGGAACTTGCACTTTTTAT 3′
PreCol P	5′RACE/PR-1, PF-2/PR-2, PF-3/3′RACE
McPR-1	5′ GGATCAGGCAGTGGCAGAGCATTTGGTGG 3′
McPR-2*	5′ GAATAACACCTGGTGCTCCT 3′
McPF-2*	5′ GAGGATTCGGTGGACCAGGTAC 3′
McPF-3	5′ GGGACCAGGAGGTGAAAGAGGAGGCCAAGG 3′

*Primers from *M. edulis*. McDR-3 and McNGR-4 are based on the 3′ UTR.
^{dg}Degenerate primer.

PCR amplification and cloning of the preCols

Primers designed for the preCols from *Mytilus edulis* (Lucas et al., 2002) and degenerate primers based on highly conserved sequence in *M. edulis* and *M. galloprovincialis* were used to screen the *M. californianus* cDNA, assuming some sequence similarity. Preliminary sequence obtained with these primers was used to create gene-specific primers used in PCR, 5′ RACE and 3′ RACE to obtain the remaining sequence. Regions of sequence where degenerate primers or primers for *M. edulis* landed were confirmed by overlap with sequence-specific primers. The universal primer mix supplied with the RACE kit (Clontech) was used for 5′ and 3′ RACE. The relevant reactions are described in Table 1.

PCR and RACE products were cloned into Pgem-T Easy plasmid (Promega, Madison, WI, USA) and transformed into One Shot chemically competent *E. coli* cells (Invitrogen, Carlsbad, CA, USA). Miniprepped plasmids were contractually sequenced (UC Davis DNA Sequencing Facility) and overlapping sequences were assembled and translated to give the deduced protein sequences. *M. californianus* preCol cDNA sequences were entered into GenBank under accession numbers EU120661, EU120662 and EU120663 for preCol D, NG and P, respectively. The deduced protein sequences from the cDNA were compared to preCol sequence from *M. galloprovincialis* and *M. edulis* in the database using ClustalW on the EBI server (Chenna et al., 2003) to determine percent identity between specific preCol variants of the three species. Alignments were performed with the identity matrix using default settings for all the other parameters.

Mechanical testing

Distal portions of fresh tank-grown threads were dissected from *M. californianus* and allowed to rest in seawater for at least 48 h prior to testing. Thread ends were secured between pieces of cardstock with cyanoacrylate glue. The thread cross-section was assumed to be slightly elliptical, and the area was calculated using the average of the long and short side as the diameter. Cardstock loaded threads were placed in the grips of a Bionix 200 tensile tester (MTS, Eden Prairie, MN, USA) and subjected to mechanical testing. Threads were hydrated by submersion in seawater or buffer immediately prior to tensile testing, which was performed within an MTS environmental chamber with the relative humidity raised to 99.9±5% in order to maintain thread hydration. Testing at high humidity was necessary to reduce slippage of the thread from the cardstock at higher strains, which was seen in test runs with submerged threads. Threads were cycled consecutively to 10, 20, 30, 40, 50, 60% strain, and finally to break, at an extension rate of 5 mm min^{-1} with no rest between, recording the extension and load on MTS Testworks 4 software. Extension was converted to engineering strain using the equation $\epsilon=(l-l_o)/l_o$ where l is the extension of the thread during tensile loading and l_o is initial length of the unloaded thread with zero load. Load data were converted to engineering stress using the equation $\sigma=F/A$, where F is the tensile force applied to the thread and A is the cross-sectional area of the thread. Young's modulus (E, stiffness) for each strain interval was determined as the maximum slope of the stress–strain curve during loading after the characteristic toe region. To assess the stress softening incurred with each successive yield, the modulus of each cycle is presented as a percentage loss of the initial stiffness ($\%\ \text{loss}=[1-E/E_o]\times100\%$) where E is the stiffness of a particular cycle and E_o is the stiffness measured during the first cycle of a previously unstrained thread.

Histidine–metal coordination cross-links are suspected to play an important role in yield and hysteresis in the distal threads. Such bonds, however, are known to be pH sensitive since they depend on the deprotonation of a histidine nitrogen, with a pK_a ~6.5 (Fig. 2). In order to investigate the effect of pH

on mechanical performance and, by extension, on histidine–metal cross-links, threads were treated in citrate-phosphate buffer at pH 3, 4, 5, 6, 7 and 8 at 4°C for at least 24 h prior to testing. A subset of threads was equilibrated in citrate-phosphate buffer at pH 5.5 for 24 h and treated with diethylpyrocarbonate (DEPC) by adding an ~tenfold molar excess (65 μmol l^{-1}) to the buffered solution (molar excess is based upon estimated concentration of His in a thread). In moderate excess, DEPC reacts with histidine residues by carbethoxylating one of the nitrogen atoms in the imidazole ring, rendering it unable to participate in metal coordination complexes (Fig. 2). The reaction was performed at pH 5.5 since DEPC targeting of histidine residues *vis-à-vis* other amino acids is highly specific there (Lundblad, 2005). Although metal-bound histidine is protected from attack by DEPC, at pH 5.5 there will be some unbound, exposed histidine residues. DEPC-treated threads and a control group of untreated threads at pH 5.5 were then re-equilibrated 24 h in pH 8 buffer prior to mechanical testing. One-way analysis of variance (ANOVA) was performed to compare differences in Young's modulus between treated threads (pH and DEPC) and untreated threads (SW ctrl) during their first strain cycle. A two-way ANOVA was performed to assess the effects of pH and strain on the Young's modulus (E) of threads during multiple cycles of increasing strain.

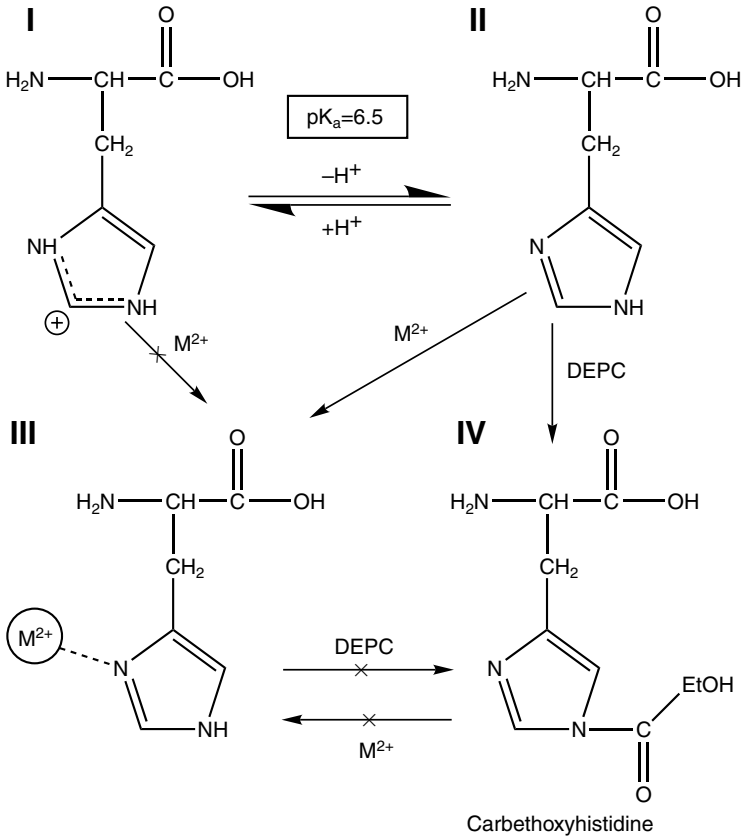

Fig. 2. Schematics of histidine chemistry concerning protonation, metal coordination and reaction with diethyl pyrocarbonate (DEPC). (I,II) The titration of a proton on the imidazole ring of His has a pK$_a$ ~6.5. The nitrogen shown losing the proton in the figure is the more likely of the two since it has slightly lower pK$_a$; however, depending on the local environment, either can be deprotonated. As shown in (III), it is possible for the deprotonated nitrogen to contribute a lone pair of electrons as a ligand in a metal coordination bond with a divalent transition metal ion such as Zn^{2+} or Cu^{2+}. When both nitrogens are protonated (I), the ring is unable to participate as a ligand in metal bonding. Reacting unbound His residues with DEPC (IV) results in a carbethoxylation of the imidazole ring, rendering it unable to participate in coordination metal chemistry. Metal bound His will not react with DEPC. DEPC treatment is specific for His at pH 5.5 and at this pH, there will be few side chains bound to metal. EtOH, ethanol.

Results

Sequence of M. californianus preCols

The complete cDNA sequence generated from PCR, 5′ RACE, and 3′ RACE was acquired for each of the three preCol variants, each consisting of 3–4 overlapping clones. Figs 3–5 show the full deduced protein sequences for preCol D, NG and P from *M. californianus,* respectively, with relevant domains labeled for each. We are confident that the assemblies represent complete cDNAs of expressed proteins. Table 2 shows the sequence identity to the preCols from *M. galloprovinicialis* and *M. edulis*. The collagen region, signal sequence and acidic patch from each preCol are highly conserved (83–100%) and are part of different overlapping constructs, giving us confidence in our overall construct composition.

In *M. californianus*, the preCol variants are all slightly larger proteins than their counterparts in *M. galloprovinicialis* and *M. edulis*. Over the length of the whole protein, each particular variant from *M. edulis* is around 90% identical to the same variant in *M. galloprovinicialis*, whereas *M. californianus* variants are only 80% identical to their counterparts (Table 2). It should be noted that the alignment scores indicated in Table 2 represent the number of identities in the best match as determined by ClustalW divided by the number of residues aligned, and therefore do not account for inserts. Domains where the score does not accurately reflect the degree of divergence between species due to inserted sequence include the flanking and His-rich domains. This will be discussed later. All three collagen domains conserve the tendency for destabilizing residues in the X and Y positions of the canonical collagen repeat.

PreCol D

Table 2 reveals that the collagen domain of preCol D is more conserved than the silk or His-rich regions. However, there is one significant difference in the collagen that will be potentially important to the function and assembly of the protein. As mentioned, the preCol collagen regions are distinct from typical fibrillar collagens in containing one to several deviations from the Gly-X-Y repeat motif, which typically lead to kinks in the characteristically rod-like collagen morphology (Kilchherr et al., 1985). In preCol D of *M. californianus* there are three aberrations in the collagen domain instead of the five seen in both *M. galloprovinicialis* and *M. edulis* (Fig. 3). Two of the aberrations are typical deletions of a single Gly also seen in *M. galloprovinicialis* and *M. edulis*, but they surround a novel aberration not seen in the other two species in which the sequence Val-Val-Gly-Gly is inserted between Gly-X-Y

Signal MVFKILTVCLVASLLETCLA

His-rich D**Y**GRK**Y**GKPS**Y**GE**Y**GGKRGGGRVSGAVA**H**A**H**A**H**A**H**ASSGADGRSRA**H**ARAVA**H**S**H**SGGGAA**H**G**H**PGFP

Silk AGSASAAARAAARAGAVGGFGGGGFGSASAAAAARAAAAGGGLGGGFGSASAAARAAAAAGGFGGGGFGSA
SANAAANAAAGGFGGFGGGFGPGF

Collagen GGP GQP GGP GGP GGP GGP GGP GMP GGP GGP GGP GMP GGP GGP GGP GTG GPG *QP
GGL GGP GGP GMP GGP GGP GGP GMP GGP GGP GGP GGP GGP VVGG GIP GIT GPA
GPP GPA GPQ GPQ GEQ GPR GRT *PA GTP GPP GNP GEP GQA GAP GAP GAP GHG GKP
GTA GAT GKA GRP GPR GQP GAS GSS GQH GAS GAP GRP GNP GST GRP GAT GDP GRP
GAT GTA GRP GPS GAP GNP GAP GAI GAP GPR GAP GFV GLP GPR GSP GEP GNQ GPI
GGP GYP GPT GPQ GPD GAM GPQ GPC GNR GAP GVP GKQ GPV GGQ GPA GPR GPR GDE
GPV GPK GEP GAK GAD GKP GDK GGD GDT GPQ GPA GPK GEV GDQ GAP GPK GET GDQ
GAR GEA GKA GEQ GPG GIQ GPK GPA GGQ GPA GPA GPL GPQ GPM GER GAQ GPT GPE
GPV GAP GPK GSV GDQ GAQ GDQ GAT GAD GKQ GEP GER GQQ GAA GPV GRP GPR GDR
GAK GIQ GSR GRP GAM GRR GNR GSQ GAV GPR GET GPD GNQ GQR GEQ GAP GVI

Acidic TLVIEDLRTAGVESPVEAFDA

Silk GAGPGGAGPFGGAGPFGGAGAGAGAGGAGPFGAGPFGAGPFGGVGGAGVGGAGPGGAAGPGGAAGPGGAAG
AGGLGGLGAGGLGGLGAGGLGGLGAGGLGGQGAGGLGGLGAGGLGGLGAGLGGGLGGGAGAAAAAQAAAAA
NGGLGGGSAAAAARAAAAANAGLGGGAVAAAQAAASAAANSGLGAGAARAAASAAARATVAGAGRGTAAAA
ASAAAQA**H**AATKAQGGS**H**A**H**AAAAAQAAA

His-rich SSVI**H**GGG**H**GG**H**GGD**YH**KPG**Y**

Fig. 3. Deduced protein sequence of preCol D from *M. californianus*. Histidine and tyrosine residues in the His-rich domains are in bold type. Breaks in the Gly-X-Y canonical repeat are underlined in the collagen domain and deletions are denoted with an asterisk.

repeats. This is the first instance of a four amino acid insert in a preCol collagen domain, but it is uncertain how the triple helix would be perturbed by such an insert.

The alignment scores of both the N- and C-terminal silk domains are low compared to the collagen domain (67% and 69% identity, respectively, *vs* 93% when compared with *M. edulis*), but still do not fully represent the degree of deviation since both are significantly longer in *M. californianus* due to inserts. The inserts come in the form of two alanine-rich regions of 24 and 10 amino acid residues, respectively, in the N terminus (Fig. 6A) and three inserts of different sizes in the C terminus. If the number of identical residues in the N-terminal silk domain of *M. edulis* and *M. californianus* is divided by the number of residues in the larger of the two (*M. californianus*=95, *M. edulis*=55, *M. galloprovincialis*=55) instead of the number of aligned residues, the percent identity drops from 70% to about 40%, indicating the large amount of deviation in the flanking domains of *M. californianus* not accounted for by the ClustalW alignment score.

The His-rich domains have a relatively low alignment score, but Fig. 6B reveals that while the sequence outside of the suspected metal binding motif is very divergent, the histidine containing sequence is highly conserved. Every single histidine in the N-terminal His-rich domain is conserved, along with Gly

Signal MVHNFLTVFLVAAIARTGLA

His-rich GSIG**Y**GKPG**Y**GSGGDFFNG**H**GG**H**GG**H**GGG**H**G**H**GGGGG

Gly-rich GSASAAAQAAAAAARALGGGGGGSASAAARAAAAAARALGGGGGGSASAAAQAAAAAARALGGVGGGSASAAAR
AAAAASALGGGGGGFGGLGGLGGGPGGIGGIGGGPGGLGGGPGGLGGGPGGFGIGPGGIGVGPGGNGGGPG
GLFGPGGALGPGGGAGAFGTGGAVGGPGGAGASASAGAFATGGGGFPLP

Collagen GAQ GPQ GPR GPA GAP GDQ GES GPP GPP GNS *PQ GPQ GSR GAP GQP GEQ GAN GNP
GQP GNA GAP GQP GAP GQA GAP GAR GPS GAA GHQ GAQ GGL GQP GSP GQQ GSA GQP
GSP GNP GAP GAP GST GQA GSV GNV GGP GEQ GPQ GNA GPR GIQ GRP GCK GLP GPK
GPD GAQ GAP GKP GAD GPA GNR GPM GPA GGK GPT GDK GAP GDV GPE GPE GKP GGP
GPK GPN GPQ GAK GSP GED GDP GAE GEP GSK GAD GLP GHG GPR GNP GPQ GPE GEV
GDK GAP GEA GGP GQP GPY GPQ GPA GEQ GDL GEI GPA GDA GAP GVP GSK GIQ GPQ
GEL GPV GKE GPA GEA GGK GRL GQK GPA GEP GQP GEE GKQ GDM GAT GTR GAT GVG
GVK GPT GFR GAR GKS GNA GRP GRP GRN GPR GPQ GPQ GLR GNQ GPD GEQ GGP GVG
GIS GTI

Acidic TIIVDDDGARYGDFADITGPNSDEVNRQLVREFIGDLDTFLSLNGPGGPAGV

Gly-rich GAGDIGIGGGLGPGGAGGIGPGGAGGPGGAGGVAGGPGGVGGFGVGPGGAAGLGGVGGLGAGLGGVGGLGA
GLGGVGGLGGGVGGLGAGLGGVGGVGGLGAGGALGGLGGLGFGGLGGGAGASAGAAAGA**H**AVSGGAGGGAS
A

His-rich **H**A**H**A**H**A**H**AASVSGGGGGVS**H**AVS**H**AVS**H**AVS**H**SVS**H**SGGGG**H**A**H**AAASA**H**A**H**AVS**H**GGSSGG**H**PL**HY**NDPF
YGKK**H**KAD**Y**

Fig. 4. Deduced protein sequence of preCol NG from *M. californianus*. Histidine and tyrosine residues in the His-rich domains are in bold type. Breaks in the Gly-X-Y canonical repeat are underlined in the collagen domain and deletions are denoted with an asterisk.

Signal MVRFSLASVLLLAVTSTAFA

His-rich GPLGD**Y**GGNGIKIVP**YH**GGGGGGGGGGGGG**H**GGSWS**H**GP**YH**R**H**GLGGGSGSGGSSA**H**A**H**SSASA**H**V**HH**FGS
GSS**H**ASAGSSANA**H**TFGGGPS**H**ASAGSSS**H**ASAS**H**SGLGGGSA**H**A**H**SSSSA**H**ALS

Elastin GGFGGFGGIGPGGIGGIGGGGLGGGPGGIGGIGGPGGIGGGGIGPGGIIFGPGGPGSASGSGSGRAFGGNG
GSSASANAAARASASGGGGF

Collagen GGP GSP GNA GAP GQP GIP GAP GVP GRP GIT *QP GRP GNA GPP GQP GNP GRP GAG
GRP GAP GRP GGP GRP GIP GKP GNL GQL GQP GSP GQP GHP GAS GQP GRN GSP GNP
GKP GTP GHS GTP GSR GSP GTP GTP GQP GVP GTT GGR GPR GPA GII GLI GPK GNP
GVA GNP GAP GIP GEG GPR GPQ GPA GGP GSS GPS GDK GSP GTP GGN GPR GAI GPQ
GPS GPP GDG GPQ GGR GTP GTP GKP GAK GPQ GSN GDV GPQ GAS GPK GPQ GPQ GKA
GVK GPA GDQ GAR GAE GTA GPP GPQ GEQ GLK GPS GGQ GPA GPA GPS GEQ GPG GER
GGQ GPQ GPE GPT GPA GPR GAS GSQ GPA GER GAP GAP GKK GPN GDR GNQ GLP GAP
GKN GAR GDR GAR GSN GSP GRT GAP GSR GKI GPQ GPH GPR GAR GSQ GPK GQR GDQ
GAP GVI

Acidic RIVIDDQRTGPEVEEF

Elastin PGFGGFPGFPGSAGAGSSSGASAFGGPGGAGGFPFGPFGGAGGPGGGPGGPGSPGGPGGPGGPGGPGGPGG
PGGPGGPGGLVGGGAGGPGGLGLGGAGGPGGLGGGAGGPGGFGGGVGPGGLGGFGGLGGGGLGGGASAGAS
SSGSASASGGGPFGVLNVGPGGGGSSSASAASRA

His-rich **H**ALALGGLGGGSASAGS**H**SSSSA**H**SFGG**H**LF**H**SVT**HH**GGPS**H**ASA**H**ATA**H**A**H**ASASGGGG**H**GG**H**GG**H**GG**H**G
GP**Y**KPG**Y**

Fig. 5. Deduced protein sequence of preCol P from *M. californianus*. Histidine and tyrosine residues in the His-rich domains are in bold type. Breaks in the Gly-X-Y canonical repeat are underlined in the collagen domain and deletions are denoted with an asterisk.

and Ala spacer residues. The C-terminal His-rich region is less conserved, but the number of His residues is mostly conserved with *M. galloprovincialis* having one extra histidine.

PreCol NG

Conservation of preCol NG domains in *M. californianus vis-à-vis* preCol NG from the other two species parallels the trends seen with preCol D. The collagen region shows the highest identity score outside of the acidic patch and the signal sequence (Table 2). The only aberration is a single deletion of a glycine in the eleventh repeat (Fig. 4), identical to that of *M. edulis*. *M. galloprovincialis* has two additional deviations in the form of substitutions of Gly with an Asp and an Arg, respectively.

The suspected His-rich metal binding motifs are less conserved in NG than in D, but are still relatively similar. The N-terminal His region of all three species has five histidine residues; however, in *M. californianus* the histidines are more closely packed. The C-terminal His-rich region is more conserved than the N-terminal, with 14 out of 17 histidine residues identical between all three species.

As with preCol D, the major differences are in the flanking domains (Table 2), especially the Gly-rich region within the C-terminal flanking domains, which contains numerous inserted regions rich in Gly, Val and Leu. Gly residues are mostly conserved between species, and the non-Gly residues exhibit variation but are always hydrophobic residues such as Ala, Leu

Table 2. *Percent identity of the preCol variants for M. californianus, M. edulis and M. galloprovincialis by domain*

	Identity (%)							
PreColvariant	Overall	Signal	N-His	N-flank	Collagen	Acidic	C-flank	C-His
D								
Mc:Me	82	85	67	70	93	95	69	66
Mc:Mg	79	85	65	52	87	95	72	100
Me:Mg	89	100	92	83	93	100	89	76
NG								
Mc:Me	83	100	80	79	89	90	71	70
Mc:Mg	81	95	83	85	89	90	68	62
Me:Mg	89	95	89	87	97	100	78	79
P								
Mc:Me	79	100	82	75	83	93	62	88
Mc:Mg	80	100	82	75	85	86	63	83
Me:Mg	91	100	84	91	94	93	91	93

Mc, *M. californianus*; Me, *M. edulis*; Mg, *M. galloprovincialis*.

Identity scores were determined by ClustalW and represent the number of identical residues in the best pairwise match, as determined by the program divided by the number of residues that were matched. This score does not take into account regions that do not overlap due to inserted sequence.

and Val. The introduction of considerably more Pro in the flanking domains of NG in *M. californianus* than in the other two species is notable since it could lead to a more amorphous secondary structure.

PreCol P

Of the three variants, preCol P has the least conserved collagen domain at 83% identity with *M. edulis* (Table 2). However, the single Gly deletion in the eleventh Gly-X-Y repeat (Fig. 5) is conserved in all three species. The N-terminal His-rich domain of P is also the most altered of the three variants in terms of the position and number of His residues, but is still mostly conserved.

The flanking domains of preCol-P are again the least conserved regions with the C-terminal end having a 20% lower identity than the N-terminal. They are not as insertionally modified as the flanking domains of D and NG, but are more substituted. The C-terminal region is especially different and reminiscent of the flanking domains of preCol-NG in certain regions, with runs rich in Gly and Leu and many PGG repeats. A BLAST of the C-terminal flank sequence against the database shows it is over 50% similar to flagelliform spider silk whereas it is only about 60% similar to the C-terminal flanks of *M. edulis* and *M. galloprovincialis*.

Mechanical testing

Cyclic quasistatic mechanical testing (Carrington and Gosline, 2004) showed that the larger the strain that distal threads are pulled to, the larger the resulting hysteresis. This study also revealed that the stiffness disparity (stress softening) between the first and second cycles increased with increasing strain of the first cycle. Both observations imply that damage is occurring at the molecular level during yield and is intimately correlated with hysteresis and reduction of modulus. The mechanical tests performed in this study were designed to probe these observations more thoroughly and to investigate more specifically the role of histidine–metal coordination complexes as important sacrificial cross-links in tensile stress.

As Fig. 7A indicates, a distal thread incubated in seawater cycled to any strain within the yield region will have a second yield point if it is taken to a strain slightly beyond the maximum strain it achieved in the previous cycle. After each successive yield, threads show a reduction in modulus from that of the previous cycle. If the softening is presented as a percentage loss of the initial stiffness (% loss=$[1–E/E_o]\times100\%$) and plotted against the strain value that caused the reduction in stiffness (Fig. 7B), a few things become apparent. First, there is little softening observed during the 10% strain cycle, which is before the yield point typically occurs. The largest reduction in the stiffness occurs during the 20% strain cycle with the initial modulus reduced by almost 40% from the first cycle. Smaller reductions are seen after the 30% and 40% strain cycles, after which there does not appear to be further significant loss in modulus.

Threads incubated in pH 8 buffer are not statistically different from seawater threads ($P<0.01$). Treating threads with buffers at acidic pH, e.g. at pH 5 and below, reduces the modulus to

A

```
Mc -AGSASAAARAAARAGAVGGFGGGGFGSASAAAAARAAAAGGGLGGGFGSA
Me VGGSASAAARAAARASAGG-------------------------LGGFG---
Mg IGGSASAAARAAARASAGG-------------------------LGGLGGFV

Mc SAAARAAAAAG-GFGGGGFGSASANAAANAAAGGFGGFGGGFGPGF
Me SAAANAAAAARAGAGFGGFGGL----------GGFGGLGGV-----
Mg SAAANAAAAANAGAGFGGFG-------------GFGGLGGV-----
```

B

```
Mc DY-GRKYGKPSYGEYGGKR--GGGRVSGAVAHAHAHAHASSGADGRSRAHA
Me DYNGNKQYGGRYGNRYGNGLGGGNGGAGAVAHAHAHAHASAGANGRARAHA
Mg DYHGNKQYGGRYGNRFG----GGIGGAGAVAHAHAHAHASAGANGRARAHA

Mc RAVAHSHSGGGAAHGHPGFP
Me RALAHAHAGGGAAHGHPGFP
Mg RALAHGHAGGSAAHGHPGFP
```

Fig. 6. ClustalW alignment of (A) the N-terminal silk domain and (B) histidine-rich domain of preCol D from *M. californianus* (Mc), *M. edulis* (Me) and *M. galloprovincialis* (Mg). Silk alignment reveals that *M. californianus* has two major inserts consisting of polyalanine runs in the N terminus. His-rich domain alignment reveals that while overall the identity between *M. californianus* and the other species is low, the regions containing histidine are more conserved. Residues in bold type are conserved between all three species.

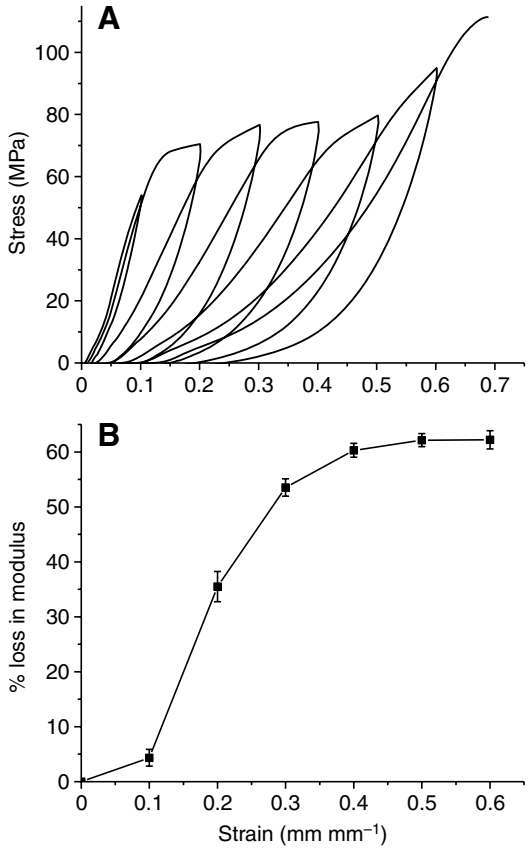

Fig. 7. (A) Representative stress–strain curve of a single distal thread cycled to incrementally increasing strain values from 10% to 70%. Yield, as seen in cycles to 20, 30 and 40%, is followed by a loss in stiffness (stress softening) in the subsequent cycle. (B) Percentage loss in modulus plotted as a function of the strain that caused the loss. Values are means ± s.e.m.; $N=10$. Most stress softening is occurring between 10% and 40% strain.

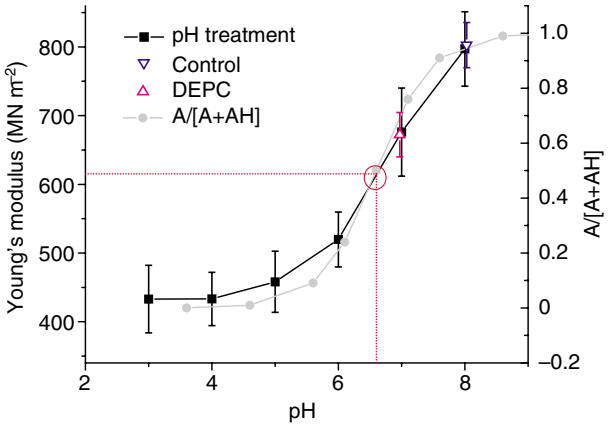

Fig. 8. Initial modulus of distal byssal threads plotted against the pH at which they were incubated. Red dotted line and circle denote the midway point of the Young's modulus and the corresponding pH of 6.6. The titration of a nitrogen proton on the histidine imidazole ring (pK_a=6.6) is plotted in light gray (A/[A+AH]) to illustrate that the modulus curve closely follows the characteristic sigmoidal shape, suggesting that histidine protonation may influence thread mechanics. The blue inverted triangle represents the initial stiffness of threads treated at pH 5.5 and brought back to pH 8 and tested, indicating that pH-induced softening is reversible. The pink triangle represents the initial stiffness of threads treated in DEPC at pH 5.5, brought back to pH 8, and then tested. DEPC treatment further supports the role of His in the mechanical properties of threads. Values are means ± s.e.m.; *N* ranged from 6 to 9 threads for each treatment.

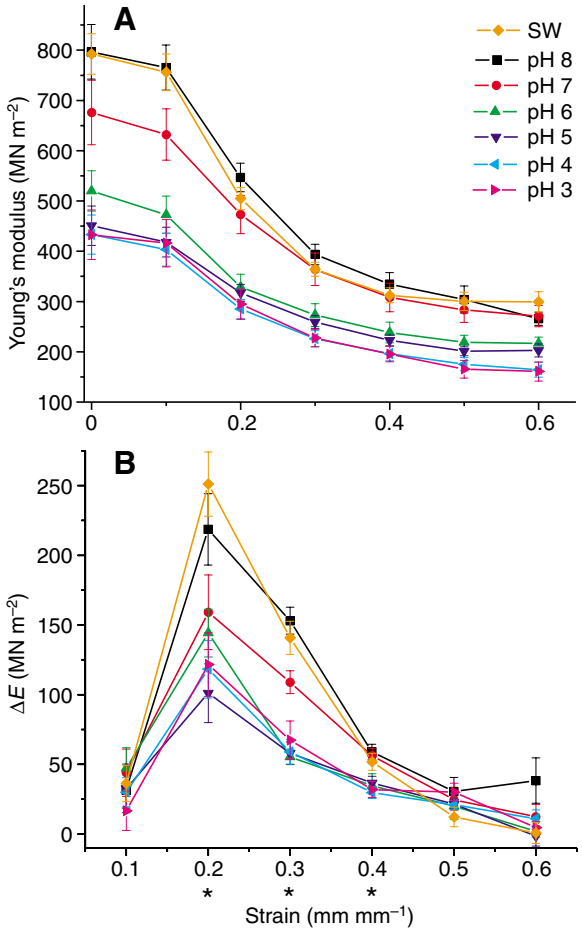

Fig. 9. (A) Young's modulus plotted against the previous strain cycle value for each pH treatment. Threads treated at pH 5 and below show a pH-induced reduction in modulus that becomes less prominent as the thread is cycled to higher strain values. (B) Reduction in modulus (Δ*E*) between strain cycles in Fig. 9A plotted against the strain that caused it for each of the pH treatments. Asterisks indicate where Δ*E* at pH 5 is significantly different from Δ*E* at pH 8 (P<0.01). Values are means ± s.e.m.; *N* ranged from 6 to 10 threads for each treatment.

almost half the value measured in seawater and pH 8. In Fig. 8, the initial modulus (mean ± s.e.m.) for each treatment is plotted against pH, revealing a roughly sigmoidal curve with a midpoint at about 615 MN m^{-2} and a pH of 6.6, as predicted in a preliminary report (Waite et al., 2006). The stiffness *vs* pH curve follows the same trend as the histidine titration curve that is plotted alongside it in light gray in Fig. 8.

The initial stiffness of threads incubated at pH 5.5 for 24 h and then re-equilibrated at pH 8 for 24 h are statistically indistinguishable from citrate-phosphate pH 8 and seawater threads (P<0.02), which demonstrates that the pH-induced reduction of stiffness is reversible (Fig. 8). Threads incubated at pH 5.5 in the presence of DEPC, and reequilibrated in citrate-phosphate buffer at pH 8 do not show this same reversibility and have a significantly different initial modulus from the control threads that went from pH 5.5 to pH 8 without DEPC treatment (P<0.02). As Fig. 8 shows, the initial modulus of DEPC-treated threads falls around 680 MN m^{-2}, which is slightly below the mean value for threads treated in buffer at pH 7.

Plotting the modulus of each strain cycle against the preceding strain value for each pH treatment (Fig. 9A) reveals that the pH-dependent stiffness variation seen in Fig. 7 becomes less prominent with increasing strain (two-way ANOVA, P<0.001). However, the stiffness values of pH 8 and pH 5 treated threads are still significantly different at all strains tested (P<0.01). Fig. 9B shows the magnitude of the modulus reduction (Δ*E*) caused by a particular strain cycle plotted against the strain that caused it, revealing that for all pH treatments the majority of damage occurs between 20–30% strain (consistent with Fig. 7B). Statistical analysis of Δ*E* values at pH 5 and pH 8 reveals that softening due to cycles of 20, 30 and 40% strain are significantly different (P<0.01) whereas values at 10, 50 and 60% are not. It should be noted that treatment at pH 5 and below does not eliminate stress softening entirely, but it does greatly reduce the magnitude.

Discussion

Comparing the primary sequence of the preCols from the mechanically superior byssal threads of *M. californianus* to those of *M. galloprovinicialis* and *M. edulis* provides a unique opportunity to examine the structure–function relationship between preCol biochemistry and the mechanical properties of the byssal thread. Byssal collagen is a modular protein with clearly defined domains resembling commonly found load-bearing protein motifs. A comparative approach reveals that the degree of sequence deviation is not uniform throughout the whole protein, but instead varies from domain to domain.

The collagen domain, signal peptide and acidic patch are the most highly conserved regions between all three species with sequence identity between 83–100%. The His-rich domains have relatively low identity scores (77.3±11.3%), but closer examination of the whole domain reveals that the putative metal binding motifs are very well preserved. For all three variants the largest sequence deviations are localized in the flanking domains (mean identity score=70.0±8.5%), which are highly substituted in *M. californianus*. In every case except for the N-terminal flank in preCol P, *M. californianus* flanking domains are also significantly longer due to multiple insert regions.

Flanking domains

As stated earlier, the gradient of the variants is believed to be responsible for the graded stiffness from distal to proximal, and the flanking domains are believed to play a major role in these mechanical differences because of their similarity to known structural proteins of varying stiffness (Waite et al., 2002). Thus, significant changes in the sequence of these domains in a particular variant between species should affect the stiffness of the threads. Spiders use a similar strategy to create a toolkit of silks with a range of mechanical properties by adjusting the presence and amount of certain modules associated with stiffness or elasticity. For example, polyalanine runs are associated with beta sheet formation and stiffening of the silk, whereas GPGXX sequences are thought to confer elasticity (Hayashi and Lewis, 1998). In this light, the extra polyalanine runs in the silk-like flanking domains of preCol D of *M. californianus* could stiffen the molecule by adding more beta sheet crystal structure, which could in turn stiffen the whole distal portion of the thread. Evidence for beta-sheet structure in byssal threads has been seen in wide angle X-rays of the distal region (Mercer, 1952; Rudall, 1955). In contrast, the increased proline content of the NG flanking domains could reduce the amount of secondary structure and add more entropic elasticity to the region.

PreCol P also shows significant changes in the C-terminal flanking domain including Gly- and Leu-rich runs reminiscent of preCol NG and a $[GGP]_{12}$ repeat motif not present in the other two species. A BLAST of the sequence reveals it has over 50% similarity to flagelliform silk rather than elastin as in *M. galloprovincialis* and *M. edulis*. Although the flanking domains of preCol P in *M. californianus* are significantly different from those of the congeners, previous studies have shown that the mechanical properties of the proximal region do not vary between the three species (Bell and Gosline, 1996). It is worth recalling that preCols make up only 66% of the proximal region (Waite et al., 2002). Consequently, contributions of proteins such as ptmp-1 (Sun et al., 2002) in the remaining 34% could significantly obscure detection of the mechanical effect of subtle variations in preCol P.

Collagen domain

The collagen domains of the three variants are highly conserved between all three species, maintaining both the presence of destabilizing residues and aberrations in the Gly-X-Y repeat sequence. Fibrillar collagens (types I, II, III, V and XI) show no tolerance for breaks in the Gly-X-Y repeat structure; however, breaks regularly occur in non-fibrillar collagen (types IV and VIII) (Soininen et al., 1987; Brazel et al., 1987; Yamaguchi et al., 1989). Although deviations are commonplace in these non-fibrillar collagens, deleterious Gly missense mutations do occur in type IV collagen (Hudson et al., 2003), suggesting that non-lethal breaks in the collagen repeat sequence are site-specific and serve a functional purpose.

Model peptides based on the Gly-X-Gly deletions of Type VIII collagen have been shown by X-ray crystallography to align into layers with the deleted regions lined up (Bella et al., 2006). This is very reminiscent of the smectic ordering of preCols seen in byssal threads by atomic force microscopy (AFM) (Hassenkam et al., 2004). Bella et al. suggest the deletion may provide a ‘registration marker’ favoring in-register parallel packing rather than the quarter stagger typical of fibrillar collagens (Bella et al., 2006). It is possible that the highly conserved breaks in the Gly-X-Y sequence of the preCols are similarly directing the smectic packing in byssal threads, which has been proposed to have important mechanical consequences for the thread as a whole.

$(Gly\text{-}Pro\text{-}Hyp)_n$ is the sequence paradigm for producing a thermally stable triple helix, and deviations in the form of less stabilizing residues in the X or Y position or breaks in the triplet will lead to a less stable conformation (Persikov et al., 2005; Bella et al., 2006). PreCols are wrought with breaks and aberrations from this stable paradigm that should obviate helix formation, yet they still form triple helices and, at low strains, exhibit mechanical properties appropriate for collagens. It is interesting that the mussel would have such an unstable collagen for its lifeline. What could the mechanical consequences of this helix destabilization be on the thread compared with traditional fibrillar collagens and what benefits might it offer for the mussel? At present, there are no answers to these questions.

Histidine-rich domain

Histidine containing sequences in the N- and C-termini resembling known metal-binding motifs (Papallardo et al., 2002) are highly conserved between all three species, implying a functional role for these sequences, possibly as cross-link participants. The conservation of His-rich domains between *M. californianus* and the other two species was unexpected in light of their mechanical differences in the distal region. Self-healing is believed to be the result of the reformation of histidine–metal coordination bonds sacrificially broken during yield, and since *M. californianus* exhibits significantly faster recovery of stiffness and hysteresis, it has been proposed that it might have significant differences in the histidine-rich regions (Lucas et al., 2002). Even where it is not entirely conserved, there is no trend that would favor *M. californianus* as a faster healer. The fact that this region is not significantly different in *M. californianus* does not exclude the His–metal interaction as the reversibly breakable bond responsible for recovery, but instead suggests that there may be a separate force driving these bonds back together and that this force is different between species, resulting in different rates of healing. This requires a separate elastomeric component in parallel with the His-rich domain and possible candidates are proposed at the end of the Discussion.

Mechanical testing

In order to correlate our sequence homology data with the mechanical data, we must make several assumptions. First, we

assume that threads are composed entirely of preCols and that they alone determine the mechanical properties. This assumption seems reasonable for the distal region considering that preCol D and NG make up 96% of the protein component. Second, we assume that the molecular packing is similar between species, and that any differences in architecture play a minor role in mechanical differences. Since the overall domain structure of the preCols has been shown to be similar in all species, it is a safe assumption that the preCols of *M. californianus* maintain a similar smectic alignment as threads of *M. galloprovincialis* (Hassenkam et al., 2004).

Mechanical testing of threads performed in this study revealed that in native seawater-treated distal threads, no modulus reduction and low hysteresis are seen between 0–10% strain (Fig. 7). Stress softening occurs once threads yield (~16% strain) and continues until about 40% strain. However, this softening is reversible in a time-dependent manner (Carrington and Gosline, 2004). Histidine–metal interactions are the ideal candidate for a sacrificial cross-link since they are known to break and reform reversibly and are significantly weaker than covalent bonds (Schmitt et al., 2000). Due to consistent conservation of abundant histidine compositions in the preCol termini, the presence of metals in elevated levels in the thread core (Coombs and Keller, 1981), and the loss of yield and recovery in EDTA-treated threads in which about 50% of the metal was removed (Vaccaro and Waite, 2001), we believe these cross-links function mechanically in mussel byssal threads.

To further probe the role of histidine in mechanical performance, threads were exposed to citrate-phosphate buffer at pH 3, 4, 5, 6, 7 and 8 prior to mechanical testing. The local pH dictates the degree of protonation of histidine within this range since the pK_a of a protein histidyl residue is ~6.5 (Sundberg and Martin, 1974). Protonation of histidine gives the side chain a positive charge, preventing it from binding metal (Fig. 2). Since pH treatment can be used to control the degree of protonation and consequently the density of metal cross-links within the thread, we predict that treatment at low pH will adversely affect the mechanical performance. Histidine is the only amino acid side chain that will undergo a complete dissociation between pH 3 and 8 so we can be reasonably confident that the acidic pH treatment is primarily targeting these residues.

As Fig. 8 reveals, the initial modulus of threads over the pH range traces a roughly sigmoidal shape with a midpoint around 615 MN m^{-2} corresponding to pH 6.6. As mentioned, this pH is significant since the pK_a of histidine usually falls in this range and at the pK_a you would expect 50% of the histidine residues protonated and unable to bind metals and 50% deprotonated and bound in metal coordination crosslinks. Since the halfway point of the modulus reduction between pH 3 and 8 falls at this particular pH, it is very suggestive that His–metal bonds are serving as cross-links against tensile stress.

The loss in stiffness from low pH treatment is completely reversible by re-equilibrating the low pH thread in buffer at pH 8. This is consistent with pH-mediated disruption of His–metal cross-links, which would be completely reversible as long as some metal remains in solution. This reversibility can be prevented by treating threads with DEPC at an acidic pH, trapping unbound, deprotonated histidine residues in a carbethoxylated state where they are unable to participate in coordinate bonds (Fig. 2), and thus maintaining threads at lower stiffness even after the pH is returned to 8.

Fig. 9A reveals that at low strain values, there is a large difference in stiffness between threads treated at pH 8 and pH 5 that becomes less prominent the more the thread is strained. Fig. 9B reveals that below yield and above 40% strain the magnitude of stress softening after each strain cycle between pH 8- and pH 5-treated threads is not significantly different, but between the onset of yield and 40% strain they are. If the only difference between threads equilibrated to pH 5 and pH 8 is their respective density of His–metal cross-links, the implication arises that these bonds are not broken before yield and are fully sacrificed by around 40% strain. It should be noted that threads treated at acidic pH still show a measure of yield and stress-softening, just to a smaller degree (Fig. 9A,B). It is possible that not all His–metal cross-links were disrupted, but this seems unlikely, especially at pH 3. More plausibly, the histidine residues are not the only source of reversibly breakable bonds in the threads and these other bonds are not affected by the pH change from 3 to 8. A possible candidate for the pH-independent

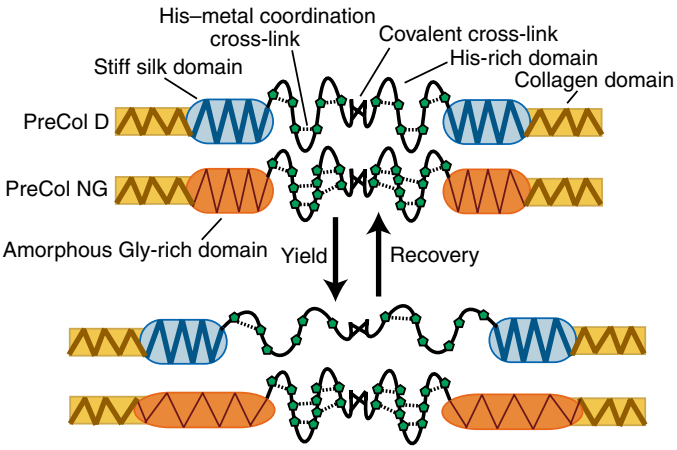

Fig. 10. Molecular model of yield and self-healing in the distal byssal thread. Our data suggests that His–metal coordinate cross-links in the termini of the preCols are bonds that are reversibly sacrificed in yield and self-healing. A difference in healing rate between species, despite high homology of the His-rich domains, suggests that there is a separate entropic elastomeric component driving recovery. Since preCol D and preCol NG are co-localized in the distal thread, and since preCols are believed to align in register, we propose that the flanking domain of NG is playing this role. In this model, the stiffness of the amorphous Gly-rich flanking domain of preCol NG is roughly the same as the His-rich domain of preCol D, and the silk domain and NG His-rich domain are somewhat stiffer. During yield, the His–metal cross-links of preCol D begin to rupture as the Gly-rich domain begins to unravel, whereas the stiffer silk domain, collagen domains, and NG His-rich domain stay folded. When the tensile force is released, there is a time-dependent entropic drive for the Gly-rich domain of preCol NG to recover its initial length. In doing so, the histidine residues are brought back within proximity of one another, allowing the metal coordination bonds to reform and initial mechanical properties to be recovered. The covalent bonds between the ends of preCols in series have been proposed to be diDOPA cross-links.

reversible elasticity is the Gly-rich flank of preCol NG. This will be amplified in the molecular model below.

These data reveal a sacrificial cross-link that is present at high pH and diminished in abundance at low pH and are consistent with our proposal that His–metal coordination mediates this crosslink. Based on these results, we propose the following molecular model of yield and self-healing in the distal threads of the mussel byssus.

Model

Prior to yield, His–metal interactions bridge the ends of a series of preCols, allowing the stiff flanking and collagen domains to dominate stress–strain behavior. At this point recovery is essentially elastic with high resilience, and initial stiffness is a function of stiffness of the flanking domains. At yield, stress on the domains of the preCol reaches a critical level that exceeds the load capacity of the His–metal coordinate bonds and they begin to rupture, dissipating the applied energy in the form of molecular friction (hysteresis). In sacrificing His–metal complexes, the covalent bonds in the backbone are spared and catastrophic failure is avoided. This may also be the point at which the non-His sacrificial bonds begin to rupture since yield, stress softening, and hysteresis are also seen in low pH threads. The best candidates for these are the NG flanks which, with a predicted amorphous structure, are likely to be more compliant than the collagen cores and the Ala-rich silk-like flanks. Beyond 40% strain, most His–metal bonds are broken, and further stress softening (which is small) is not due to the His–metal bonds breaking. Evidence for this comes from the fact that ΔE is not significantly different between high pH and low pH above 40% strain.

Since coordinate bonds are known to be reversibly breakable, His–metal could play a role in post-yield recovery if the separated ligands and metals are brought together again. His domains are not different between species, and as *M. californianus* is a faster healer, differences in recovery rate must come from another source. This suggests the presence of separate elastomeric components in parallel with the histidine-rich domains, which restore them after deformation. The flanking domains of NG are again the best candidates as the restoring elastomeric component. Since preCol D and NG are both known to be present in the distal region of the thread and since the preCols are believed to assemble in register, a composite model of yield and self-healing is outlined in Fig. 10. In this model, the stiffness of the flanking domains of NG and of the His-rich domains in D are roughly comparable. Both begin to break near the critical yield stress. When the applied load is released, the flanking regions of NG recover entropically, dragging the separated histidine residues between two preCol Ds back into mutual proximity and thus restore their metal coordination.

The assumed entropic recovery of NG flanks is an important component of our model and deserves closer scrutiny in future experimental studies. It has been proposed that Gly-rich proteins in plant cell walls homologous to those in NG flanking domains form a loose structure known as a glycine loop, reminiscent of keratin termini (Steinert et al., 1991). Glycine loops are believed to behave like molecular VELCRO™, in which the ordered structure is easily disrupted by moderate tension, but is able to recover its initial conformation after stress is released (Sachetto-Martins et al., 2000). This sort of moderately stiff, entropic elastic component would be a good fit as the component driving recovery.

This work was supported in part by the NASA University Research, Engineering and Technology Institutes on Bio-inspired Materials under award No. NCC-1-02037 and by the National Institutes of Health Grant R01 DEO 14672.

References

Bell, E. and Gosline, J. (1996). Mechanical design of mussel byssus: material yield enhances attachment strength. *J. Exp. Biol.* **199**, 1005-1017.

Bella, J., Liu, J., Kramer, R. Z., Brodsky, B. and Berman, H. M. (2006). Conformational effects of Gly-X-Gly interruptions in the collagen triple helix. *J. Mol. Biol.* **362**, 298-311.

Brazee, S. L. and Carrington, E. (2006). Interspecific comparison of the mechanical properties of mussel byssus. *Biol. Bull.* **211**, 263-274.

Brazel, D., Oberbaumer, I., Dieringer, H., Babel, W., Glanville, R. W., Deutzmann, R. and Kuhn, K. (1987). Completion of the amino acid sequence of the α1 chain of human basement membrane collagen (type IV) reveals 21 non-triplet interruptions located within the collagenous domain. *Eur. J. Biochem.* **168**, 529-536.

Brooks, A. E., Steinkraus, H. B., Nelson, S. R. and Lewis, R. V. (2005). An investigation of the divergence of major ampullate silk fibers from *Nephila clavipes* and *Argiope aurantia*. *Biomacromolecules* **6**, 3095-3099.

Carrington, E. and Gosline, J. (2004). Mechanical design of mussel byssus: load cycle and strain rate dependence. *Am. Malacol. Bull.* **18**, 135-142.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. and Thompson, J. D. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* **31**, 3497-3500.

Coombs, T. L. and Keller, P. J. (1981). Mytilus byssal threads as an environmental marker for metals. *Aquatic Toxicol.* **1**, 291-300.

Coyne, K. J., Qin, X.-X. and Waite, J. H. (1997). Extensible collagen in mussel byssus: a natural block copolymer. *Science* **277**, 1830-1832.

Fowler, S. J., Jose, S., Zhang, X., Deutzmann, R., Sarras, M. P., Jr and Boot-Handford, R. P. (2000). Characterization of Hydra type IV collagen: type IV collagen is essential for head regeneration and its expression is up-regulated upon exposure to glucose. *J. Biol. Chem.* **275**, 39589-39599.

Gosling, E. (1992). Systematic and geographic distribution of Mytilus. In *The Mussel Mytilus: Ecology, Physiology, Genetics and Culture*. Vol. 25 (ed. E. Gosling), pp. 1-20. Amsterdam: Elsevier.

Hassenkam, T., Gutsmann, T., Hansma, P., Sagert, J. and Waite, J. H. (2004). Giant bent-core mesogens in the thread forming process of marine mussels. *Biomacromolecules* **5**, 1351-1354.

Hayashi, C. Y. and Lewis, R. V. (1998). Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks. *J. Mol. Biol.* **275**, 773-784.

Hayashi, C. Y., Shipley, N. H. and Lewis, R. V. (1999). Hypotheses that correlate the sequence, structure, and mechanical properties of spider silk proteins. *Int. J. Biol. Macromol.* **24**, 271-275.

Hudson, B. G., Tryggvason, K., Sundaramoorthy, M. and Neilson, E. G. (2003). Alport's syndrome, Goodpasture's syndrome, and type IV collagen. *N. Engl. J. Med.* **348**, 2543-2556.

Kilchherr, E., Hofmann, H., Steigemann, W. and Engel, J. (1985). Structural model of the collagen-like region of C1q comprising the kink region and the fibre-like packing of the six triple helices. *J. Mol. Biol.* **186**, 403-415.

Lee, H., Scherer, N. F. and Messersmith, P. B. (2006). Single-molecule mechanics of mussel adhesion. *Proc. Natl. Acad. Sci. USA* **103**, 12999-13003.

Lucas, J. M., Vaccaro, E. and Waite, J. H. (2002). A molecular, morphometric and mechanical comparison of the structural elements of byssus from *Mytilus edulis* and *Mytilus galloprovincialis*. *J. Exp. Biol.* **205**, 1807-1817.

Lundblad, R. L. (2005). *Chemical Reagents for Protein Modification*. Boca Raton, FL: CRC Press.

Mercer, E. H. (1952). Observations on the molecular structure of byssus fibres. *Aust. J. Mar. Freshw. Res.* **3**, 199-205.

Pappalardo, G., Impellizzeri, G., Bonomo, R. P., Campagna, T., Grasso, G. and Saita, M. G. (2002). Copper(II) and nickel(II) binding modes in a histidine-containing model dodecapeptide. *New J. Chem.* **26**, 593-600.

Persikov, A. V., Ramshaw, J. A. M. and Brodsky, B. (2005). Prediction of collagen stability from amino acid sequence. *J. Biol. Chem.* **280**, 19343-19349.

Qin, X.-X. and Waite, J. H. (1995). Exotic collagen gradients in the byssus of the mussel *Mytilus edulis*. *J. Exp. Biol.* **198**, 633-644.

Qin, X.-X. and Waite, J. H. (1998). A potential mediator of collagenous block copolymer gradients in mussel byssal threads. *Proc. Natl. Acad. Sci. USA* **95**, 10517-10522.

Qin, X.-X., Coyne, K. J. and Waite, J. H. (1997). Tough tendons: mussel byssus has collagen with silk-like domains. *J. Biol. Chem.* **272**, 32623-32627.

Rudall, K. M. (1955). The distribution of collagen and chitin. *Symp. Soc. Exp. Biol.* **9**, 49-71.

Sachetto-Martins, G., Franco, L. O. and de Oliveira, D. E. (2000). Plant glycine-rich proteins: a family or just proteins with a common motif. *Biochim. Biophys. Acta* **1492**, 1-14.

Santaclara, F. J., Espineira, M., Cabado, A. G., Aldasoro, A., Gonzalez-Lavin, N. and Vieites, J. M. (2006). Development of a method for the genetic identification of mussel species belonging to *Mytilus, Perna, Aulacomya,* and other genera. *J. Agric. Food Chem.* **54**, 8461-8470.

Schmitt, L., Ludwig, M., Gaub, H. E. and Tampe, R. (2000). A metal-chelating microscopy tip as a new toolbox for single-molecule experiments by atomic force microscopy. *Biophys. J.* **78**, 3275-3285.

Sicot, F.-X., Exposito, J.-Y., Masselot, M., Garrone, R., Deutsch, J. and Gaill, F. (1997). Cloning of an annelid fibrillar-collagen gene and phylogenetic analysis of vertebrate and invertebrate collagens. *Eur. J. Biochem.* **246**, 50-58.

Soininen, R., Haka-Risku, T., Prockop, D. J. and Tryggvason, K. (1987). Complete primary structure of the α1-chain of human basement membrane (type IV) collagen. *FEBS Lett.* **225**, 188-194.

Steinert, P. M., Mack, J. W., Korge, B. P., Gan, S.-Q., Haynes, S. R. and Steven, A. C. (1991). Glycine loops in proteins: their occurrence in certain intermediate filament chains, loricrins, and single-stranded RNA binding proteins. *Int. J. Biol. Macromol.* **13**, 130-139.

Sun, C., Lucas, J. M. and Waite, J. H. (2002). Collagen-binding matrix proteins from elastomeric extraorganismic byssal fibers. *Biomacromolecules* **3**, 1240-1248.

Sundberg, R. J. and Martin, R. B. (1974). Interactions of histidine and other imidazole derivatives with transition metal ions in chemical and biological systems. *Chem. Rev.* **74**, 471-517.

Vaccaro, E. and Waite, J. H. (2001). Yield and post-yield behavior of mussel byssal thread: a self-healing biomolecular material. *Biomacromolecules* **2**, 906-911.

van Hest, J. C. M. and Tirrell, D. A. (2001). Protein-based materials, toward a new level of structural control. *Chem. Commun.* **2001**, 1897-1904.

Waite, J. H. (1992). The formation of mussel byssus: anatomy of a natural manufacturing process. *Results Probl. Cell Differ.* **19**, 27-54.

Waite, J. H., Qin, X.-X. and Coyne, K. J. (1998). The peculiar collagens of mussel byssus. *Matrix Biol.* **17**, 93-106.

Waite, J. H., Vaccaro, E., Sun, C. and Lucas, J. (2002). Elastomeric gradients: a hedge against stress concentration in marine holdfasts? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **357**, 143-153.

Waite, J. H., Lichtenegger, H. C., Stucky, G. D. and Hansma, P. (2004). Exploring molecular and mechanical gradients in structural bioscaffolds. *Biochemistry* **43**, 7653-7662.

Waite, J. H., Weaver, J. C. and Vaccaro, E. (2006). Molecular consequences of biomolecular gradients in byssal threads. In *Bionanotechnology: Proteins to Nanodevices* (ed. V. Renugopalakrishnan and R. V. Lewis), pp. 25-38. Dordrecht: Springer.

Yamaguchi, N., Benya, P., van der Rest, M. and Ninomiya, Y. (1989). The cloning and sequencing of alpha 1(VIII) collagen cDNAs demonstrate that type VIII collagen is a short chain collagen and contains triple-helical and carboxyl-terminal non-triple-helical domains similar to those of type X collagen. *J. Biol. Chem.* **264**, 16022-16029.

Yonge, C. M. (1962). On the primitive significance of the byssus in the bivalvia and its effects in evolution. *J. Mar. Biol.* **42**, 113-125.