Hyperunstable matrix proteins in the byssus of *Mytilus galloprovincialis*

Jason Sagert and J. Herbert Waite*

Marine Science Institute, University of California, Santa Barbara, CA 93106, USA

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

SUMMARY

The marine mussel *Mytilus galloprovincialis* is tethered to rocks in the intertidal zone by a holdfast known as the byssus. Functioning as a shock absorber, the byssus is composed of threads, the primary molecular components of which are collagen-containing proteins (preCOLs) that largely dictate the higher order self-assembly and mechanical properties of byssal threads. The threads contain additional matrix components that separate and perhaps lubricate the collagenous microfibrils during deformation in tension. In this study, the thread matrix proteins (TMPs), a glycine-, tyrosine- and asparagine-rich protein family, were shown to possess unique repeated sequence motifs, significant transcriptional heterogeneity and were distributed throughout the byssal thread. Deamidation was shown to occur at a significant rate in a recombinant TMP and in the byssal thread as a function of time. Furthermore, charge heterogeneity presumably due to deamidation was observed in TMPs extracted from threads. The TMPs were localized to the preCOL-containing secretory granules in the collagen gland of the foot and are assumed to provide a viscoelastic matrix around the collagenous fibers in byssal threads.

Supplementary material available online at http://jeb.biologists.org/cgi/content/full/212/14/2224/DC1

INTRODUCTION

All metazoans use collagens to build extracellular load-bearing scaffolds for living tissue. In some organisms, additional collagenous scaffolds are constructed outside the confines of living tissue: sawfly silk (Spiro et al., 1971), selachian egg case (Rusaouen et al., 1976), gorgonin (Goldberg, 1974), spongin (Aouacheria et al., 2006) and mussel byssus (Qin and Waite, 1995) are familiar examples and have attracted considerable attention for their unusual mechanical properties, rate of formation and durability (Brown, 1975). Given the ease and multiplicity of methods by which collagens can be investigated, it is not surprising that biochemical models of scaffolds are dominated by collagens next to which the properties, distribution and abundance of non-collagenous proteins have been largely neglected. This is addressed by the present study on non-collagenous proteins in the byssal threads of *Mytilus galloprovincialis*. As the relationship between collagenous and non-collagenous proteins in load-bearing scaffolds is likely to be an important and complex one, an introduction to the structure and organization of byssus and byssal collagens seems appropriate.

Superficially, mussel byssus resembles a tendon. Compared with Achilles tendon, for instance, which originates in the gastrocnemius muscle and inserts into the heel bone or calcaneum, the byssus of *Mytilus* sp. originates in the anterior and posterior byssal retractor muscles and 'inserts' onto a hard external surface such as a rock (Fig. 1). A byssus with hundreds of threads thus provides the mussel with a holdfast having adjustable tension for stable attachment in the rocky intertidal zone. Upon closer scrutiny, however, the analogy with tendons becomes less compelling. Tendons provide energy transfer whereas in byssus, energy is effectively dissipated at extensions >10% (Gosline et al., 2002). Moreover, tendons have rather uniform mechanical and compositional properties whereas byssal threads exhibit gradients in both composition and in stiffness. In *Mytilus edulis* and *M. galloprovincialis*, the proximal portion of

each thread is approximately 10 times more compliant than the distal region. *Mytilus californianus* exhibits a nearly 50-fold difference in compliance between the proximal and distal portions of the byssal thread (Bell and Gosline, 1996).

The collagenous proteins of mytilid byssal threads are known as preCOLs (from prepepsinized collagens) and account for approximately 90% and 70% of the distal and proximal threads by dry mass, respectively (Waite et al., 2002). This is comparable with the collagen content of tendons; however, preCOL structure and organization are fundamentally different from tendon type I collagens. With respect to structure, preCOLs have kinked collagen domains (Fig. 1A) and accommodate significant non-collagenous protein sequences in flanking blocks that resemble either spider dragline silk or elastin (Waite et al., 2004). For the sake of simplicity and consistent with tensile tests of isolated preCOLs (Harrington and Waite, 2008a), the silk and elastin-like blocks are here identified as 'stiff' and 'compliant' domains, respectively (Fig. 1A). 'Sticky ends' (Fig. 1A) refer to histidine and Dopa-rich N- and C-termini that mediate end-to-end cross-bridging during assembly (Harrington and Waite, 2007). PreCOL organization in the byssus never displays the quarter-stagger array of collagens observed in tendons. Transmission electron micrographs of liquid crystalline preCOLs in the collagen gland (Vitellaro-Zucarello, 1980) and atomic force micrographs of preCOLs in the thread (Hassenkam et al., 2004) reveal instead a distinct lateral register referred to as smectic alignment in liquid crystal literature (Collings, 2002).

The smectic character of preCOL organization is well adapted for making graded structures as was recently demonstrated by modeling the mechanical effects of incrementally increasing compliant blocks in the distal-to-proximal direction (Waite et al., 2002; Waite et al., 2004). Although predictions of the model were consistent with the observed Young's moduli (stiffness) of the

proximal and distal portions of byssal threads, uncertainty remains because lateral register of preCOLs in byssal threads never appears as uniform as schematically shown in Fig. 1B. Instead, a more localized smectic organization prevails with preCOLs arranged in sheet-like bundles having diameters several hundreds of nanometers wide (Fig. S1 in supplementary material). In other words, preCOLs are smectic within each bundle but there is little to no lateral preCOL register between sheets (Fig. 1C) (Hassenkam et al., 2004). Given the extreme length of preCOL fibers and the extensive potential non-covalent contact between neighboring non-aligned preCOL sequences (circled portion, Fig. 1C), molecular friction between these during tension would pose considerable resistance to deformation and thus suppress the contribution of the compliant flanking domains. In tendon, collagen fibers are separated and lubricated by proteoglycans such as decorin and cartilage oligomeric protein or COMP (Kjær, 2004). Smectic preCOL sheets also need to be isolated from each other by a matrix (Fig. 1D), and this study characterizes a peculiar matrix protein from mussel byssus.

In the present study, we investigated the hypothesis that the byssal threads of *Mytilus galloprovincialis* Lamarck contain non-collagenous proteins in addition to the preCOLs. A unique glycine (Gly)-, tyrosine (Tyr)- and asparagine (Asn)-rich protein family, collectively called the thread matrix proteins (TMPs) was identified and a thorough characterization is described. The TMPs are intimately associated with the preCOLs within the secretory granules and are distributed throughout the byssal threads. Intriguingly, TMPs are among the most deamidation-prone proteins known. Asn residues in recombinant and native TMP deamidate spontaneously at pH 8, and in doing so introduce significant new side-chain and backbone heterogeneity into the primary structure.
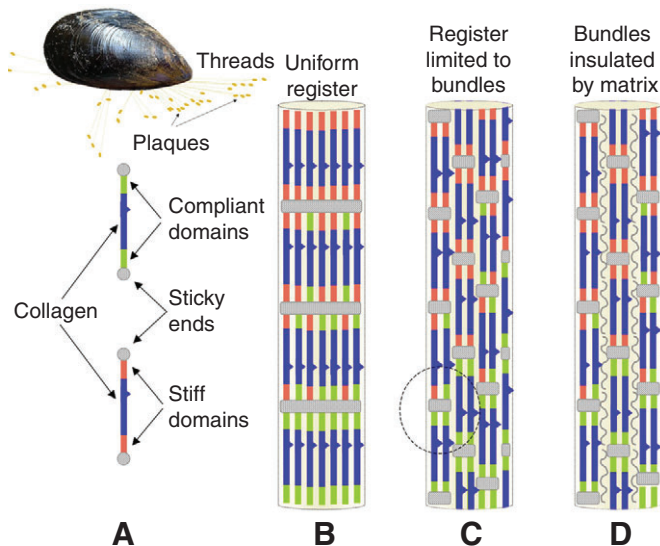

Fig. 1. Mussel (*Mytilus galloprovincialis*) attached to substratum by a byssus. The major load-bearing macromolecules in the byssus are the collagen-containing proteins (preCols) and models of their assembly in the byssus are caricatured. (A) Each preCol consists of a rigid, kinked collagen region flanked by compliant domains with histidine-rich regions at the N- and C-terminus. (B) The organization of smectic preCols as modeled in Waite et al. (Waite et al., 2004; Waite et al., 2002). (C) The smectic register is a local one and not in lateral register with other fibrils. Circled portion indicates neighboring non-aligned preCOL sequences. (D) Smectic fibrils separated and lubricated by a matrix protein.

MATERIALS AND METHODS
Mussel maintenance, mature and induced thread collections and crude protein extracts
Mussels (*Mytilus galloprovincialis* Lamarck) were collected from Goleta Pier (Goleta, CA, USA) and maintained in tanks with continuous flow seawater. Mussels were separated from each other and allowed to secrete threads for approximately three days, and then distal threads were collected, washed in distilled water and stored at –20°C until further use. Byssal proteins were induced from the foot as described in Tamarin et al. (Tamarin et al., 1976). Briefly, $0.56\,mol\,l^{-1}$ KCl was injected into the base of the foot. After approximately 30 min, byssal proteins were collected from the groove of the foot, washed in 5% acetic acid once and stored at –20°C until further use.

All samples were homogenized in 5% acetic acid, $4\,mol\,l^{-1}$ urea (extraction buffer). Approximately 500 μl of extraction buffer was used per 50 mg of threads or 25 induced threads. Samples were homogenized in a glass tissue homogenizer, spun at 15,000 g in a microcentrifuge and stored at –20°C until further use.

Partial purification of TMPs from the distal thread
Crude thread extracts were separated on a Sephacryl S-200 (16/60) (GE Healthcare, Piscataway, NJ, USA) sizing column equilibrated in 5% acetic acid, $4\,mol\,l^{-1}$ urea using a flow rate of $1\,ml\,min^{-1}$. Fractions containing the TMPs as determined by acid–urea polyacrylamide gel electrophoresis (AU PAGE) were applied to a C-4 reverse phase-high performance liquid chromatography (RP-HPLC) column (Phenomenex, Torrance, CA, USA) and eluted with an acetonitrile gradient of 0–20%, 20–40% and 40–100% for 30 min, 25 min and 15 min, respectively, with a flow rate of $1\,ml\,min^{-1}$. Five μl of $8\,mol\,l^{-1}$ urea were added to fractions containing the TMPs, as determined by AU PAGE, which were subsequently lyophilized, resuspended in 100 μl 5% acetic acid and quantified by amino acid analysis. After quantification, the protein was lyophilized until further use.

Gel electrophoresis
Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) was conducted as previously described. Briefly, samples in 5% acetic acid were run on a 10% mini-SDS PAGE denaturing gel using the Laemmli tris-glycinate electrode buffer system (Laemmli, 1970). The gel was run at a constant current of 20 mA and was stained with Coomassie Blue R-250. AU gel electrophoresis was conducted as previously described elsewhere (Panyim and Chalkley, 1969).

Amino acid analysis
Amino acid analysis was performed with a 6300 Autoanalyzer (Beckman Instruments, Fullerton, CA, USA) after the protein (5–10 μl) was hydrolyzed in $6\,mol\,l^{-1}$ HCl, 5% phenol *in vacuo* for 24 h at 110°C (Waite, 1995).

When conducting amino acid analysis from proteins separated on AU PAGE, proteins were transferred to polyvinylidene difluoride (PVDF) membrane as described in the Western blotting section of the Materials and methods. Protein bands were visualized using Coomassie Blue G-250, excised with a razor blade, destained with methanol and hydrolyzed as above.

MALDI-TOF mass spectrometry
Matrix-assisted laser desorption ionization (MALDI) mass spectrometry with time-of-flight (TOF) was performed on a Voyager DE mass spectrometer (AB Biosystems, Foster City, CA, USA).

Partially purified TMPs (~0.5 mg ml^{-1}) were diluted 1:25 in fresh sinapinic acid (Sigma Chemical Co., St Louis, MO, USA) in aqueous 30% acetonitrile and 0.1% trifluoroacetic acid. Drops of 2 μl were spotted onto the gold-plated sample plate and dried under vacuum. Mass spectrums were collected in positive ion mode and with an accelerating voltage of 25,000 V, a grid voltage of 90.0%, a guide wire voltage of 0.300%, a delay time of 1500 ns and a N$_2$ laser power of 2020 (relative units).

Lys-C digestion

Approximately 700 μg of TMP was resuspended in 200 mmol l^{-1} tris-buffer titrated to pH 7.5 with 200 mmol l^{-1} ascorbate with a final concentration of approximately 1.6 mol l^{-1} urea. Lys-C endopeptidase was added at a ratio of 1:10 and the digestion was carried out overnight at 37°C. The reaction was terminated by injection onto a C-18 RP-HPLC column. Peptides were eluted with the a gradient of acetonitrile consisting of 0–20%, 20–40% and 40–100% for 50 min, 10 min and 10 min, respectively, at a flow rate of 1 ml min^{-1}. Peak fractions (1 ml) were lyophilized, resuspended in 10 μl 5% acetic acid and subjected to N-terminal sequencing.

Protein sequencing

N-terminal gas-phase sequencing was done on a Porton 2020 protein sequencer with a dedicated inline HPLC (model 2090, Beckman-Coulter, Fullerton, CA, USA) for separating the phenol thiohydantoin (PTH) derivatives according to established procedures (Waite, 1995).

PCR and cloning of TMPs

A degenerate 5′ forward primer DP3F-5′AAATAYGG-HTAYAAYAAYAAYGGH3′ TAYGGHAAAY (Y=C and T, H=A, C and T) was made from the amino acid sequence obtained from the peptide sequence in fraction 51. Two peptides were present in fraction 51, as such, the most likely peptide sequence KYGYNNNGYG was used for primer design; however, upon cloning the TMPs, the peptide sequence was actually shown to be GSG instead of K. A vector-encoded 3′ T7 reverse primer was used to amplify the C-terminus of the TMPs. A *M. galloprovincialis* cDNA library, prepared in earlier work (Lucas et al., 2002), was used as a template.

PCR was performed using standard conditions; dNTPs, buffer and Taq polymerase were from Fisher Scientific (Waltham, MA, USA). Reaction conditions consisted of an initial denaturation step of 7 min at 94°C, followed by a 30 s denaturing step at 94°C, an annealing step of 30 s at 52°C and an elongation step of 1 min at 72°C. The reaction continued for 35 cycles and was concluded with a final elongation step of 7 min at 72°C. PCR products were separated on a 1% agarose gel and visualized with ethidium bromide. Candidate products were cloned into pCR2.1 TOPO (Invitrogen, Carlsbad, CA, USA) vector and transformed into TOP10 cells (Invitrogen). Positive transformants were identified using standard molecular biological techniques. Inserts were sequenced using a vector-encoded M13R primer at the Advanced Instrumentation Laboratory at the University of California, Santa Barbara, CA, USA.

RNA purification and 5′ RACE

Total RNA was purified using the RNeasy Plant Mini Kit from Qiagen (Valencia, CA, USA). Approximately 50 mg of mussel foot tissue was dissected and immediately frozen in liquid nitrogen. After the tissue was homogenized with a mortar and pestle in liquid nitrogen, the manufacturer's protocol was followed. Purified RNA was quantified with a UV spectrophotometer and assayed for integrity by agarose gel electrophoresis.

5′ RACE (rapid amplification to cDNA ends) was preformed using the GeneRacer kit (Invitrogen). A reverse gene specific primer (5′TGTTTGATATTTCCATAACCGTCGTTATG3′) was designed to the conserved 3′ coding region of the TMP variants identified from the cDNA library and was used in combination with the 5′ linker-specific primer. Touchdown PCR was conducted according to the manufacturer's protocol. Briefly, an initial 2 min 94°C denaturation step was followed by a 45 s 72°C annealing step and a 90 s 72°C extension step. The annealing temperature was reduced by 2°C in each of the next two cycles. The final 35 cycles were done with an annealing temperature of 54°C. Amplification products were cloned as described in the PCR and cloning of TMPs section of the Materials and methods with the exception that the vector pCR4-TOPO was used. RACE products were sequenced by Bio S&T (Montreal, Canada).

RT-PCR

Feet were dissected into three sections (distal, middle and proximal) and total RNA was purified as above. Total RNA from mantle tissue was used as a negative control. A forward primer (5′GTTA-TTCCACGATGGTCCAGTTAACAC3′) was designed to the 5′ signal peptide. The gene specific reverse primer was the same as used in 5′ RACE. The forward primer 5′GTTGTAGAAGT-CAGGGTGAGATACTC3′ and the reverse primer 5′GGGA-AGAGCGTAACCTTCGTAG3′ were designed to amplify a approximately 450 bp fragment of Actin as a positive control. The product was confirmed by sequencing. RT-PCR was conducted using the Qiagen OneStep RT-PCR kit. The reverse transcriptase reaction was conducted at 54°C, and touchdown PCR was conducted as for the 5′ RACE reactions. Amplification products were cloned as in the PCR and cloning of TMPs section of the Materials and methods.

Recombinant protein expression and purification

A partial cDNA encoding the C-terminal 155 amino acids of a variant of the TMP family was subcloned into the pRSET expression vector (Invitrogen) and subsequently expressed in BL21 Star cells (Invitrogen). Expression of the recombinant protein was induced by adding isopropyl thiogalactoside (IPTG) to a final concentration of 1 mmol l^{-1} when the bacterial cultures reached an optical density at 600 nm (OD$_{600}$) between 0.4 and 0.6. Three hours after induction, cells were harvested by centrifugation and then stored at –20°C. To lyse the bacterial cells and solubilize the recombinant protein, pellets were resuspended in 20 mmol l^{-1} phosphate buffer, 0.5 mol l^{-1} NaCl, 6 mol l^{-1} guanidine, pH 7.8, followed by a 1-min sonication. Crude cell extracts were centrifuged at 3000 g for 10 min prior to further purification.

The vector-encoded His$_6$ tag was used to purify the recombinant protein using metal affinity chromatography. The supernatant was applied to a HiTrap (1 ml) chelation column (GE Health Sciences) that was charged with Zn^{2+}. His$_6$ tagged protein was bound to the column under denaturing conditions (phosphate buffered saline, 8 mol l^{-1} urea, pH 7.8) and eluted in a stepwise fashion with phosphate buffered saline, 8 mol l^{-1} urea, pH 4.0. Affinity purified recombinant protein was further purified by HPLC on a C-8 column (Brownlee, ABI, Foster City, CA, USA), using a gradient from 0–25% acetonitrile for 10 min, followed by a linear gradient from 25–47% acetonitrile for 55 min. Fractions containing the recombinant protein were lyophilized and resuspended in 200 μl of 5% acetic acid and subsequently quantified with amino acid analysis.

Purity of the recombinant protein was determined using SDS PAGE and amino acid analysis. Protein was then lyophilized and stored at –80°C until use.

2-D electrophoresis and quantification of deamidation

Aliquots (100 µg) of lyophilized recombinant protein were aged for 0, 3 or 6 days at 37°C and for 6 days at 14°C in 20 mmol l^{-1} phosphate buffer, pH 8.0. Prior to lyophilizing, acetonitrile was added (1:1 vol/vol) and each sample was subsequently resuspended in 100 µl of 6 mol l^{-1} urea. Isoelectric focusing was performed using 7 cm pH 6–11 Immobiline Drystrip (GE Healthcare). Approximately 5 µg of recombinant protein was focused according to the manufacturer's protocol. The second dimension SDS PAGE was performed using an ExcelGel 2-D Homogenous 12.5% gel with a Multiphor II flatbed system (GE Healthcare) according to the manufacturer's protocol. To quantify the amount of deamidation, gels were stained with Coomassie Blue R-250, and the relative amount of each charge variant was quantified using the ImageJ program (National Institutes of Health, Bethesda, MD, USA) and a weighted average was calculated.

For the purposes of western blotting, 60 µg of protein extracted from induced threads with 5% acetic acid were lyophilized, resuspended in 80 µl of Destreak Rehydration Solution and focused using 7 cm pH 6–11 Immobiline Drystrip (GE Healthcare) according to the manufacturer's protocol. The second dimension was performed using a Bio-Rad (Hercules, CA, USA) criterion 10% gel and the proteins were subsequently transferred to PVDF membrane.

Western blotting

Western blots were performed using standard techniques. Antibodies to the recombinant TMP (rTMP) were produced in chickens (Aves Labs, Tigard, OR, USA) and used at a titer of 1:2000. A rabbit antichicken peroxidase-linked secondary antibody (Aves Labs) was used at a titer of 1:10,000. The blots were visualized using SuperSignal West Pico Chemiluminescent Substrate (Pierce Biotechnology, Rockford, IL, USA) followed by exposure to autoradiographic film.

Peptide mapping

rTMP was digested with Lys-C at a ratio of 1:500 in 25 mmol l^{-1} tris, 20 mmol l^{-1} methylamine, 1 mmol l^{-1} EDTA, 4 mol l^{-1} urea, pH 8.5 overnight at 37°C. Chromatography was performed on an Agilent (Wilmington, DE, USA) 1100 nanoLC system with an isocratic sample-loading pump. Approximately 30 pmol of protein was loaded onto the trap (Zorbax 300SB-C18, 5×0.03 mm, Agilent) at 100 µl min^{-1}. Solvents were as follows: A was water, 0.1% formic acid and B was acetonitrile, 0.1% formic acid. Peptides were eluted from the column (Zorbax 300SB-C18 150 mm×75 µm) with the following gradient: 5% B at 3 min, 10% B at 5 min, 20% B at 8 min, 50% B at 68 min and 90% B at 70 min.

Mass spectrometry was performed on a Micromass QTOF2 (quadrupole/time-of-flight) mass spectrometer (MicroMass UK, Ltd, Manchester, UK) with nanoflow electrospray ionization source. Flow rate was 300 nl min^{-1}. Typical source parameters were 3.5 kV capillary and 40 V sample cone. Scanning was from 500–2000 m/z, 1 s per scan.

Isoaspartate quantification

Isoaspartate (isoAsp) was quantified using the Isoquant kit (Promega, Madison, WI, USA) with the HPLC detection method with the following modifications. A Brownlee Aquapore OD-300, 220×4.6 mm column (Applied Biosciences, Foster City, CA, USA)

was used with an elution gradient of 5–25% acetonitrile in 25 min. Approximately 5 µg of rTMP was used for each Isoquant assay. Mussels were induced to secrete the byssal proteins by injecting the foot with 0.56 mol l^{-1} KCl (Tamarin et al., 1976). Induced threads, 1 day old and 7 day old threads, were homogenized in 5% acetic acid. Crude homogenates were spun at 15,000 *g* in a microcentrifuge for 10 min to obtain a crude extract (supernatant). The concentration of each extract was determined using amino acid analysis. Twenty µg
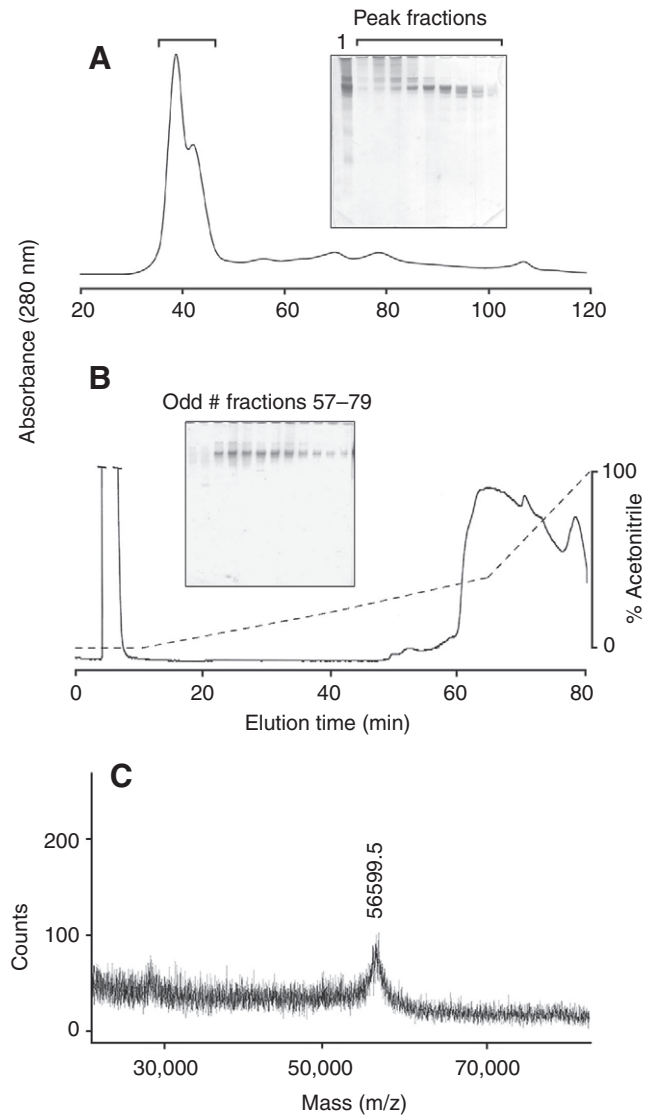

Fig. 2. Partial purification of thread matrix proteins (TMPs). (A) Size exclusion chromatography of acid extractable proteins from the distal thread. Proteins were extracted from distal threads with 5% acetic acid and separated by gel filtration on Sephacryl S-200. The solid line is the absorbance at 280 nm. The proteins in the fractions under the bracket are shown separated on a 10% acid–urea polyacrylamide gel electrophoresis (AU PAGE) gel. Lane 1 shows crude extracts. (B) Reverse-phase-high performance liquid chromatography (RP-HPLC) separation of TMPs. Fractions containing significant amounts of the TMPs were separated on a C4 RP-HPLC column. Solid line is the absorbance at 280 nm. The broken line is the gradient of acetonitrile. The TMPs eluted as a broad indistinct peak. Odd fractions between 57 and 79 min are shown separated on a 10% AU PAGE gel. (C) Matrix-assisted laser desorption ionization-time-of-flight (MALDI-TOF) mass spectrometry of partially purified TMPs. The only proteins to desorb and ionize exhibited a [M + H]$^{1+}$ of 56.5 kDa.

Table 1. Composition of distal threads, extracts and partially purified thread matrix proteins (TMPs)

Amino acid	Distal thread[a]	Acid extracts	Purified proteins	TMPs	TMPs calc.
Hyp	5.80	ND	ND	ND	0
Asx	5.17	11.50	12.01	17.27	20.6
Thr	3.09	3.62	3.09	2.82	2.7
Ser	4.75	6.02	4.17	3.92	2.3
Glx	5.21	6.64	5.45	2.80	0
Pro	7.74	5.86	4.82	1.67	0
Gly	34.17	27.97	32.69	33.01	31.2
Ala	16.21	8.58	6.51	2.92	0
Cys/2	0.13	ND	ND	ND	0
Val	2.44	2.06	2.38	2.73	3.1
Met	0.63	ND	0.48	0.22	0
Ile	1.01	1.06	2.09	3.47	6.4
Leu	2.91	1.77	2.05	1.39	0.6
DOPA	0.18	3.26	2.97	2.51	–
Tyr[b]	1.50 (1.68)	5.42 (8.68)	10.05 (13.02)	15.15 (17.66)	23.4
Phe	1.09	0.97	1.38	1.09	0.8
His	2.19	2.35	1.10	.92	0.8
Hyl	0.01	ND	ND	ND	0
Lys	3.76	4.88	4.51	4.76	5.6
Arg	2.00	4.75	4.09	3.35	2.5

Values represent mean residues per 100 residues from at least two samples. [a]Lucas et al., 2002. [b]Number in parentheses is Dopa residues + tyrosine (Tyr) residue for the total Tyr amount. Acid extracts are acetic acid soluble proteins from distal byssal threads. Purified protein refers to the RP-HPLC purified group of proteins. 'TMPs' are the most abundant group of proteins found in the partially purified TMPs. 'TMP's Calc.' is the calculated amino acid composition of the assembled full-length TMP-1 variant. The detection of Dopa was variable due to experimental procedure. In order to achieve accurate total Tyr compositions, only compositions in which Dopa was detected were included.

of each sample was digested in 5% acetic acid with pepsin at a ratio 1:10 for 48h. Urea was added to give a final concentration of $6\,mol\,l^{-1}$ after lyophilization. Samples were lyophilized, resuspended in 10μl deionized H_2O and assayed for isoAsp as above.

RESULTS
TMP from mature byssal threads
Small amounts of soluble protein (<1.0% weight) were extracted from mature distal byssal threads with 5% acetic acid, $4\,mol\,l^{-1}$ urea (Fig. 2A, inset lane 1). The most abundant of these had a composition rich in Gly, Tyr and Asx residues. When crude extracts were subjected to size exclusion chromatography on a Sephacryl S-200 column (Fig. 2A), the TMPs eluted immediately after the exclusion limit, adequately separating from larger proteins and aggregates present. On AU PAGE, the TMPs ran as a collection of several bands with similar compositions (Fig. 2A). Fractions particularly enriched in TMPs, as judged by AU PAGE, were then applied to a C4 HPLC column (Fig. 2B) from which the TMPs eluted as a broad indistinct peak. An attempt was made to better resolve the cluster of eluting bands (not shown). The TMPs, however, eluted throughout the entire peak regardless of the acetonitrile gradient used. Possible reasons for this are discussed below. The compositions of the crude extract, partially purified TMPs and cDNA-deduced TMP (calculated) are shown in Table 1. MALDI-TOF mass spectrometry conducted on the partially purified TMPs resulted in a broad peak representing a single ionized species at approximately 56 kDa (Fig. 2C). One gram of mature distal threads resulted in approximately 700μg of partially purified TMPs or 0.07 wt % as quantified by amino acid analysis. This probably represents a minimum because the proteins of byssal threads are extensively cross-linked during maturation. In densitometric scans of SDS PAGE gels of the more soluble, induced ('neonate') threads, the ratio of TMP to preCOLs ranged from 1:5 to 1:20 (J.H.W., unpublished). Given that preCOLS represent 90% of the dry mass of distal threads (Waite

et al., 2002), actual TMP abundance may be between 0.5 and 2% of dry mass.

Partial sequence of TMP and cDNA
Lys-C-generated peptides were separated with RP-HPLC (Fig. 3) and subjected to N-terminal Edman degradation. Peptides providing interpretable sequences are shown in Table 2. Using a cDNA library as a template (Lucas et al., 2002), a degenerate primer designed from peptide 51 and a vector-encoded primer was used to amplify a C-terminal sequence of a TMP. A hybridization screen of the same cDNA library using the PCR amplified sequence yielded another three C-terminal sequences, one of which was identical to the PCR amplified sequence. All three sequences consisted of similar repeats with minor differences and all possessed the same 3′ UTR as the original PCR amplified clone (Fig. S2 in supplementary material). To amplify the 5′ end, RACE was conducted with a primer designed from the conserved C-terminus of the three C-terminal
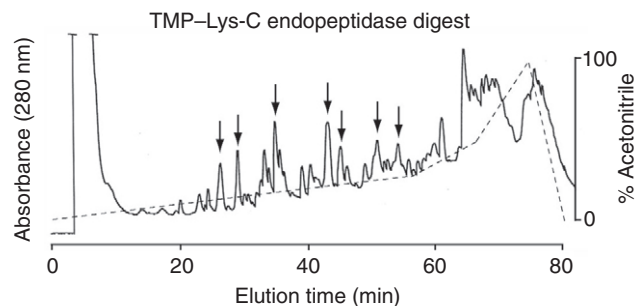

Fig. 3. Separation of Lys-C-derived peptides from partially purified thread matrix proteins (TMPs) by C18 reverse-phase-high performance liquid chromatography (RP-HPLC). The solid line is the absorbance at 280 nm. The broken line is the gradient of acetonitrile. Peak fractions denoted by arrows were subjected to N-terminal Edman degradation. Sequence results are shown in Table 2.

Table 2. Lys-C-derived peptides from partially purified thread matrix proteins (TMPs)

Fraction #	Sequence
27	IVVN
30	GYGNY
35	TNIIV
43	TNIFVN
44	GNDYGYG
51	GSGYGYNNNGYGN
54	GNIY*GYGN

Y*=Dihydroxyphenylalanine (DOPA).

clones and resulted in the amplification of two different partial transcripts designated TMP-1 and TMP-2. The 5′ UTR and signal peptide of the clones are identical but differ slightly in the initial repeat. A full-length TMP-1a sequence was assembled from clones obtained by 5′ RACE and PCR (Fig. 4). The repeats found in the TMP-1 and TMP-2 sequences are summarized in Fig. 5. The high incidence of the Asn–Gly signature is immediately evident and is suggestive of spontaneous deamidation (Robinson and Rudd, 1974). In protein-based deamidation, the amide side chain of Asn cyclizes with the backbone carbonyl of Gly to form a transient succinimide, which opens up to give both D- and L-Aspartate (Asp) and isoAsp in α and β linkages, respectively (Fig. 6).

RT-PCR reactions with RNA from the distal, middle and proximal regions of the foot resulted in the amplification of partial sequences of TMP-1 and TMP-2 (Fig. 7). The two bands observed were due to two annealing sites for the primer used within the TMP proteins. The results demonstrate that both TMP-1 and TMP-2 transcripts are expressed throughout the foot. One individual mussel yielded partial N-terminal sequence of two TMP-1 variants (Fig. S3 in supplementary material) and six TMP-2 variants; however, a gradient distribution of intra-TMP-1 and TMP-2 variants was not apparent. TMP sequences amplified from a second individual produced the same TMP-1 sequences and additional TMP-2 variants. Interestingly, the TMP-2 sequences amplified from the second individual all differed slightly in the N-terminal conserved TMP-2 sequence from that of the individual used for 5′ RACE and the first individual used for RT-PCR (partial cDNA sequences of all TMP-2 variants cloned are summarized in Fig. S4 in supplementary material), suggesting inter-individual differences in TMP transcripts as well.

Expression of recombinant TMP

The 155-residue C-terminus of a TMP variant, which is approximately one-third of a typical full-length TMP, was recombinantly expressed in *Escherichia coli* (Fig. 8). MALDI-TOF mass spectrometry and amino acid analysis all confirmed that rTMP had the expected mass and composition. The observed molecular mass of the recombinant protein was 21334.6 Da, which is consistent with the calculated mass of 21332.76 if 1.4 mass units are subtracted for the observed mean deamidation of the purified protein by 2-D gel electrophoresis (see below).

Deamidation of rTMP

The presence of numerous Asn–Gly sequences suggested that spontaneous deamidation was likely to occur within the native TMPs and rTMPs. Fig. 9A shows the charge heterogeneity in rTMP as a result of aging (non-aged and protein-aged at 37°C only). Consistent with deamidation, aging rTMP in 20 mmol l⁻¹ phosphate buffer, pH 8.0, shifted from a basic pI towards a more acidic one. In the

A

```
-14    actgttattccacgatggtccagttaacacaacattgtttggtcttgcttcttttgca
                M  V  Q  L  T  Q  H  C  L  V  L  L  L  F  A -6
                                    ↓
+46    ttggcaaatggaagtagttatggaaatggtaaaacaacaaaaattgttgtaaataaagat
        L  A  N  G  S  S  Y  G  N  G  K  T  T  K  I  V  V  N  K  D +15

+106   aatggatatggtttatgattacgacgatggttatggaaatgataaaactacgaaaattgtt
        N  G  Y  G  Y  D  Y  D  D  G  Y  G  N  D  K  T  T  K  I  V +35

+166   gtaaataaaggaaatggatatggtttatgatcacggaaattatggaaatggtaaaacaaa
        V  N  K  G  N  G  Y  G  Y  D  H  G  N  Y  G  N  G  K  T  K +55

+226   atcgtagtaaataaaggaagtggatatggttactataacaacggttatggaaacggatat
        I  V  V  N  K  G  S  G  Y  G  Y  Y  N  N  G  Y  G  N  G  Y +75

+286   ggttatggttacaacggttatagaaatggaaaaacaaatatcattgtaaataaaggaagt
        G  Y  G  Y  N  G  Y  R  N  G  K  T  N  I  I  V  N  K  G  S +95

+346   ggatatggtaatggttacaacggttatggaaatggatatggttatggaaatagatatggt
        G  Y  G  N  G  Y  N  G  Y  G  N  G  Y  G  Y  G  N  R  Y  G +115

+406   tatggaaatagatatggtttatagaaatatatatggttatggaaatagatacggttatggt
        Y  G  N  R  Y  G  Y  R  N  I  Y  G  Y  G  N  R  Y  G  Y  G +135

+466   tacaacggttatagaaatggaaaaacaaatatcattgtaaataaaggaagtggatatggt
        Y  N  G  Y  R  N  G  K  T  N  I  I  V  N  K  G  S  G  Y  G +155

+526   aatggttacaacggttatggaaatggatatggttatggaaatatatatggttatggaaat
        N  G  Y  N  G  Y  G  N  G  Y  G  Y  G  N  I  Y  G  Y  G  N +175

+587   agatacggttatggttacaacggttatagaaatggaaaaacaaatatcattgtaaataaa
        R  Y  G  Y  G  Y  N  G  Y  R  N  G  K  T  N  I  I  V  N  K +195

+646   ggaagtggatatggttacaataacaacggtatggaaatggatatggttatggaaatata
        G  S  G  Y  G  Y  N  N  N  G  Y  G  N  G  Y  G  Y  G  N  I +215

+706   tatggttatggaaatagatatggtttatggttacaacggttatagaaatggaagaacaat
        Y  G  Y  G  N  R  Y  G  Y  G  Y  N  G  Y  R  N  G  R  T  N +235

+766   atcattgtaaataaaggaagtggatatggttacaataacggcggctatggaaacggatat
        I  I  V  N  K  G  S  G  Y  G  Y  N  G  Y  R  N  G  R  T  N +255

+826   ggttatggtaatatatatggttatggttacaacggttatggaaatggaaaaacaaatatc
        G  Y  G  N  I  Y  G  Y  G  N  G  Y  N  G  Y  G  N  G  K  T  N  I +275

+886   attgtaaataaaggaagtggatatggttacaataacggcggctatggaaatggatatggt
        I  V  N  K  G  S  G  Y  G  Y  N  N  G  G  Y  G  N  G  Y  G +295

+946   tatggaaatttatatggttatggaaatatatatggttatggaaatatatacggttatggg
        Y  G  N  L  Y  G  Y  G  N  I  Y  G  Y  G  N  I  Y  G  Y  G +315
```

B

```
+1006  aatagatatggctatggttacaacggttatggaaatggaaaaacaaatatctttgtgaat
        N  R  Y  G  Y  G  Y  N  G  Y  G  N  G  K  T  N  I  F  V  N +335

+1066  aaaggaaatggatatggttatggaaatatatatggttatggaaatatatatggttacgga
        K  G  N  G  Y  G  Y  G  N  I  Y  G  Y  G  N  I  Y  G  Y  G +355

+1126  aatatatatggtcatggggaatatatatggctatggttacaacggttatggaaatggaaaa
        N  I  Y  G  H  G  N  I  Y  G  Y  G  N  G  Y  G  N  G  K +375

+1186  acaagtatctttgtaaataaaggaagtggatatggttacaataacggcggctttggaaat
        T  S  I  F  V  N  K  G  S  G  Y  G  Y  N  N  G  G  F  G  N +395

+1246  ggatatggttatggaaatctatatggttatggaaatatatatggctatggttacaatggt
        G  Y  G  Y  G  N  L  Y  G  Y  G  N  I  Y  G  Y  G  Y  N  G +415

+1306  tatggaaatatatatggctatggctacaatggttatggaaatggaaaaacaaatatcatt
        Y  G  N  I  Y  G  Y  G  Y  N  G  Y  G  N  G  K  T  N  I  I +435

+1366  gtaaataaaggaagtggatatggttacaataacgacggctatggaaatggatatggttat
        V  N  K  G  S  G  Y  G  Y  N  N  D  G  Y  G  N  G  Y  G  Y +455

+1426  ggaaatagatatggtttatggaaatatatatgtttatggtaacaacggttatggaaatgga
        G  N  R  Y  G  Y  G  N  I  Y  V  Y  G  N  N  G  Y  G  N  G +475

+1486  tatggttatggttacaacggttacagaaatggaaaaacaaatatcattgtaaacaaagga
        Y  G  Y  G  Y  N  G  Y  R  N  G  K  T  N  I  I  V  N  K  G +495

+1546  agtggatatggttatcataacgacggttatggaaatatcaaactgttttataaatattaa
        S  G  Y  G  Y  H  N  D  G  Y  G  N  I  K  L  F  Y  K  Y  * +514

+1606  tgaggagtaagactttccaattttgaaatgaaacaagaatatattggctacacagatatt

+1666  tatgtttcttttgcccaataaaacaattaaagctcaaaaaaaaaaaaaaaa
```
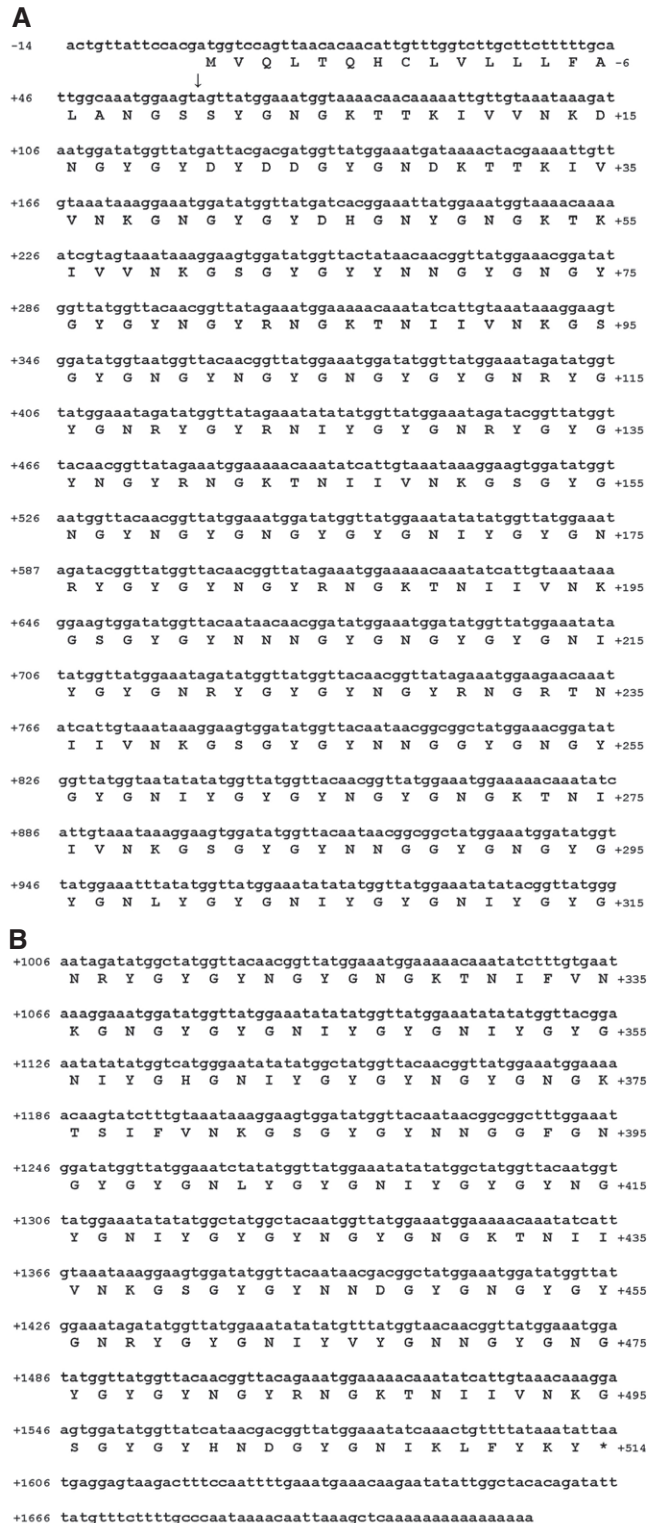
Fig. 4. Assembled full-length sequence of a thread matrix protein (TMP)-1a variant. The numbers on the left indicate the nucleotide number in reference to the start codon. The numbers on the right enumerate the amino acids in reference to the putative signal peptide cleavage site (arrow).

non-aged protein, seven charge variants were present. rTMP aged at 37°C for 3 days and 6 days showed a further decrease in the pI with 10 and 11 observable charge variants, respectively. The mean

```
TMP-1
          YGNGKTTKIVVNKDNGYGYDY
          YGNDKTTKIVVNKGNGYGYDH
          YGNGK.TKIVVNKGSGYG...

Consensus #1   YGNGKTTKIVVNKGNGYGYD.

          YGYGNIYGYGYNGYGNGKTNIIVNKGSGY.....GY.....NNG................
          YGYGNRYGYGYNGYGNGKTNIFVNKGNGY.....GY.....GNIYGY.......GNIYGYG
          YGHGNIYGYGYNGYGNGKTSIFVNKGSGY.....GYnnggfGNGYGYGN.LYGYGNIYGYG
          NGYGNIYGYGYNGYGNGKTNIIVNKGSGYgynndGY.....GNGYGYGN.RYGYGNIYVYG
          NGYGNGYGYGYNGYRNGKTNIIVNKGSGY.....GY.....HN.DGYGNiKHKL..FYKY.

Consensus #2   YGYGNIYGYGYNGYGNGKTNIIVNKGSGY.....GY.....GNGYGYGN.....GNIY.YG
               N                  F

TMP-2
          .........GY.GYGK.TTNVVVNKGS..GY...GYNN.GGYGNGYGYGNIY...
          .......DYGYNGYGNGKTNIIVNKGSryGYnngGYGNGYGYGNIYGYGNIYGYG
          YGYGNRYGYGYNGYGNGKTNIFVNKGN..GY...GYGNIYGYGN......INGIG
          YGYGNRYGYGYNGYGNGKTNIFVNKGN..GY...GYGNIYGYGNRYGFGNIYGYG
          YGYGNRYGYGYNGYGNGQTNIFVNKGN..GY...GYGNIYGY......GNVYGIG
          YGYGNRYGYGYNGYGNGRTNIFVNKGS..GY...GYNNG.GYGNGYGYGNIYGYG
          YDYGNIYGYGYNGYGNGRTNIFVNKGS..GY...GYNNA.GY............

Consensus #1   YGYGNRYGYGYNGYGNGKTNIFVNKGN..GY...GYGNIYGYGN.YGYGNIYGYG
               R              S         N            I

          GHGYGYGYNSYGNGKtTKIVINRGNGYGY
          GNGYGYGYNSYGNGK.TKIIVNKGSGYGY

Consensus #2   G.GYGYGYNSYGNGK.TKI..N.G.GYGY
```

Fig. 5. Repeats and consensus sequences found in thread matrix protein (TMP)-1a and TMP-2a variants.

number of deamidated Asn residues per peptide chain is summarized in Table 3. Protein aged for 0, 3 and 6 days averaged 1.4, 4.9 and 5.70 deamidated Asn residues per peptide chain (calculated as in Materials and methods), respectively. rTMP aged at a physiologically relevant temperature of 14°C (that of local seawater) for 6 days averaged 2.4 deamidated Asn residues per peptide chain.

To verify that the observed charge heterogeneity in recombinant protein was associated with an increase in mass expected for deamidation (1 Da), rTMP was digested with Lys-C and the peptides were analyzed with LC/MS. Three peptides were identified by mass spectrometry (Table 4). Peptides A and B

showed an increase in the isotope distribution of 2–3 and 3–4 Da, respectively, and an almost complete disappearance of the expected monoisotopic mass (Fig. 9B). By contrast, the observed isotopic distribution of peptide C was identical to that calculated. Peptides A and B have five Asn–Gly sequences each and the observed isotopic distribution is consistent with deamidation occurring at multiple sites on both peptides. The small monoisotopic peak indicates that the majority of both peptides A and B had undergone deamidation. Although these peptides are from 'non-aged' recombinant protein, the digestion protocol requires incubation at 37°C overnight, and the insolubility of rTMP requires the presence
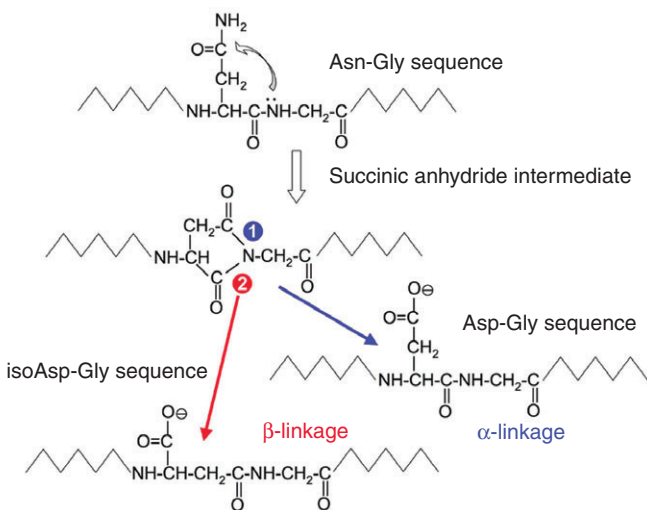

Fig. 6. Asparagine (Asn) deamidation in proteins is highly correlated with Asn–Gly (glycine) sequences. In this spontaneous, non-enzymatic reaction, the amide N of the Asn side chain attacks the peptide carbonyl of the following Gly to form a transient cyclic succinimide. The latter is hydrolyzed by water to yield both D- and L-aspartate (Asp) in both α and β linkages. A β-linked aspartate is known as isoaspartate (isoAsp).
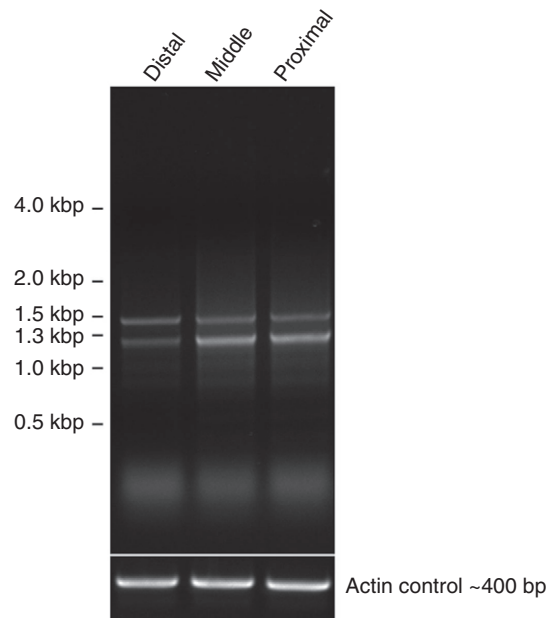

Fig. 7. RT-PCR with total RNA purified from the distal, middle and proximal regions of a mussel foot. Partial TMP-1 (upper band) and TMP-2 (lower band) sequences were amplified from all regions of the foot.

Table 3. Quantification of IsoAspartate (IsoAsp), deamidation and the IsoAsp:Aspartate (IsoAspartate:aspartate) ratio in aged recombinant thread matrix proteins (rTMP)

rTMP treatment	IsoAsp (mol/mol ±s.d.)	Deamidation (mean/peptide)	IsoAsp:Asp (approximate)
Non-aged	D[a,b]	1.4 (4.2)[c]	−
37°C 3 days	1.22±0.10[b]	4.9 (14.7)[c]	1:3
37°C 6 days	1.81±0.09[b]	5.7 (17.1)[c]	1:2
14°C 6 days	0.19±0.03[b]	2.4 (7.2)[c]	1:12

[a]None detected.
[b]$N=3$.
[c]Number in parentheses is the estimated number of deamidation events for a full-length TMP.

of urea. The presence of a denaturant probably increases the rate of deamidation during the digestion; thus, the results are difficult to quantitatively compare with those seen with 2-D electrophoresis or isoAsp quantification.

In model peptides, deamidation at Asn–Gly sites result in a ratio of isoAsp:Asp of 3:1 (Geiger and Clarke, 1987). Thus, isoAsp serves as an indirect measure of deamidation. An isoAsp detection kit was used to quantify isoAsp in non-aged and recombinant protein artificially aged at 37°C for 0, 3 and 6 days in 20 mmol l^{-1} phosphate buffer, pH 8.0. These conditions provide a means for comparing the rate of isoAsp formation in rTMP with that of other proteins aged *in vitro*. IsoAsp in approximately 5 µg of non-aged rTMP protein was below the detection limit whereas the protein aged for 3 and 6 days had 1.22±0.10 and 1.81±0.09 mol isoAsp mol^{-1} protein, respectively. In order to evaluate the rate of deamidation in rTMP at a biologically significant temperature (that of local seawater), protein was aged at 14°C for 6 days and was found to contain 0.19±0.03 pmol isoAsp pmol^{-1} protein (Table 3). The proportion of isoAsp:Asp as a result of deamidation increased with time and temperature. Protein aged at 37°C for 3 and 6 days had an approximate ratio 1:3 and just over 1:2, respectively, and approached 1:12 in rTMP aged at 14°C for 6 days (Table 3).

TMP deamidation in induced threads
To verify the biological significance of deamidation in the TMP proteins, charge heterogeneity was explored in the native TMP's using 2-D electrophoresis followed by detection with anti-rTMP antibodies. The organ that secretes the byssal proteins, the foot, can be induced to secrete the byssal precursors by injection with KCl. These proteins represent the most neonatal form of secreted byssal proteins available. Proteins extracted from induced threads with acetic acid were focused in a pH gradient of 3–11 and separated on a 10% SDS gel. As can been seen in Fig. 10, charge heterogeneity is present in several of the TMP variants immediately upon secretion. The anti-rTMP antibody (see below) does not detect TMP's extracted from mature threads probably because of the
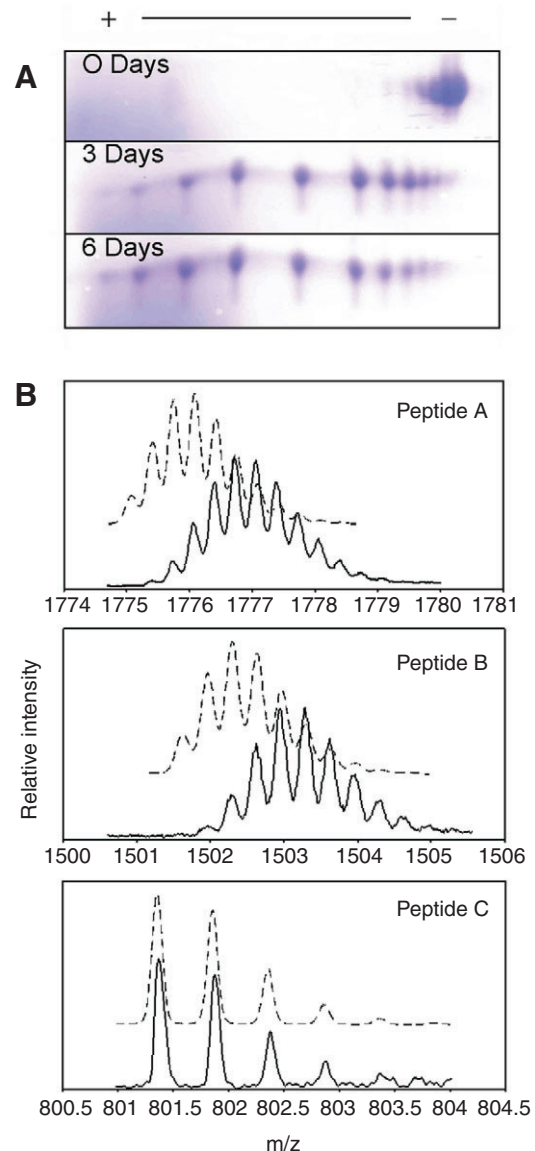

Fig. 9. Deamidation in recombinant thread matrix protein (TMP). (A) Recombinant TMP separated on a 2-D gel [12.5% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE)]. Protein was aged at 37°C in 20 mmol l^{-1} phosphate buffer, pH 8.0, for 0, 3 and 6 days. Proteins were visualized with Coomassie Blue stain. (B) Mass spectra of Lys-C-derived peptides A, B and C from recombinant thread matrix proteins (rTMP). Broken lines represent the predicted isotopic distribution of the [M+3H] ions. Solid lines are the observed isotropic distribution of the [M+3H] ions.

```
         10        20        30        40        50        60
MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDPKNSARGGNIYGHGNIYGYGYNGYGNGK
         70        80        90       100       110       120
TNIFVNKGSGYGYNNGGFGNGYGYGYGNIYGYGYNGYGNIYGYGYNGYGNGKTNIIVNKGSG
        130       140       150       160       170       180
YGYNNDGYGNGYGYGNRYGYGNIYVYGNNGYGNGYGYGYNGYRNGKTNIIVNKGSGYGYH
        190
NDGYGNIKHKLFYKY
```

Fig. 8. Amino acid sequence of recombinant thread matrix protein (TMP). The arrow indicates the end of the vector-encoded sequence. Asparagine–glycine sequences that are likely to deamidate are underlined.

Table 4. Lys-C endoproteinase-derived thread matrix protein peptides identified by mass spectrometry

Peptide	Peptide sequence	M[a]	[M+3H]$^{3+}$ [b]
A	GSGYGYNNDGYGNGYGYGNRYGYGNIYVYGNNGYGNGYGYGYNGYRNGK	5322.2	1775.08
B	GSGYGYNNGGFGNGYGYGNIYGYGYNGYGNIYGYGYNGYGNGK	4501.8	1501.63
C	GSGYGYGHNDGYGNIK	1601.6	534.57

[a]Calculated monoisotopic mass, [b][M+3H]$^{3+}$ monoisotopic mass.

accumulated post-secretory modifications or conformational changes associated with aging.

IsoAsp accumulation in byssal threads is summarized in Table 5. The assay was unable to detect any isoAsp in 40 μg of induced extracts. After 1 day, 570.4±72.8 pmol isoAsp mg^{-1} protein was detectable in crude extracts from the distal thread. IsoAsp levels rose to 1536.9±191.8 pmol isoAsp mg^{-1} of protein in crude extracts from threads that were ≤7 days old. In order to determine whether the change in isoAsp amounts was due to a change in the composition of the extractable proteins at each time point, amino acid analysis results were compared from each time point (not shown). The composition of the extracts did not vary significantly.

Polyclonal antibody generated to a rTMP sequence readily recognized the recombinant protein when transferred from both AU PAGE and SDS PAGE gels (data not shown). The antibody did not react with any proteins extracted from 1 day old or 7 day old byssal threads. The anti-rTMP antibody recognized a single band from induced thread extracts migrating similarly to the upper band on AU PAGE (Fig. 11A) that was partially purified directly from the distal thread (Fig. 2A,B). The anti-rTMP antibodies recognized a large band with a molecular mass of 100 kDa, and a group of proteins centered around 55 kDa in induced extracts separated by SDS PAGE (Fig. 11B). Variability was observed in the detection of the smaller group of proteins.

The results for both AU PAGE and SDS PAGE westerns with induced threads, although reproducible, were variable between individual extractions. When the antibody recognized protein(s), it was invariably the same protein bands; however, extracts often provided no reactivity with the antibody. Pre-immune type Y immunoglobulins (IgYs) showed no reactivity with any proteins extracted from the foot or byssal thread.

Localization of TMP in mussel foot

The lack of reactivity between the anti-rTMP antibody and proteins extracted from the distal byssal thread prompted efforts to localize TMPs and/or their precursors in the foot tissue. Two types of secretory granules are present in the portion of the foot responsible for secreting the distal byssal thread collagen granules, whose contents form the interior of the thread and accessory gland granules that coalesce to coat the thread. The TMPs were localized to the immature (or condensing) and mature collagen granules but not to the accessory gland granules (Fig. 12). Resolution was inadequate to precisely assess the relationship between preCOLs and TMP. The colloidal gold procedure used here was precisely as described earlier by Qin and Waite (Qin and Waite, 1995).

DISCUSSION
TMP

The existence of a matrix surrounding bundles of preCOL fibers in the distal thread is consistent with evidence available from scanning electron microscopy (SEM), atomic force microscopy (AFM) and compositional analysis. Indeed, a non-collagenous Asn-, Tyr- and Gly-rich byssal TMP was extracted and partially characterized (Table 1) from the distal byssal thread of *M. galloprovincialis*. The TMP variants bear no resemblance to tendon matrix proteoglycans such as decorin and COMP (Kjær, 2004); indeed, they lack significant homology with any protein sequence on the database. The anomalous elution of the TMPs from RP-HPLC is probably due to heterogeneity from alternative splicing, multiple gene copies, hydroxylation and spontaneous deamidation (further discussed below). Seven Lys-C-derived peptides compositionally related to the TMPs were sequenced (Table 2). Of these, four have exact matches with cloned TMP sequences. The unmatched peptides (#30, 44 and 54), although clearly related to the TMP proteins, presumably come from additional variants that have yet to be cloned.

TMP heterogeneity

In total, 10 TMP transcriptional variants have been partially cloned using PCR, 5′ RACE and RT-PCR. The TMP transcripts can be placed into two groups, TMP-1 and TMP-2, and are distinguishable by the initial N-terminal repeat (see Figs S3 and S4 in supplementary data). Both groups contain the same 5′ UTR and signal peptide and
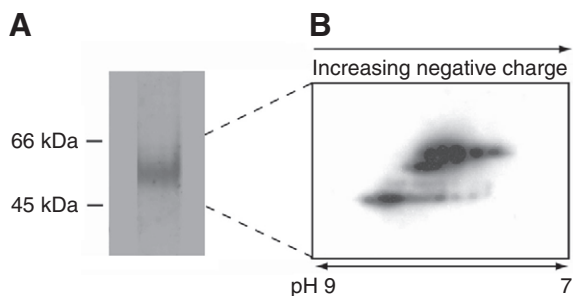

Fig. 10. Thread matrix protein (TMP)-antibody cross-reactivity at higher resolution. (A) Induced extracts separated on a one-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) gel and subsequently processed with anti-rTMP (only the relevant portion is shown). Proteins were typically not abundant enough for detection with Coomassie Blue stain. (B) Two-dimensional western blot (pH range ~7–9) of the indicated proteins with anti-rTMP. Charge heterogeneity present in TMPs extracted from induced threads is evident as a laddering of spots. The 2-D blot has been enhanced so that the individual protein spots are clearly visible.

Table 5. Quantification of IsoAspartate (IsoAsp) in proteins extracted from byssal threads

Thread age	IsoAsp[a] (pmol mg^{-1} ±s.d.)
0 days (induced thread)	ND[b,c]
1 day old	570±72.8[c]
≤7 days old	1536.9±191.8[c]

[a]IsoAsp is expressed in pmol IsoAsp mg^{-1} of protein for comparison with other values reported in the literature.
[b]None detected.
[c]N=3.
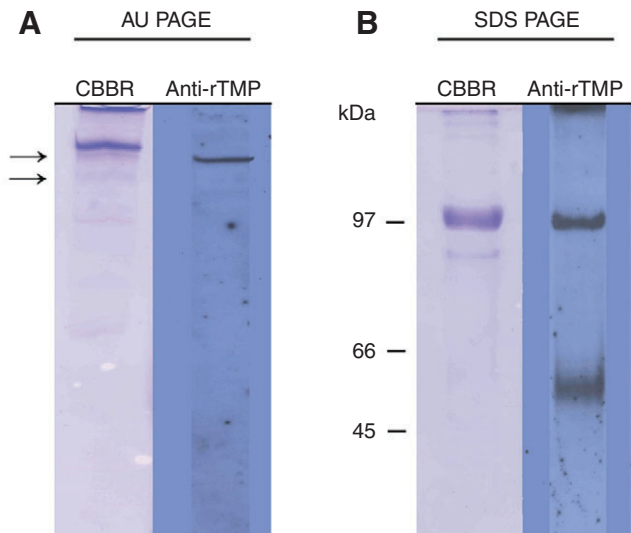
A AU PAGE

B SDS PAGE

Fig. 11. Western blotting of extracted induced threads using polyclonal anti-recombinant thread matrix proteins (rTMP) antibodies. Acetic acid extracts of induced threads (~10 µg) were separated on an acid–urea polyacrylamide gel electrophoresis (AU PAGE) gel (A) or an sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) gel (B) and processed for western blotting as described in the Materials and methods. Primarily, one band stains in AU (acid–urea) whereas there are two on SDS PAGE, of which the slower seems likely to be a dimer. The arrows in A correspond to the proteins that were partially purified in Fig. 2. CBBR, Coomassie Brilliant Blue R.
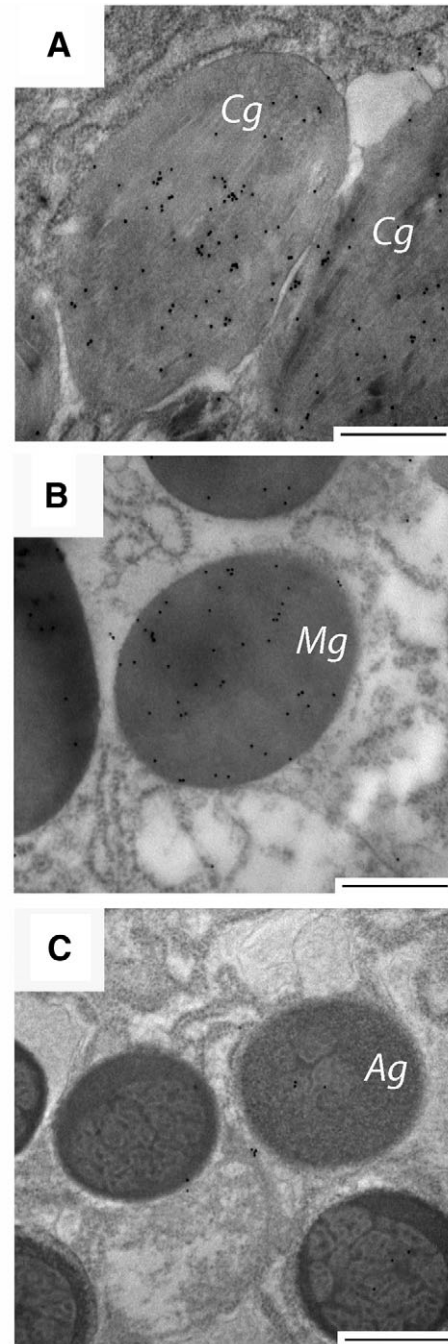

Fig. 12. Colloidal gold localization of thread matrix proteins (TMPs) in (A) condensing collagen granules (Cg) and (B) mature collagen granules (Mg). The TMPs did not localize to the (C) accessory granules (Ag). The scale bars in all micrographs are 500 nm.

the four partial C-terminal clones all contained the same 3′ UTR; however, it is not clear whether the C-terminal clones are sequences from both TMP-1 and TMP-2 variants. The TMPs are composed of partially conserved repeats (Fig. 5). An assembled full-length TMP-1 transcript (Fig. 4) encodes a protein of ~56 kDa, which is consistent with the observed MALDI-TOF mass of the partially purified proteins (Fig. 2C). A full-length clone or assemblage of TMP-2 is lacking; however, eight variants have been cloned and partially sequenced (Fig. S4 in supplementary material).

Unlike the preCOLs (Qin and Waite, 1998), RT-PCR expression analysis failed to identify gradients in the expression of the TMP-1 and TMP-2 groups of sequences. Although no gradient in intra-group variants was observed, the analysis of the differences between intra-group variants is confounded by their repetitiveness, similarity in sequence and the amount of transcriptional variation. The fact that the 5′ UTR from TMP-1 and TMP-2 clones and 3′ UTR are identical for all clones that contain these regions supports alternative splicing as a significant source of transcriptional heterogeneity in the TMPs. Furthermore, a genomic Southern blot was consistent with the presence of 1–3 gene copies (data not shown). Evidence for inter-individual variation was also found and is consistent with recent work demonstrating a high frequency of nucleotide polymorphisms in bivalves (Saavedra and Bachere, 2006).

TMP and Dopa
Peptide 54 contained the only detectable sequenced Dopa residue. Composition analysis of hydrolyzed protein suggests that at most 15% of the tyrosine residues in the TMPs are converted to Dopa and the location or pattern of post-translational hydroxylation is not known. Dopa residues in the byssal thread are capable of forming H-bonds, covalent diDopa and non-covalent metal chelate complexes (McDowell et al., 1999; Taylor et al., 1996). The Dopa

present in the TMPs may engage in similar interactions, which could be responsible for the reduced extractability of proteins as threads age.

Deamidation in TMP
The tendency of Asn residues in peptides and proteins to undergo spontaneous, non-enzymatic deamidation resulting in formation of either Asp or isoAsp has been well documented *in vivo* and *in vitro* (Robinson and Rudd, 1974). Deamidation measurably changes

affected proteins; there is racemization of the α-carbon, an increase in mass (1 Da/Asn), introduction of negative charge and, in the case of isoAsp, formation of a backbone β-linkage (Geiger and Clarke, 1987) (Fig. 6). Factors influencing rates of spontaneous deamidation include temperature, buffering ion, ionic concentration and pH (Brennan and Clarke, 1995). The TMPs have numerous Asn–Gly sequences, which appear to be the chief targets of rapid spontaneous deamidation (Robinson et al., 2004). The assembled full-length TMP-1a clone has 44 Asn–Gly sequences. If only a fraction of these sites deamidate, significant heterogeneity in the primary structure would be present. Indeed, rTMP was shown to deamidate rapidly at multiple sites.

The majority of the research on the functional significance of deamidation has been limited to cellular systems and long-lived biological materials. Long-lived proteins such as those found in acellular structures and extracellular matrices are particularly vulnerable to deamidation and aspartyl isomerization and racemization because of slow or no protein turnover. Deamidated, racemized and isomerized aspartates have been identified in a number of extracellular proteins and connective tissues including type I collagen, tooth enamel and eye lens; however, the proportion of racemized and isomerized Asp derived from Asn is not known in all cases (Fledelius et al., 1997; Helfman and Bada, 1975; Lampi et al., 1998). The discovery of an extracellular matrix protein complex in mammalian brain, high mass methyl-accepting protein (HMAP), which has naturally high levels of isoaspartate (generated from deamidation of Asn and isomerization of Asp), provided the first suggestion that spontaneous deamidation and isoAsp formation may have a functional role in extracellular matrices (David et al., 1998; Orpiszewski and Aswad, 1996). The rapid deamidation of another matrix protein sequence (TMPs) from mussel byssus and the accumulation of isoAsp in proteins extracted from the byssal threads of the marine mussel, *M. galloprovincialis* are established in the present study.

IsoAsp detection in TMP

Isoaspartic acid was below the detection limit of the Isoquant assay in freshly secreted byssal proteins induced by KCl injection; however, charge heterogeneity was present in the TMPs as detected by western blotting. Isoaspartic acid rapidly accumulated to $570.4\pm72.8\,\mathrm{pmol\,mg^{-1}}$ protein in extracts from 1 day old threads and tripled within one week.

The ratio of isoaspartic acid:Asp as a result of deamidation in model peptides is typically about 3:1 (Geiger and Clarke, 1987). When the isoaspartic acid:Asp ratio is estimated from the quantification of the total deamidation in rTMP, a ratio of 1:2 and 1:12 after 6 days at 37°C and 14°C is observed, respectively. The isoaspartic acid:Asp ratio at specific sites of deamidation within proteins was shown to vary from the 3:1 ratio observed in model peptides and can be considerably less (Johnson et al., 1989; Kossiakoff, 1988), which probably reflects secondary and tertiary structure effects. If the TMPs are the major source of isoaspartic acid in the byssus, then isoaspartic acid quantification significantly underestimates the amount of deamidation that has taken place. If the isoaspartic acid/deamidation ratio obtained from aging rTMP at 14°C is used to estimate the amount of deamidation in byssal thread extracts, the number of deamidated Asn residues in extracts from 1 week old threads is approximately $20\,\mathrm{nmol\,mg^{-1}}$ protein. This is equal to about 0.23% of the total amino acids in the extracted proteins or 2.03% of the aspartic acid and asparagine residues.

Significance of deamidation

Deamidation has been reported in numerous proteins *in vitro*, and a comprehensive list has been compiled (Wright, 1995). The effect that deamidation has on protein activity/function ranges from no measurable change to a complete loss of activity. The speed with which the rTMPs deamidate *in vitro* and the accumulation of isoAsp in the byssal thread is intriguing in that it raises the possibility that TMP sequences prone to deamidation play an adaptive role. The introduction of multiple negative charges and backbone β-linkages are likely to exert a tremendous effect on TMP structure.

An intriguing aspect of TMP deamidation concerns the staggering amount of structural heterogeneity endowed upon a single thread. Not only are there multiple TMP transcriptional variants within each mussel but spontaneous deamidation of Asn would result in the L or D forms of Asp and isoAsp. Even if only a fraction of the approximately 44 Asn–Gly sequences in a full-length TMP transcript were to undergo deamidation, the global heterogeneity in primary structure of the TMPs would be exceptional.

The group of proteins recognized by the anti-rTMP antibody that migrate at approximately 55 kDa on SDS PAGE are ostensibly the TMPs and are also the most variably detected between individual-induced extracts whereas the proteins detected as a thick band at 100 kDa are always recognized by the anti-rTMP antibody. A likely explanation is that the 100 kDa band is a dimerization or aggregation of TMPs. The results from AU PAGE showed a single band and would seem to be at odds with the results for proteins separated on SDS PAGE in that a variable number of bands are typically observed on SDS PAGE. It may be that under conditions necessary for SDS PAGE (particularly the elevated pH) the TMPs form aggregates that are not present in acetic acid and $4\,\mathrm{mol\,l^{-1}}$ urea.

The band recognized when induced thread extracts are separated on AU PAGE corresponds to the upper band purified from the mature threads (see Fig. 2). The anomalous behavior of the anti-rTMP antibody when used to detect the protein extracted from mature byssal threads, while suggestive of modifications to the TMPs, is rather confounding and currently lacks a definitive explanation.

The western blots from both the AU PAGE and SDS PAGE suggest that the chemical and structural changes undergone by TMPs during thread formation and maturation may have affected the immunoreactivity of the protein. These include hydroxylation, deamidation, intra- and intermolecular cross-links and secondary structure changes. All of these, excepting limited deamidation, would be absent from the rTMP antigen originally used to generate antibodies. Notably, the chemical and structural changes must be insensitive to acid extraction and SDS denaturation as TMPs extracted from mature threads had no antigenicity. Anderson et al. (Anderson et al., 2000) observed antibody reactivity with dpfp-1 extracted from zebra mussel foot but attempts to localize the protein within the mature thread failed. The interpretation of the results was that epitopes recognized by the antibody were shielded in the mature thread. Another example of epitope masking in a biological material is found in the eggshell proteins of the liver fluke *Fasciola hepatica*. Antibodies which were raised against the Dopa containing eggshell precursor protein vitelline protein B (vpB) recognized the protein in the eggshell protein globules within the vitelline cells. As the eggshells matured, vpB lost immunoreactivity (Robinson et al., 2001).

Collagen granules are secreted from the collagen gland and go through several identifiable steps during maturation in the foot (Tamarin and Keller, 1972). The most immature collagen granules are referred to as condensing vacuoles, in which the preCols are

loosely packed. As the granule matures, the preCols become oriented in tightly packed smectic arrays. The anti-TMP antibody localized to the condensing and mature collagen granules (Fig. 12), which provides support for the hypothesis that the TMPs form a matrix around preCol fibers. Further evidence for the association of the TMPs and preCols comes from the expression analysis; both groups of proteins are located throughout the entire length of the foot.

Recently Harrington and Waite (Harrington and Waite, 2008a; Harrington and Waite, 2008b) have studied the ability of purified preCols to form into fibers from solution. The observation that preCol fibers drawn from solutions of purified protein have remarkably similar morphology to those found in the distal end of the thread argues that the TMPs have a limited role in determining fiber morphology with the following exceptions: (1) fibers drawn from solution averaged 5 μm in diameter contrasting with several hundred nm in threads, and (2) the mechanical properties of the drawn fibers show a far less dramatic yield than that seen in whole threads. It may be that the presence of the TMPs prevents larger fibers of preCols from forming upon secretion. The discrepancy between the mechanical properties of the drawn preCol fibers and the distal byssus was argued to be due to the lack of metal ions and covalent diDopa crosslinks, both of which are known to have effects on the mechanics of the thread. We would add that a matrix composed of TMPs is likely to contribute to the observed difference in fiber diameter and mechanical properties.

The role that global heterogeneity induced by spontaneous deamidation and transcriptional heterogeneity plays in the function of the TMPs remains undetermined though in the case of deamidation some intriguing possibilities exist. The pH dependence of deamidation would predict little to no activity in TMP during storage in the secretory granules (pH 5) whereas upon secretion into the seawater environment (pH 8), the rate of deamidation would increase. Deamidation in the TMPs could thus be a pH-triggered switch involved in the rapid maturation of the byssal precursors to a functional thread. The accumulation of negative charges in TMP should dramatically change electrostatic attractions and repulsions between the TMPs and other macromolecular components of the byssal thread.

TMP deamidation has a potential to alter protein function in two ways: (1) by switching from a partnership with liquid crystals to a matrix for load bearing fibers, and (2) as a switch from function to decomposition. Moeser and Carrington (Moeser and Carrington, 2006) recently introduced the concept of a seasonally adjusted byssal 'thread quality' in New England *M. edulis*, with best and worst mechanical properties correlated with spring and autumn, respectively. Given the high sensitivity of deamidation to pH, temperature and salinity, the role that TMP instability has on byssal thread and holdfast quality deserves closer scrutiny.

LIST OF ABBREVIATIONS

|	
Asp	aspartate
Asn	asparagine
Asx	= Asn + Asp
AU	acid–urea
COMP	cartilage oligomeric protein
Gly	glycine
HMAP	high mass methyl-accepting protein
isoAsp	isoAspartate mfp-1
MALDI	matrix-assisted laser desorption ionization
PAGE	polyacrylamide gel electrophoresis
preCOLs	collagen-containing proteins
rTMP	recombinant thread matrix protein
RACE	rapid amplification to cDNA ends
SDS	sodium dodecyl sulfate
TMP	thread matrix protein
Tyr	tyrosine
TOF	time of flight
vpB	vitalline protein B

REFERENCES

Anderson, K. E. and Waite, J. H. (2000). Immunolocalization of Dpfp1, a byssal protein of the zebra mussel (*Dreissena polymorpha*). *J. Exp. Biol.* **203**, 3065-3076.

Aouacheria, A., Geourjon, C., Aghajari, N., Navratil, V., Deleage, G., Lethias, C. and Exposito, J. Y. (2006). Insights into early extracellular matrix evolution. *Mol. Biol. Evol.* **23**, 2288-2302.

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

Brennan, T. V. and Clarke, S. (1995). Deamidation and isoaspartate formation in model synthetic peptides: the effect of sequence and solution environment. In *Deamidation and Isoaspartate Formation in Peptides and Proteins* (ed. D. W. Aswad), pp. 65-90. Boca Raton, FL: CRC Press.

Brown, C. H. (1975). *Structural Materials in Animals*, p. 448. New York: Wiley and Sons.

Collings, P. J. (2002). *Liquid Crystals: Nature's Delicate Phase of Matter*, 2nd edn. Princeton, NJ: Princeton University Press.

David, C. L., Orpiszewski, J., Zhu, X. C., Reissner, K. J. and Aswad, D. W. (1998). Isoaspartate in chondroitin sulfate proteoglycans of mammalian brain. *J. Biol. Chem.* **273**, 32063-32070.

Fledelius, C., Johnsen, A. H., Cloos, P. A. C., Bonde, M. and Qvist, P. (1997). Characterization of urinary degradation products derived from type I collagen: identification of a β-isomerized Asp-Gly sequence within the C-terminal telopeptide (α1) region. *J. Biol. Chem.* **272**, 9755-9763.

Geiger, T. and Clarke, S. (1987). Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides: succinimide-linked reactions that contribute to protein degradation. *J. Biol. Chem.* **262**, 785-794.

Goldberg, W. M. (1974). Evidence of a sclerotized collagen from skeleton of a gorgonian coral. *Comp. Biochem. Physiol.* **49B**, 525-529.

Gosline, J., Lillie, M., Carrington, E., Guerrette, P., Ortlepp, C. and Savage, K. (2002). Elastic proteins: biological roles and mechanical properties. *Philos. R. Soc. Lond. B Biol. Sci.* **141B**, 121-132.

Harrington, M. J. and Waite, J. H. (2007). Holdfast heroics: comparing the molecular and mechanical properties of *Mytilus californianus* byssal threads *J. Exp. Biol.* **210**, 4307-4318.

Harrington, M. J. and Waite, J. H. (2008a). pH-Dependent locking of giant mesogens in fibers drawn from mussel byssal collagens. *Biomacromolecules* **9**, 1480-1486.

Harrington, M. J. and Waite, J. H. (2008b). How nature modulates a fiber's mechanical properties: mechanically distinct fibers drawn from natural mesogenic block copolymer variants. *Adv. Mater.* **20**, 1-5.

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**, 1351-1355.

Helfman, P. M. and Bada, J. L. (1975). Aspartic acid racemization in tooth enamel from living humans. *Proc. Natl. Acad. Sci. USA* **72**, 2891-2894.

Johnson, B. A., Shirokawa, J. M., Hancock, W. S., Spellman, M. W., Basa, L. J. and Aswad, D. W. (1989). Formation of isoaspartate at two distinct sites during *in vitro* aging of human growth hormone. *J. Biol. Chem.* **264**, 14262-14271.

Kjær, M. (2004). Role of ECM in adaptation of tendon and skeletal muscle to mechanical loading. *Physiol. Rev.* **84**, 649-698.

Kossiakoff, A. A. (1988). Tertiary structure is a principle determinant to protein deamidation. *Science* **240**, 191-194.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

Lampi, K. J., Ma, Z., Hanson, S. R. A., Azume, M., Shih, M., Shearer, T. R., Smith, D. L., Smith, J. B. and David, L. L. (1998). Age-related changes in human lens crystallins identified by two-dimensional electrophoresis and mass spectrometry. *Exp. Eye Res.* **67**, 31-43.

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.

McDowell, L. M., Burzio, L. A., Waite, J. H. and Schaefer, J. (1999). Rotational echo double resonance detection of cross-links formed in mussel byssus under high-flow stress. *J. Biol. Chem.* **274**, 20293-20295.

Moeser, G. M. and Carrington, E. (2006). Seasonal variation in mussel byssal thread mechanics. *J. Exp. Biol.* **209**, 1996-2003.

Orpiszewski, J. and Aswad, D. W. (1996). High mass methyl-accepting protein (HMAP), a highly effective endogenous substrate for protein L-isoaspartyl methyltransferase in mammalian brains. *J. Biol. Chem.* **271**, 22965-22968.

Panyim, S. and Chalkley, G. R. (1969). High resolution acrylamide gel electrophoresis of histones. *Arch. Biochem. Biophys.* **130**, 337-346.

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.

Robinson, A. B. and Rudd, C. J. (1974). Deamidation of glutaminyl and asparaginyl residues in peptides and proteins. *Curr. Top. Cell. Regul.* **8**, 247-295.

Robinson, M. W., Colhoun, L. M., Fairweather, I., Brennan, G. P. and Waite, J. H. (2001). Development of the vitellaria of the liver fluke, *Fasciola hepatica* in the rat host. *Parasitology* **123**, 509-518.

Robinson, N. E., Robinson, Z. W., Robinson, B. R., Robinson, A. L., Robinson, J. A., Robinson, M. L. and Robinson, B. A. (2004). Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides. *J. Peptide Res.* **63**, 426-436.

Rusaouën, M., Pujol, J. P., Bocquet, J., Veillard, A. and Borel, J. P. (1976). Evidence of collagen in the egg capsule of the dogfish, *Scyliorhinus canicula. Comp. Biochem. Physiol.* **53B**, 539-543.

Saavedra, C. and Bachere, E. (2006). Bivalve genomics. *Aquaculture* **256**, 1-14.

Spiro, R. G., Lucas, F. and Rudall, K. M. (1971). Glycosylation of hydroxylysine in collagens. *Nature New Biol.* **231**, 54-55.

Tamarin, A. and Keller, P. J. (1972). An ultrastructural study of the byssal thread forming system in *Mytilus. J. Ultrastruct. Res.* **40**, 401-416.

Tamarin, A., Lewis, P. and Askey, J. (1976). Structure and formation of byssus attachment plaque in *Mytilus. J. Morphol.* **149**, 199-221.

Taylor, S. W., Chase, D. B., Emptage, M. H., Nelson, M. J. and Waite, J. H. (1996). Ferric ion complexes of a DOPA-containing adhesive protein from *Mytilus edulis. Inorg. Chem.* **35**, 7572-7577.

Vitellaro-Zuccarello, L. (1980). The collagen gland of *Mytilus galloprovincialis*: an ultrastructural and cytochemical study on secretory granules. *J. Ultrastruct. Res.* **73**, 135-147.

Waite, J. H. (1995). Precursors of quinone-tanning: dopa-containing proteins. *Methods Enzymol.* **258**, 1-20.

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

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

Wright, H. T. (1995). Amino acid abundance and sequence data: clues to the biological significance of nonenzymatic asparagine and glutamine deamidation in proteins. In *Deamidation and Isoaspartate Formation in Peptides and Proteins* (ed. D. W. Aswad). Boca Raton, FL: CRC Press.

A
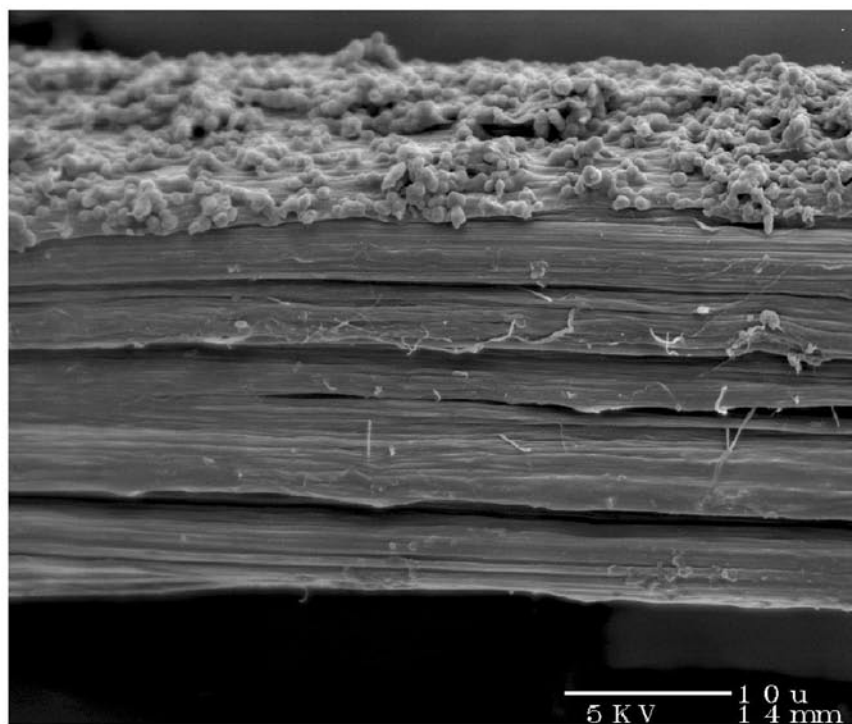
B
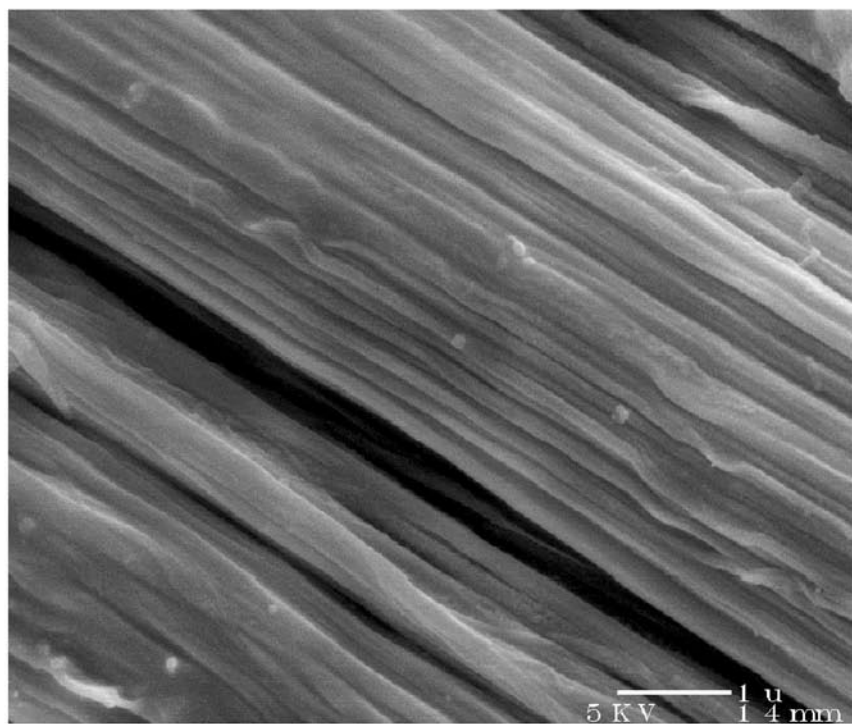
A

```
seq C1    GNIYGHGNIYGYGYNGYGNGKTNIFVNKGSGYGYNNGGFGNGYGYGN------IYGYGYN
seq C2    ------------GYNGYGNGKTNIFVNKGSGYGYNNGGFGNGYGYGNLYGYGNIYGYGYN
seq F3T7  ------------------------------------------------------------


seq C1    GYGNIYGYGYNGYGNGKTNIIVNKGSGYGYNNDGYGNGYGYGNRYGY------GNIYVYG
seq C2    GYGNIYGYGYNGYGNGKTNIIVNKGSGYGYNNDGYGNGYGYGNRYGYGNRYGYGNIYVYG
seq F3T7  ---------------------------------------GYGYGNRYGY------GNIYVYG
                                                 *********      *******


seq C1    NNGYGNGYGYGYNGYRNGKTNIIVNKGSGYGYHNDGYGNIKHKLFYKY
seq C2    NNGYGNGYGYGYNGYRNGKTNIIVNKGSGYGYHNDGYGNIKHKLFYKY
seq F3T7  NNGYGNGYGYGYNGYRNGKTNIIVNKGSGYGYHNDGYGNIKHKLFYKY
          ************************************************
```

B

```
3' UTR   taatgaggagtaagactttccaattttgaaatgaaacaagaatatattggctacacag

         atatttatgtttcttttgcccaataaaacaattcaaagcgtcaaaaaaaaaaaaaaaa
```

```
                                      ↓
TMP-1a    MVQLTQHCLVLLLFALANGSSYGNGKTTKIVVNKDNGYGYDYDDGYGNDKTTKIVVNKGN
TMP-1b        QHCLVLLLFALANGSSYGNGKTTKIVVNKDNGYGYDYDDGYGNDKTTKIVVNKGN
              ********************************************************


TMP-1a    GYGYDHGNYGNGKTKIVVNKGSGYGYYNNGYGNGYGYGYNGYRNGKTNIIVNKGSGYGNG
TMP-1b    GYGYDHGNYGNGKTKIVVNKGSGYGYYNNGYGNGYGYGYNGYRNGKTNIIVNKGSGYGNG
          ************************************************************


TMP-1a    YNGYGNGYGYGNRYGYGNRYGYRNIYGYGNRYGYGYNGYRNGKTNIIVNKGSGYGNGYNG
TMP-1b    YNGYGNGYGYGNRYGYGNRYGYRNRYGYGNRYGYGYNGYRNGKTNIIVNKGSGYGNGYNG
          *********************** ************************************


TMP-1a    YGNGYGYGNIYGYGNRYGYGYNGYRNGKTNIIVNKGSGYGYNNNGYGNGYGYGNIYGYGN
TMP-1b    YGNGYGNG------------------------------------YNGYGNGYGYGNIYGYGN
          ******                                  ****************


TMP-1a    RYGYGYNGYRNGRTNIIVNKGSGYGYNNGGYGNGY...
TMP-1b    RYGYGYNGYRNGKTNIIVNKGSGYGYNNNGYGNGY...
          *********** ***************** *****
```

```
TMP-2a   GYGNGYGYGNIYGYGNIY------DYGNIYGYGYNGYGNGRTNIFVNKGSGYGYNNAGYG
TMP-2b
TMP-2c   GYGNGYGYGNIYGYGNIYDYGNIYGYGNRYGYGYNGYGNRRTNIFVNKGSGYGYNNAGYG
TMP-2d
TMP-2e   ---------------------------IYGYGNRYGYGYNG---------------------------
TMP-2f   -------------------KIVINRGNGYGYNSNGYGNDRTHIIVNKGSGYGYNNDGYG
TMP-2g
TMP-2h


TMP-2a   HGYGYGYNSYGNGKTTKIVINRGNGYGYNSNGYGNDRTHIIVNKGSGYGYNNDGYGNGYG
TMP-2b
TMP-2c   HGYGYGYNSYGNGKTTKIVINRGNGYGYNSNGYGNDRAHIIVNKGSGYGYNYDGYGNGYG
TMP-2d
TMP-2e   ---------YGNGKTTKIVINRGNGYGYNSNGYGNDRTHIIVNKGSGYGYNNDGYGNGYG
TMP-2f   HGYGYGYNSYGNGKTTKIVINRGNGYGYNSNGYGNDRTHIIVNKGSGYGYNNDGYGNGYG
TMP-2g
TMP-2h


TMP-2a   YGYNSYGNGKTKIIVNKGSGYGY...
TMP-2b
TMP-2c   YGYNSYGNGKTKIIVNKGSGYGY...
TMP-2d
TMP-2e   YGYNSYGNGKTKIIVNKGSGYGY...
TMP-2f   YGYNSYGNGKTKIIVNKGSGYGYHNDGYGNGYGYGNRYGYGNIYVYGNNGYGNGYGY...
TMP-2g
TMP-2h

                              ↓
TMP-2a       QHCLVLLLFALANGSGYGYGKTTNVVVNKGSGYGYNNGGYGNGYGY------GNI
TMP-2b   MVQLTQHCLVLLLFALANGSGYGYGKTTNVVVNKGSGYGYNNGGYGNGYGY------GNI
TMP-2c       QHCLVLLLFALANGSGYGYGKTTNVVVNKGSGYGYNNGGYGNGYGYGYGHIYGYGNI
TMP-2d       QHCLVLLLFALANGSGYGYGKTTNVVVNKGSGYGYNNGGYGNGYGY------GNI
TMP-2e       QHCLVLLLFALANGSGYGYGKTTNLVVNKGSGYGYNNGGYGNGYGY------GNI
TMP-2f       QHCLVLLLFALANGSGYGYGKTTNVVVNKGSGYGYNNGGYGNGYGY------GNI
TMP-2g       QHCLVLLLFALANGSGYGYGKTTNVVVNKGSGYGYNNGGYGNGYGY------GNI
TMP-2h       QHCLVLLLFALANGSSYGYGKTTNVVVNKGSGYGYNNGGYGNGYGY------GNI
             *************** ******* ********************      ***


TMP-2a   YDYGYNGYGNGKTNIIVNKGSRYGYNNGGYGNGYGYGNIYGYGNIYGYGNIYGYGNRYGY
TMP-2b   YGYGYNGYGNGKTNIIVNKGSRYGYNNGGYGNGYGYGNIY-----------------GY
TMP-2c   YGYGYNGYGNGKTNIIVNKGSRYGYNNGGYGNGYGYGNIYGYG------NIYGYGNRYGY
TMP-2d   YDYGYNGYGNGKTNIIVNKGSRYGYNNGGYGSGYGYGNIYGYG------NIYDYGNIYGY
TMP-2e   YDYGYNGYGNGKTNIIVNKGSRYGYNNGGYGNGYGYGNIYGYGNIYGYGNIYGYGNRYGY
TMP-2f   YDYGYNGYGNGKTNIIVNKGSRYGYNNGGYGNGYGYGNIYGYGNIYGYGNIYGYGNRYGY
TMP-2g   YDY------------------------------GNIYDYG------NIYGYGNRYGY
TMP-2h   YGYGYNGYGNGKTNIIVNKGSRYGYNNGGYGNGYGYGNIYGYG------NIYGYGNRYGY
         *  *                             **


TMP-2a   GYNGYGNGKTNIFVNKGNGYGYGNI------YGYGNINGIGN----IYGYGNRYGYGYNG
TMP-2b   GYNGYGNGKTNIIVNKGSRYGYNNGGYGNGYGYN-GIY----------------------
TMP-2c   GYYGYGNGKTNIFVNKGNGYGYGNI------------YGIGN----IYGYGNRYGYGYNG
TMP-2d   GYNGYGNGRTNIFVNKGSGYGYNNAGYGHGYGYGYNSYGNGKTTKIVINRGNGYGYNSNG
TMP-2e   GYYGYGNGKTNIFVNKGNGYGYGNI-----YGYG-NINGIGN----IYGYGNRYGYGYNG
TMP-2f   GYNGYGNGKTNIFVNKGNGYGYGNI-----YGYG-NINGIGN----IYGYGNRYGYGYNG
TMP-2g   GYNGYGNGRTNIFVNKGSGYGYNNAGYGHGYGYGYNSYGNGITTKIVINRGNGYGYNSNG
TMP-2h   GYNGYGNGKTNIFVNKGNGYGYGNI-----YGYG-NIYGIGN----IYGYGNRYGYGYNG
         ** ***** *** ****    ***


TMP-2a   YGNGKTNIFVNKGNGYGYGNIYGYGNRYGFGNIYGYGNILYGYGNRYGYGYNGYGNGQTN
TMP-2b   ----------------------------------------GYGNTYGYG...
TMP-2c   YGNGKTNIFVNKGNGYGYGNIYGYGNIYGYGNI-------YGYGNRYGYGYNGYGNGKTN
TMP-2d   YGNDRTHIIVNKGSGYGYNN-------------------DGYGNGYGYGYNSYGNGKTK
TMP-2e   YGNGKTNIFVNKGNGYGYGNIYGYGNRYGFGNIYGYGNILYGYGNRYGYGYNGYGNGKTN
TMP-2f   YGNGRTNIFVNKGSGYGYNN-GGYGNGYGYGNIYGYGNI-YDYGNIYGYGYNGYGNGRTN
TMP-2g   YGNDRAHIIVNKGSGYGYNY-------------------DGYGNGYGYGYNSYGNGKTK
TMP-2h   YGNGKTNIFVNKGNGYGYGNIYGYGNIYGYGNI-------YGYGNRYGYGYNGYGNGKTN
                                                 *** **


TMP-2a   IFVNKGNGYGYGNIYGYGNVYGIGNIYGYGNRYGYGYNGYGNGRTNIFVNKGSGYGYNNG
TMP-2b
TMP-2c   IFVNKGNGYGYGNIYGYGN-----------RYGYGYNGYGNGRTNIFVNKGSGYGYNND
TMP-2d   IIVNKGSGYGY...
TMP-2e   IFVNKGNGYDYGNIYGYGN-------------VYG-------IGN----------------
TMP-2f   IFVNKGSGYGYNNA-GYGH-----------GYGYGYNSYGNGKTT--------------
TMP-2g   IIVNKGSGYGY...
TMP-2h   IFVNKGNGYGYGNIYGYGN-----------IYG...
```