

14. D. W. Nebert *et al.*, *DNA Cell Biol.* **10**, 1 (1991).
15. P. Su *et al.*, *ibid.* **9**, 657 (1990); E. Usui, M. Noshiro, Y. Ohyama, K. Okuda, *FEBS Lett.* **262**, 367 (1990).
16. S. Itoh *et al.*, *Biochim. Biophys. Acta.* **1264**, 26 (1995).
17. J. Picado-Leonard and W. L. Miller, *Mol. Endocrinol.* **2**, 1145 (1988).
18. K. Kawajiri *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1649 (1984).
19. S. Kato *et al.*, *Science* **270**, 1491 (1995).
20. E. B. Mawer *et al.*, *J. Clin. Endocrinol. Metab.* **79**, 554 (1994); H. Fujii *et al.*, *EMBO J.* **16**, 4163 (1997).
21. K. Takeyama *et al.*, data not shown.
22. R. T. Turner *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5720 (1980); J. S. Adams *et al.*, *J. Clin. Endocrinol. Metab.* **60**, 960 (1985).
23. G. A. Howard, R. T. Turner, D. J. Sherrard, D. J. Baylink, *J. Biol. Chem.* **256**, 7738 (1981); T. K. Gray, G. E. Lester, R. S. Lorenc, *Science* **204**, 1311 (1979).
24. T. Matkovits and S. Christakos, *Endocrinology* **136**, 3971 (1995).
25. C. Zierold, H. M. Darwish, H. F. DeLuca, *J. Biol. Chem.* **270**, 1675 (1995); Y. Ohyama *et al.*, *ibid.* **269**, 10545 (1994).
26. K. Takeyama *et al.*, *Biochem. Biophys. Res. Commun.* **222**, 395 (1996); H. Mano *et al.*, *J. Biol. Chem.* **269**, 1591 (1994).
27. U. Gubler and B. J. Hoffmann, *Gene* **25**, 263 (1983); M. Kobori and H. Nojima, *Nucleic Acids Res.* **21**, 2782 (1993).
28. Polyadenylated [poly(A)⁺] RNA was purified from the kidney of the VDR^{-/-} mice at 7 weeks (26). The total cDNA was prepared with poly(A)⁺ RNA (27). The cDNA fragments were subsequently inserted into the Hind III site of the mammalian expression vector pcDNA3 (Invitrogen). The reporter plasmid 17M2-G-lacZ was constructed by inserting yeast GAL4 upstream activating sequence x2 and β -globin promoter into a multicloning site of the Basic expression vector (Clontech). The ligand-binding domain of VDR [VDR(DEF)] fused to the GAL4 DNA-binding domain [GAL4-VDR(DEF)] (11) was used to detect ligand-induced AF-2 function. COS-1 cells cultured with Dulbecco's modified Eagle's medium containing 10% fetal bovine serum were transiently transfected with 0.5 μ g of GAL4-VDR(DEF) expression vector, 1 μ g of 17M2-G-lacZ, 0.2 μ g each of the ADX and ADR expression vectors (12), and 0.1 μ g of expression cDNA library in OPTI-MEM using lipofection agent (GIBCO BRL).
29. H. S. Tong *et al.*, *J. Bone Miner. Res.* **9**, 577 (1994).
30. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
31. H. Sasaki *et al.*, *Biochemistry* **34**, 370 (1995).
32. The solvent systems were hexane/isopropanol/methanol (88:6:6) as the mobile phase at a flow rate of 1.0 ml/min for normal phase and methanol/H₂O (80:20) at a flow rate of 1.0 ml/min for reverse phase. On both normal- and reversed-phase HPLC with other solvent systems [hexane/isopropanol (90:10), hexane/isopropanol (80:20) for normal phase, and methanol/H₂O (75:25) for reversed phase] (22), the retention times of the enzymatic product also completely matched those of authentic 1 α ,25(OH)₂D₃ (21).
33. H. Harada, R. Miki, S. Masushige, S. Kato, *Endocrinology* **136**, 5329 (1995).
34. We thank P. Chambon for critical reading of the manuscript; T. Yoshizawa, Y. Masuhiro, S. Takeda, K. Suzuki, F. Ichikawa, N. Kubota, N. Ohishi, K. Kumaki, and A. Murayama for technical assistance; Y. Nogii, T. Sakaki, and T. Matsumoto for helpful discussion and suggestions; Y. Yabusaki for the ADX and ADR cDNA plasmids; and Chugai Pharmaceuticals for vitamin D-related compounds.

Extensible Collagen in Mussel Byssus: A Natural Block Copolymer

Kathryn J. Coyne, Xiao-Xia Qin, J. Herbert Waite

To adhere to solid surfaces, marine mussels produce byssal threads, each of which is a stiff tether at one end and a shock absorber with 160 percent extensibility at the other end. The elastic extensibility of proximal byssus is extraordinary given its construction of collagen and the limited extension (less than 10 percent) of most collagenous materials. From the complementary DNA, we deduced that the primary structure of a collagenous protein (preCol-P) predominating in the extensible proximal portion of the threads encodes an unprecedented natural block copolymer with three major domain types: a central collagen domain, flanking elastic domains, and histidine-rich terminal domains. The elastic domains have sequence motifs that strongly resemble those of elastin and the amorphous glycine-rich regions of spider silk fibroins. Byssal thread extensibility may be imparted by the elastic domains of preCol-P.

Mussel byssal threads are undoubtedly among nature's most peculiar tendons. One end of each thread inserts into the byssal retractor muscles at the base of the foot; the other end is disposed outside the animal and attached to a hard surface by an adhesive plaque. In a process that resembles injection molding, the foot of the mussel is able to make a new thread in 5 min or less (1). Despite their rapid production and vulnerable location, byssal threads are durable and exquisitely engineered fibers (Fig. 1A). They contain a graded distribution of tensile molecular elements that result in a material that is strong and stiff at one end and plially elastic at the other (2, 3). Overall, byssal threads are five times tougher than Achilles tendon (4).

The collagen content of mussel byssal threads has been well established by wide-angle fiber x-ray diffraction, the presence of *trans*-4-hydroxyproline, and byssus-derived pepsin-resistant peptides with Gly-X-Y repeats (5, 6). Unlike tendon, however, byssus has a nonperiodic microstructure and shrinkage and melting temperatures in excess of 90°C (5). Any attempt to reconcile these unusual biochemical and mechanical features with what is known about collagen provokes the following molecular questions: How is byssal collagen different from tendon collagen, and what role does this difference play in the material performance of byssus?

Because of its highly cross-linked structure, there has been only one recourse for recovering collagen from byssus: acid extraction coupled with extensive treatment with pepsin. We (6) used this approach to characterize two collagen fragments, Col-P and

Col-D (both apparently homotrimers), with apparent α -chain masses of 50 and 60 kD, respectively (7). These fragments coexist in complementary gradients along the length of each byssal thread with Col-D predominating at the distal end and Col-P predominating at the proximal end. Similar gradients were found to exist in the mussel foot, which fabricates the byssus one thread at a time. Precursors to Col-P and Col-D (preCol-P and -D) were identified in foot extracts with specific polyclonal antibodies. Apparent masses of 95 and 97 kD were determined for α -chain preCol-P and preCol-D (α -preCol-P and α -preCol-D), respectively. The amino acid compositions for the NH₂- and COOH-terminal extensions consist of Gly plus Ala concentrations of 50 (preCol-P) and 60 (preCol-D) mole percent (mol%) (6). Although it was inferred that the extensions or flanking domains might resemble structural proteins such as silk fibroin or elastin or resilin, confirmation of such homology has awaited determination of the complete primary structure. Here we report the primary structure of α -preCol-P, a natural block copolymer with putative cross-linking, elastic, and collagen domains (8).

α -PreCol-P may be divided into seven domains based on its amino acid sequence (Fig. 2). The most prominent feature is a central region consisting of a 435-residue collagenous sequence identified by alignment with amino acid sequences reported previously for peptides derived from the pepsin-resistant fragment Col-P (6). The collagen domain consists of 146 Gly-X-Y repeats in which X and Y are frequently Pro. Only the Pro residues at position Y appear to be converted to *trans*-4-hydroxyproline (6). Gly also occurs in the X (10 times) and Y (once) positions. The continuity of Gly-X-Y repeats is interrupt-

College of Marine Studies and Department of Chemistry and Biochemistry, University of Delaware, Newark, DE 19716, USA.

*To whom correspondence should be addressed.

16 June 1997; accepted 22 August 1997

20 mol% of the composition or one in five residues. Other prominent residues are Gly, Ala, and Ser (Fig. 2). His-rich domains occur in several other proteins including high molecular weight kininogen (18), plasmodial His-rich protein HRP (19), blood His-rich glycoprotein HRG (20), and trematode eggshell precursors (21). Sequence homology is limited to short repeats—for example, Ala-His compared with Ala-His-His repeats in plasmodial His-rich protein, and His-Gly-Gly or Gly-Gly-His compared with Gly-His repeats in the kininogen and eggshell precursors. The function of the His-rich tracts in almost all known cases is connected with metal binding such as zinc in kininogen and HRG and heme iron in HRP. His-rich nereid jaws are high in complexed copper or zinc (22), and poly(His) tracts are routinely added to recombinant proteins for facile recovery by metalloprecipitation or metalloligand affinity chromatography (23). Of the transition metals detected in *Mytilus edulis* byssus, zinc levels are consistently the highest at about 1 mg per gram of dry weight (24). Studies are under way to determine whether the recovery of initial stiffness in distal byssal threads after yield to 0.5 extension (25) is related to the breakage and re-formation of protein-zinc complexes.

The results reported here are unusual in several respects. PreCol-P is the first known protein to contain both collagenous and elastin-like domains. Indeed, elastins were thought to be absent in invertebrates (26). In addition, the tandem array of collagenous and elastic domains is remarkably reminiscent of synthetic block copolymer designs with their hard and soft segments (27) (Fig. 1B). In this case, "blocks" of elastin (breaking strain, ≤ 1.6 ; tensile stress, $\leq 5 \times 10^6$ Pa) would improve the extensibility of the collagen (breaking strain, ≤ 0.1 ; tensile stress, $\leq 1 \times 10^8$ Pa) (28). This results in a material with breaking strength lower than that of tendon but with significantly greater toughness (3, 4). Block copolymer design has also been detected in another byssal precursor, preCol-D, which has silk fibroin-like instead of elastic flanking domains in a sequence that is otherwise homologous with that of preCol-P (29). Future studies would profit from a closer scrutiny of the mechanism of preCol fibrillogenesis as well as the interplay of hard

(collagen) and soft (elastin) segments in determining the peculiar mechanical properties of byssal threads.

REFERENCES AND NOTES

1. J. H. Waite, *Results Probl. Cell Differ.* **19**, 27 (1992).
2. L. J. Gathercole and A. Keller, in *Structure of Fibrous Biopolymers*, E. D. T. Atkins and A. Keller, Eds. (Butterworths, London, 1975), p. 153; A. Bairati, in *Form and Function in Zoology*, G. Lanzavecchia and R. Valvassori, Eds. (Mucchi, Modena, 1991), p. 163.
3. E. C. Bell and J. M. Gosline, *J. Exp. Biol.* **199**, 1005 (1996).
4. J. E. Smeathers and J. F. V. Vincent, *J. Molluscan Stud.* **49**, 219 (1979).
5. K. M. Rudall, *Symp. Soc. Exp. Biol.* **9**, 49 (1955); J. Pikkarainen *et al.*, *Eur. J. Biochem.* **4**, 555 (1968); S. C. Melnick, *Nature* **181**, 1483 (1958).
6. X. X. Qin and J. H. Waite, *J. Exp. Biol.* **198**, 633 (1995).
7. MALDI TOF mass spectrometry of preCol-P, preCol-D, and their proteolytically derived fragments indicates that apparent masses are overestimated by 20% on SDS-polyacrylamide gel electrophoresis.
8. We constructed a cDNA library in a λ ZAP expression vector (Stratagene) from mRNA extracted from foot tissue of *Mytilus edulis*. We initially screened the library with polyclonal antibodies specific to preCol-P (6, 30). We subcloned positive clones by in vivo excision into pBK-CMV phagemids for sequencing and made nested deletions with the double-stranded nested deletion kit (Pharmacia) of one clone to facilitate sequencing. We sequenced plasmids with the Taq DyeDeoxy terminator cycle sequencing kit (Applied Biosystems) and analyzed sequencing reactions on a model 373A sequencer (Applied Biosystems). Overlapping sequences in both directions were aligned by the computer program CAP (contig assembly program) (31). To obtain the 5' end of the sequence for the largest partial preCol-P cDNA, we prepared a digoxigenin-labeled RNA probe by in vitro translation with a MAXIScript in vitro RNA transcription kit (Ambion). The cDNA library was plated and overlaid with nitrocellulose membranes (Nitropure, Micron Separations). We incubated the membranes overnight at 50°C with a 1:10⁴ dilution of the RNA probe. The location of the probe on the membrane was detected with alkaline phosphatase-conjugated anti-digoxigenin (DIG) (Boehringer Mannheim, Indianapolis, IN) and visualized with nitroblue tetrazolium and 5-bromo-4-chloro-3-indolylphosphate toluidinium (NBT/BCIP). The full-length sequence for α -preCol-P is 2800 bases with an open reading frame of 2706 bases and a translated amino acid sequence of 902 residues. We verified the size of the cDNA transcript by Northern blot analysis. The absence of preCol-P mRNA in non-foot tissue confirmed its tissue-specific expression in adult mussels. Double-stranded DNA probes for Northern blots were synthesized by incorporation of DIG-11-deoxyuridine triphosphate into products during polymerase chain reaction (PCR) amplification (32). We synthesized a probe specific for preCol-P by PCR amplification of deletion plasmid P14-7aF and with T3 and T7 primers. This deletion plasmid includes the last 900 bases of the sequence for preCol-P. We synthesized a second probe for actin in the same manner from a cloned actin cDNA from the sea scallop *Placopecten magellanicus* (a generous gift from M. Patwary, National Research Council of Canada). Probes constructed from the actin clone of this sea scallop have been shown to cross-react with actin RNAs from other bivalve species, including *M. edulis* (33). We fractionated mRNA from foot and non-foot tissue on a 0.44 M formaldehyde gel in duplicate and transferred them to positively charged nylon membranes (Boehringer Mannheim) by standard techniques. We incubated the blots overnight at 42°C in separate chambers: one with the preCol-P probe and the other with the actin probe. The probe-mRNA hybrid on the membranes was localized by incubation with alkaline phosphatase-conjugated anti-DIG and visualized with NBT/BCIP.
9. E. Kilchherr, H. Hoffmann, W. Steigemann, J. Engel, *J. Mol. Biol.* **186**, 403 (1985).
10. L. Vitellaro-Zuccarello, S. De Biasi, A. Bairati, *Tissue Cell* **15**, 547 (1983).
11. K. Raju and R. A. Anwar, *J. Biol. Chem.* **262**, 5755 (1987).
12. J. Rosenbloom, W. R. Abrams, R. Mecham, *FASEB J.* **7**, 1208 (1993); D. W. Urry, *J. Protein Chem.* **3**, 403 (1984).
13. P. Robson *et al.*, *J. Biol. Chem.* **268**, 1440 (1993).
14. M. B. Hinman and R. V. Lewis, *ibid.* **267**, 19320 (1992); P. A. Guerette, D. G. Ginzinger, B. H. F. Weber, J. M. Gosline, *Science* **272**, 112 (1996).
15. B. L. Thiel and C. Viney, *MRS Bull.* **20**, 52 (1996); A. H. Simmons, C. A. Michal, L. W. Jelinski, *Science* **271**, 84 (1996).
16. P. Brown-Augsburger *et al.*, *J. Biol. Chem.* **270**, 17778 (1995).
17. X.-X. Qin and J. H. Waite, unpublished results.
18. R. A. De La Cadena and R. W. Colman, *Protein Sci.* **1**, 151 (1992).
19. D. J. Sullivan, I. Y. Gluzman, D. E. Goldberg, *Science* **271**, 219 (1996).
20. T. Koide, D. Foster, S. Yoshitake, E. W. Davie, *Biochemistry* **25**, 2220 (1986).
21. S. Johnson, D. W. Taylor, J. S. Cordingley, *Mol. Biochem. Parasitol.* **22**, 89 (1987); J. H. Waite and A. C. Rice-Ficht, *Biochemistry* **28**, 6104 (1989).
22. G. W. Bryan and P. E. Gibbs, *J. Marine Biol. Assoc. UK* **59**, 969 (1979).
23. G. Lilius *et al.*, *Eur. J. Biochem.* **198**, 499 (1991); M. C. Smith, T. C. Furman, C. Pidgeon, *Inorg. Chem.* **26**, 1965 (1987).
24. T. L. Coombs and P. J. Keller, *Aquat. Toxicol.* **1**, 291 (1981).
25. E. C. Bell, unpublished results.
26. H. Sage and W. R. Gray, *Comp. Physiol. Biochem. B* **64**, 313 (1979).
27. M. W. Denny, *Biology and Mechanics of the Wave Swept Environment* (Princeton Univ. Press, Princeton, NJ, 1988).
28. C. Hall, *Polymer Materials* (Macmillan, London, 1981).
29. X.-X. Qin, K. J. Coyne, J. H. Waite, in preparation.
30. R. A. Young and R. W. Davis, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1194 (1983).
31. X. Huang, *Genomics* **14**, 18 (1991).
32. T. Lion and O. A. Haas, *Anal. Biochem.* **188**, 335 (1990).
33. M. U. Patwary, M. Reith, E. L. Kenchington, *J. Shellfish Res.* **15**, 265 (1996).
34. Abbreviations for the amino acids are as follows: A, Ala; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; and Y, Tyr.
35. We thank A. Hunt and J. MacDonald for the use of their sequencer. M. L. Tanzer, E. C. Bell, D. Urry, and three anonymous reviewers provided useful comments for improvement of the manuscript. Supported by grants from the National Institute of Dental Research (Biomaterials) and the Office of Naval Research (MIMI Program) to J.H.W.

15 May 1997; accepted 25 July 1997