9. Rost B, Sander C: Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* 1994, 19:55-72.
10. Gekakis N, Saez L, Delahaye-Brown AM, Myers MP, Sehgal A, Young MW, Weitz CJ: Isolation of timeless by PER protein-interaction: defective interaction between timeless protein and long-period mutant PERL. *Science* 1995, 270:811-815.
11. Huang ZJ, Edery I, Rosbash M: PAS is a dimerization domain common to ***Drosophila Period*** and several transcription factors. *Nature* 1993, 364:259-262.
12. Lindebro MC, Poellinger L, Whitelaw ML: Protein–protein interaction via PAS domains: role of the PAS domain in positive and negative regulation of the bHLH/PAS dioxin receptor–Arnt transcription factor complex. *EMBO J* 1995, 14:3528-3539.
13. Edgerton MD, Jones AM: Subunit interactions in the carboxy-terminal domain of phytochrome. *Biochemistry* 1993, 32:8239-8245.
14. Hill S, Austin S, Eydmann T, Jones T, Dixon R: ***Azotobacter vinelandii*** NIFL is a flavoprotein that modulates transcriptional activation of nitrogen-fixation genes via a redox-sensitive switch. *Proc Natl Acad Sci USA* 1996, 93:2143-2148.
15. Monson EK, Weinstein M, Ditta GS, Helinski DR: The FixL protein of ***Rhizobium meliloti*** can be separated into a heme-binding oxygen-sensing domain and a functional carboxyl-terminal kinase domain. *Proc Natl Acad Sci USA* 1992, 89:4280-4284.
16. Guillemin K, Krasnow MA: The hypoxic response: Huffing and HIFing. *Cell* 1997, 89:9-12.
17. Bibikov SI, Biran R, Rudd KE, Parkinson JS: A signal transducer for aerotaxis in ***Escherichia coli***. *J Bacteriol* 1997, 179:4075-4079.
18. Baker PJ, Britton KL, Rice DW, Rob A, Stillman TJ: Structural consequences of sequence patterns in the fingerprint region of the nucleotide binding fold — Implications for nucleotide specificity. *J Mol Biol* 1992, 228:662-671.
19. Schuler GD, Altschul SF, Lipman DJ: A workbench for multiple alignment construction and analysis. *Proteins* 1991, 9:180-190.
20. Altschul SF, Gish W: Local alignment statistics. *Methods Enzymol* 1996, 266:460-480.
21. Birney E, Thompson JD, Gibson TJ: PairWise and SearchWise: finding the optimal alignment in a simultaneous comparison of a protein profile against all DNA translation frames. *Nucleic Acids Res* 1996, 24:2730-2739.
22. Eddy SR: Hidden Markov models. *Curr Opin Struct Biol* 1996, 6:361-365.

Addresses: *University of Oxford, The Old Observatory, Fibrinolysis Research Unit, South Parks Road, Oxford OX1 3RH, UK. †Department of Biology - BSBW, Texas A&M University, College Station, Texas 77843, USA.
Correspondence: Chris P. Ponting
E-mail: Ponting@bioch.ox.ac.uk

Sequence of abductin, the molluscan 'rubber' protein

Quiping Cao*, Yunjuan Wang† and Hagan Bayley†

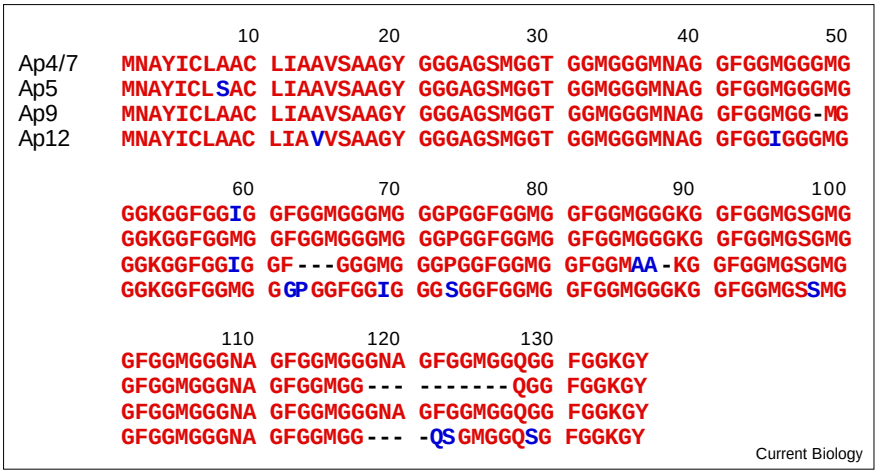
The inner hinge ligament of bivalve molluscs opposes the action of the adductor muscle. This arrangement permits the opening and closing of the shell. In scallops, the apparatus has evolved for swimming [1]. By opening and closing their shells about four times per second, scallops swim a few meters at a time to escape slow-moving predators, such as starfish. The major component of all inner hinge ligaments is the protein abductin [2]. The scallop ligament is composed almost entirely of protein and its relatively low extent of mineralization compared to the ligaments of other molluscs is thought to contribute to its high resilience (96% recovered work), which is extraordinary even by molluscan standards [3].

Here, we report amino acid sequences of abductin deduced from the sequences of cDNAs obtained from *Argopecten* (the bay scallop). Degenerate primers based on the sequences of peptides obtained from digested inner hinge ligament were used to amplify *Argopecten* genomic DNA. The PCR products were in turn used to screen an *Argopecten* cDNA library, which yielded five full length cDNAs, Ap4, Ap5, Ap7, Ap9 and Ap12. The 136 amino acid open reading frames of Ap4 and Ap7 were identical, whereas the other three (Ap5, 126 residues; Ap9, 131 residues; Ap12, 132 residues) were very closely related (Figure 1). Analysis by northern blotting and RNase protection showed that abductin transcripts are present only in the mantle tissue. RNase protection experiments were carried out on RNA from *Argopecten* eggs and various larval stages through to young adults at day 35 [4]. Weak transcription was noticeable at day 15 (prodissoconch, stage 20, 0.19 mm) and a strong signal was seen at day 35 (pre-adult scallop, stage 25, 1.67 mm). Expression of abductin continued in the adult.

Electron microscopy and X-ray diffraction experiments indicate that the inner hinge ligament is acellular and amorphous [2]. The physical properties of the ligament suggest that its resilience is a result of changes in entropy during deformation [5] and that abductin is lightly cross-linked [6,7], although the nature of the cross-links is a conspicuous problem. The abductin sequences can be divided into two domains. The first is the alanine-rich amino terminus (residues 1–20), which also contains the two conserved cysteines (residues 6 and 10) that might be involved in intermolecular or intramolecular disulfide formation. This domain also contains two of the three conserved tyrosines (residues 4 and 20), which could also be involved in cross-linking by forming 3,3′-methylene bistyrosine [6]. The first domain, however, is likely to form a signal sequence for secretion: indeed, the agreement between the amino acid analysis of the inner hinge ligament and the composition predicted from the cDNA sequences is increased when the amino-terminal residues are not included.

The second domain comprises an extraordinary glycine/methionine-rich sequence that stretches from residue 21 to the carboxyl terminus. The methionines are distributed throughout this domain and are included in the consensus sequence GGFGGMGGGX (single-letter amino acid code), which is found with the strongest compliance in the carboxy-terminal half of the polypeptide. Interestingly, it has been reported that most of the methionine in the inner hinge ligament of surf clams is actually methionine sulfoxide [8], which would modestly increase the polarity

Figure 1



Amino acid sequences of abductin deduced from the sequences of five full-length cDNAs obtained from *Argopecten*. The open reading frames of Ap4 and Ap7 were identical, whereas the other three cDNAs were very closely related and may represent alleles from the population used to construct the library. Freshly dissected inner hinge ligament from adult animals was ground in 1.0 M aqueous acetic acid to extract traces of calcium carbonate. The insoluble residue was dried and subjected to amino acid analysis, which revealed an unusual composition rich in glycine (57.3% of residues) and with a high level of methionine (14.3% of residues), as noted previously [2,3]. The extracted tissue was also digested with 15% (w/v) CNBr in 70% formic acid at room temperature for up to 48 h or with 0.1 N HCl at 95°C for 3 h. The released peptides were separated by reverse-phase HPLC and subjected to automated Edman degradation and MALDI-MS. The following sequences were obtained: NAGGFGGIGG (CNBr), GGGPGGFGGIG GGSGGFGG(M) (CNBr), GGGLGGFGGI GGFGG(M) (CNBr), GFGG (HCl) and GFGGMGG (HCl), single-letter amino acid code where (M) is homoserine lactone. Degenerate primers based on the peptide sequences were used to amplify *Argopecten* genomic DNA. Two PCR products, ApG1 and ApG17, were used to screen an *Argopecten* oligo(dT)-primed cDNA library made from young animals (3–8 mm) in which the ligament is growing rapidly. The abductin sequences contained the Edman sequences, with the exception of GGGLGGFGGIGGFG G(M), which gave a single mismatch. The cDNA sequences are available through GenBank: accession numbers AFO26845–AFO26848 inclusive.

of the polypeptide. The high glycine content is in keeping with a lack of secondary structure, although the well conserved repeat argues that 'pattern implies structure' [9], as in the case of mammalian elastin, which also has a repetitive sequence albeit different from that of abductin [10]. Our sequence data therefore call for a re-examination of the structure of the inner hinge ligament using biophysical techniques. The second domain is punctuated by lysine residues at positions 53 and 89. The third conserved lysine is two residues from the carboxyl terminus. Finally, the third conserved tyrosine is at the extreme carboxyl terminus. If the first domain is removed by a signal peptidase, these remaining lysine and tyrosine residues are probable candidates for sites of cross-linking.

Other proteins secreted by the mantle epithelium differ in composition from abductin. For example, shell proteins involved in mineralization are rich in acidic residues [11,12], whereas proteins of the periostracum (the outer surface of the shell) are rich in tyrosine [11,13]. Interestingly, two recently sequenced shell proteins contain long glycine-rich stretches of amino acids, but the characteristic abductin repeat is absent and the methionine content is low [12].

If the nature of the cross-links in abductin can be resolved, the recombinant abductin protein could be converted into resilient materials with applications in biotechnology. Hybrid materials incorporating abductin sequences are likely to be more useful than materials that simply mimic the ligament.

References

1. Vogel S: Squirt smugly, scallop. *Nature* 1997, 385:21-22.
2. Kelly RE, Rice RV: Abductin: a rubber-like protein from the internal triangular hinge ligament of ***Pecten.*** *Science* 1967, 155:208-210.
3. Kahler GA, Fisher FM, Sass RL: The chemical composition and mechanical properties of the hinge ligament in bivalve molluscs. *Biol Bull* 1976, 151:161-181.
4. Sastry AN: The development and external morphology of pelagic larval and post-larval stages of the bay scallop, ***Aequipecten irradians concentricus*** Say, reared in the laboratory. *Bull Mar Sci* 1965, 15:417-435.
5. Alexander RMcN: Rubber-like properties of the inner hinge-ligament of ***Pectinidae.*** *J Exp Biol* 1966, 44:119-130.
6. Andersen SO: Isolation of a new type of cross link from the hinge ligament protein of molluscs. *Nature* 1967, 216:1029-1030.
7. Thornhill DP: Abductin. Locus and spectral characteristics of a brown fluorescent chromophore. *Biochemistry* 1971, 10:2644-2649.
8. Kikuchi Y, Tamiya N: Methionine sulfoxide in the resilium of surf clams. *J Biochem* 1981, 89:1975-1976.
9. Vincent J: *Structural Biomaterials*. Princeton: Princeton University Press; 1991.
10. Urry DW: Molecular machines: how motion and other functions of living organisms can result from reversible chemical changes. *Angew Chem* 1993, 32:819-841.
11. Hare PE: Amino acids in the proteins from aragonite and calcite in the shells of ***Mytilus californianus***. *Science* 1963, 139:216-217.
12. Sudo S, Fujikawa T, Nagakura T, Ohkubo T, Sakaguchi K, Tanaka M, Nakashima K, Takahashi T: Structures of mollusc shell framework proteins. *Nature* 1997, 387:563-564.
13. Hunt S: Amino acid compositions of periostracal proteins from molluscs living in the vicinity of deep sea hydrothermal vents: an unusual methionine-rich structural protein. *Comp Biochem Physiol* 1987, 88:1013-1021.

Addresses: *Worcester Foundation for Biomedical Research, 222 Maple Avenue, Shrewsbury, Massachusetts 01545, USA. †Department of Medical Biochemistry & Genetics, Texas A&M Health Science Center, 440 Reynolds Medical Building, College Station, Texas 77843-1114, USA.
Correspondence: Hagan Bayley
E-mail: bayley@medicine.tamu.edu