

2

REPRODUCTION PROHIBITED BY THIS PAGE

DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

AD-A209 687

DTIC

1b. RESTRICTIVE MARKINGS
NA

FILE COPY

SELECTED

3. DISTRIBUTION/AVAILABILITY OF REPORT
Distribution unlimited

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE
NA

JUL 05 1989

4. PERFORMING ORGANIZATION REPORT NUMBER(S)
Boston University

5. MONITORING ORGANIZATION REPORT NUMBER(S)
NA

6a. NAME OF PERFORMING ORGANIZATION
Boston University

6b. OFFICE SYMBOL (if applicable)
NA

7a. NAME OF MONITORING ORGANIZATION
Office of Naval Research

6c. ADDRESS (City, State, and ZIP Code)
Chemistry Department
Boston University
590 Commonwealth Avenue, Boston, MA 02215

7b. ADDRESS (City, State, and ZIP Code)
800 N. Quincy St.
Arlington, VA 22217-5000

8a. NAME OF FUNDING/SPONSORING ORGANIZATION
Office of Naval Research

8b. OFFICE SYMBOL (if applicable)
ONR

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
N00014-86-K-0217

8c. ADDRESS (City, State, and ZIP Code)
800 N. Quincy St.
Arlington, VA 22217-5000

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
61153N	RR04106	441c014	

11. TITLE (Include Security Classification)
(U) Characterization of Marine Bioadhesive Proteins

12. PERSONAL AUTHOR(S)
Laursen, Richard A.

13a. TYPE OF REPORT
Final

13b. TIME COVERED
FROM 4/1/86 TO 6/30/89

14. DATE OF REPORT (Year, Month, Day)
1989 July 1

15. PAGE COUNT
11

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Adhesive; Proteins; MusseL; Amino Acid Sequences; Gene Sequence; cDNA; (KT)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)
The objective of this research is to elucidate the amino acid sequences, via gene sequencing, of the adhesive proteins from several species of mussel and of other organisms, with the aim of understanding how these organisms attach themselves to wet surfaces. During this project period, we have cloned and sequenced fragments of the adhesive protein genes from three species of mussel. Two classes of protein are now apparent: they are similar in their content and location of lysine and tyrosine (or DOPA) residues, but different in repeat length and content of other amino acids. *Keywords: Bioadhesives;*

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
 UNCLASSIFIED/UNLIMITED SAME AS RPT. DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
(U)

22a. NAME OF RESPONSIBLE INDIVIDUAL
M. Marron

22b. TELEPHONE (Include Area Code)
(202) 696-4760

22c. OFFICE SYMBOL
ONR

FINAL REPORT ON CONTRACT NO0014-86-K-0217

PRINCIPAL INVESTIGATOR: Richard A. Laursen

CONTRACTOR: Boston University

CONTRACT TITLE: Characterization of Marine Bioadhesive Proteins

START DATE: 1 April 1986 (Initial funding start date: 1 July 1986)

RESEARCH OBJECTIVE: Our primary initial objective has been to clone and sequence adhesive protein genes for marine organisms such as barnacles and mussels with the aim of understanding what common (if any) structural features give these proteins their adhesive properties. It is hoped that this knowledge will lead to the development of adhesives that will have medical and other applications.

SUMMARY OF PROGRESS (3-YEAR): At the beginning of this project, we set out to clone and sequence genes from both barnacles and mussels. As work progressed, however, it became evident that the barnacle project was going to be very difficult because of problems in identifying a unique barnacle adhesive protein. For this reason we have concentrated our efforts on mussel adhesive protein genes, particularly those of the blue mussel (*Mytilus edulis*) and the ribbed mussel (*Geukensia demissa*), but also of the Pacific blue mussel (*Mytilus californianus*) and the horse mussel (*Modiolus modiolus*).

A problem that plagued us in the early stages of this work was that of getting clones large enough (ca. 3 kb) to code for the entire gene. In the initial stages of our work, we typically could isolate clones of only a few hundred bases. Furthermore, after sequencing many of these fragments and comparing our data with that of investigators at Genex Corp. (who also have sequenced *Mytilus edulis* gene fragments), we find that we have much more sequence data than is needed to code for a protein containing around 1100 amino acids, yet almost none of the fragments overlap completely. We have attributed this problem to recombination, which is frequently noted for repetitive proteins, and have more recently been using rec⁻ cloning strains, which have enabled us to obtain clones which may code for the full-length strains. However, preliminary sequencing results on these large clones suggest another possible explanation for the occurrence of so many non-overlapping peptides: the existence of multiple genes for the adhesive protein, or of species variations within populations of mussels.

Now that nearly full-length protein sequences are available it is possible to begin analyzing them with the aim of correlating structure with function. Some preliminary ideas are presented here.

DETAILED REPORT

Mytilus edulis gene sequence; *Mytilus edulis* gene fragments were isolated by screening a λ gt10 cDNA library with oligonucleotide probes synthesized based on peptide sequences originally determined by Waite (U. Delaware). Sequencing of these clones, which account for around 900 amino acids, showed



LA

odes
or

only variations on the previously observed theme of decapeptide and hexapeptide repeats:

xx1-Lys-xx2-xx3-Tyr-Pro-Pro-Thr-Tyr-Lys

where xx1 is usually Pro, Ser or Ala; xx2 is Pro, Ser, Leu, Ile or Lys; and xx3 is Thr or Ser. The hexapeptide arises by deletion of four residues from the middle of the decapeptide and is interspersed in no yet discernible pattern among the decapeptide repeats.

Subsequently, a cDNA library was fractionated on an agarose gel, and hybridizing fragments were cloned into the EcoRI site of puc18 cloning vector. The *E. coli* host strain CES201 was selected as the host strain for the library. Two large (ca. 2.8 kb) clones were obtained and sequenced from both ends by plasmid sequencing, using exonuclease III to obtain progressively deleted cDNA inserts. One of these clones has been almost entirely sequenced, except for a few gaps in the middle that are now being filled in (Fig. A1). Several observations can be made from these data:

1. The protein consists primarily of approximately 80 repeats (with small variations) of decapeptide (80%) and hexapeptide (20%), which are not arranged in any obvious pattern.
2. The consensus peptide is AKPSYPPTYK, but at the carboxyl- and amino-termini, the sequences become increasingly degenerate, with the peptide PKXTYPPTYK, and more unusual variations, becoming frequent.
3. There is a non-repeat sequence at the N-terminus (5' end), but we have not yet found a start codon. At the 3' end (C-terminus) the repeat sequence continues to the stop codon in the polyA tail.

We have sequenced the 5' and 3' ends of two large clones (clone 412 and clone 26). Their sequences are nearly, but not exactly identical, differing in 6 bases out of about 450 bases. This suggests that there may be some genetic variability among populations of mussels. This inference has been strengthened by comparison of our data with a partial genomic sequence and other data from Genex Corporation. Between us, we have essentially four sets of data, representing different sources of mussels (Chesapeake Bay or New England) and different cloning methods, and none of the data sets overlaps for long stretches with any of the others. The overall picture of repeating deca- and hexa-peptides is the same, but the arrangements and details are different. Our conclusion is that while recombination was probably our primary problem earlier, underlying that is species variation or possibly multiple genes within a single organism.

The *G. demissa* gene: A λ gt10 library was initially constructed for this species, but screening with *Mytilus* probes was unsuccessful because (as we now know) of the significant sequence differences. For this reason, a λ gt11 library was constructed and immunoscreened with antibodies raised against the protein. As with *Mytilus edulis*, we initially isolated only small clones, although we have recently succeeded in isolating large ones. Based on

sequence data for several hundred residues (Fig. A2), we find that the *G. demissa* protein is significantly different from *Mytilus edulis* in that it contains repeats of from 11 to 13 amino acids, e.g.,

Gly-Lys-Pro- Thr-Thr-Tyr-Asp-Ala-Gly-Tyr-Lys-
Gly-Gln-Gln-Lys-Gln- Thr-Gly-Tyr-Asp-Thr-Gly-Tyr-Lys-
Gly-Gly-Val-Gln-Lys- Thr-Gly-Tyr-Ser-Ala-Gly-Tyr-Lys-,

and contains large amounts of glycine and glutamine, but relatively little proline. Also there is considerably more variability at most positions than is seen in *Mytilus edulis* (Tables A1 and A2). With this species, the repeat pattern is more complex. There is an 8-amino acid repeat, each of which is separated by a tripeptide or one of two pentapeptides (see above and Fig. A2). It is also apparent from the pattern of repeats that the *G. demissa* protein arose by gene duplication (Fig A2).

Other species: We have also looked at two other species of mussel, *Mytilus californianus* and *Modiolus modiolus*. Cloning of *M. californianus* and of *M. modiolus* genes was carried out as for *M. edulis* by construction of a λ gt10 cDNA library and screening with probes from *M. edulis*. The sequence of a clone from *M. californianus* (Fig. A3) was very similar to that of *M. edulis*, except for the occurrence of Arg (50% of the time) at position 1 and about a 50% occurrence of Ser and Ala at position 7. We have one clone from *Modiolus modiolus*, but have not sequenced it yet.

A conformational model for the Mytilus protein: It has been argued that the adhesive proteins may have a "random coil" structure, because a structureless protein would have better access to surfaces. Furthermore, spectroscopic studies (in other laboratories) have not turned up any evidence of a regular structure. Nevertheless, we believe that the adhesive proteins probably have some sort of regular folded structure, either in solution or in a condensed state, as in cured adhesive. In the first place, Nature does not (as far as the writer knows), make unstructured proteins. Even portions of proteins that used to be called "random coil" can now be classified as α -loops [J.F. Leszczynski and G.D. Rose, *Science*, 234, 849-855 (1986)]. Also it is difficult to understand why the decapeptide repeat has been conserved as faithfully as it has if folding to a regular structure were unimportant. And finally the invariability of Tyr and Lys and certain other residues and the patterns of posttranslational modification of Tyr and Pro (to Dopa and hydroxyproline, respectively) residues suggest some sort of regular structure. Given the large amount of proline, a structure with turns or loops seems more likely than a regular helical or sheet structure. Given the propensity for Tyr and Thr residues to occur in β -sheets, for Pro-Pro sequences not to be found in β -turns, but to cause a 90° bend in the peptide backbone, we have postulated a β -sheet- β -turn (β -hairpin) model (Fig A4) to serve as a working hypothesis for spectroscopic studies. Such structures have been studied recently by Fehrentz et al. [*Biochemistry*, 27, 4071-4078 (1988)] (Fig. A5).

This model, though speculative, has some attractive features. It puts all the polar groups on the faces of the β -sheet loop, where they could interact with surfaces. In addition the Tyr and Lys residues are on both faces in

pairs, in a symmetrical arrangement, where they might pair up with corresponding pairs in another chain to form interchain crosslinks. If one considers several of these β -hairpins linked together, they might form a sort of superhelix with prolines at the core and all the lysines on the outer arms where repulsive forces would be minimized.

The major failing of this model is that one cannot make a similar model for the *Geukensia* protein, which contains little proline and has a less regular repeat structure. Of course *Geukensia* could have a completely different structure, but one would think, given the relatively constant placement of the critical Tyr and Lys residues, that there might be some conformational similarities. This dilemma can be resolved only by experiment.

The degeneracy or tendency to depart from the consensus sequence at the N- and C-termini, as noted above (see Fig. A1) is reminiscent of the collagen molecule (Fig. A6), which exists as a triple helix, except at the termini (teleopeptide regions), where the regular -Gly-X-X-sequence breaks down. Perhaps in the adhesive protein strand, the terminal regions actually do have a random structure making the ends "sticky" and more able to adhere to surfaces and also to collagen fibers in the byssal thread.

Attempted isolation of the barnacle adhesive. Our efforts in this area are based on the report of Cheung et al. [Marine Biol. 43, 157-163 (1977)] that a proteinaceous material, which hardens over a period of a few hours, could be isolated from the bases or undersides of barnacles. After some abortive attempts to isolate adhesive protein from local barnacles *Balanus balanoides*, which have soft, membranous bases, we switched to the barnacle originally studied by Cheung et al., *Balanus eburneus*, which has a hard calcified base. This species can be obtained from the Marine Biology Laboratories at Woods Hole, Mass. We try to obtain animals that have been growing on styrofoam blocks since they usually have flat bases. Following the method of Cheung et al., we mounted the barnacles in holes bored in plastic Petri dishes, which were then floated in aquaria in such a way that the animal's base was oriented upwards and its movable plates were immersed in seawater, allowing it to feed.

Initially we did not observe the beads of exudate reported by Cheung et al. However, by blotting the base with glass filter paper and staining the filter, we could detect proteinaceous material near the outer edges of the barnacle. Furthermore, by shaving the base lightly with a razor blade (as described by Cheung et al.), we were able to obtain rather substantial amounts of an exudate, which did appear to solidify after some time. SDS-polyacrylamide gel electrophoresis of this material showed 7 major protein bands. The major band (ca. 50% of the total) has an apparent molecular weight of 35 kD, and the minor proteins range in size from 50 kD to >200 kD. The exudate turns opaque and appears to polymerize or gel upon standing 3 to 4 hours at room temperature. The gelling occurs in concert with, or as a result of, an apparently random proteolytic degradation of all the major proteins present in the exudate. Since the gel can be readily solubilized by the addition of SDS and mercaptoethanol, this material may not be involved in the process of barnacle adhesion. At the present time, we do not know whether these proteins originate from the cement cells or are simply body fluids. It is possible that the gelling protein observed by us and by Cheung et al. is

fibrinogen, which is found in hemolymph of other crustaceans, such as lobster. The function of fibrinogen is to plug up holes in the vascular system by forming gelatinous fibrin clots. However, this material has a molecular weight of about 450,000, which is substantially higher than what we saw on electrophoresis gels.

We also initiated experiments with a Pacific Ocean goose barnacle, *Pollicipes polymerus*, which unlike *B. eburneus*, grows on a long stalk such that its base is separated from its mantle by 2-4 cm. The advantage of this species is that its cement gland or cells are anatomically distinct from other organs. We have extracted protein from this part of the animal using the acid extraction method developed by Waite for the mussel adhesive protein. Gel electrophoresis shows that the extracts contain mainly (about 50%) a single protein with a molecular weight of about 95,000, which is in a size range comparable to the mussel proteins. The protein does not react with the nitrous acid-molybdate dopa stain, and amino acid analysis did not reveal a distinctive composition as might be expected for a protein with tandem repeats.

We have not yet tried to characterize this protein further, because other projects have taken priority.

INVENTIONS

None

PUBLICATIONS AND REPORTS

We have delayed publication until we could sort our recombination problems. Furthermore, a characteristic of sequencing projects is that until one has the entire sequence, one frequently does not know what the whole picture is. Nevertheless, we plan to submit one paper on comparative aspects of the mussel adhesive protein before September 1981, and two others in the near future on the total sequences of two of the proteins.

Lectures and presentations of some of this work have been made at:

ONR Contractors Conference: "Bioengineering for Materials Applications", Bethesda, MD June 15-16, 1987

State University of New York at Plattsburg, Department of Chemistry, 12/21/87.

Schering Corporation, Bloomfield, NJ, 12/23/87.

Massachusetts Centers of Excellence, Polymer Science and Plastics Technology Symposium, Sturbridge, MA, 3/3/88.

14th International Congress of Biochemistry (Prague, Czechoslovakia, July 10 -15, 1988).

Society for Industrial Microbiology Meeting (Chicago, IL, August 7-12, 1988).

New England Society for Industrial Microbiology Meeting (Wayland, MA,
1/5/89)

U.S. Army Research, Development and Engineering Center, (Natick, MA,
1/11/89).

Department of Chemistry, Wellesley College (Wellesley, MA, 5/5/89)

TRAINING ACTIVITIES:

This work has been performed by the following individuals, who received support from this contract:

Michael J. Connors (Postdoctoral Research Associate)
Jung-Jung Ou (Grad Student, Chinese, Republic of China citizen)
Xiao-Tong Shen (Grad Student, Chinese, People's Republic of China citizen)

In addition, Dr. Leszek Rzepecki, a postdoctoral research associate from Dr. Herbert Waite's lab, spent a month (February, 1988) with us learning cloning techniques from Mike Connors.

AWARDS/FELLOWSHIPS:

R.A. Laursen was the first recipient of the Pehr Edman Award (for contributions to methodology in protein sequence analysis). It was awarded at the Seventh International Conference on Methods of Protein Sequence Analysis, Berlin, July, 1988.

EGIKNLCLLCIFTFDVLGFSNGNIYNAHVSSYAGASAGAYKK
LPNAYPYGTXPEPVYKPVKTSYSAPYKPPTYQ <gap> PVKTSYS
SAPYKPPTYQPLKKSTYRPTKSYPPPTYGSKTNYLPLAKKLS
YKPIKTTYN AKTNYPPVYK PKNTYPPTYK PKPSYPPTYK
SKP TYK PKITYPPTYK AKPSYPPTYK PKKTYPPTYK
PKLTYPPTYK PKPSYPPTYK SKP TYK PKITYPPTYK
AK <gap> PTYK PKLTYPPTYK PKPSYPPSYK PKKTYPPTYK
PKLTYPPTYK AKPSYPPTYK AKPSYPPTYK AKPSYPPTYK
AKP TYK AKPTYPSTYK AKPTYPPTYK AKPSYPPTYK
AKPSYPPTYK AKL TYK AKPSYPPTYK AKPSYPPTYK
AKPS <gap> PSYPPTYK AKPSYPPTYK AKPSYPPTYK
AKPSYPPTYK VKP TYK AKPTYPSTYK AKPSYPPTYK
AKPSYPPTYK AKPSYPPTYK AKPSYPPTYK AKPTYPSTYK
AKP TYK AKPSYPPTYK AKPSYPPTYK AKPSYPPTYK
AKP TYK AK <gap> TYK AKPSYPPTYK AKPSYPPTYK
AKPSYPPTYK AKPSYPPTYK AKPSYPPTYK AKPSYPPTYK
AKP TYK AKPTYPSTYK AKPSYPPTYK AKPSYPPTYK
AKP TYK AKP TYK AKPTYPSTYK AKPSYPPTYK
AKP AYK <gap> PSYPPTYK PKISYPPTYK
AKPSYPPSTYK AKSSYPPTYK AKP TYK AKPTYPSTYK
AKP TYK AKPTYPPTYK AKPSYPPTYK PHPSYPPTYK
SKSSYPSSYK PKKTYPPTYK PKLTYPPTYK PKPSY <gap>
LKPSYPPTYK SKTSYPPTYN KKISYPPSSYK AKTSYPPAYK
<gap> RY

Figure A1. Partial amino acid sequence of *M. edulis* adhesive protein based on DNA sequence data from clone 412. "Gap" refers to unsequenced regions.

G K P		T G Y S A D Y K
	C Q Q K Q	S S Y D P G Y K
	C Q Q K Q	T G Y D T G Y K
		T A Y D P G Y K
	G G V K K	T G Y S A D Y K
G K P		S S N V P G Y K
	C Q Q K Q	T G Y V P G Y K
	C Q Q K Q	T G Y D K G Y K
		S G Y T A G Y K
	G G V Q K	S P Y V P G Y K
A K P		T G H V P G Y K
	C Q Q K K	T G Y S A G Y K
		S P Y V P G Y K
	A G V Q K	T G Y L P G Y K
A K P		T A Y D P G Y K
	C Q Q K Q	N G Y V P G Y K
	C Q Q K Q	T G Y S A G Y K
	C Q Q K K	S P Y V P G Y K
		T G Y L P G Y K
	G G V Q K	T A Y D P G Y K
A K P		T A Y N P G Y K
	C Q Q K Q	T G N S A G Y K
	C Q Q K Q	S P Y V P G Y K
	C Q Q K Q	T G Y V P G Y K
		T G Y D L G Y K
	G G V Q K	T A Y D P G Y K
A K P		S G Y S A D Y K
	C Q Q K Q	S S N V P G Y K
	C Q Q K Q	T G Y D P G Y K
	C Q Q K Q	T G Y D K G Y K
		T G Y D K G Y K
G K P		
	C Q Q K Q	
	C Q Q T Q	
	A Q Q K Q	

Figure A2. Partial amino acid sequence of *G. demissa* adhesive protein from cDNA clone XTS-Gd-6.10.

Tyr Lys
Arg Lys Ile Ser Tyr Pro Pro Thr Tyr Lys
Thr Lys Pro Ser Tyr Pro Ala Thr Tyr Lys
Arg Lys Ile Ser Tyr Pro Pro Thr Tyr Lys
Thr Lys Pro Ser Tyr Pro Ala Ser Tyr Lys
Arg Lys Thr Ser Tyr Pro Ser Thr Tyr Lys
Arg Lys Thr Ser Tyr Pro Pro Thr Tyr Lys
Pro Lys Ile Ser Tyr Pro Ser Thr Tyr Lys
Thr Lys Pro Ser Tyr Pro Pro Thr Tyr Lys
Pro Lys Pro Ser Tyr---

Figure A3. Derived amino acid sequence from the largest clone from *M. californianus*. Note the great similarity to the *M. edulis* sequences except for the frequent occurrence of Arg in position 1 and Ala or Ser in position 7 of the decapeptide repeat.

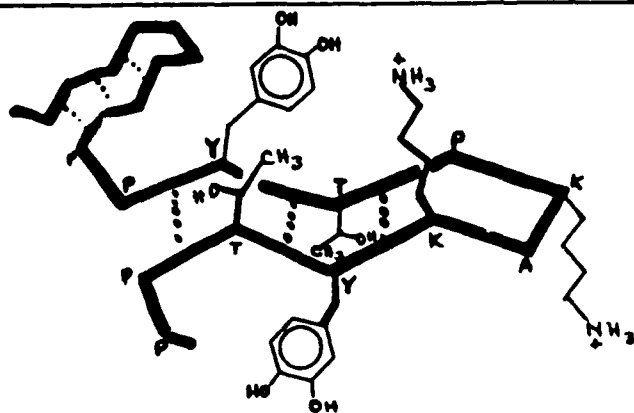


Figure A4. Hypothetical β -hairpin model for *M. edulis* repeat sequence.

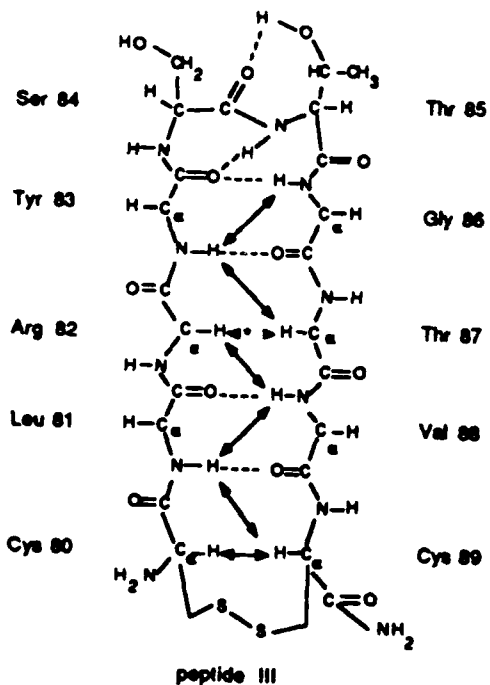


Figure A5. β -Hairpin structure of renin flap peptide as shown by NMR studies. Arrows indicate NOEs; dashed lines are hydrogen bonds. (from Fehrentz et al. 1988).

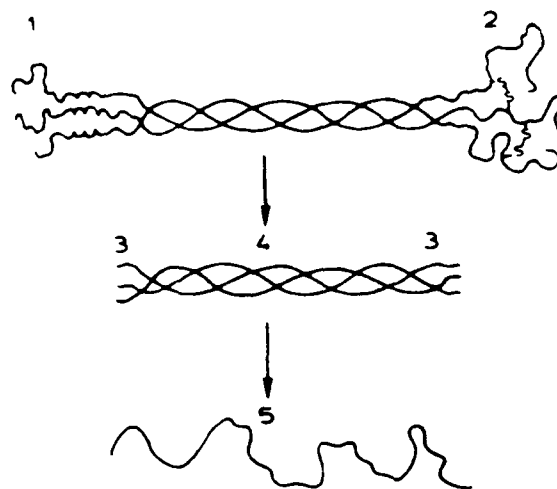


Figure A6. Schematic representation of procollagen (top) and collagen (center). 3) NH_2 - and COOH -terminal teleopeptides; 4) triple helical region. (from Kuhn, 1987).

Table A1
M. edulis Amino Acid Frequencies
 (based on 73 repeat units)

Position	Amino acid (% abundance)
1	Ala (71) Pro (19) Ser,Val,Lys,Leu (10)
2	Lys (100)
3	Pro (53) Thr,Met,Ile,Lys,Leu,Ser, (27)
4	Ser (68) Thr (26) Asn (2)
5	Tyr (100)
6	Pro (100)
7	Pro (85) Ser (15)
8	Thr (92) Ser,Val,Ala (8)
9	Tyr (100)
10	Lys (99) Asn (2)

Table A2
G. demissa Amino Acid Frequencies
 (based on 29 repeat units)

Position	Amino acid (% Abundance)
1	Thr (65) Ser (31) Asn (4)
2	Gly (59) Ser,Pro,Ala (41)
3	Tyr (86) Asn,His (14)
4	Val (34) Asp (31) Ser (20) Leu,Thr,Asn (15)
5	Pro (65) Ala (24) Thr,Lys,Leu (11)
6	Gly (93) Asp (7)
7	Tyr (100)
8	Lys (100)