Reference: *Biol. Bull.* **186:** 349–355. (June, 1994)

The Adhesive Protein cDNA of *Mytilus galloprovincialis* Encodes Decapeptide Repeats but No Hexapeptide Motif

KOJI INOUE AND SATOSHI ODO

Kamaishi Institute, Marine Biotechnology Institute, Heita, Kamaishi, Iwate 026, Japan

Abstract. A mussel is attached to hard surfaces by its byssus, which consists of a bundle of threads, each with a fibrous collagenous core coated with adhesive proteins. We constructed a cDNA library from RNA isolated from the foot of the mussel *Mytilus galloprovincialis* sampled in Japan. The library was probed with a nucleotide sequence corresponding to a part of the decapeptide repeat motif in the major adhesive protein of the closely related species *M. edulis*, and a clone including the whole coding region of the same adhesive protein of *M. galloprovincialis* was isolated. The sequences of the signal and nonrepetitive regions of the protein of *M. galloprovincialis* were homologous to those of *M. edulis*, despite several substitutions and a deletion of 18 amino acids. The repetitive region included a tetradecapeptide sequence and 62 repeats of the same decapeptide motif as in *M. edulis*, but hexapeptide sequences present in *M. edulis* were absent in the protein of *M. galloprovincialis*. In the decapeptide motif, two tyrosine residues, two lysine residues, and one of the two proline residues were highly conserved, but other residues were frequently substituted. In some residues in the decapeptide motif, specific codon usages were observed, suggesting that the nucleotide sequence itself has a function.

Introduction

Mussels in the genus *Mytilus* are distributed globally in temperate marine intertidal zones. They attach themselves to solid intertidal surfaces by means of the byssus. The byssus is a bundle of threads each consisting mainly of a fibrous collagenous core coated by adhesive proteins. The protein components of the byssus have been extensively studied in *M. edulis* (Waite, 1987, 1992, for reviews). A major adhesive protein is a 130 kDa protein containing a high proportion of 3,4-dihydroxyphenylalanine (DOPA) residues (Waite and Tanzer, 1981; Waite, 1983). The protein is reported to be largely composed of tandem repeats of the decapeptide Ala-Lys-Pro-Ser-Tyr-Hyp-Hyp-Thr-DOPA-Lys, where Hyp is 3- or 4-hydroxyproline (Waite *et al.*, 1985). Other mussel species have similar proteins, each with a unique repeat motif (Waite, 1986; Waite *et al.*, 1989; Rzepecki *et al.*, 1991). Partial sequences of cDNA and genomic DNA encoding the adhesive protein were reported in *M. edulis* (Strausberg *et al.*, 1989; Filpula *et al.*, 1990). The complete amino acid sequence of the adhesive protein from *M. edulis* has been deduced from its cDNA (Laursen, 1992). These studies showed that this adhesive protein contains more than 80 tandem repeats, of which more than 70 are decapeptides and others are hexapeptides.

In this study, we isolated a cDNA clone containing the whole coding region of the adhesive protein from another major species of mussel, *M. galloprovincialis*, which is closely related to *M. edulis* (Gosling, 1984; Gardner, 1992; Geller *et al.*, 1993). We have found that the cDNA encodes a polypeptide containing 62 repeats of the decapeptide found in *M. edulis*, as well as a tetradecapeptide, but no hexapeptide repeat.

Received 10 January 1994; accepted 28 March 1994.

Materials and Methods

Isolation of mRNA

Mussels (*M. galloprovincialis*) about 4 cm in shell length were sampled at Miyako Bay, Iwate prefecture, Japan. The foot was isolated from 12 mussels and the total RNA

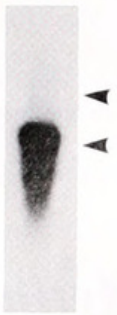


Figure 1. Northern blot analysis of RNA extracted from the foot of *Mytilis galloprovincialis*. One microgram of RNA was electrophoresed on a 1% agarose gel, transferred onto a nylon membrane, and hybridized with the oligonucleotide probe corresponding to a part of the decapeptide sequence. Allowheads indicate the position of 18S and 28S rRNA.

was extracted using the Total RNA Separator Kit (Clontech Laboratories, Palo Alto, CA). Poly(A)$^+$RNA was isolated using the mRNA separator (Clontech Laboratories, Palo Alto, CA).

Northern blot hybridization

Poly(A)$^+$RNA was electrophoresed on a 1% agarose gel, transferred onto a nylon membrane, and hybridized with a [^{32}P]ATP-labeled oligonucleotide probe,

ATA(T,A)GTTGGAGGATAA(C,G)TTGGCTT,

that corresponds to a part of the antisense sequence of the decapeptide repeats of *M. edulis* (Strausberg *et al.*, 1989).

Screening of the cDNA library

cDNA was synthesized using the cDNA Synthesis Kit Plus (Amersham). A cDNA library was constructed using the cDNA cloning system lambda gt10 (Amersham). The library was screened using the same probe used for the northern blotting. Ten positive clones were picked up and the size of inserts was determined by excising with *Eco*RI. The longest insert was subcloned into a plasmid vector, BluescriptIISK+ (Stratagene). Restriction analysis of the BluescriptIISK+ subclone was performed using *Apa*I, *Bam*HI, *Eco*RI, *Hinc*II, *Hind*III, *Kpn*I, *Not* I, *Pvu*II, *Pst*I, *Sac*I, *Sal*I, *Sca*I, *Sma*I, *Spe*I, *Xba*I and *Xho*I.

Sequencing

To determine the whole sequences of both strands of the insert, the plasmid containing the insert was digested with *Apa*I/*Hin*dIII or *Sac*I/*Xba*I, and deletion derivatives were produced using the Kilo-Deletion Sequence Kit (Takara, Kyoto, Japan). The original subclone, 28 *Apa*I/*Hin*dIII-generated clones, and 17 *Sac*I/*Xba*I-generated clones were sequenced using a 373A DNA sequencer (Applied Biosystems Inc.).

Results

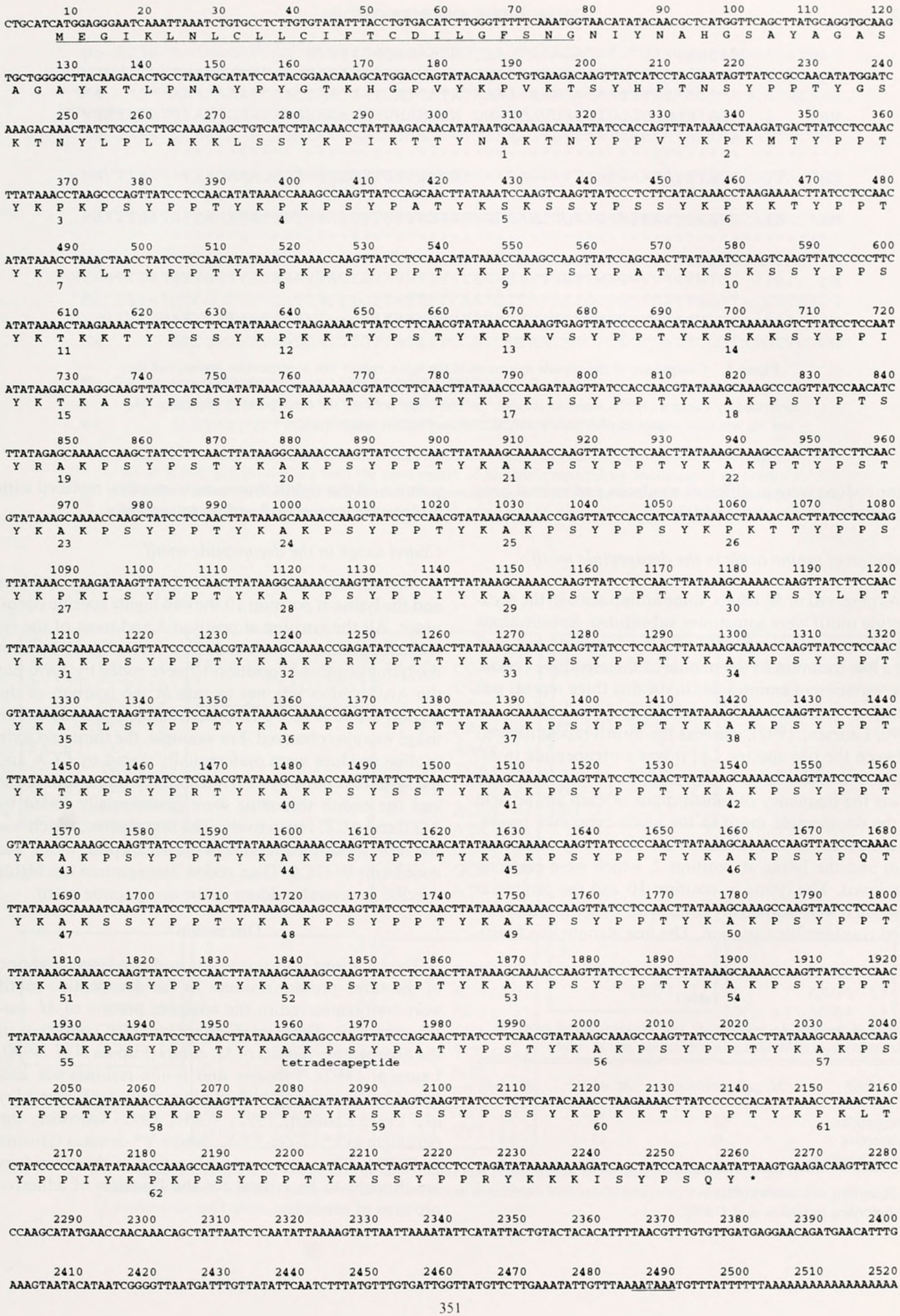
Northern blot hybridization

To examine the efficiency of the probe and to obtain information about the length of the target, northern blot hybridization was carried out. As shown in Figure 1, an intense signal was detected at a position slightly higher than 18S rRNA. This result indicates that the probe is applicable to the screening of the adhesive protein of *M. galloprovincialis*. It also indicates that the target mRNA is expressed in the foot and its length is more than 2.4 kb.

Outline of the structure of the adhesive protein cDNA

About 5×10^4 clones were screened, and more than 50 positive plaques were detected. Of 10 randomly selected clones, 2 were found to have inserts of about 2.5 kb. The longer clone was chosen for further analysis because the shorter one lacked the first several nucleotides (data not shown). Because no restriction site for 16 different enzymes could be found on the insert, deletion derivatives were generated for nucleotide sequence determination. The determined sequence was 2/520 bp, as shown in Figure 2. The coding region determines 751 amino acids, which consist of three distinct parts: the signal peptide of 24 residues, a nonrepetitive region of 76 residues, and a long repetitive region. The amino acid sequence of the signal peptide was similar to that of the *M. edulis* adhesive protein: 22 of 24 residues were conserved between the two species. The amino acid sequence of the nonrepetitive region was also conserved, but several substitutions and a deletion of 18 amino acids were observed (Fig. 3). The repetitive region included 62 repeats of the same decapeptide motif found in *M. edulis*. Although the hexapeptide motif characteristic of *M. edulis* was not observed in the repeats, an irregular tetradecapeptide was seen between the 55th and 56th repeats (Fig. 2, Table I). The sequence of the 3′-untranslated region was also conserved between *M. galloprovincialis* and *M. edulis*, although the termi-

Figure 2. Nucleotide and deduced amino acid sequences of the adhesive protein of *Mytilis galloprovincialis*. Underlined sequence indicates the signal peptide. Also underlined is the polyadenylation signal. Numbers under the amino acid sequence indicate numbers of decapeptide repeats. The asterisks represent the termination codon.


tetradecapeptide

```
            Signal                          Non-repetitive
Mg     1    MEGIKLNLCLLCIFTCDILGFSNG NIYNAHGSAYAGASAGAYKTLPNAYPYGTKHGPVYK
            *************** * ****** ****** * ********** **********  ****
Me     1    MEGIKLNLCLLCIFTFDVLGFSNG NIYNAHVSSYAGASAGAYKKLPNAYPYGTKPEPVYK

Mg    61    PVKTSY------------------HPTNSYPPTYGSKTNYLPLAKKLSSYKPIKTTYN
            ******                   ** ******************************
Me    61    PVKTSYSAPYKPPTYQPLKKKVDYRPTKSYPPTYGSKTNYLPLAKKLSSYKPIKTTYN

            Repetitive
Mg   101    AKTNYPPVYKPKMTYPPTYKPKPSYPPTYKPKPSYPATYKSKSSYPSSYKPKKTYPPTYK
            ********************************* **    *** *  **  ** *  ******
Me   119    AKTNYPPVYKPKMTYPPTYKPKPSYPPTYKSKP----TYKPKITYPPTYKAKPSYPPTYK
            1          2          3          4          5          6
```

Figure 3. Comparison of the peptide sequences of the signal region, the nonrepetitive region, and the first part of the repetitive region of adhesive protein from *Mytilis galloprovincialis* with those of *M. edulis* described by Laursen (1992). Asterisks indicate the homology between the corresponding sequences. Me and Mg represent sequences of *M. edulis* and *M. galloprovincialis*, respectively.

nator codons were in different positions and several base substitutions were observed (Fig. 4).

Variation of amino acids in the decapeptide motif

As observed in *M. edulis*, some amino acids in the decapeptide motif were sometimes substituted. Substitutions were frequent in the first 17 and last 5 repeats, but they were less common in the middle of the repetitive region. The variation of amino acids in the first three repeats was identical with that of *M. edulis* (Fig. 3) (Filpula *et al.*, 1990; Laursen, 1992), whereas the fourth repeat differed between the two species; *i.e.*, it was a decapeptide in *M. galloprovincialis* but a hexapeptide in *M. edulis*. Figure 5 lists the frequency of substitutions of each amino acid in the decapeptide motif in the whole repetitive region. The most conserved residues were the two tyrosine residues and the lysine at position 2, which were perfectly conserved. The lysine at position 10 and the proline at position 6 were also highly conserved. Other residues suffered considerable variation. The first alanine, the fourth serine, and the eighth threonine were often replaced with proline, threonine, and serine, respectively.

Table I

Number of decapeptide, hexapeptide, and tetradecapeptide motifs in adhesive proteins of Mytilus galloprovincialis *and* M. edulis

Motif	*M. galloprovincialis*	*M. edulis*[1]	*M. edulis*[2]
Decapeptide	62	71	72
Hexapeptide	0	13	14
Tetradecapeptide	1	0	0

[1] According to Laursen (1992).
[2] According to Filpula *et al.* (1990).

Codon usage in the decapeptide motif

Among the conserved residues, two tyrosine residues and the lysine at position 10 showed highly specific codon usage. All the tyrosine at position 5 and most of the tyrosine at position 9 were coded by TAT (Fig. 5). Most of the lysine residues at position 10 were coded by AAA, but the AAG codon was not as rare at the position of the second lysine (Fig. 5). In other residues, specific codon usage was also observed. For example, the third and sixth proline residues were preferentially coded by CCA and the seventh proline by CCT. In addition, the fourth serine and the eighth threonine were preferentially coded by AGT and ACT, respectively. The first alanine, which was the most substituted residue in the decapeptide motif, was coded only by GCA. Thus, codon usage pattern was highly specific in several residues in the decapeptide motif.

Discussion

The locations of two tyrosine and two lysine and one of the three proline residues in the decapeptide motif were well conserved in the adhesive protein of *M. galloprovincialis*. These residues are also well conserved in the decapeptide repeats of *M. edulis* (Filpula *et al.*, 1990; Laursen, 1992). Tyrosine and lysine residues are also found in the repeat motifs of other mussels (Rzepecki *et al.*, 1991; Laursen, 1992; Waite, 1992) following the paradigm x-Y* -x-x-x-Y*-K, where Y* denotes tyrosine or DOPA. The presence and location of these residues are thought to be critical for the function of adhesive proteins of mussels.

```
Mg 2234 AAAAAGATCAGCTATCCATCACAATATTAAGTGAAGACAAGTTATCCCCAAGCATATGAA
        ********************  **** ***  ********************* ******* **
Me      AAAAAGATCAGCTATCCATCATCATATAAAGCTAAGACAAGTTATCCCCCAGCATATAAA

Mg 2294 CCAACAAACAGCTATTAATCTCAATATTAAAAGTATTAATTAAAATATTCATATTACTGT
        ********** ***************************** ********** ********
Me      CCAACAAACAGATATTAATCTCAATATTAAAAGTATTAACTAAAATATTCACATTACTGT

Mg 2354 ACTACACATTTTAACGTTTGTGTTGATGAGGAACAGATGAACATTTGAAAGTAATACATA
        ********************* ****************************************
Me      ACTACACATTTTAACGTTTGTATTGATGAGGAACAGATGAACATTTGAAAGTAATACATA

Mg 2414 ATCGGGGTTAATGATTTGTTATATTCAATCTT--TATGTTTGTGATTGGTTATGTTCTTG
        ********************************  ************* ************
Me      ATCGGGGTTAATGATTTGTTATATTCAATCTTAATATGTTTGTGATTTGTTATGTTCTTG

Mg 2474 AAATATTGTTTAAAATAAATGTTTATTTTTT(Poly A)
        ** ******** ******* ******* ***
Me      AAGTATTGTTTCAAATAAAAGTTTATTCTTTTCTGGT(Poly A)
```

Figure 4. Comparison of the nucleotide sequence from the last part of the repetitive region to the poly(A) tail of *Mytilis galloprovincialis* with the same region of *M. edulis* reported by Strausberg *et al.* (1989). Asterisks indicate the homology between the corresponding sequences. The underlined sequence indicates the stop codon. Me and Mg represent sequences of *M. edulis* and *M. galloprovincialis*, respectively.

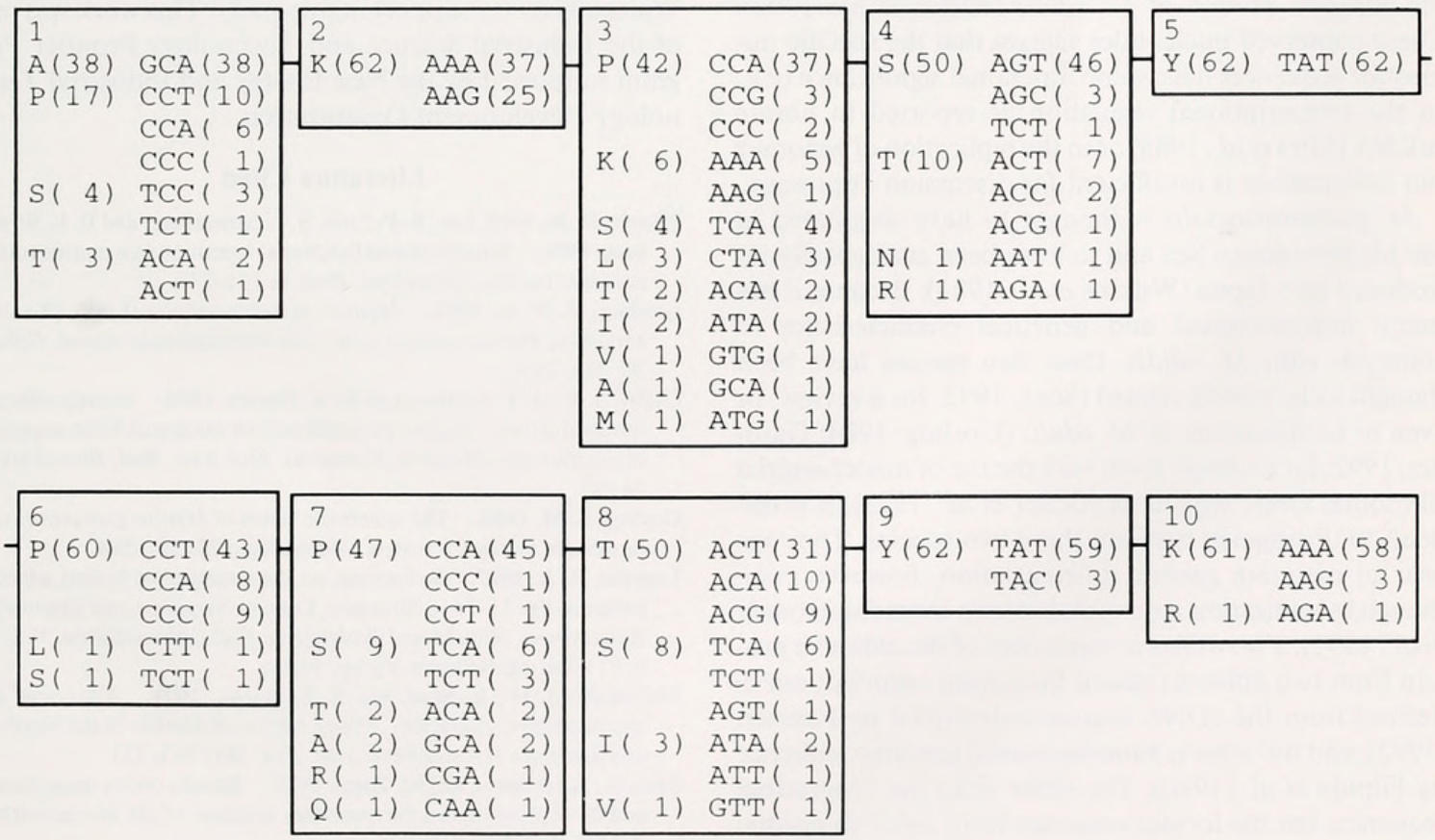


Figure 5. Frequencies of amino acid substitutions and codon usages in the decapeptide motif of the adhesive protein of *Mytilis galloprovincialis*. Numbers in parentheses indicate the frequencies of each of the amino acids and codons, respectively.

In the previous and present studies, three motifs—decapeptide, hexapeptide, and tetradecapeptide—were observed. We noticed that these motifs can be divided into two submotifs, (Y)KAKPSY (submotif A) and (Y)PPTY (submotif B), according to the position of the tyrosine residues. The hexapeptide, decapeptide, and tetradecapeptide motifs corresponded to A, A + B, and A + B + B, respectively, though minor variations were observed in the positions of Ala, Pro, Ser, and Thr residues. The decapeptide motif is obviously the basic motif of the adhesive protein. The hexapeptide motif is apparently also a functional unit, or at least it does not prevent the function, because it is not rare in the repetitive region of *M. edulis*. The tetradecapeptide may also be a functional unit because it is composed of the same submotifs.

In the nucleotide sequence of the repetitive region, specific codon usage was observed. It is interesting that the codon usage in one of the two lysine residues was highly specific but the other was not. In addition, the alanine residues were often replaced with proline or other amino acids, but all the alanine residues at this position were coded only by GCA. The third proline was also frequently substituted by various amino acids, but the CCA codon was preferentially used for proline, and the codons whose third bases are A were preferentially used for other amino acids. The same tendencies were observed in the gene for the adhesive protein of *M. edulis* (Filpila *et al.*, 1990). These conserved nucleotides suggest that the specific nucleotide sequences have some functional significance (*e.g.*, in the transcriptional regulation as reported in fibroin mRNA (Mita *et al.*, 1988) or in the replication of genome); but information is insufficient for discussion at present.

M. galloprovincialis is thought to have originated in the Mediterranean Sea and to have been accidentally introduced into Japan (Wilkins *et al.*, 1983). Because it has many morphological and genetical characteristics in common with *M. edulis*, these two species have been thought to be closely related (Seed, 1992, for a review) or even to be subspecies of *M. edulis* (Gosling, 1984; Gardner, 1992; for reviews). Even with the use of mitochondrial ribosomal DNA sequences (Geller *et al.*, 1993), it is difficult to distinguish between these two species. They appear to maintain genetic differentiation, however, even though hybridization is possible between them (McDonald *et al.*, 1991). Two different sequences of the adhesive protein from two different strains have been reported; one is derived from the cDNA sequence described by Laursen (1992), and the other is from the partial genomic sequence by Filpula *et al.* (1990). The latter lacks the N-terminal sequence, but the former sequence from the 53th residue to the end of the nonrepetitive region was identical with the corresponding sequence of the latter at the amino acid level. The signal region and the nonrepetitive region of *M. galloprovincialis* had amino acid sequences similar to but not identical with those of *M. edulis*. It has been reported that two *M. edulis* sequences in the repetitive region were identical in the first nine and last five repeats and only the distribution pattern of hexapeptides in the middle of the repetitive region was different. The hexapeptides that exist in the repetitive region of *M. edulis* were not found in the sequence of *M. galloprovincialis;* a tetradecapeptide was found instead. We determined partial sequences of other cDNA clones of *M. galloprovincialis*, but also failed to find any hexapeptides (data not shown). However, more information is required before we can discuss the correlation of sequence differences with diversity among populations and species. We are now using polymerase chain reaction to look for interspecific and intrapopulation variation in adhesive protein sequences. The sequence of the adhesive protein may offer a key to understanding not only the function of this protein but also the genetic diversity among different populations and species of mussels.

Acknowledgments

The authors express their sincere thanks to Drs. J. Herbert Waite and Shigeaki Harayama for critical reading of the manuscript, and to Dr. Nobuhiko Takamatsu for helpful advice. We also thank Drs. Shigetoh Miyachi and Wataru Miki for support in this study. This work was part of the Industrial Science and Technology Frontier Program supported by the New Energy and Industrial Technology Development Organization.

Literature Cited

Filpula, D. R., S-M. Lee, R. P. Link, S. L. Strausberg, and R. L. Strausberg. 1990. Structural and functional repetition in a marine mussel adhesive protein. *Biotechnol. Prog.* **6:** 171–177.

Gardner, J. P. A. 1992. *Mytilus galloprovincialis* (Lmk) (Bivalvia, Mollusca): the taxonomic status of the Mediterranean mussel. *Ophelia* **35:** 219–243.

Geller, J. B., J. T. Carlton, and D. A. Powers. 1993. Interspecific and intrapopulation variation in mitochondrial ribosomal DNA sequences of *Mytilus* spp. (Bivalvia: Mollusca). *Mol. Mar. Biol. Biotechnol.* **2:** 44–50.

Gosling, E. M. 1984. The systematic status of *Mytilus galloprovincialis* in western Europe: a review. *Malacologia* **25:** 551–568.

Laursen, R. A. 1992. Reflections on the structure of mussel adhesive proteins. Pp. 55–74 in *Structure, Cellular Synthesis and Assembly of Biopolymers. Results and Problems in Cell Differentiation,* Vol. 19, S. T. Case, ed. Springer-Verlag, Berlin.

McDonald, J. H., R. Seed, and R. K. Koehn. 1991. Allozymes and morphometric characters of three species of *Mytilus* in the Northern and Southern Hemispheres. *Mar. Biol.* **111:** 323–333.

Mita, K., S. Ichimura, and M. Zama. 1988. Specific codon usage pattern and its implications on the secondary structure of silk fibroin mRNA. *J. Mol. Biol.* **203:** 917–925.

Rzepecki, L. M., S.-S. Chin, J. H. Waite, and M. F. Lavin. 1991. Molecular diversity of marine glues: polyphenolic proteins from five mussel species. *Mol. Marine Biol. Biotechnol.* **1:** 78–88.

Seed, R. 1992. Systematics, evolution and distribution of mussels belonging to the genus *Mytilus:* an overview. *Am. Malac. Bull.* **9:** 123–137.

Strausberg, R. L., D. M. Anderson, D. Filpula, M. Finkelman, R. Link, R. McCandliss, S. A. Orndorff, S. L. Strausberg, and T. Wei. 1989. Development of a microbial system for production of mussel adhesive protein. Pp. 453–464 in *Adhesives from Renewable Resources.* ACS Symposium Series 385, R. W. Hemingway and A. H. Conner, eds. American Chemical Society, Washington, DC.

Waite, J. H. 1983. Evidence for a repeating Dopa and hydroxyproline containing decapeptide in the adhesive protein of *Mytilus edulis. J. Biol. Chem.* **258:** 2911–2915.

Waite, J. H. 1986. Mussel glue from *Mytilus californianus* Conrad: a comparative study. *J. Comp. Physiol. B* **156:** 491–496.

Waite, J. H. 1987. Nature's underwater adhesive specialist. *Int. J. Adhesion Adhesives* **7:** 9–14.

Waite, J. H. 1992. The formation of mussel byssus: anatomy of a natural manufacturing process. Pp. 55–74 in *Structure, Cellular Synthesis and Assembly of Biopolymers. Results and Problems in Cell Differentiation,* Vol. 19. S. T. Case, ed. Springer-Verlag, Berlin.

Waite, J. H., and M. L. Tanzer. 1981. Polyphenolic substances of *Mytilus edulis. Science* **212:** 1038–1040.

Waite, J. H., T. J. Housley, and M. L. Tanzer. 1985. Peptide repeats in a mussel glue protein: theme and variations. *Biochemistry* **24:** 5010–5014.

Waite, J. H., D. C. Hansen, and K. T. Little. 1989. The glue protein of ribbed mussels (*Geukensia demissa*): a natural adhesive with some features of collagen. *J. Comp. Physiol. B* **159:** 517–525.

Wilkins, N. P., K. Fujino, and E. M. Gosling. 1983. The Mediterranean mussel *Mytilus galloprovincialis* Lmk. in Japan. *Biol. J. Linn. Soc.* **20:** 365–374.