Polyphosphoprotein from the Adhesive Pads of *Mytilus edulis*†

J. Herbert Waite*

Marine Science Institute and MCDB Department, University of California at Santa Barbara, Santa Barbara, California 93106

Xiaoxia Qin

Department of Cell Biology, Duke University, Durham, North Carolina 27706

Received November 28, 2000; Revised Manuscript Received January 5, 2001

ABSTRACT: Achieving a satisfactory biochemical explanation for the opportunistic underwater adhesion of marine invertebrates such as mussels and barnacles requires a detailed characterization of proteins extracted from holdfast structures produced by these organisms. Mefp-5 is an adhesive protein derived from the foot of the common mussel, *Mytilus edulis*, and deposited into the byssal attachment pads. Purification and primary structure of mefp-5 was determined by peptide mapping and cDNA sequencing. The protein is 74 residues long and has a mass of about 9500 Da. Mefp-5 composition shows a strong amino acid bias: aromatic amino acids, lysine, and glycine represent 65 mol % of the composition. More than a third of all the residues in the protein are posttranslationally modified by hydroxylation or phosphorylation. The conversion of tyrosine to 3, 4-dihydroxyphenyl-L-alanine (DOPA) and serine to *O*-phosphoserine accounts for the hydroxylation and phosphorylation, respectively. Neither modification is complete since variations in the extent of phosphorylation and hydroxylation can be detected by mass spectrometry. More than 75% of the DOPA is adjacent to basic residues, e.g., Lys-DOPA and DOPA-Lys. Phosphoserine occurs in sequences strikingly reminiscent of acidic mineral-binding motifs that appear in statherin, osteopontin, and others. This may be an adaptation for adhesion to the most common substrata for mussels, i.e., calcareous materials.

Sessile marine invertebrates that live along rocky wave-swept shores depend on the design and performance of their attachment strategies for survival. Because they can never be certain of the nature of the available substrata, many have developed biochemical attachment strategies that are opportunistic in the sense that they adhere to almost any hard surface. This is especially true of two mussel genera (*Mytilus*, *Dreissena*, etc.) and makes them attractive models for studying the chemical requirements for adhesive bonding to wet surfaces (*1*, *2*).

Mussels attach to hard surfaces by producing a large number of byssal threads. These function as shock absorbers (*3*, *4*) and, at their distal tips, as adhesive holdfasts. The distal tip of each thread in *Mytilus edulis* is expanded and flared into a flattened adhesive pad with a diameter of 2−3 mm. Thousands of adhesive pads have been analyzed in an attempt to identify the proteins that specifically mediate adhesion. The analysis of byssal proteins is made difficult by their limited solubility, but also because they undergo extensive posttranslational modification and irreversible maturational cross-linking. Most of these problems are related to the redox chemistry of a single functionality in the proteins: 3,4-dihydroxyphenylalanine (DOPA). To date, five DOPA-containing protein families have been identified from the adhesive pads of *Mytilus edulis*: byssal collagen (preCol-D), mefp-1,[1] mefp-2, mefp-3, and mefp-4 (*2*). Recently, a number of random copolymers of DOPA and other amino acids were synthesized and tested in adhesion (*5*, *6*). Results demonstrated that adhesive strength is directly correlated to DOPA content and that DOPA oxidation beyond 5−10 mol % of total DOPA compromises adhesion. This suggested for the first time that DOPA oxidation is not the only fate for DOPA-containing proteins (*7*): indeed, unoxidized DOPA-rich proteins might be critical for good surface-coupling (*8*). In apparent accordance with this, mefp-3, which contains 20 mol % DOPA, predominates near the interface between the pad and substratum (*9*, *10*).

This study had two aims: to characterize other DOPA-rich adhesive proteins in the byssal pads and to determine whether adhesive motifs other than DOPA can be found in these pad proteins. The isolation and characterization of a novel mefp formerly associated with the mefp-3 family of variants is described. In addition to DOPA, mefp-5 contains phosphoserine residues in sequences that resemble those from known mineral-binding proteins.

EXPERIMENTAL PROCEDURES

Protein Extraction from Adhesive Pads. Byssal pads were collected from mussels maintained in circulating seawater

† Funded by Grants N00014-99-1-0774 from the Office of Naval Research and DE10042-7A1 of the Biomaterials Program of the National Institute of Dental and Craniofacial Research. Sequence submitted to GenBank with accession no. AF333078.

* To whom correspondence should be addressed. Fax: (805) 893-7998. E-mail: waite@lifesci.ucsb.edu.

[1] Abbreviations: AU PAGE, acetic acid−urea polyacrylamide gel electrophoresis; EDTA, ethylenediamine tetraacetic acid; mefp, *Mytilus edulis* foot protein; MALDI TOF, matrix-assisted laser desorption ionization time-of-flight; NBT, nitroblue tetrazolium; PCA, perchloric acid; PCR, polymerase chain reaction; RT-PCR, reverse transcription with PCR; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis.

10.1021/bi002718x CCC: $20.00
Published on Web 02/08/2001

at 4−7 °C. In general, mussels were kept tethered to sheets of plexiglas (100 × 100 cm) so that they had no choice but to attach to Plexiglas sheets. Every other day, new pads were harvested by scraping from the sheets using a clean single-edge razor blade. These pads were washed in two changes of milli-Q water and frozen at −80 °C. When 1−2 g (wet weight) of adhesive pads were accumulated, they were then triturated using frosted glass tissue grinders (Kontes) and extracted for protein. This process consisted of homogenization in 10 mL of 5% (v/v) acetic acid with 8 M urea. The resulting puree was microfuged at 15K rpm for 10 min. After removing the supernantant, the pellet was rehomogenized with 5 mL of 5% acetic acid with 6 M guanidine-HCl and 20 mM tris-(2-carboxyethyl)-phosphine (Pierce Chemicals.). This was stirred in the dark at room temperature for 3 h prior to microcentrifugation at 15K for 10 min. The supernatant was collected for dialysis against 4 L of 0.1 M sodium borate, pH 8.2, using tubing with a molecular weight cutoff of 1000 (Spectrum Industries). This resulted in a protein precipitate which was collected by another round of microcentifugation.

Protein Extraction from Mussel Feet. The mussel foot contains stockpiled reserves of all byssal precursors. Thus, foot tissue is a convenient resource when larger amounts of byssal precursor proteins are required. Mussel feet (*Mytilus edulis*) were obtained in bulk on dry ice from NorthEast Transport (Walpole, ME). Feet in lots of 200−300 were thawed, arrayed ventral-side up on 6 × 15 cm glass plates and frozen at −80 °C. The top 0.5 mm of pigmented integument was carefully flayed from each frozen foot, thus revealing the underlying phenol gland (*12*). Dissected phenol glands were pooled, and lots of 19−25 g of glands were successively extracted with (A) 5% acetic acid with protease inhibitors (10 μM leupeptin and pepstatin and 10 mM EDTA), (B) 5% acetic acid with 8 M urea, and finally, (C) 5% acetic acid with 4 M guanidine-HCl. Treatment A was used at a volume to initial weight ratio of 10 mL/g tissue, treatment B at 5 mL/g, and treatment C at 2.5 mL/g. All homogenizations of the feet or pellets were done on ice using 50 mL tissue grinders from Kontes, Inc (Vineland, NJ) and followed by centrifugation at 5 °C at 40000*g* for 45 min. The supernatant for treatment A was enriched in mefp-1 and -2, hence set aside. The supernatant of treatment B contained mostly mefp-3 (*13*). The supernatant of the last extraction, treatment C, was decanted into a beaker, ammonium sulfate was slowly added to a final concentration of 30% (w/v), and the mixture was stirred for at least 40 min at room temperature. The precipitate was removed by centrifugation at 5 °C at 40000*g* for 45 min, and the supernatant was carefully collected for dialysis against 4 L of 0.1% perchloric acid (PCA) using tubing having a molecular weight cutoff of 1000 (Spectrum Industries, Los Angeles), followed by centrifugation and another dialysis against 4 L of 0.1 M sodium borate, pH 8.2. Both dialyses resulted in protein precipitates which were collected by centifugation at 40000*g* for 45 min.

Protein Purification. Precipitates formed during dialysis against PCA and borate and harvested by centrifugation were redissolved in 1−2 mL of 5% acetic acid with 8 M urea. Residual insoluble material was microfuged 15K rpm for 10 min (Eppendorf). The clear supernatant was loaded onto a Brownlee C_8 RP-300 column (0.4 × 25 cm) (Perkin-Elmer) eluted with aqueous 0.1% trifluoroacetic acid and a linear gradient of 5 to 25% (v/v) acetonitrile over 50 min period.

Peptide Mapping. Peptide digests of mefp-5 were prepared by treatment with Lys-C and Arg-C endoproteinase and pepsin (Boehringer-Mannheim). Conditions were as follows for digestion: for pepsin, E:S (weight:weight) 1:200 in 5% acetic acid for 1−5 h at 18−20 °C. Buffer and activator for Arg-C endoproteinase were used according to the specifications of the supplier (Boehringer-Mannheim). For Lys-C, digests were done in 50 mM Tris ascorbate, pH 7.8, at room temperature for 6 h. Peptides were separated by C_8 HPLC using a linear gradient of 1 to 20% acetonitrile over 20 min. Alkaline phosphatase (2000 units/mL, Roche) digestion was done in 5 mM Tris-ascorbate. Positive and negative controls were done with phosvitin as substrate and heat-denatured enzyme, respectively.

Routine Methods. Acid−urea polyacrylamide gel electrophoresis was used to assess the number and mobility of extracted proteins (*14*). Gels were stained for protein with coomassie blue R-250 (Serva Fine Chemicals) and for quinoproteins by a redox cycling method based on nitroblue tetrazolium (NBT) in glycinate buffer (*12*). Isoelectric focusing was done by PAGE in 8 M urea and 6% Triton X-100 in the pH range 3−10 (Sigma ampholyte mix 3−10 and 7−9) (*15*). The direction of electrophoresis was from the anolyte (0.02 M acetic acid) to the catholyte (1 N NaOH) at 1000 V and 3 h using methyl red (pI 3.8), myoglobin (pI 6.8, 7.2), and cytochrome *c* as markers. Phosphoserine was detected using the Ames test with malachite green (*16*) and by amino acid analysis (*12*). Amino acid compositions of proteins/peptides were determined on a 6300 Autoanalyzer (Beckman Instruments, Fullerton, CA) following 1 h (6 M HCl with 10% trifluoroacetic acid and 5% phenol in vacuo at 155 °C) or 24 h (6 M HCl with 5% phenol in vacuo at 110 °C) hydrolyses. The latter was repeated at 6, 12, and 18 h to correct the Ser for losses due to hydrolysis. N-Terminal sequencing was done on a Porton 2020 gas-phase protein sequencer (Beckman-Coulter) with the model 2090 in-line HPLC (*12*). In some samples, phosphoserine was modified to *S*-ethylcysteine with ethanethiol prior to sequencing (*17*).

MALDI TOF Mass Spectrometry. Matrix-assisted laser desorption ionization mass spectrometry with time-of-flight (MALDI TOF) was done on a PerSeptives Voyager DE instrument (Perkin-Elmer). A fresh solution of sinapinic acid (Aldrich) in aqueous 30% acetonitrile and 0.1% trifluoroacetic acid was prepared daily for use as matrix. Sample protein was diluted 1:20 with matrix solution and 1−2 μL was spotted onto the gold-plated sample plate. Typical operating conditions in positive ion mode included an accelerating voltage of 25 000 V, grid voltage at 96%, delay time of 2000 ns, guide wire voltage at 0.10% and N_2 laser power at 1900 to 2500 (arbitrary units). Internal calibrants were bovine insulin with $[M + H]^+$ at 5734.59 and $[2M + H]^+$ at 11468.18 and horse apomyoglogin $[M + 2H]^+ =$ 8476.7

PCR and cDNA Sequencing. We used a cDNA library made from the foot of *M. edulis* prepared in an earlier study (*3*) using the ZAP expression vector (Stratagene, La Jolla, CA). This was screened by PCR with a degenerate sense primer synthesized by Operon Technologies (Alameda, CA): 5′-GARTAYAARGGIGGITAYTAYCC-3′ based on a sequence (EYKGGYYP) near the N-terminus of the

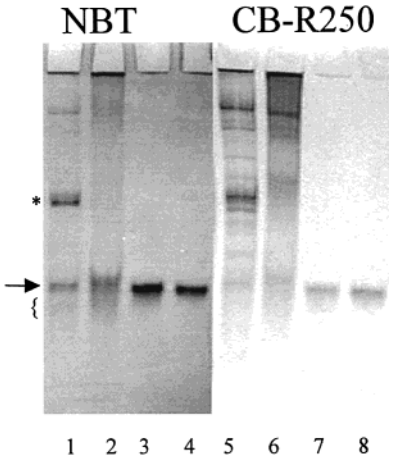


FIGURE 1: Acid−urea polyacrylamide gel electrophoresis of pad and foot extracts containing mefp-5. Urea-insoluble tissue (8 M) was reextracted with 4 M GuHCl; Proteins soluble in 30% (w/w) ammonium sulfate were precipitated by dialysis against 0.5% perchloric acid (foot) or 0.1 M borate (pad): In the left-hand panel (lanes 1−4), the stain was nitroblue tetrazolium (NBT stain). (1) PCA dialysis ppt (foot); (2) borate dialysis ppt, (pad); (3) HPLC fraction 37 (foot); (4) HPLC fraction 21 (pad). The right-hand side (lanes 5−8) was stained with coomassie blue R-250 (CB-250). (5) Same as 1; (6) same as 2; (7) same as 3; and (8) same as 4. Arrow denotes mefp-5. Bracket indicates position of mefp-3 while the asterisk (*) marks mefp-2.

protein. The antisense primer was based on sequence in the universal T7 primer of the ZAP vector 5′-GTAATAC-GACTCACTATAGGGC-3′ (Operon). The PCR reaction was carried out in 50 μL of 1× universal buffer with 100 pmol and 10 pmol of degenerate primer and universal primer, respectively, 250 μmol of each dNTP, 3 μL of the cDNA library (4.4 × 108 plaque forming units) and 2.5 units of Taq polymerase (Stratagene) for 30 cycles on a Robocycler (Stratagene). Each cycle consisted of 30 s at 94 °C, 30 s at 44 °C, and 30 s at 72 °C, with a final extension of 2 min. A 470-bp product with a 370-bp insert was obtained. The products were subcloned into a PCR TA vector (TA Cloning Kit, Invitrogen, Carlsbad, CA) and transformed into the competent bacterial cells (TA Cloning Kit) for amplification, purification, and further restriction and sequence analysis as described previously (*18*). The insert encodes most of the protein and poly-A tail.

To obtain more 5′ sequence, the cDNA foot library was rescreened with an antisense gene specific primer: GSP1, 5′-TGATAACTACCACCTGAATG-3′ (HSGGSYH) and sense universal T3 primer, 5′-AATTAACCCTCACTAAAG-GG-3′ (Operon). A 204 bp product with a 120 bp insert was obtained. The PCR products were gel-purified to remove the primers, salts and enzymes, and then sequenced. The resulting sequence encodes the rest of N-terminus of the protein and partial signal peptide.

5′RACE. To obtain the complete signal sequence, RNA was extracted from fresh foot tissue, reverse transcribed (RT), and 5′ RACE (rapid amplification of cDNA ends) was performed using 5′ RACE System kit (GIBCO, BRL, Bethesda, MD) as follows: amputated feet from freshly collected mussels were placed on metal sheets and fast-frozen with liquid nitrogen. Pigmented epithelium was removed by scraping, and the distal tip of each foot was severed and pooled. Total RNA was extracted and purified as previously described (*18*). One microliter of total RNA was used for RT and 5′ cDNA amplification. Two nested PCR reactions were performed in order to enhance the specificity of PCR. The paired PCR primers were the sense primer provided in the kit (5′ RACE Abridged Anchor primer) and two antisense gene-specific primers (GSPs made by Operon) corresponding to mefp-5 sequences: GSP1, 5′ TGATAACTACCACCT-GAATG 3′ located in the mature protein sequence as described above, and GSP2, 5′ TAAGCTGCTAAGGTCA-CAAG 3′ in the signal sequence. The resulting product was subcloned into a TA vector, amplified, and sequenced.

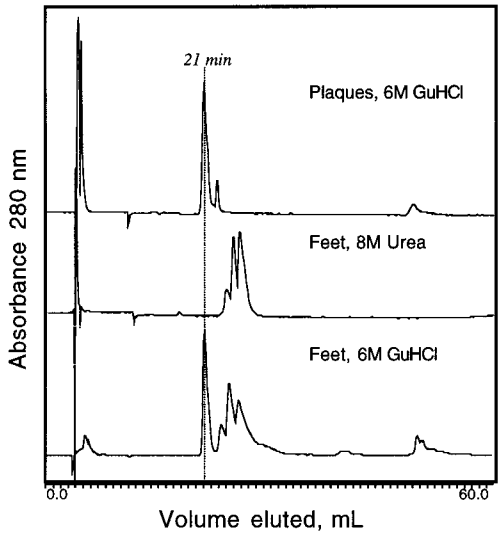


FIGURE 2: Separation of proteins in pad and foot-extracted tissues by C_8 reversed phase HPLC. (a) Guanidine-HCl extracted byssal pads; (b) Guanidine-HCl extracted phenol glands dissected from mussel feet; (c) urea-extracted phenol glands dissected from mussel feet. Dashed vertical line at 21 mL marks position of mefp-5. Gradients are as described in the Experimental Procedures: *Protein Purification*. Flow is 1 mL/min.

Table 1: Amino Acid Compositions of Mefp-5 (mol % or mol/mol)[a]

residue	mefp-5 hydrol (mol %)[b]	mefp-5 hydrol (mol/mol)	mefp-5 cDNA (mol/mol)	mefp-3F (mol/mol)
Asx	2.8	2	2	8
Thr	0	0	0	0
Ser	10.8	8	8	0
Glx	2.8	2	2	0
Pro	1.4	1	1	3
Gly	20.7	15	15	12
Ala	4.7	3	3	1
Cys/2	0	0	0	0
Val	0	0	0	0
Met	0	0	0	0
Ile	0	0	0	0
Leu	1.5	1	1	0
Dopa	25.5	19	0	9
Tyr	0.7	<1	20	1
Phe	0	0	0	0
His	6.5	5	5	0
Lys	19.5	15	15	2
Trp	0	0	0	3
HOArg	0.5	0	0	7
Arg	2.6	2	2	2
Total	100	~74	74	48

[a] A representative of mefp-3, variant F, is added for comparison (*13*). [b] Ser and DOPA were corrected for losses due to hydrolysis.

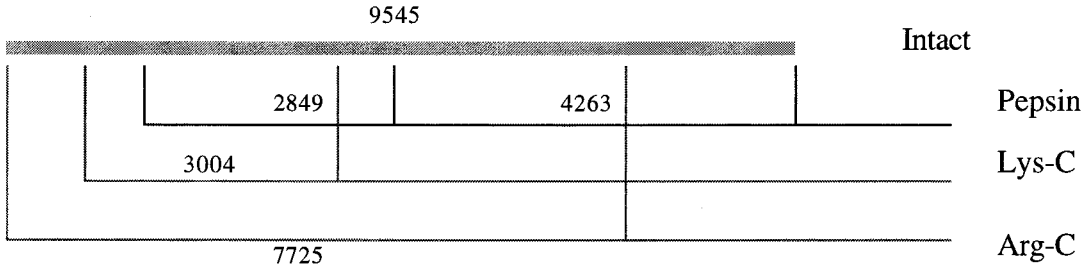
RESULTS

Extraction and Purification. Extraction of mefp-5 from foot and pad tissues requires unusually aggressive conditions,

Table 2: N-Terminal Sequences Determined by Gas-Phase Automated Edman Sequencing[a]

Peptide	*[M+H]$^+$*	*M_{calc}*	*Sequence*
Intact (foot)	9545	-	**SS**EE**Y**KGG**YY**PGNA**Y**H**Y**H**S**GG**SY**HG**S**G**Y**..
		-	*SS*EEXK…. .
Intact (pad)	9545	-	**SS**EE**Y**KGG**YY**PGNA**Y**H**Y**H**S**GG. . . . .
Pep-A	4263.0	4261.36	**YY**K**Y**KN**S**GK**Y**X**Y**LK. . . . .
Pep-B	2840.7	2839.60*	A**Y**H**Y**H…. .
Arg-C I	7727.1	7724.71	**SS**EE**Y**KGG**YY**PGNA**Y**…. .
Lys-C I	3005.8	3003.81	GG**YY**PGNA**Y**H**Y**H**S**GG**SY**HG**S**G**Y**HGG**Y**K…

* Calculated mass assumes one of two Ser phosphorylated



[a] [M + H]$^+$ values were determined by MALDI TOF while those for M_{calc} were estimated from the cDNA sequence and amino acid compositions. Bold S and Y denote detection of PTH−dehydroalanine and DOPA, respectively. Italicized *S* denotes phosphoserine derivatized to *S*-ethylcysteine.

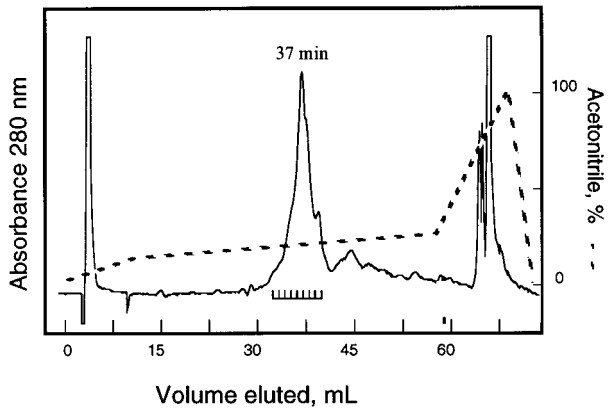


FIGURE 3: Purification of mefp-5 by C_8 reversed-phase HPLC. Fractions under the 37 mL peak (nos. 35−41) were analyzed by mass spectrometry and amino acid analysis. Flow is 1 mL/min.

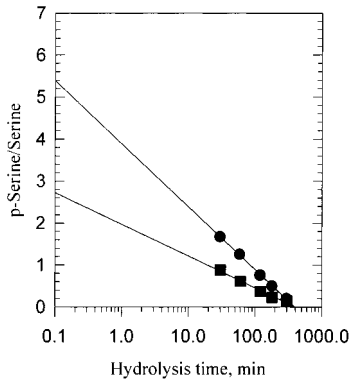


FIGURE 4: Phosphoserine-to-serine ratios in mefp-5 following different hydrolysis times. Both were detected by amino acid analysis. Closed circles denote HPLC fraction 37 (Figure 3) while the closed squares denote no. 39. Moles of phosphorylated residues per mole of mefp5 (P) in each fraction is derived from P = (ratio)-(8-P).

i.e., acidic 4 M guanidine hydrochloride. Earlier studies showed that mefp-3 is extracted by acidic solutions containing 8 M urea (*13*). Present results using AU PAGE suggest that an additional protein, i.e., mefp-5 is extracted from the pellets by 4 M guanidine following the urea treatment (Figure 1). Like mefp-3, this protein exhibits a bathochromic shift when stained with Coomassie Blue R-250 and reduces nitroblue tetrazolium (NBT) in the redox-cycling assay for DOPA (Figure 1). Although mefp-5, is often obscured by mefp-3 on AU PAGE gels, it has a distinct elution time (21 min) on C8 reversed-phase HPLC (Figure 2). The mefp-3 variants all elute later in the gradient as a cluster of peaks (Figure 2). We have exploited this differential extractability by extracting tissue twice with acidic 8 M urea before reextracting the insoluble pellets with 4 M GuHCl. Although the peak at 21 min gave a single sharp band by AU PAGE and a single N-terminal sequence (Table 2), subtle heterogeneity

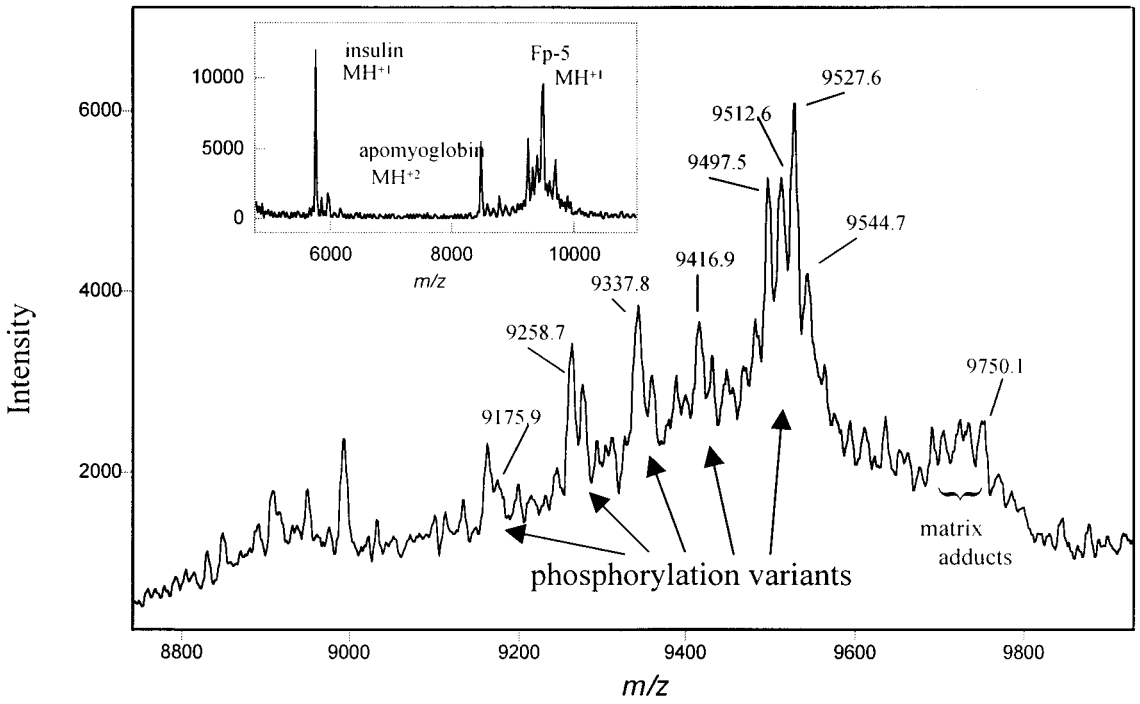


FIGURE 5: MALDI TOF mass spectrometry of HPLC purified mefp-5 (fraction 36). Inset shows the *m/z* distribution relative to two internal standards, monoprotonated human insulin (5734.00) and doubly protonated horse apomyoglobin (8475.00). The expanded spectrum of mefp-5 highlights the hydroxylation variants ($\Delta \approx 16$ Da), the phosphorylation variants ($\Delta \approx 80$ Da), and sinapinic matrix adduct (+222.1 Da) at 9750.1.

was revealed by lowering the acetonitrile gradient in reversed phase HPLC. Figure 3 shows a broadened peak centered at 37 min with several shoulders. Fractions under the peak (36−40) all showed the same N-terminal sequence, however, fractions near the leading edge of the peak contained higher levels of phosphoserine (Figure 4) and DOPA (not shown) than those in the trailing edge which were more serine- and tyrosine-rich. Fraction 37 exhibited an isoelectric point at 8.3 ($\pm$ 0.1, $n = 3$) in 8 M urea; this is significantly lower than the pI> 10 for mefp-3 (*13*).

Composition and Sequence. Amino acid analysis of mefp-5 showed it to have the highest DOPA content of any known protein at 27 mol % (Table 1). Unlike mefp-3 variants, which are Arg-rich, mefp-5 is Lys-rich at 20 mol %. Other abundant residues are Gly (21 mol %) and His (6.5 mol %). Adjusting for losses during hydrolyis, Ser levels are estimated to be 11 mol % or 8 residues/mefp-5. The feature of mefp-5 that most sets it apart from other pad precursors is its phosphate content which averages 6.5 equiv (SD $\pm$ 1.1; $n = 4$) of phosphate/mole protein for fraction 38 as measured by the Ames test (*16*). The phosphate appears to occur entirely as O-phosphoserine: A timed series of mefp-5 (fractions 37 and 39) hydrolyses using 1 M HCl in vacuo at 100 °C followed by cation-exchange chromatography shows early release of an amino acid with the same elution time (3.2 min) as a phosphoserine standard. By 360 min, p-Ser recovery is greatly reduced (Figure 4). Extrapolation of p-Ser/Ser ratios to short hydrolysis times indicates that at least 6.8 Ser residues are phosphorylated in fraction 37 and 5.8 in fraction 39.

The partial sequence of mefp-5 was determined by proteolysis and Edman degradation (Table 1) in combination with mass spectrometry. Pepsinization of the phosphoprotein results in two stable fragments with $[M + H]^+$ masses of 4263 (Pep-A) and 2841 (Pep-B) as detected by MALDI TOF mass spectrometry. Partial N-terminal sequences suggest that Pep-A includes the C-terminus, while Pep-B is an internal sequence. Lys-C and Arg-C-endoproteinase digests of mefp-5 were less useful because of their low mass and extensive DOPA oxidation at pH 7.5, but two larger peptides were partially characterized (Table 2).

Table 3: Comparative Calculated and Observed Masses for Purified Posttranslationally Modified Variants of mefp-5[a]

$[M + H]^{+1}$ (obsd)	mass (calcd)[b]	DOPA[c]	P-Ser[c]
	8664.5	0	0
	9624.3	20	8
9544.7	9544.3	20	7
9527.6	9528.3	19	7
9412.6	9416.3	17	6
9337.8	9336.4	17	5
9258.7	9256.4	17	4
9175.9	9176.4	17	3
9750.1 (matrix adduct)	9528.3 + 224.1	19	7

[a] Matrix is sinapinic acid (mass 224.1). [b] On the basis of mature primary sequence deduced from cDNA and posttranslational modifications as determined by amino acid analysis. [c] Res/mefp-5 molecule.

Mass Spectrometry. Mass spectrometric analysis of fractions corresponding to HPLC-purified mefp-5 reveals a complicated pattern: There is no sharp single peak for the purified protein (fraction 37); rather there is a broad cluster spread out over a mass range of 8.8 to 9.5 kDa (Figure 5). Two periodicities are evident within the clusters. The smaller of these is 16 Da; the larger is 79−80 Da (Figure 5). The 16-Da separation is common in mefps and reflects variability in the degree of posttranslational hydroxylation, e.g., Tyr to DOPA. Peak intervals of about Δ80 Da are suggestive of the presence of phosphate or sulfate groups (Table 3). To probe further the possibility of phosphorylation, the 9.5 kDa component was subjected to alkaline phosphatase treatment.

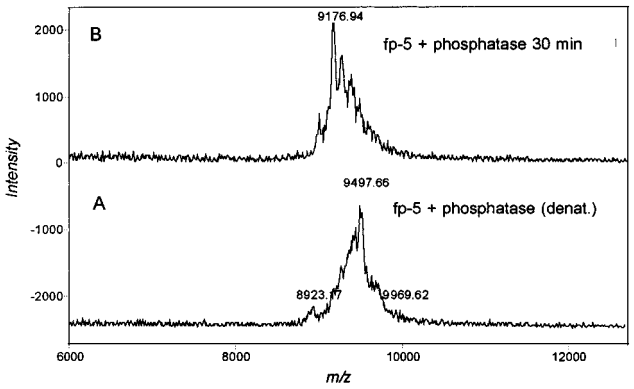


FIGURE 6: MALDI TOF mass spectrometry of HPLC purified mefp-5 after a 30-min treatment with native and boiled alkaline phosphatase. The highest peak at *m*/*z* 9497.66 (panel A) shifts by −321 Da, which approximates the removal of 4 phosphates (320 Da).

MALDI TOF analysis exhibits a progressive mass decrease in decrements of 80−90 Da—the mass change associated with dephosphorylation (Figure 6).

PCR and cDNA. A nearly complete cDNA sequence for Mefp-5 was achieved by PCR amplification of a foot-specific cDNA library using a degenerate primer prepared to the N-terminal sequence in combination with a downstream nondegenerate vector-based primer. The complete sequence including the remainder of the mature N-terminus and most of the signal sequence was obtained by 5′ RACE (Figure 7A). According to SignalP (www.expasy.ch), the predicted signal peptide cleavage site is ambiguous: the highest score (*Y* score = 0.663) occurs after Ala-16; the second highest at Ser-25 (*Y* = 0.580) occurs amid a cluster of serines and can be recognized as the N-terminal sequence of the purified protein. The deduced mature sequence of 74 residues resembles the purified protein in composition (Table 1) except that DOPA is replaced by Tyr, its precursor, and much of Ser is replaced by p-Ser. The dominant residues are Gly (15 mol/mol), Lys (15 mol/mol), and Tyr (20 mol/mol) together accounting for 67% of the total (Table 1). Although there are several incidences of K−Y and G−Y, there is no evidence of repeated sequences beyond this. K and Y are preferentially concentrated in the C-terminal half of the molecule, while H and E are located in the N-terminal half.

The cDNA-derived sequence sans signal peptide corresponds to a mass of 8664.54 and a theoretical pI of 10.03 (Swiss-Prot). Adding to this the mass of nineteen hydroxylations (Tyr→Dopa, 15.9994 × 19) and seven phosphorylations (79.9799 × 7), an expected mass of 9528.3 and pI of 8.3 for the mature mefp-5 is obtained. This matches the measured pI and differs from the measured $[M + H]^+$ of 9527.7 by 1.6 Da (Table 3). A comparable accuracy is evident for other variants of posttranslational modification in Table 3. While better accuracy has been observed with MALDI TOF in other proteins, it is quite good in the present instance given the tendency of each DOPA to lose 2 Da by oxidation. A similar application of the cDNA-deduced sequence to the mass and known sequence of peptic, Lys-C and Arg-C endopeptidase fragments (Table 2) confirms that, allowing for mass increases due to DOPA and P−Ser, the deduced sequence is likely to be the correct one.

A.

```
5′ GGCCACGCGTCGACTAGTACGGGGGGGGGGGGAAA

AGATTAGCCAACATGAAGCTTAGTTGTATAGTCTTGGTGCTCTTT
            M  K  L  S  C  I  V  L  V  L  F
       1as
CTTGTGACCTTAGCAGCTTATAGCGATGTAGGCTCATCGAGTTCT
 L  V  T  L  A  A  Y  S  D  V  G  S  S  S  S
              2s ⇑                       ⇑
GAAGAATACAAGGGGGGGTATTACCCAGGCAATGCGTACCACTAT
 E  E  Y  K  G  G  Y  Y  P  G  N  A  Y  H  Y
          3as
CATTCAGGTGGTAGTTATCACGGATCCGGCTATCATGGAGGATAT
 H  S  G  G  S  Y  H  G  S  G  Y  H  G  G  Y

AAAGGAAAGTATTACGGAAAAGCAAAGAAATACTATTATAAATAT
 K  G  K  Y  Y  G  K  A  K  K  Y  Y  Y  K  Y

AAAAACAGCGGAAAATACAAGTATCTGAAGAAGGCTAGAAAATAC
 K  N  S  G  K  Y  K  Y  L  K  K  A  R  K  Y

CATAGAAAGGGTTACAAGTATTATGGAGGTAGCAGTTAGACACTC
 H  R  K  G  Y  K  Y  Y  G  G  S  S  ♠

TTGATAGAATGACCACTACATGAAGATGAATTCTAAAGAGTACA
ATGGTATTGTTCAGCAACTATTCCAAATATGGTCATCTTTAATC
GACAGGTTGTCATAAAAAATACAAATAAACAAATATATGTATAAA
AAAAAAAAAAAAAAAAAAAAAA 3′
```

B.

```
1         10          20          30          40
SSEEYKGGYY PGNAYHYHSG GSYHGSGYHG GYKGKYYGKA

41        50          60          70
KKYYYKYKNS GKYKYLKKAR KYHRKGYKYY GGSS
```

FIGURE 7: (A) cDNA-derived protein sequence of mefp-5. Residues confirmed by peptide sequencing are denoted by a dotted underline. Spade marks the termination codon. Vertical double upward arrows indicate the ambiguous predicted signal peptide cleavage sites, and in the second case, the beginning of mature fp-5 transcript. Polyadenylation signal is doubly underlined. Singly underlined cDNA sequences denote primers used to determine the complete sequence: 1as, GSP-2 nested antisense primer for RACE; 2s, degenerate oligo primer for PCR of the cDNA; 3as, GSP-1 antisense primer for RACE. (B) Deduced mature sequence of mefp-5, bold **S** and **Y** denote phosphoserine and DOPA, respectively. Underlined S in position 27 may be underphosphorylated relative to the other serines.

DISCUSSION

Mefp-5 is an ingredient of the mussel's adhesive pad and, at over one in every 4 amino acids, contains the highest level of DOPA reported in a protein. DOPA and the abundance of basic residues are reminiscent of other characterized byssal precursors, particularly mefp-3, but detection of a significant number of phosphoserine residues distinguishes it. Typically, seven of eight Ser residues in mefp-5 are phosphorylated. The precise location of the nonphosphorylated Ser is unknown, however, peptic peptide B and phosphorylation identification programs (ref *19*; NetPhos) suggest it to be Ser-26. Since the predicted signal peptide cleavage site is ambiguous at Ala-16 or Ser-24, we are left with the following processing quandary: if the former pertains, then removal of propeptide AYIDVGSS is required prior to secretion; if the latter, then signal peptide cleavage results in the mature sequence.

Mefp-5 joins a growing number of phosphorylated extracellular proteins with a structural role. These include salivary statherins (*20*), osteopontins (*21*), dentin phosphoryns (*22*,

Table 4: Sequences at or near the N-Terminus of mefp-5 and Mineral-Binding Proteins

sequence (range)[a]	protein	reference
*pSp*SEEYK (1−6)	mefp-5	present results
*pSp*SEEKF (2−7)	statherin	*20*
*pSp*SEEKQ (10−15)	osteopontin	*21*
*pSp*SEESI (18−23)	β-casein	*28*

[a] Mature protein sequence.

23), epidermal precursor profilaggrin (*24*), and certain proteins from mollusk shells (*25*). In most of these, the phosphoserines are presumed, or have been shown, to bind calcium or a calcareous mineral surface (*26*, *27*). Of the seven phosphorylation sites in mefp-5, the N-terminus, i.e., pSpSEE, is the most revealing. This closely matches similar sequences in statherin (*20*), β-casein (*28*) and osteopontin (*21*) that are mineral and calcium ion-binding (Table 4). In statherin, binding of the hexapeptide DpSpSEEK to hydroxyapatite was investigated by ^{13}C solid-state NMR (*26*). These results and others (*26*−*27*, *29*) suggest that the peptide is bound to the mineral in an open extended conformation and that binding is much diminished by dephosphorylation of Ser (*30*). Mineral binding by the sequence is such that it has been included to confer mineral-binding properties on recombinant peptides (*31*).

Since mussel byssus is devoid of living cells and not mineralized, it is unlikely that phosphoserines in mefp-5 are intended for regulatory pathways or for calcification of extracellular matrix. Instead, p-Ser may be an ideal adhesive adaptation for the mussel's most probable substratum: another mussel shell made of calcium carbonate (*32*). Given the binding affinity phosphoproteins have for calcareous materials (*20*−*21*, *26*), it seems plausible that the inclusion of phosphoserine into mefp-5 serves to enhance the byssal adhesive repertory for calcareous surfaces such as shells.

Recent studies by Yu and Deming (*5*, *6*) with synthetic DOPA-containing random copolymers have revealed that high interfacial concentrations of DOPA are critical for wet adhesion. Perhaps, the high mol % of DOPA in the precursors, mefp-3 and mefp-5, is an adaptation to their interfacial role. DOPA is not, however, the only functionality exploited for adhesion. Phosphoserine would be an ideal modification for adhesion to those substrata, e.g. calcareous, to which DOPA -rich proteins show only moderate binding.

ACKNOWLEDGMENT

We thank the Office of Naval Research for the PerSeptives Voyager DE mass spectrometer.

REFERENCES

1. Anderson, K. E., and Waite, J. H. (1998) *Biol. Bull. 194*, 150−160.
2. Vreeland, V., Waite, J. H., and Epstein, L. (1998) *J. Phycol. 34*, 1−8.
3. Coyne, K. J., Qin, X. X., and Waite, J. H. (1997) *Science 277*, 1830−1832.
4. Waite, J. H., Qin, X. X., Coyne, K. J. (1998) *Matrix Biol. 17*, 93−108.
5. Yu, M., and Deming, T. J. (1998) *Macromolecules 31*, 4739−4745.
6. Yu, M., Hwang, J., and Deming, T. J. (1999) *J. Am. Chem. Soc. 121*, 5825−5826.
7. McDowell, L. M., Burzio, L. A., Waite, J. H., and Schaefer, J. (1999) *J. Biol. Chem. 274*, 20293−20295.
8. Ooka, A. A., and Garrell, R. L. (2000) Biopolymers *57*, 92−102.
9. Floriolli, R., von Langen, J., and Waite (2000) *Mar. Biotechnol. 2*, 1−12.
10. Warner, S. C., and Waite, J. H. (1998) *Mar. Biol. 134*, 729−734.
11. Diamond, T. V. (1993) Masters Thesis, University of Delaware.
12. Waite, J. H. (1995) *Methods Enzymol. 258*, 1−20.
13. Papov, V. V., Diamond, T. V., Biemann, K., and Waite, J. H. (1995) *J. Biol. Chem. 270*, 20183−20192.
14. Waite, J. H., and Benedict, C. V. (1986) *Methods Enzymol. 107*, 26−46.
15. Martensen, T. M. (1984) *Methods Enzymol. 107*, 1−23.
16. Rzepecki, L. M., Hansen, K. M., and Waite, J. H. (1992) *Biol. Bull. 183*, 123−137.
17. Meyer, H. E., Hoffmann-Posorske, E., and Heilmeyer, L. M. G. (1991) *Methods Enzymol. 201*, 169−185.
18. Qin, X. X., Coyne, K. J., and Waite, J. H. (1997) *J. Biol. Chem. 272*, 32623−32627.
19. Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) *Protein. Eng. 10*, 1−6.
20. Oppenheim, F. G., Offner, G. D., and Troxler, R. F. (1982) *J. Biol. Chem. 257*, 9271−9282.
21. Kohri, K., Suzuki, Y., Yoshida, K., Yamamoto, K., Amasaki, N., Yamate, T., Iguchi, M., Sinohara, H., and Kurita, T. (1992) *Biochem. Biophys. Res. Commun. 184*, 859−864.
22. Butler, W. T. (1998) *Eur. J. Oral Sci. 106* (Suppl), 204−210.
23. Veis, A. Wei, K., Sfeir, C., George, A., and Malone, J. (1998) *Eur. J. Oral Sci. 106* (Suppl 1), 234−238.
24. Resing, K. A., Dale, B. D., and Walsh, K. A. (1985) *Biochemisty 24*, 4167−4175.
25. Marsh, M. E. (1986) *Biochemistry 25*, 2392−2396.
26. Long, J. R., Dindot, J. L., Zebroski, H., Kiihne, S., Clark, R. H., Campbell, A. A., Stayton, P. S., and Drobny, G. P. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 12083−12087.
27. Meisel, H., and Olieman, C. (1998) *Anal. Chim. Acta 372*, 291−297.
28. Ribadeau-Dumas, B., Brignon, G., Grosclaude, F., and Mercier, J.-C. (1971) *Eur. J. Biochem. 20*, 264−268.
29. Raj, P. A., Johnsson, M., Levine, M. J., and Nancollas, G. H. (1992) *J. Biol. Chem. 267*, 5968−5976.
30. Jono, S., Peinado, C., and Giachelli, C. M. (2000) *J. Biol. Chem. 275*, 20197−20203.
31. Gilbert, M., Shaw, W. J. Long, J. R., Nelson, Kjell, Drobny, G. P., Giachelli, C. M., and Stayton, P. S. (2000) *J. Biol. Chem. 275*, 16213−16218.
32. Falini, G., Albeck, S., Weiner, S., and Addadi, L. (1996) *Science 271*, 67−69.

BI002718X