THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 281, NO. 16, pp. 11090–11096, April 21, 2006

Probing the Adhesive Footprints of *Mytilus californianus* Byssus*[S]

Received for publication, October 3, 2005, and in revised form, December 16, 2005 Published, JBC Papers in Press, February 22, 2006, DOI 10.1074/jbc.M510792200

Hua Zhao, Nicholas B. Robertson, Scott A. Jewhurst, and J. Herbert Waite[1]

From the Department of Molecular, Cellular and Developmental Biology and Department of Chemistry and Biochemistry, Marine Science Institute, University of California, Santa Barbara, California 93106

California mussels *Mytilus californianus* owe their tenacity to a holdfast known as the byssus, a fibrous extracellular structure that ends distally in flattened adhesive plaques. The "footprints" of freshly secreted plaques deposited onto glass coverslips were shown by matrix-assisted laser desorption ionization time of flight mass spectrometry to consist chiefly of proteins ranging in mass from 5200 to 6700 Da. These proteins, variants of a family known as mcfp3 (*M. californianus* foot protein 3), were purified from acetic acid/urea extracts of plaques and foot tissue. Mcfp3 appears to sort into fast and slow electrophoretic variants. Both are rich in Gly and Asn and exhibit post-translational hydroxylation of Tyr and Arg to Dopa and 4-hydroxyarginine, respectively, with the fast variant containing more than twice as much Lys + Arg. Both the slow and fast variants were partially sequenced from the N terminus, and the complete sequences of 12 variants were deduced from cDNA using degenerate oligonucleotides, PCR, and rapid amplification of cDNA ends. Mcfp3s are highly polar molecules and contain up to 28 mol % Dopa, which remains intact and may be crucial for adhesion to metal and mineral surfaces.

The fabrication of strong and durable adhesive bonds between macromolecules and metal or mineral surfaces in a wet environment is one of the most persistent challenges for modern adhesive technology (1). Ironically, this is "business as usual" for the marine mussels who inhabit rocky wind-swept seashores. Mussels (*Mytilus*) thrive despite persistent surf and tides thanks in part to a robust holdfast structure called the byssus, which consists of a bundle of threads each of which is tipped distally by an adhesive plaque that bonds to mineral and metal surfaces. The adhesive strategies of mussels and other sessile marine invertebrates are increasingly envisioned as paradigms for designing tough water-resistant adhesives (2, 3).

How is water a problem for adhesion? There are many complicating factors, but one of the most fundamental involves the dielectric constant of water. Most known cases of practical and biological adhesion are mediated by noncovalent interactions between the adhesive and the adherend surface. The interaction energy of a charge-charge interaction, for example, is governed by Coulomb's law as

$$E = -(Q_1Q_2)/[4\pi\varepsilon r] \quad \text{(Eq. 1)}$$

where Q_1 and Q_2 are the charges on two interacting molecules, ϵ is the dielectric constant of the medium, and r is the interatomic distance (4). The dielectric constant of water ($\epsilon = 80$) is much higher than that of vacuum ($\epsilon = 1$) or nonpolar solvents ($\epsilon = 2$–3), and this, consequently, greatly diminishes the magnitude of interaction energies that are possible in water. Ligand-receptor adhesion typically gets around this limitation by conformation-dependent protein-ligand interactions, in which ligand binding usually occurs within hydrophobic protein pockets having a much lower dielectric constant than the surrounding bulk water (5). The extent to which mussels can exploit this approach in byssal adhesion seems limited given how polar the mineral and metal surfaces to which they stick are.

The biochemistry of mussel byssus has been investigated in primarily *two* species, *Mytilus edulis* and *Mytilus galloprovincialis*, both relatively sheltered species (6). Not surprisingly, all of the proteins characterized so far in the two species show a very high degree of sequence homology. The California mussel, *Mytilus californianus*, in contrast to its congeners, dominates the most wave-swept and exposed habitats along the Pacific coast from Baja California to Vancouver Island. Its mean dislodgement tenacity was determined to be 250 N/mussel, forty times greater than that of *M. edulis* (7). Despite this superiority, almost nothing is known about the comparative biochemistry of adhesion in *M. californianus*.

At least six different proteins have been characterized from freshly secreted adhesive plaques of *M. edulis*. These are mefp1, 2, 3, 4, and 5 and various preCols (6). Extensive loss of solubility and epitopes caused by rapid protein processing has greatly impeded the precise immunochemical localization of each of these in the byssus (8).

Our hypothesis in the present study is that one of these proteins is strategically localized near the interface for adhesion. We have interrogated the footprints of *M. californianus* byssal adhesive plaques by direct laser desorption ionization mass spectrometry and identified several new Dopa-rich mcfp3 proteins. The polar and highly modified nature of these proteins suggests that adhesion may be governed by novel interactions in addition to the usual noncovalent types.

MATERIALS AND METHODS

Protein Purification from Mussels—Mussels were collected locally from rocks and jetties around Campus Point and Goleta Pier in Santa Barbara, CA and immediately transferred to shallow holding tanks with circulating raw seawater at 15 °C. About 60 mussels were tethered to each 18 × 25-cm plate made of acrylic or glass (thickness, 1 cm), from which byssal plaques were harvested daily using a clean single-edge razor blade, briefly rinsed with 200 volumes of MilliQ water, and stored at −80 °C. About 1000 accumulated plaques were thawed and extracted in a small volume (5 ml/200 plaques) of 5% acetic acid (v/v) containing

* This work was funded in part by grants from the University Research Engineering and Technology Institute on Bio-Inspired Materials of NASA under Award NCC-1-02037 (to J. H. W.) and by National Institutes of Health Grant DE015415. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains a supplemental figure.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AY960607, AY960608, AY960609, AY960610, AY960611, AY960612, AY960613, DQ165553, DQ165554, DQ165555, and DQ165556.

[1] To whom correspondence should be addressed. Tel.: 805-893-2817; Fax: 805-893-7998; E-mail: waite@lifesci.ucsb.edu.

8 M urea and 0.1 mM tri(carboxyethyl)-phosphine by homogenization on ice using a small hand-held tissue grinder (Kontes, Vineland, NJ). The homogenate was centrifuged for 30 min at 20,000 × *g* and 4 °C. The supernatant was collected (10 ml), and half was dialyzed against 4 liters of MilliQ water using membrane tubing with a 1,000-dalton cut-off (Spectrum Industries). Dialysis resulted in a turbidity, which was clarified by centrifugation (30 min at 20,000 × *g* and 4 °C), and sedimented material was redissolved in 5% acetic acid with 8 M urea, whereas the supernatant was freeze-dried at −80 °C.

The acetic acid/urea plaque extracts were subjected to reverse phase HPLC[2] using a 260 × 7-mm RP-300 Aquapore (Applied Biosciences Inc., Foster City, CA) column eluted with a linear gradient of aqueous acetonitrile (9). Eluant was monitored continuously at 280 nm, and collected 1-ml fractions were assayed by amino acid analysis and electrophoresis following freeze-drying.

Electrophoresis—Routine electrophoresis was done on polyacrylamide gels (7% acrylamide and 0.2% *N*,*N*′-methylenebisacrylamide) containing 5% acetic acid with 8 M urea (10). This system is ideal for basic Dopa-containing proteins because it can be processed for protein or Dopa staining with equal facility. The proteins were stained with Serva Blue R (Serva Fine Chemicals, Westbury, NY), whereas Dopa was stained with either the Arnow reagent (10) or nitro blue tetrazolium redox cycling (11).

Amino Acid Analysis and Sequencing of Peptides—The peptides and proteins were hydrolyzed in 6 N HCl with 5% phenol *in vacuo* at 110 °C for 12–48 h to correct for the losses of certain amino acids (12). Recovery of tryptophan required hydrolysis in 4 N methanesulfonic acid (13) or in 6 N HCl with 30% phenol *in vacuo* for 20 and 40 min at 165 °C (14). Routine amino acid analysis was done by ion exchange and a ninhydrin-based detection system (Beckman System 6300 Auto Analyzer) using a previously described step elution program (15). Hydroxyarginine eluted at 80 min and was quantified using the molar color yield of arginine. Because of its coelution with ammonia, tryptophan was separately quantitated on System 6300 using a 40-min program of NaD (5% sodium chloride and 1.9% sodium citrate at pH 6) at a column temperature of 70 °C. The amino acid sequence of protein and peptides was derived by automated Edman degradation using a Porton Instruments 2090 Microsequencer (Beckman-Coulter, Fullerton, CA). Phenylthiohydantoin derivatives of amino acids were chromatographically separated according to a gradient program specified earlier (15). The elution time for phenylthiohydantoin-4-hydroxyarginine (8.2 min) was determined in a previous study by Papov *et al.* (9).

Adhesive Plaque Footprints—To obtain adhesive plaque footprints, the mussels were tethered over glass coverslips, and when one or more plaques had been deposited onto the substrates, the connecting threads were severed, and the coverslips were removed. The coverslips were thoroughly rinsed with two changes of MilliQ water. The underside of the glass coverslip was dried so that the locations of the adhering plaques could be marked with a black marker pen. The plaques were then removed with a clean single-edge razor. The footprints were imaged by stereo light and scanning electron microscopy.

MALDI-TOF Mass Spectrometry—MALDI-TOF experiments were performed using a Voyager DE (linear delayed extraction) mass spectrometer (Applied Biosystems, Foster City, CA). The MALDI matrix was prepared by dissolving sinapinic acid (10 mg/ml) in 70% acetonitrile. The Mcfp3 protein or peptides derived thereof were dissolved in this matrix solution to give a final concentration between 1 and 10 pmol/μl. About 1 μl of this solution was applied to the target plate and allowed to evaporate. The sample spots were irradiated using an N_2 laser (LSI, Inc., Cambridge, MA) with a wavelength of 337 nm and a pulse width of 8 ns and operated at a repetition rate of 5 Hz. MALDI ionization generates protonated singly and doubly charged ions and dimers for the Mcfp3 protein (mostly singly charged ions for peptides) that were accelerated using either 20- or 25-kV accelerating voltage. The resolution was about 1:500, which was sufficient to allow mass assignment of the major peaks caused by the different hydroxylation states of the Mcfp3 protein.

Footprints of byssal plaques were screened for proteins by MALDI-TOF mass spectrometry. The technique as currently practiced involves the following: glass coverslips were collected shortly after plaque deposition. The coverslips were cleaned of slime and debris by washing thoroughly with MilliQ water, after which the plaques were scraped off using a clean single-edge razor. The glass surface with the footprint residue was dried and mounted onto a MALDI sample plate with double-sided tape. Two microliters of matrix mixture (30% sinapinic acid) were applied to the footprint and air-dried before subjecting to pulsed laser irradiation. With respect to matrix type, α-cyano-4-hydroxycinnamic acid can be substituted for sinapinate in fp3 analysis although with a lower signal-to-noise ratio. Singly and doubly charged peaks for bovine insulin and thioredoxin (5734.59 and 5837.74, respectively, for average masses) were used as internal calibrants. To compensate for possible surface effects on MALDI by different materials such as glass and the gold sample plate, internal standards were always recalibrated on each test surface.

RNA Extraction, Reverse Transcription-PCR, and 5′-RACE—Total RNA was extracted from a foot tip of *M. californianus* using a RNA plant mini kit (Qiagen). Qiagen's protocol was followed after disrupting one freshly dissected mussel foot tip in liquid nitrogen with a mortar and pestle.

First strand cDNA was synthesized from total RNA using superscript II reverse transcriptase with an Adaptor Primer (5′-GGC CAC GCG TCG ACT AGT ACT T_{16}-3′) from Invitrogen. The product of the reverse transcription reaction was used in subsequent PCR. The first strand was PCR-amplified with the degenerate oligonucleotides (sense, 5′-GGN GGN AAY TAY TAY CCN AAR TAY AA-3′ and 5′-GGN TAY GAY TAY TAY AAR GGN TA-3′), which correspond to the N-terminal sequence of two variants of Mcfp3, *i.e.* GGNY*Y*PKYN and GY*DY*Y*KGY*(Y* denotes DOPA), respectively, and an abridged universal amplification primer (antisense, 5′-GGC CAC GCG TCG ACT AGT AC-3′ from Invitrogen). The PCR was performed in 25 μl of 1× Buffer B (Fisher) and 5 pmol of each primer, 5 μmol of each dNTP, 1 μl of first strand reaction, and 2.5 units of *Taq* DNA polymerase (Fisher) for 32 cycles on a Robocycler (Stratagene). Each cycle consisted of 30 s at 94 °C, 30 s at 52 °C, and 20 s at 72 °C, with a final extension of 5 min. The PCR products were subjected to 2% agarose gel electrophoresis, purified, cloned into a PCR TA vector (TOPO TA cloning kit; Invitrogen), and transformed into competent Top 10 cells (Invitrogen) for amplification, purification, and sequencing. The insert encodes the entire sequence of mature Mcfp3 with the 3′-untranslated region.

The gene-specific primer, antisense 5′-ACT AAT AGT TAT GGA GAT AGC TTG CTT C-3′, and 5′-GTT ATT CCA GCC TTT ATT CCA GCC ATA-3′, which prime the C terminus of two variants of Mcfp3, respectively, and GeneRacer 5′ primer were used to clone the 5′ end of Mcfp3 by a 5′-rapid amplification of cDNA ends (5′-RACE) protocol using a GeneRacer kit from Invitrogen. PCR conditions, cloning, and sequencing were the same as described above.

[2] The abbreviations used are: HPLC, high pressure liquid chromatography; MALDI-TOF, matrix-assisted laser desorption ionization time of flight; NBT, nitro blue tetrazolium; RACE, rapid amplification of cDNA ends; CB, Coomassie Blue.

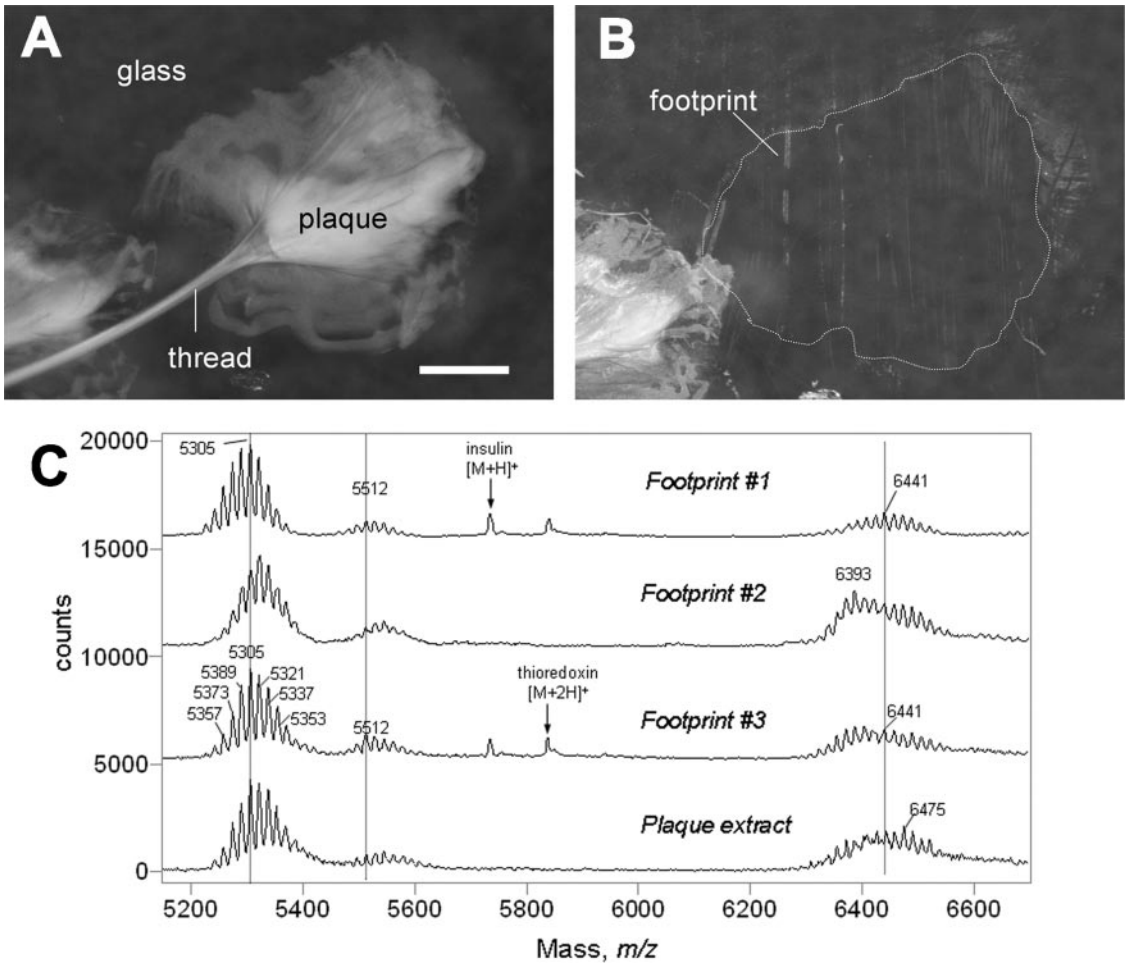

FIGURE 1. **Light micrograph of a byssal adhesive plaque (*A*) and footprint (*B*) on glass following removal with a clean single-edge razor blade.** Outline (*dotted line*) denotes the original contact area between glass and plaque. *Scale bar*, 1 mm. *C*, MALDI-TOF mass spectrometry of the footprint proteins in three different footprints compared with proteins extracted from plaques using 5% acetic acid and 8 M urea. The calibration standards included in footprints 1 and 3 are insulin ($M+H^+ = 5734.5$) and thioredoxin ($M+2H^+$ 5837.7).

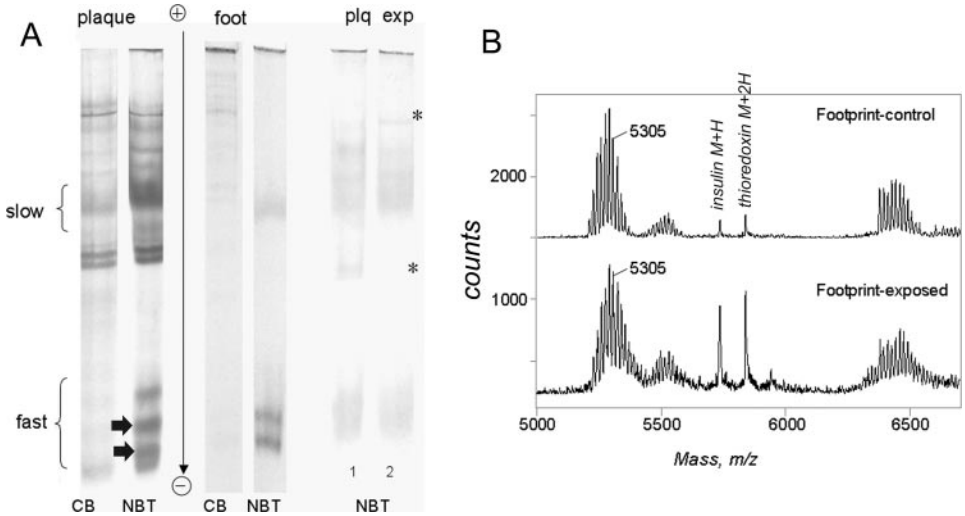

FIGURE 2. **Mussel fp3 adhesive plaque proteins separated by acid-urea polyacrylamide gel electrophoresis.** *A*, plaques extracted with 5% acetic acid and 8 M urea. Foot extract used the same solvent. The *arrows* indicate the positions of fast fp3 variants. Fresh (*lane 1*) and 24-h exposed (*lane 2*) plaque extracts (10 plaques each) were stained with NBT. Proteins denoted by an *asterisk* are those most affected by exposure; the *lower* disappeared, whereas the *upper* accumulated. *B*, fresh and exposed footprints interrogated by MALDI. Peaks at 5734.5 and 5837.7 represent internal calibrants as in Fig. 1.

RESULTS

Plaque Footprint Proteins—MALDI-TOF mass spectrometry was used to interrogate plaque footprints after plaque removal from glass coverslips so that little to no residue could be observed by light microscopy or scanning electron microscopy (Fig. 1, *A* and *B*). MALDI mass spectroscopy reproducibly detected small proteins ranging from 5 to 7 kDa in the footprints (Fig. 1*C*), which are 1–2 orders of magnitude more abundant than any other detectable proteins in the range of 1,000–50,000 Da (supplemental material). All of the footprint proteins exhibit laddering. For example, the peptide centered at 5.3 kDa is a cluster of 8–9 peaks separated from one another by 16-Da steps (Fig. 1*C*). Two other less prominent families were clustered at 5.5 and 6.4 kDa with

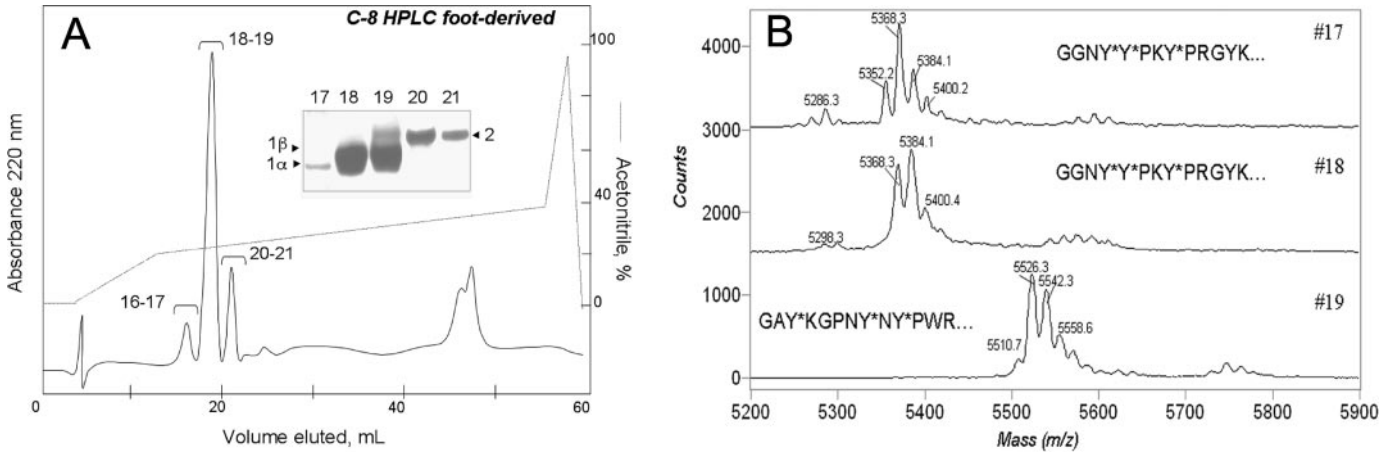

FIGURE 3. **Mcfp3s isolated from mussel feet.** *A*, acid-urea extracted proteins separated by C8 HPLC. *A*, *inset*, selected fractions subjected to acid-urea PAGE. *B*, selected fractions subjected to MALDI-TOF mass spectrometry. Edman-based N-terminal sequences for each fraction are as indicated. *Y** denotes more Dopa detected than tyrosine, whereas *Y* denotes the converse. *R* denotes more arginine than hydroxyarginine.

TABLE 1

Amino acid composition of whole plaque, foot extract, fast and slow variant plaque extracts, and cDNA-deduced variants in residues/100 residues

	Plaque	Foot		HPLC fast		HPLC slow				cDNA variants				
		16	*20*	*30*	*34*	*47*	*49*	*51*	*53*	*3-1*	*3-2*	*3-3*	*3-5*	*3-7*
Asn	10	11	12	10	10	18	18	17	18	11	11	20	21	15
Thr	3	0	0	1	1	1	1	1	1	0	0	0	1	0
Ser	7	0	0	2	1	3	2	3	3	0	0	4	0	0
Glx	4	0	0	2	1	1	1	3	3	0	0	0	0	0
Pro	8	7	4	8	6	7	8	8	8	7	7	6	7	7
Gly	17	24	24	24	25	29	29	28	29	25	25	26	30	27
Ala	1	0	4	2	2	1	1	2	2	0	2	0	0	0
Cys	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Val	4	0	0	1	1	0	0	0	0	0	0	0	0	0
Met	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ile	2	0	0	1	0	0	1	0	0	0	0	0	0	0
Leu	5	0	0	1	0	2	2	3	3	0	0	2	2	2
Dopa	4	17	20	19	19	13	8	14	8	0	0	0	0	0
Tyr	12	6	4	2	1	16	19	12	16	23	23	28	24	24
Phe	4	0	0	1	1	0	0	0	0	0	0	0	0	0
His	4	0	0	1	1	3	3	2	3	0	0	2	2	0
Lys	7	15	14	15	15	4	4	5	4	16	12	4	2	13
Trp		8	8	4	6					7	9	7	9	9
HArg[a]	1	2	2		1	0	0	0	0	0	0	0	0	0
Arg	7	9	7	8	9	2	2	2	2	11	9	2	2	3
Total	100	99	100	100	100	100	99	100	100	100	98	100	99	99

[a] HArg, hydroxyarginine.

similar 16 Da laddering, although more superimposition and overlap is evident at the higher masses (Fig. 1*C*).

Collected plaques were extracted with acetic acid plus 8 M urea, and the soluble portion was subjected to MALDI mass spectroscopy and acid-urea PAGE (note that SDS-PAGE was avoided because many plaque proteins formed insoluble complexes with SDS). The soluble proteins exhibited *m/z* values in the range of 5–7 kDa that were essentially identical with those of the footprints (Fig. 1*C*). Acid-urea PAGE revealed many additional proteins, which when stained in parallel for protein (CB) and Dopa (NBT) showed strong differential staining (Fig. 2, *first plaque panel*). In the lane stained for protein (*Plaque CB*), the most prominent bands appeared in the upper half of the gel, whereas in the lane stained for Dopa (*NBT*), there are very strong bands in the lower half as well. In general, the more mobile proteins were poorly fixed by CB regardless of how they were isolated.

The harvest of soluble proteins from plaque extracts was meager at best. From about 0.5 g (dry weight) of collected plaques, only 4 mg of soluble protein was obtained, indicating that less than 0.1% (g/g of dry weight) of the plaques was solubilized.

Protein Extraction and Purification—To better characterize the 5–7-kDa footprint proteins, purifications from both foot tissue and plaques as starting materials were attempted using methodology previously developed for mefp3 (9). In extractions of *M. californianus* feet two or three fast proteins could be detected by acid-urea PAGE and a redox cycling assay (Fig. 2, *foot*, *NBT*); the latter assay was crucial for detection because of poor fixation by CB. Two of these were purified by C8 HPLC, partially sequenced, and analyzed by MALDI-TOF (Fig. 3). The most abundant of these, 1β (Fig. 3*B*, *#18*), has an *m/z* centered at 5368; 1α had the same N-terminal sequence to R12 but was scarcer, with *m/z* at 5286. Protein 2 had a unique N terminus (Fig. 3*B*) and was centered at *m/z* 5545. Dopa approached 20 mol % in the composition of these proteins (Table 1). Proteins with masses greater than 5.5 kDa could not be detected in the foot extractions.

Plaque-extracted proteins run on reversed phase HPLC (Fig. 4*A*) resulted in elution of sharp clustered peaks followed by a broad peak with shoulders. Peak constituent fractions were subjected to acid-urea PAGE (Fig. 4*A*, *inset*) and analysis for amino acid composition and sequence. In general, the fast and slow eluting proteins on HPLC corre-

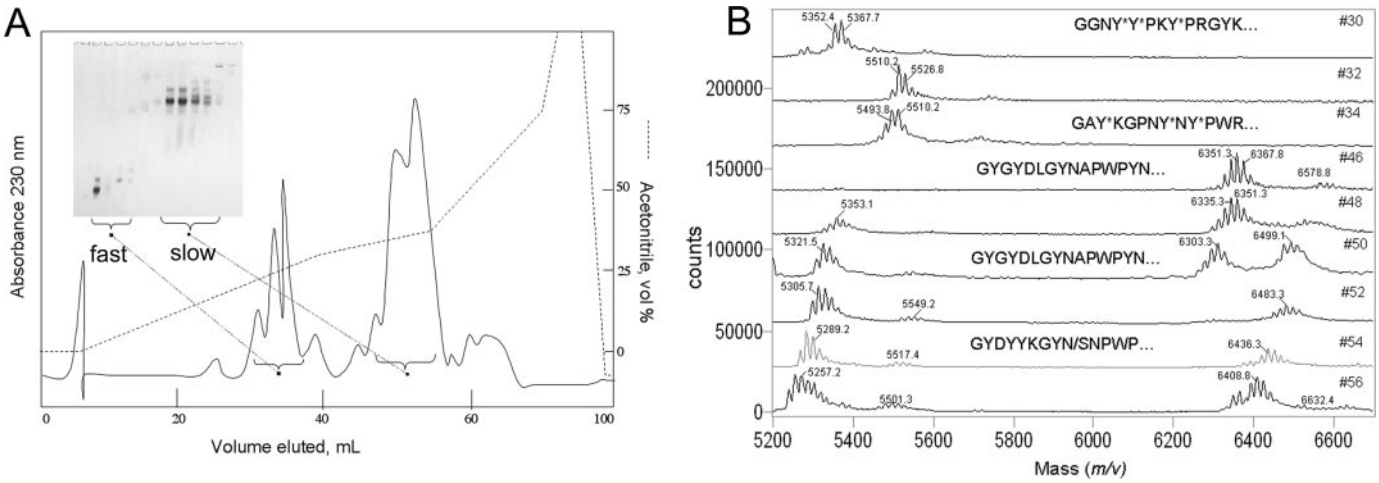

FIGURE 4. **Mcfp3s isolated from byssal plaques.** *A*, extracted proteins separated by C8 HPLC. The *inset* shows mobility on acid-urea PAGE. *B*, selected fractions subjected to MALDI-TOF mass spectrometry. Edman-based N-terminal sequences for each fraction are as indicated.

FAST	*Signal*	*MK/RSLSIVVLVALVLIGSFAVQSDA*
3-1α	GGNYYPKYKY	PRGYKGGYNG YPRGNYGWNK GWKKGRWGRK YY
3-1β	GGNYYPKYNY	PRRYKGGYNG YPRGNYGWNK GWKKGRWGRK YY
3-2	GAYKGPNYNY	PWRYGGKYNG YKGYPRGYGW NKGWNKGRWG RKYY
SLOW	*Signal*	*MNKFSVTVLLALVLIGF/LFAVQSDA*
3-3α	GYDYYKGYNS	PWPYNNGYYG YNGYNGYHGR YGWNKGWNNG PWGGSYYGNK GYLY
3-3β	GYDYYKGYNN	PWPYNNGYYG YNGYNGYHGR YGWNKGWNNG PWGGSYYGNK GYLY
3-4	GYDYYKGYNN	PWPYNNGYYG YNGYNGYHGR YGWNKGWNNG PWGGY
3-5α	GYGYDLGYNN	PWPYNNGYYG YNGYNGYHGR YGWNKGWNNG PWGGY
3-5β	GYGYDLGYNT	PWPYNNGYYG YNGYNGYHGR YGWNKGWNNG PWGGY
3-5γ	GYGYDLGYNA	PWPYNNGYYG YNGYNGYHGR YGWNKGWNNG PWGGY
3-6	GYGYDLGYNA	PWPYNNGYYG YNGYNGYHGR YGWNKGWNNG PWGGSYYGNK GYLY
3-7	GYGYDLGYDY	YKGYNNPWPY GGKYNGYKGY PRGYGWNKGW NKGWNKGRWG RKYY
3-8	GYGYDLGYDY	YKGYNNPWPY NNGYYGYNGY NGYHGRYGWN KGWNNGPWGG Y

FIGURE 5. **cDNA-deduced sequences of seven mcfp3 variants.** Mcfp3-1β and 1α differ by a single residue at position 13 (Arg and Gly). The apparent heterogeneity in the signal sequence is due to the presence of Arg in mcfp3-1 and Lys and mcfp3-2. The GenBank™ accession numbers are: AY960607 (fp3-1β), AY960608 (fp3-2), AY960609 (fp3-3α), AY960610 (fp3-3β), AY960611 (fp3-6), AY960612 (fp3-7), and AY960613 (fp3-8), DQ165553 (fp3-4), DQ165554 (fp3-5α), DQ165555 (fp3-5β), and DQ165556 (fp3-5γ).

sponded to the fast and slow electrophoretic fp3 variants, respectively (Figs. 2*A* and 4A). The proteins in the sharp initial peaks resembled the fast and foot-extracted proteins in both mass and N-terminal sequence (Fig. 4*B*) and contained significantly more Dopa than fractions under the broad peak eluting after 45 min. Amino acid analysis of fast fractions 30 and 34 indicated proteins especially rich in Gly (24 mol %), Lys (17 mol %), Arg + hydroxyarginine (12 mol %), Asp/Asn (10 mol %), Dopa + Tyr (≥24 mol %), and Pro and Trp (each 7 mol %) (Table 1).

Fractions (fractions 46–56) under the broad, shouldered peak beginning at 50 ml in the gradient elution from the reversed phase HPLC (Fig. 4*A*) resembled the foot-derived and fast plaque fractions in having compositions dominated by Gly and Asn (Table 1). A distinguishing property of these fractions, however, is that they are precipitated en masse from the rest by dialysis against MilliQ water ("Materials and Methods"; results not shown). Other biochemical distinctions include 75% less Lys and Arg and a lower average conversion of Tyr to Dopa. N-terminal sequencing of slow fractions proved to be distinct from the fast ones as well (Fig. 4*B*). MALDI mass spectra exhibit both similar and different

TABLE 2

Summary of the calculated masses for each variant based on the cDNA-deduced sequence without signal peptide (M_{min}) as well as the fully hydroxylated version with regard to hydroxyarginine and Dopa, respectively (M_{max})

The *pI* values are estimated by Expasy.

Variant	No. of amino acids	*pI*	M_{min}	M_{max}	Tyr + Arg
Mcfp3-1α	42	10.15	5105.7	5329.7	10 + 4
Mcfp3-1β	42	10.21	5190.8	5446.8	10 + 5
Mcfp3-2	44	10.10	5348.9	5572.9	10 + 4
Mcfp3-3α	54	9.10	6407.7	6663.7	15 + 1
Mcfp3-3β	54	9.1	6434.7	6690.7	15 + 1
Mcfp3-4	45	8.9	5388.6	5560.6	12 + 1
Mcfp3-5α	45	8.2	5267.5	5459.5	11 + 1
Mcfp3-5β	45	8.2	5254.5	5446.5	11 + 1
Mcfp3-5γ	45	8.2	5224.5	5416.5	11 + 1
Mcfp3-6	54	8.84	6270.6	6510.6	14 + 1
Mcfp3-7	54	9.73	6554.1	6794.1	13 + 3
Mcfp3-8	51	8.11	6057.3	6281.3	14 + 1

mass ranges (Fig. 4*B*). Fraction 52, for example, contained protein ladders similar to the 5.3- and 5.5-kDa masses of the 1β and 2 proteins in the fast fractions. Closer scrutiny, however, reveals a peak offset of about 1 Da in the first cluster (note 5352 in fraction 30 and 5353 in fraction 48) and 7 Da in the second (note 5510 in fraction 34 and 5517 in fraction 54).

At $m/z > 6000$, the slow fractions contained additional protein ladders that were unique and present in the footprints. The most abundant were centered from *m/z* 6350, 6500, and 6600 (Fig. 4*B*, *scans 46–56*). As in other hydroxylated proteins, the variants with higher Dopa content eluted before those with lower Dopa content. Thus, when inspecting consecutive fractions, the downward shift in the apparent centroid reflects lower Dopa content.

Cloning and Sequencing of Mcfp3—Degenerate primers based on the four unique detected N termini were used to amplify a set of partial cDNAs by reverse transcription-PCR. Complete sequences were obtained using nondegenerate sequences from the partial cDNAs coupled with 3′,5′-RACE. Twelve complete sequences have been determined and are listed in Fig. 5. It is interesting to compare the calculated masses of these variants with the MALDI spectra of the footprints and purified proteins. Consistent with the variability in hydroxylation of Arg and Tyr in the fp-3s in *M. edulis* and *M. galloprovincialis* (9, 16, 17), the masses for each mcfp3 variant are predicted to range from no hydroxylation to 100% hydroxylation of the target residues (Table 2).

The 12 sequences are classified into eight subgroups within mcfp3. Mcfp3-1 (α and β) and Mcfp3-2 correspond to the major components of

TABLE 3

Comparison of fp3s in three species of *Mytilus*

Sequence 1, from Papov *et al.* (9); sequence 2 from Warner and Waite (17); sequence 3 from Inoue *et al.* (22); sequences 4 and 5 from the present study.

Mytilus edulis	
1	---AD**Y**Y**G**PN**Y**G**P**PRR**Y**G**G**GN**Y**NRYNRY**G**RR---Y**G**GY**K**GWNNGWNRG---RRGK**Y**W-
2	---SD**Y**Y**G**PN**Y**G**P**SRR**Y**R**G**GN**Y**NRYNRY**G**GGK**G**GY**G**GY**K**GWNNGWRRGSWGRR-K**Y**NY
Mytilus galloprovincialis	
3	---AD**Y**Y**G**PN**Y**G**P**PRR**Y**G**G**GN**Y**GR–RYG-----YY**G**GY**K**GWNNGWK**R**GRWGP--K**Y**Y
Mytilus californianus	
4	--GGN**Y**Y-PK**Y**K**Y**PRG**Y**K**G**G-**Y**NGYPR--------GNY-GWNK**G**WK**K**GRWGR--K**Y**Y
5	GYGYDLGYNAPWPYNNGYYG-**Y**NGYNGYH------GR**Y**-GWNK**G**WNNGPWG–GS**Y**YGNKGYLY

foot-extracted and fast HPLC fractions and to the masses in Figs. 3 and 4 when hydroxylation efficiency is high. The *m/z* 5368 corresponds to mcfp3-1β replete with 11 of 15 possible hydroxylations. Similarly, *m/z* 5526 denotes mcfp3-2 with 8 of 14 Tyr and/or Arg positions hydroxylated.

Mcfp3-1α and 1β differ only by a Gly to Arg substitution at position 13. Likewise mcfp3-3α and mcfp3-3β differ only in an Ser to Asn substitution at position 10. Mcfp3-5 has three polymorphs based on substitutions at position 10: α has Asn, β has Thr, and γ has Ala. Mcfp3-1 and Mcfp3-2 have 4 or 5 Arg and 10 Tyr each, whereas mcfp-3 to mcfp-6 have a single Arg and 14–15 Tyr each. This is expected to increase the mass in steps of 16 Da to a maximum of at least 224 Da for 14 hydroxylation sites. In actuality, completely unhydroxylated and hydroxylated forms were rare. For the two most abundant fp3 variants in footprints, for example, a mass range of 5190.8 to 5446.8 and 6270.6 to 6510.6 is expected for mcfp3-1β and mcfp3-5γ, respectively. These masses are fairly consistent with *m/z* values of fp-3s detected in footprints and of the plaque-extracted and purified proteins.

Dopa in free or peptide-bound form is very prone to oxidation in seawater. Given this, we speculated that Dopa detected in footprint fp3s might remain intact for one of two reasons: 1) it is stabilized by the interface or 2) it is stabilized by factors present in the plaque itself. To test the first hypothesis, we extracted 20 plaques with acid-urea. Ten of these were carefully scraped from the glass slips and incubated in seawater for 24 h before extraction; the other ten were extracted immediately after scraping (Fig. 2). Exposing the interface to seawater prior to extraction does not decrease the abundance of the fast or slow fp3s. One protein (marked with an *asterisk*), however, is noticeably diminished by exposure (Fig. 2, *plq*), whereas another slower band is enhanced. Interrogation of control and exposed plaque footprints by MALDI-TOF revealed a similar trend: mcfp3s are not detectably changed by a 24-h exposure of footprints (Fig. 2*B*).

DISCUSSION

Our results suggest that the California mussel exploits two electrophoretically distinct Dopa-rich variants called mcfp3s for securing adhesion of the byssal plaques to hard substrata. These are the fast and slow variants. A previous study using *M. galloprovincialis* established in principle that MALDI-TOF mass spectrometry could be used to interrogate the protein composition on the underside of adhesive plaques (16); however, two technical difficulties limited interpretation of these results. First, the UV laser is known to penetrate up to 1 μm into porous organic membranes (18), of which the plaque is certainly one, and second, matrix imbibed by the plaque might cause percolation of selected proteins to the surface.[3] It was thus not unequivocally possible to say whether the detected proteins were from the interface or from micrometers deeper within the plaque interior. Present MALDI-TOF analyses of *M. californianus* plaque footprints on glass show that a variety of fp3s is present on the substratum and that they are not rendered insoluble by cross-linking.

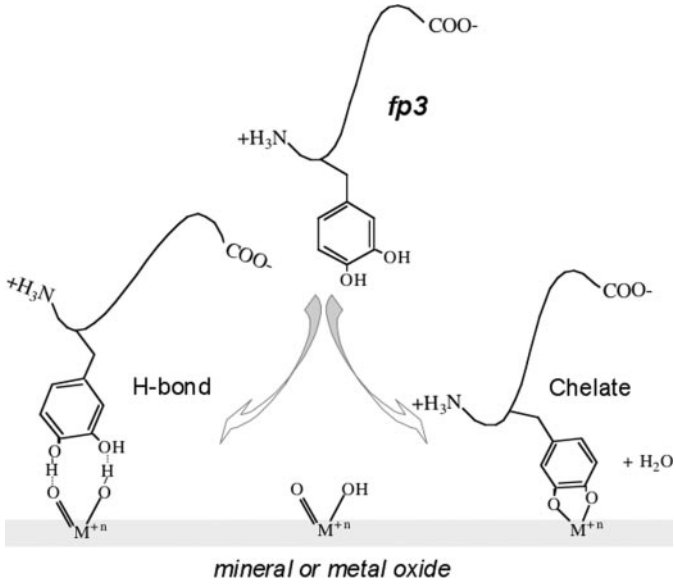

FIGURE 6. **Surface-coordinated Dopa on metals and minerals.** On the right is the chelate proposed by Dalsin *et al.* (25). For the highest affinity between surfaces and peptidyl Dopa, *M* in the oxide would likely be Si, Fe(III), Ti (IV), or Al (III), all with predicted log stabilities of >35. On the *left* is the H-bond proposed by Chirdon *et al.* (26).

If the calibrated MALDI spectra of footprints are representative ones (Fig. 1*C*), some fp3s are more abundant than others. The most abundant would appear to be variant fp3-5γ, whose calculated mass range of 5224.5–5416.5 (Table 2) is nearly fully realized in footprints as well as in extracted purified protein (Fig. 4*B*, *fractions 48–56*). The highest peaks in the cluster (*m/z* 5305 and 5321) correspond to the seventh and eighth hydroxylations. Another potential contributor to the cluster at 5305 is mcfp3-1β with a hydroxylation-dependent mass range of 5191–5431 and an *m/z* of 5304 for the seventh hydroxylation. The finding that purified fp3-1β rarely contains forms with less than 10 hydroxylations (Figs. 3*B* and 4*B*) argues against its abundance.

The smaller cluster at *m/z* 5500 is likely to contain two variants, mcfp3-2 and mcfp3-4 with ranges of 5348–5557 and 5388–5564, respectively (Table 2). At *m/z* > 6000, there is extensive variant overlap, but all belong to the electrophoretically slow group. MALDI-TOF analysis of HPLC fractions in Fig. 4*B* suggests the presence of mcfp3-6 in fractions 46–50, and *m/z* 6351 corresponds to five hydroxylations. Mcfp3-3α is also prominent in fractions 50–56, and peak 6499 represents four hydroxylations.

Mcfp3s resemble fp3s from other mytilid mussels in their small size and in having compositions dominated by Gly, Asn, and Dopa/Tyr, but specific sequence identities were spotty at best (Table 3). The only obviously persistent theme in the sequences involves GWN*X*GW*X* near the C terminus. Curiously, this corresponds to a prominent motif in a retroviral envelope glycoprotein (19). Arg and its hydroxylated derivative (4-hydroxyarginine) in mcfp3s are at less than half the levels of those detected in the fp3s from *M. edulis* (9), which suggests that 4-hydroxyarginine may not have a bearing on tenacity *per se*.

Perhaps the most distinguishing feature of mcfp3s is the high level of Dopa, which could approach 28 mol % in some variants *e.g.* mcfp3-3α/β. This level is exceeded in only one other plaque protein, mefp5 (20) whose role in adhesion is not yet clear because of its erratic desorption/ionization behavior in MALDI. The presence of dopa in plaque mcfp3s is particularly intriguing for two reasons: dopa formation from Tyr, particularly in the slow variants, is adjustable and wide ranging, and dopa persists in fp3s from

[3] F. Hillenkamp, personal communication.

plaques despite the latter having been in seawater for several days. The unoxidized condition of Dopa is astonishing, given the susceptibility of the Dopa/O_2 redox couple to oxidation ($E'_o = +0.60$, pH 7) (21) and the high pH and redox potential of seawater (www.gso.uri.edu/~dkester/eh/ehph1.htm). Because it is shielded from oxidation even in footprints following removal of the plaques, the protection of dopa would seem to come from factors built into the plaque. There are several possible protective factors: first, that the sequence around Dopa may effectively lower the E'_o; second, that a reductant may be present in the plaque mixture; or third, that Dopa groups may be protected by weak chelation of a group such as borate, for example. At this time, all seem to be possible candidates.

Why concentrate Dopa residues at the interface and shield them from oxidation? Given the bidentate structure of the catechol moiety of Dopa, one plausible reason is that it is needed for surface coordination (23, 24) (Fig. 6). In a thorough analysis of pegylated Dopa on titanium oxide, Dalsin *et al.* (25) have shown beyond reasonable doubt that Dopa is engaged by forming a coordination chelate of surface metal oxides. Alternatively, on hydroxyapatite, catechols are thought to be bound by hyperstable H-bonding (26). If these interactions also pertain to byssal adhesion in mussels, then it is of fundamental importance for understanding and engineering water-resistant adhesive bonds. Chelate bonds such as those between Dopa and TiO_2 differ from noncovalent charge-charge and van der Waals' interactions in that they are not dependent on the dielectric constant of the medium (27). On the other hand, unlike covalent bonds, chelates can reform reversibly once broken (22). If mimickable, these two features, indifference to water and reversibility, augur a new generation of bio-inspired adhesives that should be both stronger and tougher than what is presently available.

Acknowledgments—We thank Dong Soo Hwang for helping prepare a cDNA library from M. californianus foot tissue. Sandra Morrison and Chengjun Sun assisted with protein purification. The NASA URETI group is acknowledged for useful discussions.

REFERENCES

1. Comyn, J. (1981) *Developments in Adhesives* (Kinloch, A. J., ed) pp. 279–313, Applied Science Publishing, Barking, UK
2. Deming, T. J. (1999) *Curr. Opin. Chem. Biol.* **3,** 100–105
3. Dalsin and Messersmith (1990) (2005) *Materials Today* **8,** 53–57
4. Israelachvili, J. (1985) *Intermolecular and Surface Forces*, p. 22, Academic Press, London, UK
5. Chen, B. N., Piletsky, S., and Turner, A. P. F. (2002) *Combinatorial Chem. High Throughput Screening* **5,** 409–427
6. Waite, J. H. (2002) *Integr. Comp. Biol.* **42,** 1172–1180
7. Witman, J. D., and Suchanek, T. H. (1985) *Mar. Ecol. Prog. Ser.* **16,** 259–268
8. Anderson, K. E., and Waite, J. H. (2000) *J. Exp. Biol.* **203,** 3065–3076
9. Papov, V. V., Diamond, T. V., Biemann, K., and Waite, J. H. (1995) *J. Biol. Chem.* **270,** 20183–20192
10. Waite, J. H., and Benedict, C. V. (1984) *Methods Enzymol.* **107,** 397–413
11. Paz, M., Flückinger, R., Boak, A., Kagan, H. M., and Gallop, P. M. (1991) *J. Biol. Chem.* **266,** 689–692
12. Tsugita, A., Uchida, T., Mewes, H. W., and Ataka, T. (1987) *J. Biochem.* (*Tokyo*) **102,** 1593–1597
13. Simpson, R. J., Neuberger, M. R., and Liu, T.-Y. (1976) *J. Biol. Chem.* **251,** 1936–1940
14. Muramoto, K., and Kamiya, H. (1990) *Anal. Biochem.* **189,** 223–230
15. Waite, J. H. (1991) *Anal. Biochem.* **192,** 429–433
16. Floriolli R. Y., von Langen, J., and Waite, J. H. (2000) *Mar. Biotechnol.* **2,** 352–363
17. Warner, S. C., and Waite, J. H. (1999) *Mar. Biol.* **134,** 729–734
18. Strupat, K., Karas, M., Hillenkamp, F., Eckerskorn, C., and Lottspeich, F. (1994) *Anal. Chem.* **66,** 464–470
19. Garvey, K. J., Oberste, M. S., Elser, J. E., Braun, M. J., and Gonda, M. A. (1990) *Virology* **175,** 391–409
20. Waite, J. H., and Qin, X. X. (2001) *Biochemistry* **40,** 2887–2893
21. Proudfoot, G. M., and Ritchie, I. M. (1983) *Aust. J. Chem.* **36,** 885–894
22. Inoue, K., Takeuchi, Y., Miki, D., Odo, S., Haruyama, S., and Waite, J. H. (1996) *Eur. J. Biochem.* **239,** 172–176
23. Pettit, L. D. (1984) *Pure Appl. Chem.* **56,** 247–292
24. Waite, J. H. (1990) *Int. J. Biol. Macromol.* **12,** 139–144
25. Dalsin, J. L., Lin, L., Tosatti, S., Vörös, J., Textor, M., and Messersmith, P. M. (2005) *Langmuir* **21,** 640–646
26. Chirdon, W. M., O'Brien, W. J., and Robertson, R. E. (2003) *J. Biomed. Mat. Res. B-Appl. Biomat.* **66B,** 532–538
27. Schmitt, L., Ludwig, M., Gaub, H. E., and Tampe, R. (2000) *Biophys. J.* **78,** 3275–3285