trans-2,3-*cis*-3,4-Dihydroxyproline, a New Naturally Occurring Amino Acid, Is the Sixth Residue in the Tandemly Repeated Consensus Decapeptides of an Adhesive Protein from *Mytilus edulis*

Steven W. Taylor* and J. Herbert Waite

Department of Chemistry and Biochemistry and College of Marine Studies, University of Delaware Newark, Delaware 19716

Mark M. Ross,[1] Jeffrey Shabanowitz, and Donald F. Hunt

Department of Chemistry, University of Virginia Charlottesville, Virginia 22901

Received August 8, 1994

Marine mussels depend on an extraorganismic holdfast called byssus for secure attachment to solid surfaces in the turbulent environment of the rocky shore. These bundles of silky threads consist of a variety of adhesive proteins that undergo quinone tanning following secretion from holocrine glands in the foot. The first of these proteins to be characterized from the mussel *Mytilus edulis* was found to consist predominantly of tandemly repeated decapeptide sequences.[2] The protein, designated *M. edulis* foot protein 1 (Mefp1), is a polyphenolic protein (M_r 125 000) with extensive hydroxylation of tyrosine to (3,4-dihydroxyphenyl)-L-alanine (DOPA) and of proline to hydroxyproline (HYP).[2] The most frequently repeated sequence is Ala-Lys-Pro-Ser-Tyr-Hyp-Hyp-Thr-DOPA-Lys. *trans*-4-Hydroxyproline is invariably found at position 7 and occasionally at position 3. The amino acid in the sixth position was tentatively identified as *trans*-3-hydroxyproline with occasional substitution by *trans*-4-hydroxyproline.[2] Along with elastin[3] and geographotoxins,[4,5] Mefp1 is one of the few known animal proteins to contain hydroxyprolines in noncollagenous sequences. The hydroxyprolines of the decapeptide were identified by coelution with authentic standards of *trans*-3- and *trans*-4-hydroxyproline during amino acid analysis[2] and subsequent tandem mass spectrometry (MS/MS).[6] Recently, we subjected peptides isolated from tryptic digests of Mefp1 to electrospray ionization (ESI) on a triple-quadrupole mass spectrometer[7−9] and were surprised to find that all decapeptides yielded ions that indicated a molecular mass 16 Da greater than that calculated from the proposed structures. In addition, ESI MS/MS results suggested that the hydroxyproline in the sixth position was, in fact, a dihydroxyproline (diHYP). For example, the mass spectrum resulting from collisional activation decomposition (CAD) of the doubly-protonated molecule of peptide B (most hydroxylated of the Mefp1 decapeptides) is illustrated in Figure 1. Of particular importance is the mass of the y_5 ion (*m/z* 669), which indicates that the unexpected, additional 16 Da must be associated with the residue in the sixth position, which must have a mass of 129 Da. The initial suggestion that this residue could be a diHYP was supported by additional mass spectrometry experiments involving chemical derivatization (acetylation, methyl esterification) and MS/MS/MS on an ion trap mass spectrometer. Given these data, we set out to isolate the amino acid from acid hydrolysates of Mefp1 in sufficient quantities for further characterization.

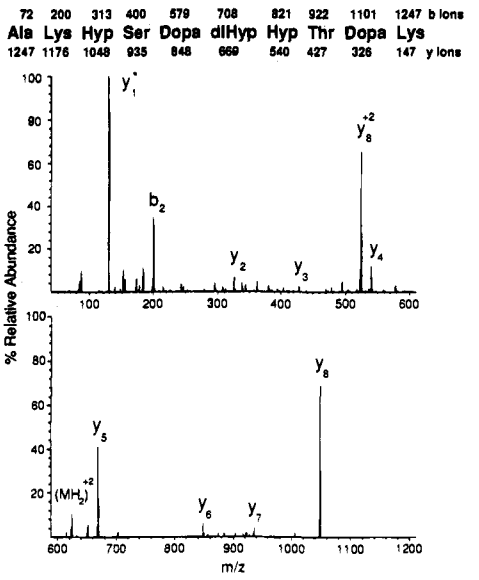


Figure 1. Sequence and mass spectrum obtained from the CAD of the doubly-protonated peptide B molecule labeled $(MH_2)^{2+}$. The labeled fragment ions correspond to b and y ions, as noted in the sequence.

Mefp1 was partially purified from mussel feet (approximately 200 g wet weight) by perchloric acid extraction and acetone precipitation.[10] Precipitated material was extracted into 5% acetic acid (20 mL) and spun and the supernatant lyophilized. The crude freeze-dried protein (0.39 g) was then resuspended in 15 mL of 6 N HCl, 1.5 mL of phenol was added, and then the protein was hydrolyzed *in vacuo* at 105 °C for 24 h. The acid was subsequently removed by repeated flash evaporation at 60 °C. Chromatography was performed on an IR-120 PLUS resin (ground and sieved, 100−200 wet mesh, Sigma Chemical Co.) using a water-jacketed column (90 × 1.6 cm) thermostated at 37 °C and equilibrated with 0.2 M formic acid−NH_4OH buffer pH 3.1 at a flow rate of 54 mL/h.[11] A 50 μL aliquot of each 4 mL fraction collected was analyzed using the acid−ninhydrin method B of Piez et al.[12] (scaled down accordingly). Those fractions containing higher absorbances at 440 nm were further analyzed on a Beckman 6300 amino acid analyzer. Fractions that gave a peak with a retention time identical to that of *trans*-3-hydroxyproline (9.7 min, $A_{440} > A_{570}$) were pooled and lyophilized (V_e = 539−544 mL). Final purification was achieved using the recently developed technique of hydrophilic interaction chromatography (HILIC).[13] This was necessary due to the high polarity of the amino acid, which made it inseparable from ammonium formate and other polar impurities by conventional reversed-phase HPLC. A PolyHYDROXYETHYL Aspartamide column (200 × 9.4 mm, 300 Å, PolyLC Inc., Columbia, MD) was employed using a linear 80−60% acetonitrile gradient over 60 min, flow rate = 2 mL/min, with

(1) On sabbatical leave from Chemical Division, Naval Research Laboratory, Washington, DC 20375.

(2) Waite, J. H.; Housley, T. J.; Tanzer, M. L. *Biochemistry* **1985**, *24*, 5010−5014.

(3) Gray, W. R.; Sandberg, L. B.; Foster, J. A. *Nature* **1973**, *246*, 461−466.

(4) Sato, S.; Nakamura, H.; Ohizumi, Y.; Kobayashi, J.; Hirata, Y. *FEBS Lett.* **1983**, *155*, 277−280.

(5) Olivera, B. M.; McIntosh, J. M.; Cruz, L. J.; Luque, F. A.; Gray, W. R. *Biochemistry* **1984**, *23*, 5087−5090.

(6) Kassel, D. B.; Biemann, K. *Anal. Chem.* **1990**, *62*, 1691−1695.

(7) The electrospray ionization mass spectra were obtained with a Finnigan MAT TSQ-7000 triple-quadrupole mass spectrometer. The peptide sample were infused as 10 pmol/μL solutions in (0.1% acetic acid) at a flow rate of 0.6 μL/min. A liquid sheath consisting of a 70/30 methanol/0.1 M acetic acid mixture was employed. CAD mass spectra were obtained using methods that have been described.[8,9]

(8) Hunt, D. F.; Yates, J. R., III; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 6233−6237.

(9) Hunt, D. F.; Alexander, J. E.; McCormack, A. L.; Martino, P. A.; Michel, H.; Shabanowitz, J.; Sherman, N.; Moseley, M. A.; Jorgenson, J. W.; Tomer, K. B. *Techniques in Protein Chemistry II*; J. J. Villafranca: New York, 1991; pp 441−454.

(10) Rzepecki, L. M.; Chin, S.-S.; Waite, J. H.; Lavin, M. F. *Mol. Mar. Biol. Biotechnol.* **1991**, *1*, 78−88.

(11) Ogle, J. D.; Arlinghaus, R. B.; Logan, M. A. *J. Biol. Chem.* **1962**, *237*, 3667−3673.

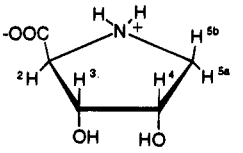
(12) Piez, K. A.; Irreverre, F.; Wolff, H. L. *J. Biol. Chem.* **1956**, *223*, 687−697.

(13) Alpert, A. J. *J. Chromatogr.* **1990**, *499*, 177−196.

0002-7863/94/1516-10803$04.50/0

50 mM formic acid−NH_4OH buffer pH 2.8 as the mobile phase. The lyophilized IR-120 PLUS fractions were dissolved in 70/30 mobile phase/acetonitrile and loaded on the column. The target amino acid eluted from 20 to 22.5 min (detected by the acid−ninhydrin method B[12] after the elution volume of *trans*-4-hydroxyproline had been determined at higher concentration by its absorbance at 220 nm) and was lyophilized, yielding 0.4 mg of pure compound (single peak on analyzer).

A MH^+ peak at *m/z* 148.0603 was obtained from fast-atom bombardment mass spectrometry (FABMS), consistent with the calculated MH^+ of *m/z* 148.0610 (Δmmu = 0.7) for $C_5H_{10}NO_4$. A proton NMR spectrum[14] of the compound was consistent with *trans*-2,3-*cis*-3,4-dihydroxyproline (**1**), the D-isomer of which has been synthesized and characterized by Dho and co-workers.[15,16] The *cis* configuration of the hydroxyls was



1

confirmed by periodate oxidation of the purified amino acid, which resulted in the loss of its chromatographic peak during amino acid analysis. The same experiment performed on the hydrolysate of purified Mefp1 had a similar result. The absolute configuration of the compound remains unknown due to the limited quantities we could recover and the fact that the reported $[\alpha]^{20}_D$ is very small (for the D-isomer it is −6.8°).[15] This compound has been reported to have an elution time identical to that of *trans*-3-hydroxyproline from an amino acid analyzer.[17]

While this is the first report of *trans*-2,3-*cis*-3,4-dihydroxyproline as a naturally occurring amino acid, two other isomers have been found in biological systems. *cis*-2,3-*trans*-3,4-Dihydroxy-L-proline has been isolated from acid hydrolysates of the proteinaceous material which encases the siliceous shell of the cell wall of the diatom *Navicula pelicullosa*,[18] and *trans*-2,3-*trans*-3,4-dihydroxy-L-proline has been isolated from acid hydrolysates of a mixture of toxic peptides from the mushroom *Amanita virosa*.[19]

Significantly, this the first reported case in which the position of a dihydroxyproline in the primary sequence of a protein has been unequivocally determined. The fact that no dihydroxyproline is found in the third or seventh positions suggests that the proline dihydroxylating enzyme system is specifically targeted to the sixth position. The detection of traces of *trans*-4-hydroxyproline in the sixth position of some of the decapeptides indicates that either the dihydroxyproline is formed from previously hydroxylated proline or that the 4-prolylhydroxylase is in competition with a dihydroxylating enzyme system.

The consequences of this discovery are that any models for the secondary and tertiary structure of Mefp1, which has 75−80 of these sequences, most containing *trans*-2,3-*cis*-3,4-dihydroxyproline in the sixth position,[2] have to be completely reevaluated.[20,21] Such reevaluation should also have an impact on bioengineering strategies to produce biomimetic analogues as biocompatible cell and tissue adhesives.[22,23]

Acknowledgment. We would like to thank Dr. Andrew Alpert for use of his PolyHYDROXYETHYL Aspartamide HPLC column and valuable discussions. NMR measurements by Martha Bruch and FABMS measurements by Gordon Nicol are gratefully acknowledged as is the technical assistance of Nathan Yates and Nicholas Sherman. J.H.W. acknowledges financial support from the Office of Naval Research and the National Institutes of Health.

(14) **1**: ^1H NMR (D_2O, 250 MHz, DSS internal reference) δ (ppm) 3.29 (dd, $J_{5a,5b}$ = 12.4 Hz, $J_{5a,4}$ = 4.3 Hz, 1 H, H-5_a), 3.55 (dd, $J_{5b,4}$ = 4.9 Hz, 1 H, H-5_b), 3.96 (d, $J_{2,3}$ = 4.7 Hz, 1 H, H-2), 4.36 (m, 2 H, H-3,4).

(15) Dho, J. C.; Fleet, G. W. J.; Peach, J. M.; Prout, K.; Smith, P. W. *Tetrahedron Lett.* **1986**, *27*, 3203−3204.

(16) While the absolute configuration of the naturally occurring amino acid remains to be experimentally determined, we believe that we are justified in drawing an L-configuration in structure **1**. The parent amino acid, peptidyl proline, is most certainly in the L-configuration (all eucaryotic prolyl residues are), and there is no reason to suppose that the addition of two hydroxyls to carbons 3 and 4 should alter this.

(17) Hudson, C. B.; Robertson, A. V.; Simpson, W. R. J. *Aust. J. Chem.* **1968**, *21*, 769−782.

(18) Nakajima, T.; Volcani, B. E. *Science* **1969**, *164*, 1400−1401.

(19) Baku, A.; Faulstich, H.; Wieland, T.; Dabrowski, J. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 2370−2371.

(20) Nagy, P. I.; Patel, H. C.; Dreyer, W.; Hopfinger, A. J. *Int. J. Pept. Protein Res.* **1991**, *38*, 84−89.

(21) Olivieri, M. P.; Baier, R. E.; Loomis, R. E. *Biomaterials* **1992**, *13*, 1000−1008.

(22) Strausberg, R. L.; Link, R. P. *Trends Biotechnol.* **1990**, *8*, 53−57.

(23) Notter, M. F. D. *Exp. Cell Res.* **1988**, *177*, 237−246.