Application of Mass Spectrometric Sequencing Techniques to Naturally Occurring and Synthetic Histidine-Containing Peptides[1]

M. L. Polan, W. J. McMurray, S. R. Lipsky, and S. Lande*

Contribution from the Department of Molecular Biophysics and Biochemistry, the Section of Physical Sciences, and the Section of Dermatology, Yale University, New Haven, Connecticut 06520. *Received March 13, 1971*

Abstract: Products suitable for mass spectrometric studies were obtained by controlled permethylation of histidine-containing peptides. The technique allowed successful mass spectrometric sequence determination of a histidine-containing tripeptide obtained from naturally occurring α-melanocyte stimulating hormone by cyanogen bromide and chymotryptic cleavage. The sequence of a synthetic histidine-containing octapeptide was also determined by this method. Knowledge of the amino acid composition of peptides formed by tryptic digestion of α-melanocyte stimulating hormone, coupled with mass spectrometry of certain suitably derivatized cyanogen bromide and chymotryptic cleavage products of the hormone, permitted the determination of the entire sequence of the parent tridecapeptide without recourse to classical, stepwise techniques.

Although the Edman degradation[2] remains the method of choice for sequence determination of protein, this technique is not infallible because of the possibility of nonspecific backbone chain cleavages or rearrangements. In fact, the necessity for revision of reported sequences of proteins because of unexpected side reactions seems almost the rule rather than the exception; see ref 3 for a recent example. Therefore, substantiation of sequence data by independent means seems almost necessary. In addition, there are many naturally occurring N^{α}-blocked[4] or cyclic peptides,[5] as well as those incorporating nonpeptide bonds in the backbone chain;[6] such substances are not amenable to structure determination by the Edman technique. For these reasons, interest in alternative methods of sequence determination remains high and the application of mass spectrometric techniques in this regard has received considerable attention in the past few years.[7] Although relatively expensive instrumentation is required, sequence determination of suitably derivatized peptides by mass spectrometry can be a rapid and simple process, due to its nonrepetitive nature,[7] in comparison with the classical techniques of stepwise degradation.

A most important criterion for general utility of mass spectrometry for sequence determination of peptides is applicability to minute quantities of material. The problem is not one of sensitivity: 10–20 μg of properly derivatized material is sufficient for a clear spectrum.[8] However, application of standard enzymatic or chemical cleavage methods to small quantities of large precursors in order to obtain derivatives of a size suitable for mass spectral sequencing often introduces impurities in quantities as great or greater than that of the desired products. Impurities, arising from the production or purification of smaller peptides and from side products formed during permethylation of peptide fragments, may complicate and obscure desired peptide bond sequence cleavages in the mass spectrum. Although the number of mammalian peptides partially sequenced by mass spectrometry is growing, only in the study of gastrin was a cleavage employed before sequencing to obtain peptides of more suitable size.[9] Even in this case, unequivocal identification of the amino acid sequence of one of the products required comparison with the spectrum of a known synthetic subunit.

A second important criterion for general utility is that mass spectrometry be applicable to all amino acids commonly found in proteins. Although permethylation has been extremely useful both in enhancing peptide volatility and in reducing the complexity of spectra,[10] the technique introduces certain problems. For example, previous permethylation techniques transformed cysteine into a methylsulfonium iodide, which resulted in diminished volatility and complex spectra caused by extensive random backbone chain cleavage. This problem was solved by utilizing a controlled permethylation technique, recently described.[11] Although sequence determining spectra have been obtained from histidine-containing peptides,[12–15] mass spectra of permethylated histidine-containing peptides have not been reported. The formation of *N*-methylimidazolium iodides during permethylation[7] probably results in non-sequence-type fragmentation in the mass spectrometer. It was hoped that controlled permethylation would prevent such onium ion formation and allow

(1) This study was supported by USPHS Grants 2T01 GM 00053-12, CA 04679, FR 00356, NASA Grant No. NGL 07-004-008, and American Cancer Society Grant No. P 168.

(2) W. A. Schroeder, *Methods Enzymol.*, **11**, 445 (1967).
(3) H. D. Niall, *Nature (London), New Biol.*, **230**, 90 (1971).
(4) R. Burgus, T. F. Dunn, D. Desiderio, and R. Guillemin, *C. R. Acad. Sci.*, **269**, 1870 (1969).
(5) R. Consden, A. H. Gordon, A. J. P. Martin, and R. L. M. Synge, *Biochem. J.*, **41**, 596 (1947).
(6) T. Willand and W. Schon, *Justus Liebigs Ann. Chem.*, **593**, 157 (1955).
(7) E. Lederer, *Pure Appl. Chem.*, **17**, 489 (1968).
(8) J. Lenard and P. M. Gallop, *Anal. Biochem.*, **34**, 286 (1970).
(9) K. L. Agarwal, G. W. Kenner, and R. C. Sheppard, *J. Amer. Chem. Soc.*, **91**, 3096 (1969).
(10) D. W. Thomas, B. C. Das, S. D. Géro, and E. Lederer, *Biochem. Biophys. Res. Commun.*, **32**, 199 (1968).
(11) M. L. Polan, W. J. McMurray, S. R. Lipsky, and S. Lande, *ibid.*, **38**, 1127 (1970).
(12) Yu. A. Ovchinnikov, A. A. Kiryushkin, E. I. Vinogradova, B. V. Rozynov, and M. M. Shemyakin, *Biokhimiya*, **32**, 427 (1967).
(13) M. Senn, R. Venkataraghavan, and F. W. McLafferty, *J. Amer. Chem. Soc.*, **88**, 5593 (1966).
(14) J. van Heijenoort, E. Bricas, B. C. Das, E. Lederer, and W. A. Wolstenholme, *Tetrahedron*, **23**, 3403 (1967).
(15) R. Burgus, T. F. Dunn, D. Desiderio, and R. Guillemin, *C. R. Acad. Sci.*, **269**, 1870 (1969).

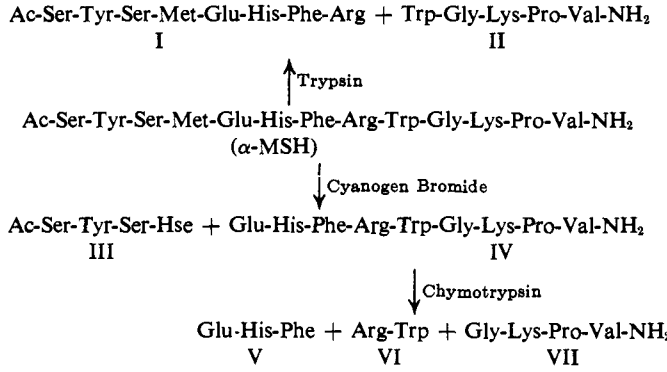


Figure 1. Structure of α-MSH showing peptides obtained by treatment with trypsin, cyanogen bromide, and chymotrypsin.

permethylated histidine-containing peptides to be sequenced by mass spectrometry.

In view of the difficulty in producing purified subunits from minute quantities of parent peptide and in sequencing permethylated peptides containing histidine, an attempt was made to determine the total sequence of α-melanocyte stimulating hormone[16] without recourse to phenylisothiocyanate procedures. The sequence of α-MSH, an acetyltridecapeptide amide, is known.[17] It is a complex molecule, containing 12 different amino acids including histidine, which seemed to offer a uniquely useful model upon which to test the various manipulations required to obtain useful mass spectra. In separate experiments, α-MSH was treated with trypsin and cyanogen bromide; the latter produced an N^{α}-amino blocked tetrapeptide and a nonapeptide which were separated and purified by electrophoresis. The nonapeptide was exposed to chymotrypsin and the products were resolved electrophoretically. Acetylated chymotryptic peptides and the N^{α}-amino blocked tetrapeptide resulting from cyanogen bromide treatment were permethylated under controlled conditions and the final products introduced directly into the mass spectrometer. Techniques described above, coupled with quantitative amino acid analyses of the chemically and enzymatically generated subunits, allowed the complete sequence of the parent polypeptide hormone to be determined.

A synthetic renin substrate, Pro-Phe-His-Leu-Leu-Val-Tyr-Ser,[18] was studied as an independent test of the applicability of controlled permethylation for sequence determination of histidine-containing peptides by mass spectrometry. By acetylation, controlled permethylation, and mass spectrometry the sequence of this histidine-containing octapeptide was deduced.

Experimental Section

The isolation of α-MSH was as previously described.[19] Cyanogen bromide treatment was carried out according to a standard procedure.[20] Trypsin or chymotrypsin (Worthington, recrystallized three times) was incubated at a 1:50 weight ratio with peptide in NH_4HCO_3, 0.06 *M*, pH 8.1, at 37°. Tryptic digestion was terminated after 4 hr. After 8 hr of chymotryptic digestion, a second aliquot of enzyme, containing one-half the original amount, was added and the reaction allowed to continue overnight. Enzymatic digestions were terminated by evaporation of the reaction mixtures to dryness.

Peptides were separated electrophoretically on Whatman 3 mm filter paper in a buffer of pyridine–acetic acid–water, 200:8:1800, pH 6.5, at 4000 V. Areas containing peptide were cut out, eluted with 5% acetic acid, and lyophilized to dryness. Amino acid analyses were performed according to the method of Moore and Stein[21] on a Beckman Model 120 B amino acid analyzer. This gave quantitative amino acid compositions and also indicated the amount of peptide present for further chemical steps.

Acetylation and controlled (equimolar) permethylation were carried out as described previously.[11] An oil–sodium hydride preparation (50 mg) was rinsed three times with dry ether, suspended in dry dimethyl sulfoxide (2 ml), and heated at 65° in an atmosphere of argon until cessation of evolution of gaseous hydrogen. All remaining steps up to the extraction procedures were also carried out in an atmosphere of argon. The resultant solution, 0.5 *M* in sodium methylsulfinylmethide, was treated at room temperature with peptide employing quantities such that a fivefold excess of methide ion was present relative to the number of peptide "reactive equivalents," *i.e.*, the molar quantity of sites in the peptide reactive to permethylation. After 2 min of stirring, the solution was cooled and a molar quantity of methyl iodide was added equivalent to the initial amount of methide ion present. The mixture was left at room temperature for 1 hr, after which water (2 ml) was added and the product was extracted with chloroform (5 ml). The lower phase was extracted four times with water (1 ml) and evaporated to dryness *in vacuo*. The mass spectra were obtained with an Associated Electronics Industries MS-9. Without further purification, the samples were introduced *via* the direct introduction probe. The spectra were recorded at 70 eV, trap current 100–500 μA, and at a resolution of 1000.[22]

Careful preparation and storage of dimethyl sulfoxide is necessary to obtain reproducible spectra. Dimethyl sulfoxide was dried over calcium hydride for several days, redistilled at a pressure of 4 mm (bp 58–60°), and stored in brown glass bottles in desiccators containing calcium sulfate. It is vital to exclude water to prevent decomposition of sodium hydride into sodium hydroxide; otherwise sodium methylsulfinylmethide ion will not be formed in sufficient quantity to allow permethylation to proceed to completion.

Results and Discussion

Tryptic digestion of α-MSH, 0.2 mg, resulted in two components which could be resolved by electrophoresis; one was ninhydrin and Ehrlich positive, the other ninhydrin negative and Pauly positive. Amino acid analyses and color tests established the peptide order shown in Figure 1 as I and II; the specificity of trypsin placed arginine at the C terminus of I.

Cyanogen bromide cleavage of 5 mg of α-MSH followed by electrophoresis produced the pattern shown in Figure 2a. Peptide III, ninhydrin negative, Pauly positive, was identified as the amino-terminal moiety of the parent compound; the action of cyanogen bromide

(16) α-Melanocyte stimulating hormone is abbreviated α-MSH.
(17) A. B. Lerner and T. H. Lee, *J. Amer. Chem. Soc.*, 77, 1066 (1955).
(18) We are grateful to R. H. Mazur of G. D. Searle and Co. for the synthetic peptide used in these studies.
(19) G. V. Upton, A. B. Lerner, and S. Lande, *J. Biol. Chem.*, **241**, 5585 (1966).
(20) E. Gross, *Methods Enzymol.*, **11**, 238 (1967).
(21) S. Moore and W. H. Stein, *ibid.*, **6**, 820 (1963).
(22) The technical assistance of Mr. Peter Arneson in obtaining mass spectra is gratefully acknowledged.

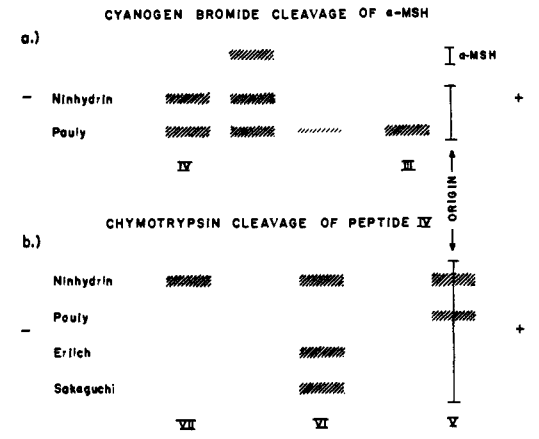


Figure 2. (a) Electrophoretic pattern showing α-MSH standard and peptides obtained after treatment of the hormone with cyanogen bromide. (b) Electrophoretic pattern obtained after digestion of peptide IV by chymotrypsin.

places homoserine at its carboxyl terminus. Quantitative amino acid analysis of III gave the composition Ser_2, Tyr, Hse;[23] other amino acids were present to an extent of 10% or less. The other major component, peptide IV, Figure 2a, was eluted and subjected to the action of chymotrypsin. A third, minor component of the cyanogen bromide cleavage contained all the amino acids of the starting material and probably consisted of the parent peptide with methionine in the sulfoxide or sulfone form, which accounts for its resistance to the action of cyanogen bromide. This material was not investigated further.

Electrophoresis of products of chymotryptic action gave the pattern shown in Figure 2b; response to ninhydrin, Pauly, Ehrlich, and Sakaguchi color reagents is indicated in the figure. The three products, peptides V, VI, and VII, Figure 2b, were eluted from the paper; one-tenth of each product was analyzed for amino acid composition. Quantities of amino acids obtained from this procedure ranged from 5 to 10 nmol, amounts too small for quantitative analyses although qualitative results were unequivocal. The major amino acids were: V, Glu, His, Phe; VI, Arg; and VII, Gly, Lys, Pro, Val. Peptide VII, which contained no aromatic amino acids, was identified as the carboxyl-terminal portion of IV on the basis of chymotrypsin specificity. Peptide VI, which was Ehrlich positive but contained only arginine in the acid hydrolysate, must have had the structure Arg-Trp; again tryptophan can be placed at the dipeptide carboxyl terminus because of the specificity of chymotrypsin. Knowledge of the amino acid composition of tryptic peptides I and II placed peptide VI adjacent to VII (Figure 1) and peptide V (containing phenylalanine at its carboxyl end) adjacent to III. The order of the cyanogen bromide and chymotrypsin-derived fragments must, therefore, be III, V, VI, VII and the sequences of peptides III, V, and VII determine the total sequence of the parent hormone molecule.

Peptides V and VII were acetylated and permethylated. Peptide III was permethylated directly. After extraction, neutral, dry products were introduced directly into the mass spectrometer, without further purification.

(23) Homoserine is abbreviated Hse.

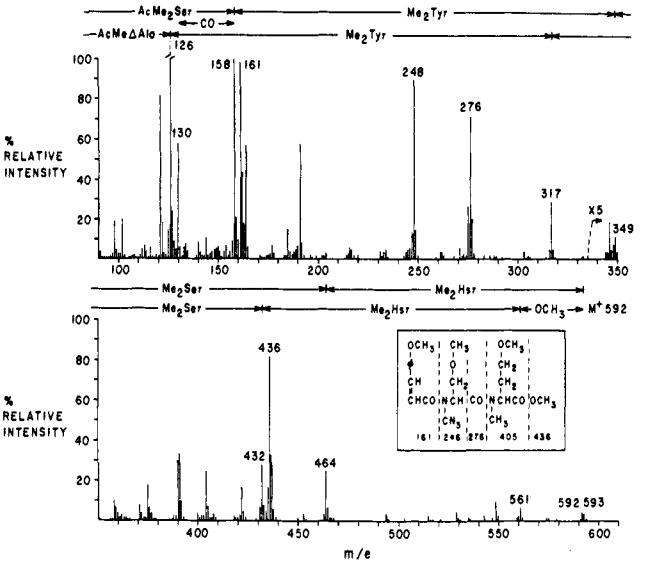


Figure 3. Mass spectrum at 210° of peptide III, Ac-Ser-Tyr-Ser-Hsr, after permethylation.

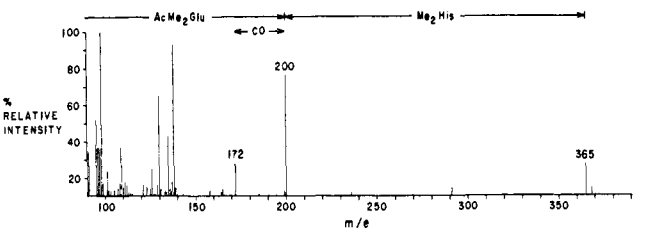


Figure 4. Mass spectrum at 210° of peptide V, Glu-His-Phe, after acetylation and permethylation.

The spectrum of peptide III, shown in Figure 3, yielded the sequence Ac-Ser-Tyr-Ser-Hse. The permethylated peptide cleaved to give sequence determining peaks at *m/e* 158, 349, 464, and 593. These fragments also served to identify the amino-terminal blocking group of the parent hormone as an acetyl moiety. The molecular ion, *m/e* 624, was not observed. Elimination of methanol at the N-terminal serine residue, either by electron impact or thermal decomposition in the mass spectrometer, formed a dehydroalanine peptide.[24] Sequence peaks for the dehydroalanine peptide appear in the spectrum at *m/e* 126, 317, 432, and 561 with a molecular ion at *m/e* 592. A typical cinnamoyl cleavage of the tyrosine residue[12,25] formed a derivative with the structure shown in the insert in Figure 3. This cinnamoyl product also underwent normal peptide bond sequence cleavage to produce peaks at *m/e* 161, 248, 276, and 436.

The structure of peptide V can be determined as Glu-His-Phe from the spectrum in Figure 4. Sequence peaks for the first two amino acids appeared at *m/e* 200 and 365. Fragments containing the C-terminal phenylalanine were not observed. Similar results have been obtained with peptides containing C-terminal tyrosine.[9] However, assignment of the first two residues placed phenylalanine at the C-terminal position of the peptide in accord with the specificity of chymotrypsin.

(24) Dehydroalanine is abbreviated ΔAla.

(25) A. A. Kiryushkin, V. A. Gorlenko, Ts. E. Agadzhanyan, B. V. Rosinov, Yu. A. Ovchinnikov, and M. M. Shemyakin, *Experientia*, **24**, 883 (1968).

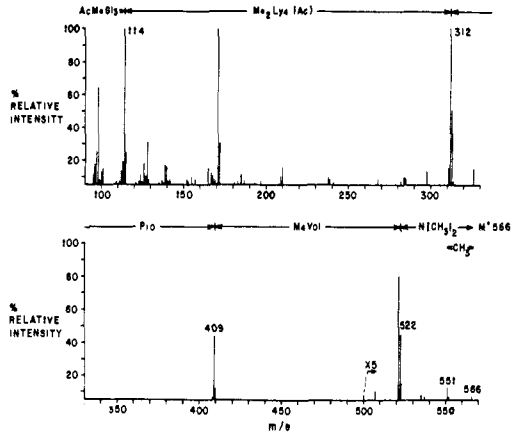
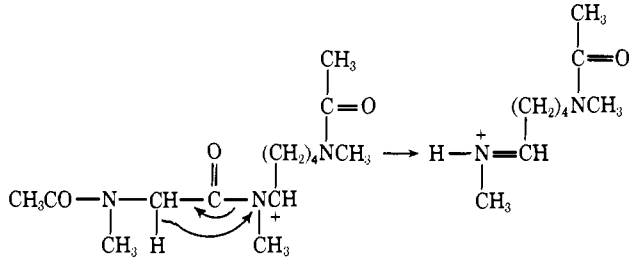


Figure 5. Mass spectrum at 210° of peptide VII, Gly-Lys-Pro-Val-NH_2, after acetylation and permethylation.

The amino acid sequence of peptide VII was inferred to be Gly-Lys-Pro-Val-NH_2 from the spectrum in Figure 5. Sequence cleavage peaks occurred at *m/e* 114, 312, 409, and 522 with a molecular ion at *m/e* 566. The large peak at *m/e* 171 was measured at high resolution and found to be *m/e* 171.1497 which is compatible with the elemental structure $C_9H_{19}N_2O$ (calcd, 171.1497). The elemental composition of the *m/e* 171 fragment indicates it is derived from lysine and the following pathway is suggested.



From the structures of peptides III, V, and VII and the order of peptides obtained from chemical and enzymatic cleavages, the complete sequence of the parent molecule could be determined (Figure 1).

Since spectra of permethylated histidine-containing peptides have not been reported previously, it was felt that further study of such peptides would be valuable.

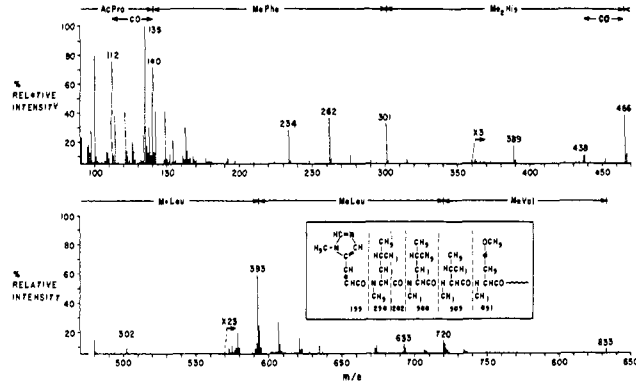


Figure 6. Mass spectrum at 310° of Pro-Phe-His-Leu-Leu-Val-Tyr-Ser, after acetylation and permethylation.

Accordingly, the synthetic renin substrate Pro-Phe-His-Leu-Leu-Val-Tyr-Ser[18] was acetylated and permethylated under controlled "equimolar" conditions and the mass spectrum of the product obtained, Figure 6. Sequence peaks were evident for the first six residues at *m/e* 140, 301, 466, 593, 720, and 833. In addition, histidine appears to have undergone a cinnamoyl-like cleavage to produce the derivative shown in the insert in Figure 6. Rather than complicating the spectrum, such a derivative provides additional sequence information since the product, containing a transformed histidyl amino terminus, fragments in the usual manner at peptide bonds to form the mass ion peaks at *m/e* 135, 234, 262, 389, 502, and 693 (Figure 6). Such cleavage allows sequence determination of the first seven amino acids of the parent molecule. The eighth residue, serine, can then be positioned by difference at the carboxyl terminus of the peptide, yielding the entire octapeptide structure.

By unfortunate coincidence, the peak observed at *m/e* 693 also could have been formed by cleavage of an acetylprolyl radical from the *m/e* 833 fragment. Therefore, peak 693 cannot be identified unequivocally; attempts at identification by peak matching failed because of the small quantities of material available coupled with the low sensitivity inherent in the technique. However, the relatively high yield of peaks at *m/e* 135, 234, 262, 389, and 502 argues in favor of our initial interpretation.