Review Letter

THE RATIONAL USE OF MASS SPECTROMETRY FOR AMINO ACID SEQUENCE DETERMINATION IN PEPTIDES AND EXTENSION OF THE POSSIBILITIES OF THE METHOD

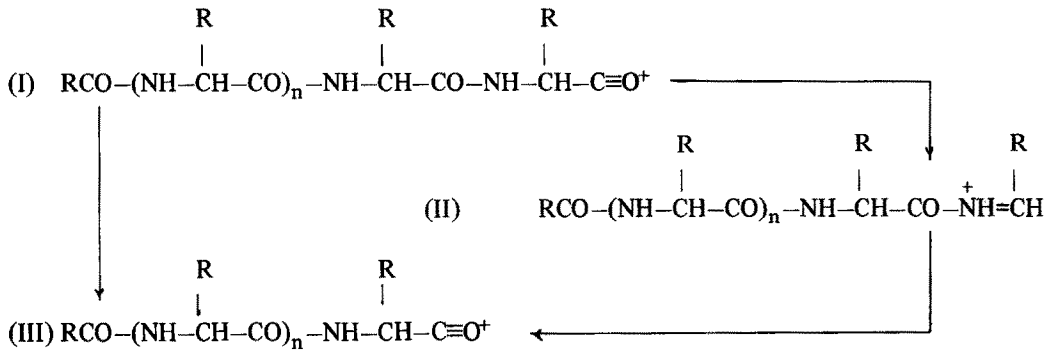
M.M.SHEMYAKIN, Yu.A.OVCHINNIKOV, E.I.VINOGRADOVA, A.A.KIRYUSHKIN, M.Yu.FEIGINA, N.A.ALDANOVA, Yu.B.ALAKHOV, V.M.LIPKIN, A.I.MIROSHNIKOV, B.V.ROSINOV and S.A.KAZARYAN

Institute for Chemistry of Natural Products, USSR Academy of Sciences, Moscow, USSR

Received 23 January 1970

It has been shown earlier by us and by others that mass spectrometry can be used for the elucidation of the primary structure of peptides containing residues of all the amino acids commonly found in proteins, *N*-acylpeptide esters being most suitable for this purpose (see [1, 2] and references therein). Mass spectrometric determination of amino acid sequences is based on fragmentation of *N*-acylpeptide esters involving rupture of amide (ester) bonds and localization of the positive charge at the *C*-terminus of the amino acid fragments of type (I). These fragments decompose further, consecutively eliminating amino acid residues either in one (I → III) or in two steps (I→II→III), in the latter case via aldimine fragments (II). This pattern of fragmentation provides amino acid and/or aldimine fragment ions arising from the partial cleavage of every amide bond, the difference in *m/e* of the fragments allowing the determination of the amino acid sequence in the peptide being investigated. The peaks due to amino acid fragments are usually more intense than those of aldimine fragments, though the reverse may be observed with di- and tripeptide derivatives especially trifluoroacetylpeptides. Amino acid type of fragmentation of *N*-acylpeptide esters is usually accompanied by dehydration of amide groups of the peptide chain; the dehydration is greater with peptides of low volatility and is enhanced by the presence of aromatic and heterocyclic amino acid residues and glycine.

A number of specific processes, due to the nature of the side chain of the constitutive amino acid residues, take place along with the general fragmentation of the peptide chain described above. Of these the most common is the loss of the side chain or of a part of it. Loss of the entire side chain is typical of methionine, tryptophan, and tyrosine and is also often observed with valine, leucine, isoleucine, phenylalanine, histidine, serine and threonine residues. In side



 North-Holland Publishing Company – Amsterdam

chains, partial cleavage can occur at different C-C bonds, but more typical is the elimination of functional groupings: e.g., hydroxyamino acids (Ser, Thr, Hyp, Homoser) losing water, asparagine and glutamine eliminating NH_3, while β-methyl aspartate and γ-methyl glutamate give up CH_3OH and carboxymethylcysteine eliminates $HSCH_2COOMe$ and CH_2COOMe. The presence of aromatic and heterocyclic amino acid and carboxymethylcysteine residues leads to specific cleavage of the peptide chain (N-C_α bond of the residue ruptures with subsequent amino acid type fragmentation of the ion formed). Histidine residues and to some degree N^δ-pyrimidylornithine and tryptophan residues under mass spectrometric conditions undergo intermolecular thermal *N*-methylation. Amino acid type as well as specific fragmentations are often accompanied by hydrogen atom migrations.

Amino acid type of fragmentation together with specific fragmentations due to the nature of the amino acids provide sufficient material for the mass spectrometric determination of amino acid sequence in peptides [1]. However, the method cannot always be used directly and if its application is to be successful the experiment must be carefully planned to avoid possible pitfalls. In many cases, its use is most feasible in combination with other analytical techniques of peptide and protein chemistry. In the following paragraphs a number of procedures elaborated to meet the requirements of different types of peptides are discussed.

Prior to mass spectrometry, the amino acid composition of the peptide must be quantitatively determined, since this 1) gives information about the volatility of the peptide derivative and consequently the usefulness of the mass spectrometric approach; 2) determines the type of peptide derivative that would be suitable for mass spectrometry and the conditions for its preparation; and 3) is indicative of the general and specific fragmentation patterns that have to be taken into account in the interpretation of the spectrum.

The next step is the conversion of free peptides into *N*-acylpeptide methyl esters [1]. Acylation of peptides is best accomplished with *N*-hydroxysuccinimide esters of carboxylic acids of medium molecular weight. The reaction proceeds smoothly in aqueous dioxan (1:1). In a typical experiment, 10 to 20% excess of *N*-hydroxysuccinimide ester in dioxan is added to a solution of peptide in water (pH 8; $NaHCO_3$) and after 5–10 hr (20°) the mixture is acidified, extracted with AcOEt, the organic layer evaporated and the residue washed with hot hexane and dried. Under these conditions α- and ω-amino groups of all amino acid residues are acylated, while the guanidino group of arginine is not changed. Esterification of the *N*-acylpeptides obtained is effected with MeOH in the presence of catalytic amounts of SO_2Cl_2 (20°, 48 hr), the residue resulting on evaporation is washed by decantation with 5% $NaHCO_3$ followed by water and it is then dried. Arginine residues in arginine-containing peptides should after acylation be converted into N^δ-pyrimidylornithine residues with simultaneous esterification of the *N*-acylpeptides (condensation with 10 to 20% excess of 1,1,3,3-tetramethoxypropane in MeOH saturated with HCl, 20°, 15 hr) or alternatively the free peptide may be hydrazinolysed (1–5 μmole of peptide refluxed in 0.2 ml of 20% aqueous NH_2NH_2 for 30 min) to an ornithine-containing peptide, acylated and esterified (for details see [3]). It is noteworthy that the procedures described work well with micro amounts of substances and the *N*-acylpeptide esters obtained usually do not need any purification before mass spectrometry.*

Interpretation of mass spectra of peptide derivatives is based principally on amino acid type of fragmetation. However, unequivocal interpretation usually can be achieved only if the general as well as the whole variety of specific fragmentation routes are taken into consideration [1]. The assumption [4, 8, 9] that the amino acid fragmentation pattern alone is sufficient for amino sequence determination does not generally hold, even using computer interpretations of high resolution mass spectral data. According to our data [1] the mass spectrum does not necessarily exhibit every possible sequence peak; moreover on fragmentation some amino acid residues lose their side chains, either partially (e.g. leucine) or completely (e.g. tryptophan), giving rise to alanyl or glycyl residues. Using our scheme of analysis we have elucidated, for instance, the struture of the following peptides isolated from hydroly-

* To prepare *N*-acylpeptide esters it was suggested [4] that the free peptide be esterified with HCl in MeOH with subsequent acylation by $(CF_3CO)_2O + CF_3COOH$ or Ac_2O + AcOH, however, these conditions of esterification sometimes cause splitting of peptide bonds [5]. Cases have been described [6, 7], when free lysine and *S*-β-aminoethylcysteine-containing peptides were acylated with Ac_2O in MeOH.

sates of pig pepsin (IV) and pig heart aspartate aminotransferase (V, VI):

(IV) Ile–Leu–Gly–Asp–Val–Phe
(V) Leu–Glu–Ala–Leu–Lys
(VI) Val–Glu–Glu–Arg

Mass spectrometry was successfully used by Lederer and coworkers [2] for structural investigations or confirmation of earlier suggested formulae in a series of natural peptidolipids (VII–XI) and also by Wolstenholme and Vining [10] for the elucidation of the amino acid sequence in methyl isariate (XII) in establishing the structure of isariin.

Usually peptides higher than decapeptides do not give interpretable mass spectra. However, sometimes even pentapeptides do not give complete mass spectra, if they are of low volatility or thermally unstable (for example Decanoyl–Phe–Trp–Phe–His–Leu–OMe). These two factors impose the strongest limitations on mass spectrometry of peptides. Recently there have been several attempts to extend the possibilities of mass spectrometry for studying peptides. Lederer suggested *N*-permethylation of *N*-acylpeptide esters, since such derivatives are more volatile and often produce simpler mass spectra [2]. Permethylation was accomplished using MeI and reaction was most effectively achieved with sodium derivatives of dimethylsulfoxide [11] or dimethylacetamide [12], as in this way it is possible to avoid some side reactions occurring with other permethylation techniques [13]. However, even then there are cases when interpretation of mass spectra of methylated acylpeptide esters becomes baffling, as it is for instance for peptides containing lysine, ornthine, tryptophan and histidine which, according to our data, often undergo incomplete *N*-methylation at ω-amide groups (cf. [13]) and heterocyclic rings; on the other hand, in tryptophan-, histidine- and glycine-containing peptides, *N*-methylation is often accompanied by *C*-methylation. As a result, mass spectra of these peptide derivatives exhibit fragments differing by a CH_2-group thus impairing discrimination between such amino acid pairs as glycine and alanine, serine and threonine, etc. However, in spite of these shortcomings, permethylation does indeed increase the scope of mass spectrometry in peptide analysis. Lederer and Thomas successfully applied the method to the investigation of the amino acid sequence in the heptapeptide (XIII) from the zymogen of phospholipase A [14], stendomycin (XIV) [2], and esperinic acid (XV) [15] and also to elucidate the partial sequences (XVI) and (XVII) in the octadecapeptide and the docosapeptide obtained from hydrolysates of the λ-chain of pig immunoglobulin [6, 14]; see also [12].

(VII) $CH_3(CH_2)_nCO$–Phe–Ile–Ile–Phe–Ala–OMe
n = 14, 16, 18, 20

(VIII–X) $CH_3(CH_2)_nCH(O{-}\cdots{-})CH_2CO$–X–Val–Ala–Pro–alle–Y–Thr–(…)

(VIII) X = Thr, Y = Ala, n = 16

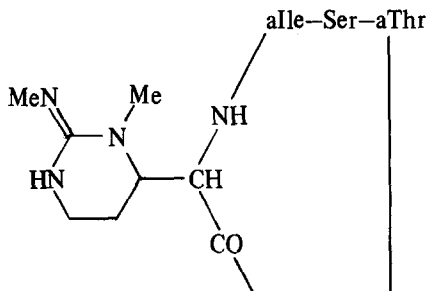
(IX) X = Thr, Y = Val, n = 16, 17, 18

(X) X = Abu, Y = Ala, n = 16, 17, 18

(XI) $CH_3(CH_2)_nCO$–Val–MeLeu–Val–Val–MeLeu–Thr(Ac)–Thr(Ac)–Ala–Pro–OMe
n = 18, 20

(XII) $CH_3(CH_2)_8CH(OH)CH_2CO$–Gly–Val–Leu–Ala–Val–OMe

(XII) PyroGlu–Glu–Gly–Ile–Ser–Ser–Arg



(XIV) $C_{13}H_{27}CO$–Pro–MeThr–Gly–Val–aIle–Ala–Abu–aThr–Val–Val

(XV) $CH_3(CH_2)_nCH(OH)CH_2CO$–Glu–Leu–Leu–Val–Asp–Leu–Leu(Val)
n = 9, 10, 11

(XVI) Ala–Thr(Ala)–Leu–Thr–Ile–Gly–Ala–Gln–Ala–...

(XVII) PyroGlu–Thr–Val–Leu(or Ile)–Gln–Glu–...

Volatility of sulfur-containing peptides is considerably enhanced by their desulfuration; this operation is likewise advantageous as a pretreatment for the permethylation technique (the latter leads to sulfonium derivatives, the mass spectra of which are rather difficult to interpret). According to our data desulfuration is effectively accomplished by reduction of *N*-acylpeptide esters using a tenfold excess of Raney nickel in dimethylformamide (20°, 48 hr) [16]. Methionine residues are thereby converted into α-aminobutyric acid residues, while cysteine derivatives and cystine are changed to alanine; no other amino acid residue is changed. Reduction with Raney nickel in boiling MeOH, as suggested by Lederer [7], brings about reduction of tyrosine and tryptophan residues to the corresponding hexa- and octahydro derivatives, thus complicating the mass spectra greatly; also in these conditions considerable adsorption of the peptide on the catalyst occurs [16].

Chemical or enzymic shortening of a peptide chain prior to mass spectrometry provides a new approach allowing the application of mass spectrometry to a wider range of peptides. Our experiments have shown that the Edman procedure allows removal of three to five amino acid residues from the *N*-terminus of the peptide, the resulting peptides can then be directly converted into *N*-acylpeptide esters and then subjected to mass spectrometry [17]. Such a combination of the Edman procedure and mass spectrometry is more advantageous when the *N*-terminus of the peptides contains amino acids which reduce volatility or require some modification before mass spectrometry, or otherwise complicate the mass spectra (these are histidine, tryptophan, tyrosine, phenylalanine, lysine, arginine and sulfur-containing amino acids). Some peptides (XVIII–XXIII) analysed by such a combined technique are as follows:

(XVIII) $\overrightarrow{\text{Met}}$–$\overrightarrow{\text{Phe}}$–$\overrightarrow{\text{Gly}}$–Ala–Phe–Val–Gly

(XIX) $\overrightarrow{\text{Phe}}$–$\overrightarrow{\text{Leu}}$–$\overrightarrow{\text{Phe}}$–Arg–Leu–Leu–Gly–Pro

(XX) $\overrightarrow{\text{Val}}$–$\overrightarrow{\text{Tyr}}$–$\overrightarrow{\text{Phe}}$–$\overrightarrow{\text{Trp}}$–Phe–His–Leu

(XXI) $\overrightarrow{\text{Arg}}$–$\overrightarrow{\text{Pro}}$–$\overrightarrow{\text{Pro}}$–$\overrightarrow{\text{Gly}}$–$\overrightarrow{\text{Phe}}$–Ser–Pro–Phe–Arg

(XXII) $\overrightarrow{\text{Phe}}$–$\overrightarrow{\text{Phe}}$–$\overrightarrow{\text{Met}}$–Phe–Gly–Leu–Ala

(XXIII) $\overrightarrow{\text{Phe}}$–$\overrightarrow{\text{Pro}}$–Pro–Phe–Phe–Val–Pro–Pro–Ala–Phe

Stepwise Edman degradation was carried out as follows: condensation with PhNCS in triethylamino acetate buffer (argon 40°, 1 hr), cyclization in a (2:1)

mixture of AcOH saturated with HCl-water (argon, 40°, 1 hr) (cf. [18]). In compounds (XIX) and (XXI) the arginine residue was converted into N^{δ}-pyrimidylornithine after acylation of the shortened peptides. The combined use of mass spectrometry and Edman degradation only becomes somewhat difficult for lysine-containing peptides (e.g. Leu–Gly–Phe–Ala–Lys) as in the Edman degradation the lysine ω-amino group gives a phenylthiocarbamoyl (PTC) derivative (Gly–Phe–Ala–(PTC)Lys), which impairs mass spectrometry. However, we have found [17] that the PTC-group is easily removed by hydrazinolysis in the conditions used to convert arginine residues into ornithine (see above) resulting in a free peptide (Gly–Phe–Ala–Lys), which is subsequently *N*-acylated, esterified and subjected to mass spectrometry without purification.

Another way of adapting mass spectrometry to longer peptides is to combine it with methods which release amino acids from the *C*-termini of peptides. The most effective of these is the use of carboxypeptidase B, as it removes arginine and lysine which are cumbersome for mass spectrometry, yet must be very often dealt with, as in tryptic digests, when they are always present at the *C*-terminus. Carboxypeptidase B degradation is convenient since this enzyme does not affect other amino acids. It is advisable to remove lysine and arginine residues with 0.04 mg carboxypeptidase B for each μmole of arginine or lysine-containing peptide in 0.2 M *N*-ethylmorpholine acetate (pH 8.5, 37°, 1.5–2 hr). Digestion is stopped by acidification with acetic acid (to pH 3–4), and the shortened peptide after acylation and esterification is subjected directly to mass spectrometry. The following compounds have been analysed by mass spectrometry combined with carboxypeptidase B degradation [19]:

Ser–Pro–Phe–Arg ←
Leu–Glu–Ala–Leu–Lys ←
Phe–Trp–Ala–Gly–Leu–Gly–Arg ←

Thus the possibilities of amino acid sequence determination of peptides are considerably extended by using a combination of different methods.

A detailed paper on the application of mass spectrometry to the elucidation of the primary structure of peptides and proteins will be published shortly [20].

References

[1] M.M.Shemyakin, Pure Appl. Chem. 17 (1968) 313.
[2] E.Lederer, Pure Appl. Chem. 17 (1968) 489.
[3] M.M.Shemyakin, E.I.Vinogradova, Yu.A.Ovchinnikov, A.A.Kiryushkin, M.Yu.Feigina, N.A.Aldanova, Yu.B. Alakhov, V.M.Lipkin, B.V.Rosinov and L.A.Fonina, Tetrahedron, in press.
[4] M.Senn, R.Venkataraghavan and F.W.McLafferty, J. Am. Chem. Soc. 88 (1966) 5593.
[5] R.T.Aplin, I.Eland and J.H.Jones, Org. Mass Spectrometry 2 (1969) 795.
[6] F.Franěk, B.Keil, D.W.Thomas and E.Lederer, FEBS Letters 2 (1969) 309.
[7] D.W.Thomas, B.C.Das, S.D.Gero and E.Lederer, Biochem. Biophys. Res. Commun. 32 (1968) 519.
[8] K.Biemann, C.Cone, B.R.Webster and G.P.Arsenault, J. Am. Chem. Soc. 88 (1966) 5598.
[9] M.Barber, P.Powers, M.J.Wallington and W.A.Wolstenholme, Nature 212 (1966) 784.
[10] W.A.Wolstenholme and L.C.Vining, Tetrahedron Letters (1966) 2785.
[11] D.W.Thomas, Biochem. Biophys. Res. Commun. 33 (1968) 483.
[12] K.L.Agarwal, G.W.Kenner and R.C.Sheppard, J. Am. Chem. Soc. 91 (1969) 3096.
[13] D.W.Thomas, FEBS Letters 5 (1969) 53.
[14] G.H.de Haas, F.Franěk, B.Keil, D.W.Thomas and E. Lederer, FEBS Letters 4 (1969) 25.
[15] D.W.Thomas and T.Ito, Tetrahedron 25 (1969) 1985.
[16] A.A.Kiryushkin, V.A.Gorlenko, B.V.Rosinov, Yu.A. Ovchinnikov and M.M.Shemyakin, Experientia 25 (1969) 913.
[17] N.A.Aldanova, E.I.Vinogradova, S.A.Kazaryan, B.V. Rosinov and M.M.Shemyakin, Biokhimiya, in press.
[18] J.Sjöquist, Arkiv Kemi 14 (1959) 291.
[19] V.M.Lipkin, Yu.B.Alakhov, N.A.Aldanova, M.Yu.Feigina, A.A.Kiryushkin, A.I.Miroshnikov, Yu.A.Ovchinnikov, B.V.Rosinov, M.M.Shemyakin and E.I.Vinogradova, in: Peptides, Proceedings of the Xth European Peptide Symposium, ed. E.Scoffone (North–Holland, Amsterdam) in press.
[20] M.M.Shemyakin, Yu.A.Ovchinnikov and A.A.Kiryushkin, in: Mass spectrometry, ed. G.W.A.Milne (Wiley, New York, 1970) in press.