Determination of amino acid sequences in oligopeptides by mass spectrometry. I. The structure of fortuitine, an acyl-nonapeptide methyl ester.[*]

M. BARBER[1], P. JOLLÈS[2], E. VILKAS[3] and E. LEDERER[3]

1) Scientific Apparatus Department, Associated Electrical Industries, Manchester.
2) Laboratoire de Chimie Biologique, Faculté des Sciences, Paris.
3) Institut de Chimie des Substances Naturelles, C.N.R.S., Gif-sur-Yvette (Seine et Oise).

Received December 18, 1964

Preliminary studies on the use of mass spectrometry for the determination of amino acid sequences have been published by Biemann et al.(1959, 1960), who reduced the peptide bonds and the terminal carboxyl by $LiAlH_4$ or $LiAlD_4$, to obtain polyaminoalcohols or the corresponding polyamines. Stenhagen (1961) and Weygand et al. (1963) have obtained mass spectra of N-trifluoro-acetyl methyl esters of oligopeptides up to pentapeptides and Manusadzhyan et al. (1964) have used N-acetyl oligopeptide ethyl esters. With a direct inlet system, Heyns and Grützmacher (1963) have studied N-acetyl peptides and have obtained a fairly satisfactory sequential splitting, the highest peptide measured being a pentapeptide derivative of mass 539.

In the following, we describe the results obtained with fortuitine, a natural peptidolipid (M = 1359) where the whole structure and the sequence of nine amino acids could be deduced by mass spectrometry.

Fortuitine, m.p. 199-202°, $[\alpha]_D$ = - 72° ($CHCl_3$) isolated from Mycobacterium fortuitum is a peptidolipid containing approximately equimolecular quantities of eicosanoic and docosanoic acids linked to an oligopeptide chain reported to contain $Val_3 Thr_2 Ala_1 Pro_1$. The carboxyl terminal of the oligopeptide was shown to be esterified by methanol. The fatty acids are combined by an amide linkage to one of the valine molecules, thus giving the preliminary structure (I) (Vilkas et al., 1963) :

[*] Supported, in part, by Grant AI-02838 of the National Institute of Allergy and Infectious Diseases (U.S. Public Health Service).

$CH_3(CH_2)_nCO$-Val-(Val$_2$, Thr$_2$, Ala$_1$, Pro$_1$)OMe n=18,20 (I)

From the results of partial acid and enzymatic hydrolysis of fortuitine the C-terminal sequence, -Thr-Thr-Ala-Pro-OMe, was derived and from a determination of acetyl groups it was thought that one of the hydroxyl groups of the two threonine molecules was acetylated. This led to structure (II) for fortuitine (Lederer, 1963, 1964 a) :

$CH_3(CH_2)_nCO$-Val-Val-Val-$\overset{\overset{Ac}{\frown}}{\text{Thr-Thr}}$-Ala-Pro-OMe (II)

A closer examination of the results of quantitative determinations of the amino acids of a total hydrolysate of fortuitine, by the method of Moore and Stein (as modified by Piez and Morris, 1960), showed, however, that a nitrogenous portion of the molecule was not yet accounted for. Mass spectrometry has led to the identification of this "unknown constituent" as two molecules of N-methyl leucine and has allowed their place in the nonapeptide chain to be determined without ambiguity.

The mass spectrum of fortuitine (Fig. 1) [1] shows two parent peaks

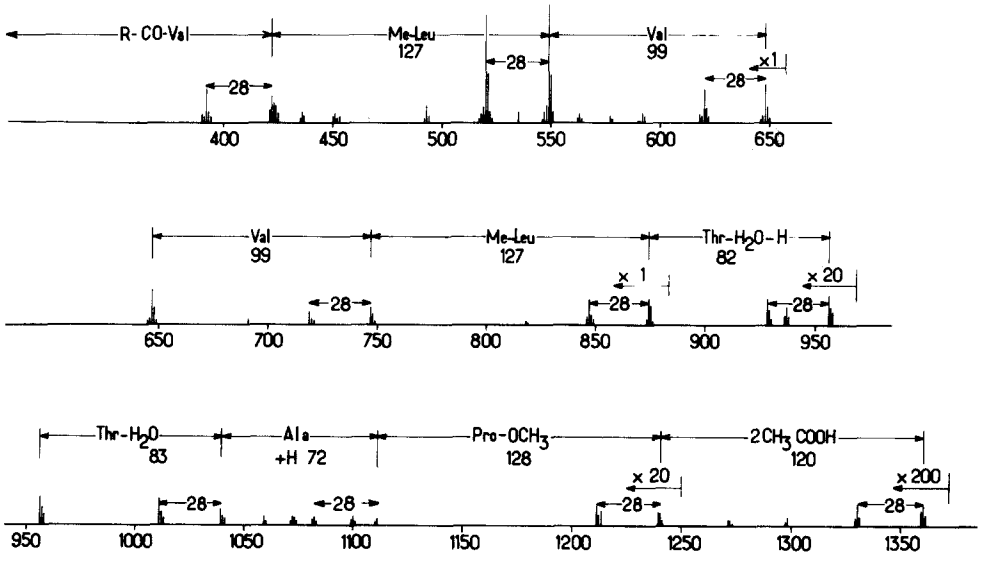

Fig. 1 Line spectrum of Fortuitine

1. Fortuitine was introduced into the MS9 double focussing mass spectrometer by the direct evaporation technique described previously (Barber et al., 1964 a, 1964 b) and a satisfactory spectrum obtained with an evaporating temperature of 200° and less than 10 μg of the substance.

at 1331 and 1359, due to it being a mixture of two homologs containing a C_{20} and C_{22} fatty acid respectively. All ions containing the lipid moiety will therefore give sets of peaks 28 mass numbers apart, thus facilitating interpretation.

The parent peaks, at 1331 and 1359, each lose 2 x CH_3COOH = 120 to give peaks at 1211 and 1239. This shows the presence of two acetyl groups : both threonine molecules are O-acetylated.

Two small peaks at 1083 and 1111 are apparently due to the loss of-Pro-OMe (128).

A prominent set of peaks occurs at 1011 and 1039. This is 72 less than 1083 and 1111, due to the loss of Ala + H (migration of one H atom).

The next set of peaks occurs at 928 and 956. This loss of 83 from 1011 and 1039 corresponds to the loss of one molecule of anhydrothreonine, again after migration of one H atom. This so far sets the sequence R-Thr-Ala-Pro-OMe.

Very large peaks occur at 846 and 874 and can be thought of as a straight cleavage between a second anhydrothreonine and the remainder of the peptide, i.e. a loss of 82 from 928 and 956, now giving the sequence R-Thr-Thr-Ala-Pro-OMe, in agreement with the results of chemical degradation.

The next characteristic grouping of two peaks 28 mass numbers apart occurs at 719 and 747. This is 127 (=MeLeu) less than 846 and 874 [2]. The sequence is now R-MeLeu-Thr-Thr-Ala-Pro-OMe.

The 719 and 747 group then loses 99 to give peaks at 620 and 648 which then again lose 99 to give the 521 and 549 peaks. (There are "metastable" ions at m/e 437.8 and 465.1 to confirm this). The mass difference of 99 indicates the loss of valine, so that the sequence becomes R-Val-Val-MeLeu-Thr-Thr-Ala-Pro-OMe.

The next group of large peaks is at 394 and 422 corresponding

---

2. N-methyl-leucine was identified in acid hydrolysates of fortuitine by paper chromatography (H_2O-saturated tert. amylalcohol). R_f= 0.39, identical with authentic N-methyl leucine, whereas N-methyl-isoleucine had an R_f of 0.31. The configuration of this N-methyl leucine has not yet been established ; all other amino acids of fortuitine are L (Vilkas et al., 1963).

again to a loss of 127 (=MeLeu). The sequence is now

R-MeLeu-Val-Val-MeLeu-Thr-Thr-Ala-Pro-OMe.

As the molecular weight of the long chain acyl groups is known (295 and 323), we can conclude that the peaks at 394 and 422 are acyl + 99 = RCO-Val-. This now gives the whole structure of fortuitine as (III):

$$CH_3(CH_2)_{18-20}CO\text{-Val-MeLeu-Val-Val-MeLeu-}\overset{\overset{Ac}{|}}{Thr}\text{-}\overset{\overset{Ac}{|}}{Thr}\text{-Ala-Pro-OMe} \quad \text{(III)}$$

corresponding to $C_{70}H_{125}N_9O_{15}$ = 1331 and $C_{72}H_{129}N_9O_{15}$ = 1359 respectively.

The following table gives mass measurements of five peaks (all corresponding to the lower fortuitine homolog containing the C_{20} fatty acid, with the exception of the 648 peak), which are in excellent agreement with the proposed structure. (Part of these results has already been mentioned (Lederer, 1964 b ; Barber, Elliott and Merren, 1964).

Table

The atomic composition of the prominent fragment peaks in the mass spectrum of fortuitine

Nominal mass	Formula	Measured mass	Theoretical mass	Error (ppm)
m/e 521	$C_{32}H_{61}N_2O_3$	521.4669	521.4682	- 2
648	$C_{39}H_{74}N_3O_4$	648.5661	648.5679	- 2
719	$C_{42}H_{79}N_4O_5$	719.6046	719.6054	- 1
846	$C_{49}H_{92}N_5O_6$	846.7001	846.7046	+ 5
928	$C_{53}H_{96}N_6O_7$	928.7327	928.7340	- 1

Previous papers, quoted above, have shown that the main splitting of oligopeptides occurs at both sides of the peptide bond $R\{CO\{NH-$ and that interpretation is often difficult because several amino acids lose parts of their side chains.

We now show, that, in certain cases, the splitting of the peptide bond can be the most prominent one, thus allowing the determination of the amino acid sequences.

Future papers will give further examples of sequential analysis of natural and synthetic acyl-peptide esters. Using the present model MS9, the upper limit of usefulness of mass spectrometry for the ana-

lysis of sequences in peptides is probably about 12 amino acids ; this estimation is based on the fact that fortuitine with its nine amino acids seems not yet to be the upper limit, whereas gramicidin A (a formyl pentadecapeptide with a terminal ethanolamine group ; Sarges and Witkop, 1964), introduced as its C-terminal acetate, gave no spectrum at all, being apparently insufficiently volatile.

References

Barber, M., Elliott, R. M. and Merren, T. O., Institute of Petroleum/ ASTM Mass Spectrometry Symposium, Paris (1964 a).

Barber, M., Merren, T. O. and Kelly, W., Tetrahed. Lett., 1063, (1964 b)

Biemann, K., Gapp, F. and Seibl, J., J. Am. Chem. Soc., 81, 2274 (1959).

Biemann, K. and Vetter, W., Biochem. Biophys. Res. Comm., 3, 578 (1960).

Heyns, K. and Grützmacher, H. F., Tetrahed. Lett., 1761 (1963).

Heyns, K. and Grützmacher, H. F., Ann. 669, 189 (1963).

Lederer, E., Pure and Applied Chemistry, 7, 247 (1963).

Lederer, E., Angew. Chem., 76, 241 (1964 a) ; Angew. Chem. Intern. Ed. 3, 393 (1964 a).

Lederer, E., I. U. B. Symposium Series, (1964 b), in press.

Manusadzhyan, V. G., Zyakoon, A. M., Chuvilin, A. V. and Varshavskii, Ya. M., Izv. Akad. Nauk Arm. SSR Khim, 17, 143 (1964).

Piez, K. A. and Morris, L., Anal. Biochem., 1, 187 (1960).

Sarges, R. and Witkop, B., J. Am. Chem. Soc., 86, 1862 (1964).

Stenhagen, E., Z. Anal. Chem., 181, 462 (1961).

Vilkas, E., Miquel, A. M. and Lederer, E., Biochim. Biophys. Acta, 70, 217 (1963).

Weygand, F., Prox, A., König, W. and Fessel, H. H., Angew. Chem., 75, 724 (1963).

Acknowledgement. - We thank Dr. B. Witkop for several samples of gramicidin A and its derivatives.