AUTOMATIC AMINO-ACID-SEQUENCE DETERMINATION IN PEPTIDES

by Martin Senn and F. W. McLafferty

Department of Chemistry, Purdue University

Lafayette, Indiana 47907

Received March 9, 1966

In the elucidation of protein structures, the classical determination of amino acid sequences in oligopeptides by degradation and amino acid analysis is time-consuming and requires relatively large samples. An alternate source of structure information which can utilize much smaller samples is indicated by recent publications on the mass spectra of derivatives of linear and cyclic peptides and depsipeptides (Bricas, et al., 1965; Weygand, et al., 1965; Millard, 1965; Wulfson, et al., 1965; Heyns and Grützmacher, 1963). Weygand and his coworkers report that the low resolution mass spectra of 150 N-trifluoroacetyl peptide derivatives are consistent with their known structures, and describe how sequence information can be obtained from such spectra. Barber, Lederer, and their coworkers (Bricas, et al., 1965) similarly use N-fatty acid derivatives, whose mass spectra have the advantage of providing unequivocal structure information for peptides whose molecular weight is less than the fatty acid moiety. Routine application of the method to protein structure determinations, especially with utilization of computer techniques, requires unambiguous mass spectral data.

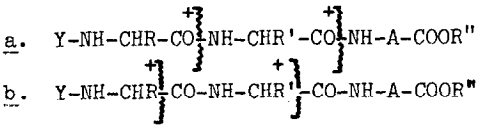
The determination of the structure of a linear molecule from its mass spectrum involves reassembling the ionic fragments observed into a logical sequence. For long chain molecules, especially those composed of units of similar structure, reassembly of the many possible complex fragments is generally difficult and ambiguous. This can be simplified by recognition of the fact that, barring rearrangements, the structure of a linear molecule is determined unequivocally by using only the possible fragments which contain

one end of the chain. The fundamental advantage of the Barber and Lederer method is not just that the N-terminal fatty acid moiety raises the masses of these ions above the masses of other fragment ions in the spectrum (Bricas, et al., 1965), but that by doing so it provides unequivocal identification of all fragments containing one end of the peptide chain.

To mark the chain end, a number of derivatives of higher volatility can be used, which thus permit the analysis of larger peptides. For example, comparison of the spectra of the N-acetyl and the N-trideuterioacetyl derivatives should indicate the fragments containing the N-terminal grouping. Another possibility is an N-derivative containing an element with characteristic natural isotopes; N-CF_2BrCO-containing fragments would all show the typical Br-79:Br-81 isotopic abundance ratio of 1:1.

High-resolution mass spectrometry provides a general method of chain-end identification. For almost any peptide derivative of sufficient volatility the presence of the terminal functional group makes a unique mass contribution to the exact mass of a particular fragment ion. This approach will be illustrated by the oligopeptide spectra of Tables 1, 2, and 3. The exact masses measured are recorded for the principal ions only.

The published fragmentation patterns of a large variety of peptides show that cleavage of the chain involves two main pathways:



The relative probabilities of these reactions depend on the particular peptide used, and the cleavage may be accompanied by loss or gain of a hydrogen atom (Weygand, et al., 1965).

The computer program devised to analyze such spectra first searches for the N-terminal amino acid. Combinations of the exact mass of the N-terminal group (CH_3CO, 43.01839; CD_3CO, 46.03722; CF_3CO, 96.99012; etc.) and the mass of each of the possible amino acid fragments (Table 4) are checked against the

Table 1. N-Trifluoroacetyl-leu-gly-phe-OMe

scheme a measured masses	calculated sequence	scheme b calculated masses	scheme b measured masses
	96.99012		
	113.08406 (leu)		
210.0726 [1.6]	210.07418	182.07927	182.0781 [1.1]
	57.02146 (gly)		
267.0956 [0.0]	267.09564	239.10073	239.0991 [1.6]
	147.06841 (phe)		
414.1647 [0.7]	414.16405	386.16914	386.1682 [0.9]
	31.01839 (OMe)		
445.1829 [0.5]	445.18244		

Table 2. N-Acetyl-pro-gly-phe-gly-OMe

scheme a measured masses	calculated sequence	scheme b calculated masses	scheme b measured masses
	43.01839		
	97.05276 (pro)		
140.0712 [0.0]	140.07115	112.07624	112.0743 [1.9]
	57.02146 (gly)		
197.0908 [1.8]	197.09261	169.09770	169.0958 [1.9]
	147.06841 (phe)		
344.1605 [0.5]	344.16102	316.16611	316.1648 [1.3]
	57.02146 (gly)		
401.1826 [0.1]	401.18248	373.18757	373.1893 [1.7]
	31.01839 (OMe)		
432.2004 [0.5]	432.20087		

masses of ions in the spectrum.* Correspondence of mass within experimental error (2 millimass units; actual errors are shown in brackets in Tables 1 - 3) indicates the N-terminal acid; only one such match should be found. A check is alsc made for the corresponding ion from path b utilizing masses obtained by subtracting 27.99491 (CO) from each of the first combinations. The search for each additional amino acid of the chain follows the same pattern, starting with the combined theoretical masses of the terminal functional group and the

* Spectra were recorded on photoplates using a high-resolution mass spectrometer of Mattauch-Herzog geometry (CEC 21-110), with direct introduction of the samples into the ion source.

Table 3. N-Trifluoroacetyl-ala-leu-ala-val-val-val-OCH_3

scheme a measured masses	calculated sequence	scheme b calculated masses	scheme b measured masses
	96.99012		
	71.03711 (ala)		
168.0287 [1.5]	168.02723	140.03232	140.0339 [1.6]
	113.08406 (leu)		
281.1124 [1.2]	281.11129	253.11638	253.1164 [0.0]
	71.03711 (ala)		
352.1474 [0.6]	352.14840	324.15349	324.1539 [0.5]
	99.06841 (val)		
451.2180 [1.2]	451.21681	423.22190	423.2205 [1.4]
	99.06841 (val)		
550.2859 [0.7]	550.28522	522.29031	522.2868 [1.5]
	99.06841 (val)		
649.3525 [1.1]	649.35363		
	31.01839 (OMe)		
680.3718 [0.2]	680.37202		

Table 4. Amino Acid Fragments

gly	57.02146	met	131.04048
ala	71.03711	hist	137.05891
ser	87.03203	$gluOCH_3$	143.05824
pro	97.05276	phe	147.06841
val	99.06841	tyr	163.06332
thr	101.04768	try	186.07931
leu	113.08406		

identified acids. If two different amino acids are both indicated as a particular chain member (ordinarily due to an error in the measured spectrum), the ambiguity is usually resolved by searching for a fragment mass to correspond to the addition of another acid group to either possibility; an acceptable match is usually found for only one. When the search for an additional chain acid finds no suitable match for either path a or b, a check is made for the molecular ion by adding mass units corresponding to the C-terminal derivative -OR". For peptides which do not contain a molecular ion the amino acid sequence from the C-terminal end can be determined by using the same procedure utilizing the "ester peaks" $H_2\overset{+}{N}=CR-COOR"$ and $H_3\overset{+}{N}-CHR-COOR"$. The relative abundance of these usually decreases rapidly with chain length.

As an additional check it appears useful to include in Table 4 the masses of acid groups in which a hydrogen atom is added or lost by rearrangement.

Checks for particular amino acids can be made through masses of characteristic ions. Other details of the computer program will be reported in a full paper.

A concurrent research program in this laboratory is the development of a fully automatic system for data reduction of high resolution mass spectra (McLafferty, 1966). In this, photoplate-recorded spectra of submicrogram samples are automatically measured for line position and density using a Grant-Datex comparator-densitometer, requiring 15 minutes per spectrum to record this data on magnetic tape. Although this system is still under development, computer calculation of ion masses appears capable of substantially improving the accuracy reported here. Further automation of this technique and application to a wider variety of peptide derivatives is under investigation in our laboratory.

Acknowledgements

The authors are indebted to Dr. A. Light of this department for helpful discussions; to the following for valuable supplies of peptide samples: Drs. G. Anderson, Lederle Laboratories; F. M. Bumpus, Cleveland Clinic; K. Hofmann, Univ. of Pittsburgh; P. G. Katsoyannis, Brookhaven National Laboratory; C. H. Li, Univ. of California, Berkeley; E. L. Smith, Univ. of California, Los Angeles; B. Witkop, National Institutes of Health; and to the National Aeronautics and Space Administration and to the National Institutes of Health for generous financial support.

References

E. Bricas, J. van Heijenoort, M. Barber, W. A. Wolstenholme, B. C. Das, and E. Lederer, Biochemistry 4, 2254 (1965).

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

F. W. McLafferty, Science 151, 641 (1966).

B. J. Millard, Tetrahedron Letters 1965, 3041.

F. Weygand, A. Prox, H. H. Fessel and K. K. Sun, Z. f. Naturforsch. 20b, 1169 (1965).

N. S. Wulfson, V. A. Puchkov, B. V. Rozinov, A. M. Zyakin, M. M. Shemyakin, Yu. A. Ovchinnikov, A. A. Kiryushkin and V. T. Ivanov, Tetrahedron Letters 1965, 2793.