DETERMINATION OF AMINO ACID SEQUENCES IN PEPTIDE MIXTURES BY MASS SPECTROMETRY.*

F. W. McLafferty, R. Venkataraghavan, and Philip Irving

Department of Chemistry, Cornell University
Ithaca, New York 14850

Received March 5, 1970

A critical limitation in the present general procedure for the determination of the amino acid sequence of polypeptides is the requirement that the oligopeptides resulting from the degradation steps must be of relatively high purity for the Edman sequence determination. It is shown that at least partial sequence information can be obtained on oligopeptide mixtures by mass spectrometry utilizing exact mass, partial vaporization, and metastable ion data.

Classical methods for the determination of the primary structure of polypeptides have been extended to large and complex molecules. The recently-determined human γ G immunoglobulin contains two chains of 214 and two chains of 446 amino acid residues.[2] The amino acid sequence of such an antibody determines its specificity;[3] the similar biological importance of sequence in many other kinds of polypeptides has justified the man-years of effort required for such a determination. In the classical method[2,3] the polypeptide chain is cleaved into several fragment peptides, each of these is isolated in pure form, and these steps are repeated until all of the peptide fragments are short enough (< ca. 25 residues) for determination of their amino acid sequence by stepwise Edman degradations. Purity of the oligopeptide is a key factor; failure to purify an end product of a particular degradation scheme can compromise the usefulness of all information from that scheme. To determine the positions of these identified units in the original polypeptide, it can be hydrolyzed to yield oligopeptides containing the overlap sequence informa-

* Mass Spectrometric Studies of Peptides. IV; part III, see ref. 1.

tion; these usually must be identified in the presence of many other product oligopeptides whose sequences provide no useful information.*

Although mass spectrometry is a promising alternative for sequence determination of 50 μg quantities of pure oligopeptides containing up to approximately ten residues,[4] little use has been made of it to date for polypeptide sequences.[5] Its value would be greatly increased by providing sequence information on mixtures of oligopeptides;[6] we show here that this is possible by utilizing exact mass, partial vaporization and metastable ion data.

In an example illustrating several aspects of the proposed method, a mixture of oligopeptides was methyl-esterified and acetylated,[1,4,7] and the exact masses of 396 peaks in its spectrum were measured with an AEI MS-902 mass spectrometer.[8] The computer program for determining the possible sequences consistent with these peaks follows the approach[1,7] utilized for the pure peptides.** The N-terminal amino acid of each component should produce peaks corresponding to $CH_3CONHCHR^+$ and $CH_3CONHCHRCO^+$; in this spectrum glycine (72.0461, 100.0377), valine (114.0933, 142.0870), methionine (146.0620, 174.0559), and alanine (86.0606, but none at 114.0561) were indicated by the peaks in parenthesis. In similar fashion a search is made with each possible terminal amino acid for the possible dipeptide fragments $CH_3CONHCHRCONHCHR'^+$ and $CH_3CONHCHRCONHCHR'CO^+$ and molecular ions $CH_3CONHCHRCOOCH_3$. The process is continued with polypeptide fragments, and a similar search is made from the C-terminal end, until no further sequence possibilities are found. The 11 possible sequences found (Table I) involved 5 different molecular compositions, of which only 1 and 11 had a single sequence assignment. Only 10 can account for the N-

* The authors are deeply indebted to Drs. G.M. Edelman, W.E. Gall, and M.J. Waxdal of Rockefeller University for pointing out these and other important problems.

** The ambiguity of the sequence information increases rapidly with the number of components and the proportion of each component's sequence peaks which are not observable; however, these can be indicated by the sample's prior history, amino acid analysis, behavior in purification, or peak abundances.

Table I. Computer-postulated sequences for test mixture

	Sequence	molecular weight	N-sequence peaks found/expected
1	Ac-gly-OMe[a]	131.0582	3/3
2	Ac-gly-gly-val-OMe	287.1479	5/7
3	Ac-gly-ala-gly-OMe	259.1166	7/7
4	Ac-gly-ala-leu-OMe[a]	315.1791	6/7
5	Ac-gly-val-gly-OMe	287.1479	5/7
6	Ac-ala-gly-gly-OMe	259.1166	5/7
7	Ac-ala-gly-leu-OMe	315.1791	5/7
8	Ac-ala-ala-ala-OMe	287.1479	5/7
9	Ac-ala-ala-val-OMe	315.1791	5/7
10	Ac-val-gly-gly-OMe[a]	287.1479	5/7
11	Ac-met-phe-gly-OMe[a]	409.1667	6/7

a - This test mixture was made by mixing 4, 10, and 11 in approximately equimolar amounts. 1 was found to be an impurity in 4.

terminal valine, but the other possibilities for MW 287.1479 could conceivably be present also. Causes of these multiple possibilities include the elemental identity of amino acid combinations such as gly-leu and ala-val, and peaks from pathways other than cleavages of the polyamide chain. Thus many of the same peaks are used to assign the sequences of isomers 4 and 7, and rearrangement loss of C_4H_8 from the leucine side chain in 4 and 7 should produce all of the peaks which indicate isomers 3 and 6, respectively.[1,4,7]

It would obviously be helpful to determine which peaks originate from the same oligopeptide. If fractional vaporization of the sample occurs in the ion source, the ratios of the abundances of only those peaks formed from a common oligopeptide should not change as the sample is depleted. Vaporization rates of the components producing the peaks of postulated sequences 2

through 10 were too similar to make any distinctions, but the behavior of those peaks expected from 1 and 11 showed these components to be more and less volatile, respectively, than the rest, confirming the presence of 1 and 11.

If a metastable is identified corresponding to a particular ion decomposing to yield another ion, both of these ions must arise (at least in part) from the same component of the mixture.[9] Nine metastables from the test mixture were consistent only with the presence of 11. Eight metastables were consistent with 10, of which 287 → 114 would not be expected from 1 - 9 or 11.* Fourteen metastables were consistent with the presence of 4, although only the very weak metastables 315 → 100 and 284 → 100 would be unique for 4. No metastables were detected which would be unique for other postulated sequences (e.g., 287 → 100 for 2 and 5).

Thus the presence of components 1, 4, 10, and 11 were correctly predicted. The method does not, however, place an upper limit on the concentration of other possible components, and anomalous peaks (e.g., m/e 86.0606 in this example) can arise. These ambiguities can be reduced by utilizing mass spectra of alternative derivatives,[1,4,7] relative abundances expected from correlation studies[7] (e.g., sequences 6 - 9 should yield a detectable m/e 114.0561 as well as 86.0606 peak), or information from conventional techniques (e.g., amino acid analyses). A great many aspects of this method obviously need further study, such as techniques for the preparation of stable volatile derivatives of all possible peptide combinations;[4,5,10] in some mixtures containing higher peptides the present method did not eliminate all ambiguities. However, the ability to obtain even partial sequence information on some components of a degradation mixture should be of substantial importance in a number of phases of the classical method for amino acid sequence determination in polypeptides.[2,3,6]

* A detectable metastable for 287 → 114 would not be expected from 8, and 171 → 114 from 6 and 7, as the 114.0561 peak is not detectable in the spectrum.[10]

Acknowledgments. Dr. M. McCamish has contributed to further studies on other peptide mixtures, and Dr. J. Hribar assisted with the operation of the mass spectrometer. We are indebted to the Public Health Service (NIH Grant GM-16609) for generous support of this research.

REFERENCES

1. M. Senn, R. Venkataraghavan, and F. W. McLafferty, J. Amer. Chem. Soc., 88, 5593 (1966).
2. G.M. Edelman, B.A. Cunningham, W.E. Gall, P.D. Gottlieb, U. Rutishauser, and M.J. Waxdal, Proc. Natl. Acad. Sci. U.S., 63, 78 (1969).
3. G.M. Edelman and W.E. Gall, Ann. Rev. Biochem., 38, 415 (1969).
4. For recent reviews see: (a) E. Lederer, Intern. Sympos. Chem. Natural Prod., 5th, London, July, 1968; (b) G.E. Van Lear and F. W. McLafferty, Ann. Rev. Biochem., 38, 289 (1969).
5. K.L. Agarwal, G.W. Kenner, and R.C. Sheppard, J. Amer. Chem. Soc., 91, 3096 (1969).
6. Ref 4b, p 303.
7. K. Biemann, C. Cone, B.R. Webster, and G.P. Arsenault, J. Amer. Chem. Soc., 88, 5598 (1966).
8. On-line computer data reduction described in R.J. Klimowski, R. Venkataraghavan, F.W. McLafferty, and E.B. Delaney, submitted for publication.
9. K. Biemann, "Mass Spectrometry. Organic Chemical Applications," McGraw-Hill, New York, 1962, p 155; F.W. McLafferty and W.A. Bryce, Chem. Commun., 1215 (1967).
10. V. Bacon, E. Jellum, W. Patton, W. Pereira, and B. Halpern, Biochem. Biophys. Res. Commun., 37, 878 (1969).