Mass Spectrometric Studies of Peptides. V.[1] Determination of Amino Acid Sequences in Peptide Mixtures by Mass Spectrometry

Hans-Kaspar Wipf,[2a] Philip Irving,[2b] Malcolm McCamish,[2c] Rengachari Venkataraghavan, and F. W. McLafferty*

Contribution from the Department of Chemistry, Cornell University, Ithaca, New York 14850. *Received August 8, 1972*

Abstract: Amino acid sequence information on the individual components of mixtures of oligopeptides can be obtained by mass spectrometry using a combination of low- and high-resolution, chemical ionization, metastable ion, and fractional vaporization data. The proposed method yields valid, although not necessarily complete, sequence information on components in 5–10% concentration and 10–20 nmol quantity for mixtures with at least four components. Such a capability should make possible a simpler approach to the routine sequencing of polypeptides.

Classical chemical and enzymatic methods[3] for the determination of amino acid sequences in proteins and other polypeptides have become very sophisticated, as evidenced by successful studies of peptide chains containing several hundred amino acid residues. Despite notable progress, these determinations are still very laborious, utilizing the basic approach of early studies.[4] This approach involves a series of selective cleavages of the polypeptide chain, each followed by separation of the resulting oligopeptides in high purity, with individual sequence determination of each product whose chain length has been sufficiently reduced. Despite the development of sophisticated separation processes,[3] oligopeptide purity requirements for chemical sequencing procedures, such as the Edman degradation,[5] are a major limitation in this basic approach.[6] If separation cannot be achieved for one oligopeptide mixture, the sequence information for the corresponding portion of the chain cannot be obtained, necessitating a new strategy for degradation of the original polypeptide.

Mass spectrometry has been shown to be a promising tool for the sequencing of pure oligopeptides,[7-9] and has been used to identify closely similar peptides in binary mixtures.[10,11] Unfortunately, the usual submicrogram sensitivity of mass spectrometry cannot be utilized in this application; chemical conversion of the sample to a volatile derivative is necessary, requiring 0.1–1 μmol of the peptide. The sensitivity of chemical sequencing methods is now at least comparable, so that mass spectrometry is used mainly in special cases, such as for an unusual amino acid residue or when the N-terminal amino group is blocked.[7-9]

A general advantage of mass spectrometry is the unusual amount and specificity of analytical information it can provide. In a preliminary communication,[1] we showed that this made possible the sequence determination for individual components of peptide mixtures. We have now extended this to a variety of complex mixtures, and report here that several types of mass spectral data, including low- and high-resolution, chemical ionization, metastable ion, and fractional vaporization measurements, can provide complementary sequence information of high reliability.[12,14]

Experimental Section

The most suitable chemical derivatization techniques found in this study are similar to those of other workers.[7-9] N-Acetylation and combined O- and N-methylation of the peptide mixture give satisfactory derivatives for all but arginine-containing peptides; special techniques are necessary for these.[7-9] Mass spectra of N-permethylated samples have greater relative abundances of structurally significant peaks; for these derivatives the initial methyl esterification used for earlier samples[1] is unnecessary.

Acetylation. The peptide sample (0.2–1 μmol) is allowed to stand in 1 ml of AcOH–Ac_2O (1:1 v/v) at room temperature for 4 hr, after which the reagents are removed *in vacuo*. With this procedure the N-terminal and ω-amino groups are acetylated. A separate perdeuterioacetylation provides useful complementary information for complex samples.

(1) Part IV: F. W. McLafferty, R. Venkataraghavan, and P. Irving, *Biochem. Biophys. Res. Commun.*, **39**, 274 (1970).

(2) (a) Postdoctoral Fellow, 1970–1972; (b) abstracted in part from the Ph.D. Thesis of P. I., Cornell University, 1971; (c) Postdoctoral Fellow, 1969–1970.

(3) S. Blackburn, "Protein Sequence Determination," Marcel Dekker, New York, N. Y., 1970.

(4) F. Sanger in "Nobel Lectures—Chemistry 1942–1962," Elsevier, Amsterdam, 1964, p 544.

(5) P. Edman and A. Begg, *Eur. J. Biochem.*, **1**, 80 (1967).

(6) We are greatly indebted to Drs. G. M. Edelman, B. A. Cunningham, and W. E. Gall of Rockefeller University for many helpful discussions of polypeptide sequence determination, including strategies involving the use of mass spectrometry. Further work in this area is in progress.

(7) E. Lederer, *Pure Appl. Chem.*, **17**, 489 (1968).

(8) M. M. Shemyakin, Yu. A. Ovchinnikov, and A. A. Kiryushkin in "Mass Spectrometry—Techniques and Applications," G. W. A. Milne, Ed., Wiley, New York, N. Y., 1971, p 289.

(9) K. Biemann in "Biochemical Applications of Mass Spectrometry," G. R. Waller, Ed., Wiley, New York, N. Y., 1972, p 405.

(10) S. Takeuchi, M. Senn, R. W. Curtis, and F. W. McLafferty, *Phytochemistry*, **6**, 287 (1967).

(11) G. H. de Haas, F. Franek, B. Keil, D. W. Thomas, and E. Lederer, *FEBS (Fed. Eur. Biochem. Soc.) Lett.*, **2**, 309 (1969).

(12) Since the appearance of our original communication, Williams and coworkers have made a similar proposal.[13] However, they take issue with our method in one regard, stressing that few ambiguities arise for which high-resolution data are useful, although they present data on only one mixture to support this conclusion. We will attempt to show that the high specificity of high-resolution information increases the reliability of most determinations and is especially useful for mixtures that are more complex and contain more kinds of amino acids (specifically, histidine and those containing sulfur) than those examined by these workers.[13]

(13) H. R. Morris, D. H. Williams, and R. P. Ambler, *Biochem. J.*, **125**, 189 (1971).

(14) Biemann and coworkers have recently proposed[9] a promising scheme for polypeptide sequencing involving degradation to mixtures of di- and tripeptides whose derivatives can be separated and sequenced by gas chromatography–mass spectrometry.

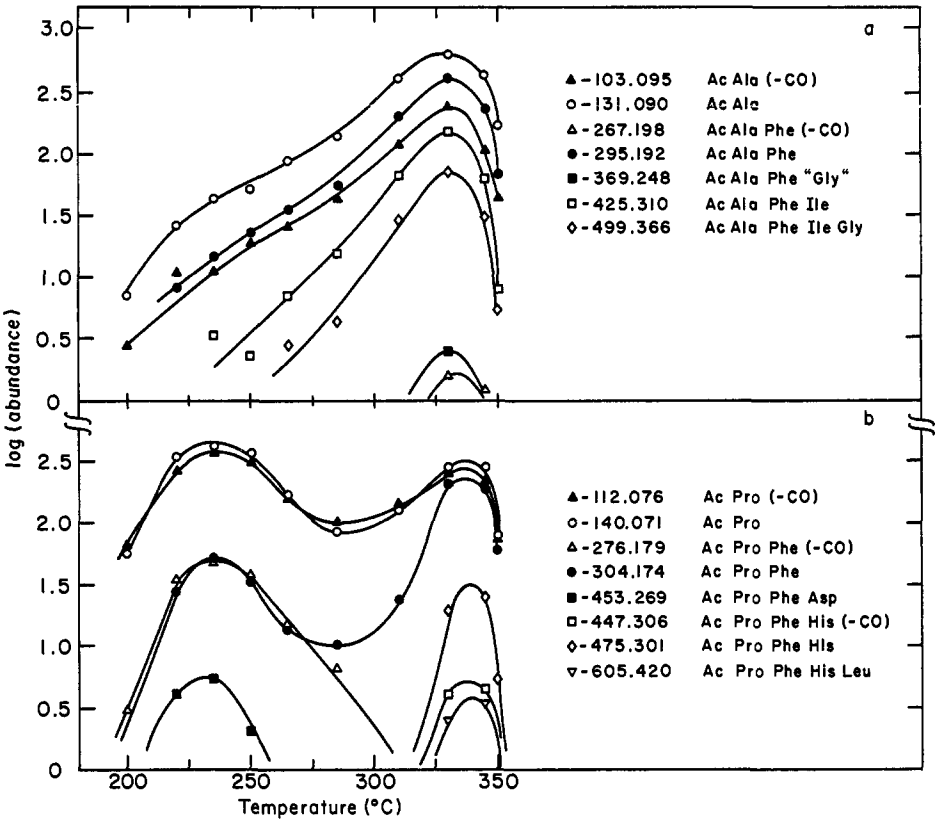


Figure 1. Variation of ion abundance with sample temperature for mixture 3.

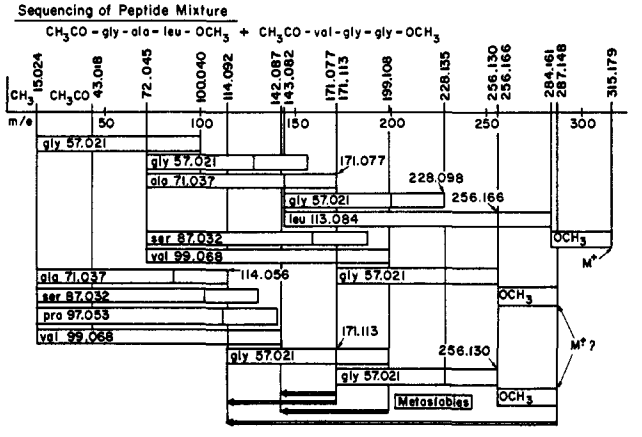


Figure 2. Sequence information available from mass spectral data on mixture 1.

Permethylation. The Hakomori method[15] employed requires scrupulous elimination of water. Baked glassware is used, and the dimethyl sulfoxide (DMSO) is redistilled over CaH_2 under reduced pressure. The permethylating base is prepared by reacting 0.2 g of NaH (50% oil dispersion washed repeatedly with hexane under Ar) with 10 ml of DMSO at 80° until H_2 evolution ceases. The acetylated peptide is dissolved in 0.5 ml of DMSO and reacted with 0.1 ml of cooled base followed by 0.05 ml of CD_3I. After 1 hr the reaction is stopped by the addition of 2 ml of H_2O, the product extracted with $CHCl_3$ (4 × 1 ml) and washed with H_2O (3 × 1 ml), and the solvent removed *in vacuo*.

Under these conditions all hydrogens on heteroatoms (including sulfur) are replaced by CD_3. Histidine derivatives are converted to quaternary salts, but these lose CD_3I thermally in the mass spectrometer to give the CD_3–histidine derivative. Esters are transesterified to the CD_3–ester. Some C-methylation can occur, such as with glycine and methionine, lowering the overall sensitivity; CD_3I is used so that the new peaks thus formed can be distinguished from those of higher homologs.

Esterification. Mixtures 1–3 were also esterified before acetylation by dissolving the peptide in 1 ml of HCl–CH_3OH (1:4 w/w). After 5 min at room temperature the reagents are removed *in vacuo*. All carboxylic acid groups (C-terminal and side chain) are esterified.

Mass Spectral Data. Approximately 0.1 μmol of the crude permethylated product is introduced into an AEI MS-902 mass spectrometer by means of a direct insertion probe having separate temperature control. For chemical ionization (CI) data a Chemspect CIS-2 source from Scientific Research Instruments, Inc., was added to the MS-902; CH_4 is used as the ionizing reagent. Abundance data are obtained for CI as well as electron ionization (EI) by scanning the complete spectrum at 2-min intervals (Figure 1) while heating to about 350°. The high-resolution system utilizing a DEC PDP-9 computer is similar to that previously described;[16] it can provide ion abundance and exact mass data (±0.003 amu) on as many as 600 peaks for each 2-min interval. Perfluorokerosene is used as the mass reference for both EI and CI. Low-resolution and metastable data are obtained in a separate scan with oscillographic recording.

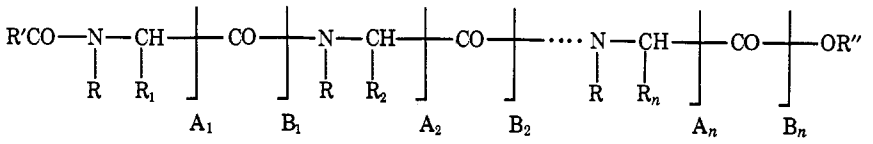
Sequencing. Sequence information is obtained from the high-resolution data by an extension of the method used for pure peptides.[17] This is illustrated in Figure 2 with selected data from mixture 1, a binary mixture of tripeptides. Possible sequence peaks are calculated using an off-line computer from a list of exact masses of the possible amino acids (*vide infra*) in their derivatized form.[18] A search is carried out for N-terminal amino acids by

(15) D. W. Thomas, *Biochem. Biophys. Res. Commun.*, **33**, 483 (1968).

(16) R. J. Klimowski, R. Venkataraghavan, F. W. McLafferty, and E. B. Delaney, *Org. Mass Spectrom.*, 4, 17 (1970).

(17) M. Senn, R. Venkataraghavan, and F. W. McLafferty, *J. Amer. Chem. Soc.*, **88**, 5593 (1966).

(18) For serine and threonine the corresponding dehydro amino acids are included too because they are known to lose their functional group under mass spectrometric conditions. For glycine and methionine the homologous amino acids containing an extra perdeuteriomethyl group (overmethylated) are also listed. No distinction is made between leucine and isoleucine.



1

R_i = side chain of amino acid i $\quad$ $R' = CH_3$ or CD_3

$R = H$ or CD_3 $\quad$ $R'' = CH_3$ or CD_3

checking the combined high-resolution data set for sequence peaks within experimental error of the expected masses (*vide infra*). Then possible dipeptide sequence peaks are calculated for the N-terminal acids found and a similar check is made. The procedure is repeated, searching stepwise for peptide fragments of higher order until no new sequence possibilities are found. For every partial sequence found a molecular ion $M\cdot^+$ is also sought. For the identification of tripeptide and larger fragments the computer also examines the additional low resolution peaks for which the value of abundance times $(m/e)^2$ is above an arbitrary minimum. Otherwise the abundance values are not used in sequence determination except for the fractional vaporization studies described below. Exact mass data on low-resolution peaks selected as possible sequence ions can sometimes be obtained by manual peak-matching, for which the sensitivity is 2–3 times that of exact mass data from the automated system.

For all but the simplest mixtures this procedure finds sequences in addition to those actually present. For example, in Figure 2, the sequence AcGlyValGlyOMe is indicated by a combination of sequence peaks from the components AcGlyAlaLeuOMe and AcValGlyGlyOMe. Other misleading peaks can arise through other mass spectral degradations and impurities. Complementary mass spectral information, as described below, is incorporated to reduce the possible sequences; abundance thresholds (*vide infra*) have been chosen to eliminate peptide components present in less than 5–10% concentration.

Amino Acid Analysis. The amino acids considered for possible sequences are restricted to those whose presence is indicated by conventional amino acid analysis or by characteristic fragment ions. The most important of the latter group are the immonium ions I_i of structure $N^+HR{=}CHR_i$, where R_i is the side chain of the amino acid in question. These have been reported for acylated and esterified derivatives[8,19] ($R = H$); we find that they are also a reliable indication for the presence of a corresponding amino acid residue in permethylated peptide derivatives ($R = CD_3$). Tyrosine and tryptophan are exceptions; fortunately, for these and other aromatic amino acid residues the characteristic benzylic ions originating from the cleavage of the side chain[8,19] (*e.g.*, Tyr gives $HOC_6H_4CH_2^+$) are abundant and so are used as an alternative verification.

Sequence Peaks. The prominent N-terminal sequence ions, whose use is illustrated in Figure 2, arise through cleavage of the peptide chain on either side of a carbonyl group with retention of the charge on the N-terminal fragment (**1**). For our sequencing procedure the presence of both the A_1 and B_1 peaks is required to identify an N-terminal amino acid; the presence of either the A or B peak is a sufficient condition for dipeptide and higher fragments. Retention of charge on the corresponding C-terminal fragments is also possible giving a similar C-terminal ion series; these ions are generally of much lower abundance, and are used only for confirmational purposes.

A different type of cleavage of the main peptide chain can occur adjacent to aromatic amino acids (Phe, His, Tyr, Try)[8,19] to give an odd-electron ion (**2**) through loss of the N-terminal portion of

$$[Ar{-}CH{=}CH{-}CO{-}(NR{-}CHR_i{-}CO)_x{-}OR'']\cdot^+$$

2

the peptide. Such peaks are most abundant when $x = 0$. Fragments corresponding to further decompositions of **2** at the carbonyl groups, analogous to the fragmentation of **1**, can also be observed;[8] we will call these pseudoterminal sequence ions. A separate search of these is made to determine the neighboring amino acids on the C-terminal side of such aromatic residues.

Metastable Transitions (m*). If a so-called metastable peak corresponds to the decomposition of a particular ion to yield another ion, both of these ions must arise, at least in part, *from the same component* of the mixture.[1] Thus for the binary mixture of Figure 2 the presence of metastables corresponding to $289^+ \rightarrow 114^+$, $199^+ \rightarrow 142^+$, $171^+ \rightarrow 142^+$, and $171^+ \rightarrow 114^+$ confirm the presence of the postulated AcValGlyGlyOMe. Such "sequence → sequence" metastables correspond to transitions from one sequence ion (**1**) to any other sequence ion. The most frequently observed, especially in permethylated derivatives, arise from the loss of one complete amino acid residue from an ion of type B ($B_i \rightarrow B_{i-1}$). The program requires the presence of the transition $B_2 \rightarrow B_1$ to confirm a postulated dipeptide fragment.

"Sequence → immonium" metastables involve decomposition of a sequence ion A_k with rearrangement of one hydrogen to yield an immonium ion corresponding to the last amino acid residue (**3**).

$$R'CO{-}(NR{-}CHR_i{-}CO)_{k-1}{-}\overset{+}{N}R{=}CHR_k \longrightarrow \overset{+}{N}HR{=}CHR_k$$

3

This transition is particularly abundant for $k = 1$ (loss of ketene). Our sequencing procedure requires the presence of either $m^*(A_1 \rightarrow I_i)$ or the sequence-sequence transition $B_1 \rightarrow A_1$ as verification of a proposed N-terminal amino acid.[20]

Fractional Vaporization. Direct evaporation in the ion source of a mixture of peptides of differing volatilities requires that a series of spectra be taken to acquire full information on all components. However, these data can help to sort out peaks arising from different components in the mixture (Figure 1).[1] All of the peaks arising from a common component should reach their maximum abundance in the spectrum simultaneously; plotting the logarithm of the absolute intensities of peaks *vs.* probe temperature should yield parallel plots for peaks from a common component. A fractional vaporization plot with more than one maximum indicates that more than one component contributes to that particular peak (Figure 1b). Such data are especially useful to confirm sequences suggested by low-abundance peaks of high mass for which the corresponding metastable transitions cannot be observed.

Results

Mixture 2. A synthetic mixture containing the three tripeptides GlyAlaLeu, ValGlyGly, and MetPheGly was esterified and acetylated only. The preliminary sequencing procedure utilized in this case employs only the high-resolution spectral data, and requires that the molecular ions of each component be present; this procedure can only be used with simple mixtures, such as mixture 1, or with highly sensitive data collection techniques.[2,3] Following the requirement that both sequence ions A_1 and B_1 must be found for N-terminal amino acids, and either A_i or B_i for the others, seven possible sequences were proposed by the computer

(19) K. Heyns and H. F. Grützmacher, *Fortschr. Chem. Forsch.*, **6**, 536 (1966).

(20) The intensities of sequence → sequence metastables can be increased considerably by the introduction of a target gas into the field-free drift region of the mass spectrometer ("collisional activation").[21] This method also appears to be valuable for specific identifications, such as distinguishing leucine and isoleucine,[22] and for ions generated by chemical ionization. This and other techniques of increasing m* ion sensitivities, such as the Barber-Elliott defocusing method, were not employed for these mixtures, as they require a separate measurement step.

(21) R. A. Kornfeld, Ph.D. Thesis, Cornell University, 1971; F. W. McLafferty, P. F. Bente III, R. A. Kornfeld, S.-C. Tsai, and I. Howe, *J. Amer. Chem. Soc.*, **95**, 2120 (1973).

(22) K. Levsen, H.-K. Wipf, and F. W, McLafferty, *Org. Mass Spectrom.*, submitted for publication.

(23) F. W. McLafferty, R. Venkataraghavan, J. E. Coutant, and B. G. Giessner, *Anal. Chem.*, **43**, 967 (1971); F. W. McLafferty, J. A. Michnowicz, R. Venkataraghavan, P. Rogerson, and B. G. Giessner, *ibid.*, **44**, 2282 (1972).

Table I. Computer Postulated Sequence for Mixture 2

Sequences	Sequence ions (A_n, B_n)[a]			$M\cdot^+$, *m/e*
AcGlyOMe	72/100			131.058
AcGlyGlyValOMe	72/100	(129)/157	(228)/256	287.148
AcGlyAlaGlyOMe	72/100	143 /171	200 /228	259.117
AcGlyAlaLeuOMe	72/100	143 /171	256 /284	315.179
AcGlyValGlyOMe	72/100	(171)/199	(228)/256	287.148
AcValGlyGlyOMe	114/142	(171)/199	(228)/256	287.148
AcMetPheGlyOMe	146/174	293/321	(350)/378	409.167

[a] Those ions which appear in parentheses were present in low-resolution spectra but were not found at the correct exact mass in high-resolution spectra.

(Table I). On the basis of these data, the component AcMetPheGlyOMe must be present in the mixture; this sequence is the only one consistent with a molecular ion at *m/e* 409.167 and the five lower mass sequence ions shown. Similarly AcGlyAlaLeuOMe is the only sequence consistent with the molecular ion at *m/e* 315.179 and the sequence peaks at *m/e* 284.161 and 256.166. The peak at *m/e* 131.058 is unique for AcGlyOMe; this was also found in the spectrum of the derivatized pure peptide GlyAlaLeu, and presumably is due to an impurity or alcoholysis during esterification.

serious ambiguities in identifications based solely on low resolution data.[13]

Unimolecular metastables were sought for the sequence → sequence transitions of the proposed peptides (Table II). The three positively indicated tripeptides are confirmed by a number of unique metastables; although several metastables could conceivably originate from one of the other sequences of Table I, not one was found that would be unique for such a sequence.

Thus there is complementary mass spectral evidence for the presence of all of the actual components of the mixture. Although the coincidental presence of the three incorrect sequences of Table I is possible, none is required by the mass spectral data.

Mixture 3. This mixture, consisting of equimolar amounts of AlaPheIleGlyLeuMet, ProPheHisLeuLeu, and ProPheAsp, was esterified, acetylated and perdeuteriomethylated. The composite low- and high-resolution spectral data are given in Table III. The data analysis procedure used, described in the Experimental Section, assumes that spectral information leading to complete sequences will not necessarily have been obtained for each component.

The presence of both sequence peaks A_1 and B_1 and the characteristic ions (immonium and benzyl) indicate

Table II. Sequence → Sequence Metastables Found for Mixture 2

AcGlyAlaLeuOMe		AcValGlyGlyOMe		AcMetPheGlyOMe	
Transition	*m/e*	Transition	*m/e*	Transition	*m/e*
315 → 284[a]	256.1	287 → 228	181.1	409 → 321[a]	251.9
315 → 256[a]	208.1	287 → 171	101.9	409 → 174[a]	74.0
315 → 171[a]	92.8	287 → 114[a]	45.3	378 → 321[a]	272.6
284 → 256[a]	230.8	256 → 199	154.7	378 → 174[a]	80.1
284 → 143[a]	72.0	256 → 171	114.2	350 → 293[a]	245.3
256 → 171	114.2	199 → 142[a]	101.3	350 → 146[a]	60.9
256 → 100	39.1	171 → 142[a]	117.9	321 → 146[a]	66.4
171 → 143	119.6	171 → 114[a]	76.0	293 → 174[a]	103.3
143 → 100	69.9			174 → 146[a]	122.5
100 → 72	51.8				

[a] Unique for this sequence.

A component AcGlyAlaGlyOMe would explain a molecular ion at *m/e* 259.117 and the fragment ions at *m/e* 228.098 and 200.104. However, all these ions can be explained by rearrangement loss of C_4H_8 from the leucine side chain of AcGlyAlaLeuOMe; this is substantiated by a metastable at *m/e* 183.1 corresponding to the transition 284 → 228.

Val is found as an N-terminal amino acid; AcValGlyGlyOMe is the only postulated component that can account for this. The selection of the other two isomers with molecular weight 287.148 is based on peaks that should arise from this peptide[24] and AcGlyAlaLeuOMe.

Fractional vaporization data confirm AcMetPheGlyOMe as the component with the lowest volatility and AcGlyOMe as that of highest volatility; the vaporization rates of the other components are too similar to allow distinction. Note that these five remaining possibilities (Table I) give rise to isobaric sequence ions (*e.g.*, *m/e* 256.130 and 256.166) which would cause Ala, Pro, Leu, and Phe as N-terminal amino acids. However, the required characteristic metastable is found only for Ala and Pro,[25] and so only these are verified as terminal amino acids. Using these there are sequence peaks corresponding in mass to 11 possible dipeptide fragments: AcAlaGly, AcAlaAla, AcAlaPhe, AcAlaMet, AcAlaTyr, AcProGly, AcProSer, AcProMet, AcProAsp, AcProPhe, and AcProAsn. There are no immonium ions for Ser and Asn; the requirement of a sequence → sequence metastable transition $B_2 \rightarrow B_1$ is met only by AcAlaPhe, *m/e* 58.2 (295 → 131), and AcProPhe, *m/e* 64.5 (304 → 140). A search of the combined data set (Table III) employing these dipeptide termini yields ten larger sequences, but the requirement for the observation of a characteristic ion for each residue in the sequence reduces the list to: *AcAlaPhe*GlyLeu, *AcAlaPheLeu*Gly, *AcProPheAsp*-OCD_3, and *AcProPheHis*Leu.

Verification by metastables is not required for tripeptide and larger fragments due to sensitivity consid-

(24) *m/e* 157.061 can be formed from the dipeptide fragment of AcValGlyGlyOMe by rearrangement of C_3H_6 from the valine side chain.

(25) Actually both the $B_1 \rightarrow A_1$ and $A_1 \rightarrow I_1$ transitions are observable: Ala, *m/e* 81.1 (131 → 103) and 36.2 (103 → 61); and Pro, *m/e* 89.7 (140 → 112) and 43.8 (112 → 70).

Table III. Mass Spectral Data for Mixture 3[a]

69.036	97.104	117.074	140.091	170.143	214.136	275.174	396.252
69.059	98.064	118.064	141.073	171.122	216.176	276.179	417.307
69.072	98.079	118.103	141.096	172.131	216.187	276.193	423.298
70.056	99.068	119.075	141.121	173.135	217.186	277.185	424.303
70.063	99.086	119.109	142.078	174.126	219.123	277.210	425.310
71.069	99.013	120.095	142.106	174.134	220.133	277.225	426.315
71.088	100.094	121.087	142.125	174.147	221.139	278.232	427.325
72.078	100.111	121.102	143.089	174.169	222.166	282.213	433.355
73.085	101.080	122.109	143.131	175.141	225.141	289.214	434.362
74.039	101.101	123.116	144.093	175.171	225.154	292.210	447.306
74.064	101.115	124.108	144.137	176.181	227.142	293.181	448.315
74.092	102.049	125.105	145.109	177.112	232.157	293.215	453.269
75.065	102.088	126.113	145.143	177.149	232.171	294.188	474.298
76.081	102.124	128.068	146.095	179.129	233.175	294.221	475.301
77.038	103.049	128.124	146.152	183.132	234.183	295.192	476.302
77.056	103.056	129.075	147.087	184.137	235.179	295.231	477.314
78.015	103.095	129.113	147.103	185.125	235.191	296.200	481.353
78.046	103.129	130.083	147.177	185.139	237.154	297.200	487.306
78.064	104.063	131.049	148.090	185.151	237.182	298.243	498.364
78.081	104.100	131.090	148.114	186.126	238.160	300.233	499.366
79.056	104.136	131.110	148.125	187.080	239.192	301.169	500.367
80.059	105.035	131.126	149.116	187.152	240.199	301.242	501.379
81.046	105.071	132.057	152.090	190.138	242.202	302.164	516.398
81.067	105.093	132.093	153.097	191.128	242.211	302.247	539.454
83.048	105.140	132.116	154.103	191.166	243.225	303.169	547.428
83.071	108.094	132.132	156.122	192.136	244.173	303.238	605.420
83.086	109.100	133.097	157.106	193.105	245.179	304.122	606.428
84.078	110.081	133.129	157.128	196.119	252.151	304.174	
85.051	111.069	134.094	157.143	197.124	252.175	305.175	
85.085	111.087	134.108	158.136	199.175	254.213	306.180	
86.092		135.057	158.154	202.159	255.219	314.185	
87.064	112.076	135.100	159.123	202.186	257.192	315.191	
87.099	112.094	136.062	159.163	203.165	260.155	301.200	
88.050	112.114	136.107	159.176	203.189	261.169	322.218	
89.040	113.080	137.068	160.118	204.102	262.175	322.269	
90.047	113.103	137.114	161.122	204.174	263.181	323.219	
90.066	114.077	138.059	161.140	205.105	264.214	323.264	
91.054	114.088	138.073	163.097	205.182	265.181	334.223	
92.061	114.111	138.120	164.104	206.146	267.198	335.229	
93.070	115.060	139.065	164.158	206.114	268.196	336.234	
94.094	115.097	139.081	165.112	206.188	269.197	337.241	
95.088	115.133	139.124	166.061	207.123	270.204	369.248	
96.046	116.044	139.144	166.176	207.144	272.215	375.231	
96.061	116.054	140.046	169.125	209.158	273.215	391.321	
97.066	117.054	140.071	170.124	212.113	274.186	392.320	

[a] Additional peaks were detected at *m/e* 344, 347, 349, 354, 358, 374, 393, 404, 411, 470, 726, 727, and 729 whose abundance was insufficient for exact mass measurement.

erations; however, confirmatory metastable data are observed for the italicized residues. The observed pseudoterminal sequences are also in agreement: PheHisLeu, PheLeuGly, PheAsp, HisLeu, and PheGlyLeu. Peaks corresponding in mass to the C-terminal sequences PheAspOCD$_3$, LeuLeuOCD$_3$, and MetOCD$_3$ are also present.

The fractional vaporization data from the high-resolution data set are given in Figure 1. All peaks associated with AcProPheAspOMe show a common maximum intensity at 235°. However, the peaks belonging to the partial sequence AcProPhe show a second maximum at about 335°, consistent with a second component with this starting sequence. The other peaks exhibiting maximum intensities at 335° are associated with AcProPheHisLeu, confirming it as a second component. All peaks attributed to the last two possibilities (AcAlaPheLeuGly and AcAlaPheGlyLeu) have their maxima near 330°. AcAlaPheLeuGly must be present because it alone accounts for the peak at *m/e* 425.310. On the other hand, the only unique sequence peak for AcAlaPheGlyLeu (*m/e* 369.249) can be explained by the expected rearrangement loss of C_4H_8 from the leucine side chain of the tripeptide fragment of AcAlaPheLeuGly.

Thus the method has identified correctly the tripeptide sequence plus the first four residues for the two larger peptides (except that it does not differentiate between Leu and Ile). The method could not eliminate the possibility of one additional sequence (AcAlaPheGlyLeu), as the data used for its identification is expected from another component. Although the two larger peptides are poorly separated by fractional vaporization, the high-resolution data have minimized the ambiguity in identifying them.

Mixture 4. A mixture containing the four peptides ThrLysTyr, TryGlyGly, PheAspAlaSerVal, and ValAlaIlePheAsn was acetylated and perdeuteriomethylated. The computer search of the combined high-resolution spectral data for N-terminal amino acids gives Gly, Phe, Thr and Thr($-H_2O$), Try, and Val; all except Gly are verified by metastables. The search for corresponding dipeptide fragments yields AcThrPro, AcThrLys and AcThr($-H_2O$)Lys, AcThr($-H_2O$)Gln, AcPheAsp, AcTryGly, and AcValAla. No characteristic ions for Pro and Gln are observed,

eliminating their dipeptides; all the remainder except AcThrLys are confirmed by sequence → sequence metastables. Extension of the search based on the confirmed dipeptide fragments indicates AcThr($-H_2O$)Lys, AcPheAspAlaSer, AcTryGly, AcValAlaLeuPhe, AcValAlaGlu, AcValAlaThr, and AcValAlaThr($-H_2O$). There is no immonium ion for Glu, eliminating AcValAlaGlu. Note that the exact mass of AcThr($-H_2O$)Lys is identical with that of AcValAlaThr($-H_2O$) due to the identity in elemental composition of Lys and (Val + Ala). A choice can be made between these possibilities[26] based on the fractional vaporization data. All peaks from AcThrLys show their maxima at 220°, while the peak AcValAla as well as all other peaks from AcValAlaLeuPhe give maxima at 250°. The remaining possible sequences *AcThrLys*, *AcTryGly*, *AcPheAspAla*Ser, and *AcValAla*LeuPhe are therefore confirmed. Unimolecular metastable transitions are present which correspond to the italicized residues. C-Terminal peaks corresponding to Gly-OMe and Val-OMe are found, but no molecular ion or sequence peaks were of sufficient intensity to complete any of the sequence assignments. The method has given sequence data on all components for all but the C-terminal amino acids, and no additional or incorrect assignments have been proposed. These results illustrate our general observation that samples which have been acetylated and perdeuteriomethylated give fewer additional assignments than those derivitized by our earlier procedures.[1]

Chemical Ionization (CI). Recently, two laboratories have utilized CI to obtain low-resolution mass spectra of pure peptides, and cite advantages over electron impact ionization (EI) of increased sensitivity for higher mass ions and especially for C-terminal sequence peaks of the type H_2-(NR-CHR$_i$-CO)$_n$OR''.[27] A combination CI–EI ion source was recently installed on the high resolution instrument used in this work. Exact mass data obtained on <0.1 μmol of mixture 4 using CI supported these conclusions. Although $(M + 1)^+$ ions were not found, ions corresponding in elemental composition to the following fragments (among others) were identified: AcThrLys, H_2LysTyrOMe, AcTryGly, H_2GlyGlyOMe, AcPheAspAlaSerVal, H_2AspAlaSerValOMe, AcValAlaLeuPheAsn, and H_2LeuPheAsnOMe. These data obviously complete the identifications of the two pentapeptide components, and give valuable new information on the identities of the other components.

Applications to Unknown Mixtures. In our experience the specificity of the high-resolution and metastable ion data is especially valuable when unexpected numbers or kinds of components are present, a not uncommon situation for actual unknowns. For example, mixture 5 which supposedly contained GluValPhe, TryGlyGly, ThrLysTyr, and PheAspAlaSerVal was acetylated, esterified, and perdeuteriomethylated. Peaks at more than 500 masses are present in the resulting high-resolution spectra. At the dipeptide level alone ten possibilities are found by our method: AcThr($-H_2O$)Lys, AcThr($-H_2O$)Gly, AcValPhe, AcValSer($-H_2O$), AcThrTyr, AcTryGly, AcPheGlu, AcPheAsp, AcPheThr, and AcPheThr($-H_2O$). To see if any of these possibilities were due to coincident peaks of nonpeptide impurities, an alternative derivative of the same mixture (deuterioacetylated instead of acetylated) was prepared. The mass spectral data of this sample give eight dipeptide possibilities: AcThr($-H_2O$)Lys, AcValSer, AcValPhe, AcThrPhe, AcGluVal, AcGluGlu, AcPheAsp, and AcTryGly. Only four of the possibilities are found from both data sets; the remainder must be due to peptides present in marginal concentration (*ca.* 5%) or to nonpeptide impurities. The complete analysis of both derivatives gives the following common peptide fragments: AcThr($-H_2O$)Lys, AcValPheValPhe, AcPheAspAlaSer, and AcTryGly, three of which correspond to the expected components. Evidence for AcGluValPhe, corresponding to the fourth supposed component, is found, but is of marginal intensity and confirmed only in the second derivative, while the indicated component AcValPheValPhe is unexpected. Examination of the supposedly pure peptide GluValPhe by thin layer chromatography demonstrated the presence of at least seven components, the major one of which was identified as dicyclohexylurea. Two other spots corresponded to peptides (ninhydrin-positive) and the rest could not be identified. An N-terminal amino acid analysis[28] of this sample showed the presence of glutamic acid and valine, supporting the identifications of AcGluValPhe and of AcValPheValPhe.

Conclusion

The specificity and variety of mass spectral data which can be obtained from peptide mixtures yield reliable and extensive, although not necessarily complete, information on the amino acid sequences of components of peptide mixtures. For example, nearly complete information was obtained on mixture 4 which contained four components and 12 different amino acids; useful information was obtained from mixture 5 on components present in 5–10% abundance in the presence of at least five other compounds.

The operation of the derivatization and data collection procedures used is amenable to routine analysis. Although chemical ionization and automated high-resolution capabilities are not widely available, this situation is improving rapidly. It should be emphasized that the various types of mass spectral data used are complementary to a substantial extent, greatly increasing the reliability of the analytical conclusions. The extent of data collection efforts should thus depend on the complexity of the mixture; in difficult cases it may be necessary to obtain data from additional derivatives (we find the substitution of deuterioacetylation for acetylation especially helpful), from a quantitative Edman degradation[5] on the same mixture,[29] or from defocused or collisional-activation metastables.[21] We feel that the specificity of high-resolution data is especially valuable for complex mixtures. In our ex-

(26) These can also be distinguished by using the trideuterioacetyl derivatives, for which the elemental compositions are not identical.

(27) W. R. Gray, L. H. Wojcik, and J. H. Futrell, *Biochem. Biophys. Res. Commun.*, **41**, 1111 (1970); A. A. Kiryushkin, H. M. Fales, T. Axenrod, E. J. Gilbert, and G. W. A. Milne, *Org. Mass Spectrom.*, **5**, 19 (1971).

(28) We are indebted to Dr. B. A. Cunningham and Mr. J. Wang of Rockefeller University for this analysis.

(29) As an example, the Edman degradation of a mixture might give the following residues in successive positions: (A, B, G); (B, E, F); (A, C, G); (D); (E); (F). Even if mass spectrometry gives only the partial sequences ABCD, BEG and GF the correct and complete sequence information ABCDEF, BEG and GFA would be obvious.

perience the spectra of these can contain 300–600 peaks, making the identification of low molecular weight peptides from low-resolution spectra especially difficult. A further disadvantage of low resolution data of complex mixtures is that its interpretation must depend heavily on predicted relative abundances as well as masses. We also find that computer interpretation techniques are necessary for the data from complex mixtures to be certain that all sequence possibilities are considered; even for high-resolution data unexpected coincidences can arise, such as the exact mass identity cited in mixture 4 for derivatized Lys and (Val + Ala). For a thorough analysis every peak in the spectrum must be considered for its possible structural significance; this is a relatively trivial assignment for the modern computer.

This capability of analyzing mixtures makes mass spectrometry a more attractive tool in polypeptide sequencing by reducing the number of chemical derivatization operations necessary and the required quantities of individual oligopeptides. Techniques for routine chemical derivatization of smaller samples are needed; mixture samples of $\ll 0.1$ μmol should be sufficient using chemical ionization[27] and/or improved methods for collection of high-resolution data,[23] and would represent a substantial advantage over present chemical methods. Chemical ionization also promises to make possible more complete sequence information on larger peptides. Finally, this method should make possible entirely new strategies for polypeptide sequencing,[6] with polypeptide degradations and separations tailored to give the most suitable oligopeptide mixtures.

Acknowledgments. We are very grateful to Professors V. du Vigneaud and G. M. Edelman for extensive consultation, to Drs. J. Hribar and J. W. Serum of the Cornell University Mass Spectrometry Facility (NIH Grant RR355) for help in obtaining high-resolution mass spectra, and to the U. S. Public Health Service (NIH Grant GM16609) for generous support of this research.

Nuclear Magnetic Resonance Study of the Binding of Glycine Derivatives to Hemocyanin

C. H. Ke,*[1a] J. Schubert,[1a] C. I. Lin,[1b] and N. C. Li[1b]

Contribution from the Department of Radiation Health, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, and the Department of Chemistry, Duquesne University, Pittsburgh, Pennsylvania 15219. *Received July 11, 1972*

Abstract: Hemocyanin is a non-heme, copper-containing oxygen-carrying protein found in many molluscs and arthropods. The existence of cupric ions in oxyhemocyanin has not been proved directly, but is inferred from results of physicochemical studies. Optical absorption spectra do not indicate any interaction between oxyhemocyanin and added ligands. We have utilized nuclear magnetic resonance line broadening techniques to investigate the interactions of glycine derivatives with the copper of oxyhemocyanin. Evidence is provided which demonstrates directly the presence of paramagnetic Cu(II) ion in oxyhemocyanin and the binding of ligands to the protein-bound Cu(II) at the active sites. The line-width measurements obtained with glycine, glycine ethyl ester, acetylglycine, and glycylglycine lead us, depending on the ligand, to conclude that the amino, carboxyl, and amide groups of the ligand are involved in the binding of cupric ions of oxyhemocyanin with formation constants ranging from 2 to 1500 M^{-1}.

Hemocyanin is a non-heme, oxygen-carrying copper protein found in many molluscs and arthropods. One oxygen molecule binds two copper atoms in hemocyanin.[2] The protein, a deep blue color when oxygenated, becomes colorless when deoxygenated. It appears that the copper in deoxyhemocyanin is Cu(I), since only cuprous compounds regenerate active hemocyanin from apohemocyanin.[3] But, the oxidation state of copper in oxyhemocyanin is uncertain.[4] Oxyhemocyanin does not display a Cu(II) electron spin resonance (esr) signal unless the protein is denatured.[5,6] The existence of paramagnetic Cu(II) in oxyhemocyanin is inferred from the absorption spectra,[7,8] chemical reactions,[9,10] and radiation chemical studies.[11] The presence of Cu(II) in oxyhemocyanin has never been proved directly.

(1) (a) University of Pittsburgh; (b) Duquesne University.
(2) A. C. Redfield, T. Coolidge, and H. Montgomery, *J. Biol. Chem.*, **76**, 197 (1928).
(3) (a) F. Kubowitz, *Biochem. Z.*, **299**, 32 (1938); (b) R. Lontie, V. Blaton, M. Albert, and B. Peeters, *Arch. Int. Physiol. Biochim.*, **73**, 150 (1964).
(4) G. Morpurgo and R. J. P. Williams, "Physiology and Biochemistry of Hemocyanin," G. Ghiretti, Ed., Academic Press, New York, N. Y., 1968, p 113.
(5) T. Nakamura and H. S. Mason, *Biochem. Biophys. Res. Commun.*, **3**, 297 (1960).
(6) C. Manwell, "Oxygen in the Animal Organism," F. Dickens and E. Niel, Ed., Macmillan, New York, N. Y., 1964, p 49.
(7) E. Frieden, S. Osaki, and H. Kobayashi, *J. Gen. Physiol.*, **49**, No. 1, Part 2, 213 (1965).
(8) K. E. Van Holde, *Biochemistry*, **6**, 93 (1967).
(9) G. Felsenfeld and M. P. Printz, *J. Amer. Chem. Soc.*, **81**, 6259 (1959).
(10) R. Lontie and R. Witters, "The Biochemistry of Copper," J. Peisach, A. Aisen, and W. E. Blumberg, Ed., Academic Press, New York, N. Y., 1966, p 455.
(11) (a) J. Schubert, E. R. White, and L. F. Becker, Jr., *Advan. Chem. Ser.*, **No. 81**, 480 (1968); (b) C. H. Ke and J. Schubert, *Radiat. Res.*, **49**, 507 (1972).