Vol. 167, No. 2, 1990 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
March 16, 1990 Pages 686-692

ELECTROSPRAY IONIZATION MASS SPECTROMETRIC PEPTIDE MAPPING: A RAPID, SENSITIVE TECHNIQUE FOR PROTEIN STRUCTURE ANALYSIS

S.K. Chowdhury, V. Katta, and B.T. Chait

The Rockefeller University, 1230 York Avenue, New York, NY 10021

Received January 26, 1990

SUMMARY: Electrospray ionization mass spectrometric peptide mapping is demonstrated to be a useful new technique for protein structure analysis. The procedure involves the digestion of the protein with trypsin and subsequent analysis of the total unfractionated digest by electrospray ionization mass spectrometry. The utility of the technique for investigating protein structure is illustrated by a peptide mapping analysis of human apolipoprotein AI (M_r = 28 kDa). The technique is rapid, sensitive, and requires no prior separation of the peptides. The discrimination effects observed in other mass spectrometric methods are less important in the present procedure.

Mass spectrometric peptide mapping has become an established and powerful structural tool for the analysis of proteins (1). The procedure involves the enzymatic or chemical degradation of the protein into small peptides, followed by direct mass spectrometric molecular weight determination of the resulting peptides (2-7). The measured molecular weights of the peptides are then compared with the molecular weights calculated for a proposed protein primary structure. Agreement between the observed and the calculated molecular weights provides a rapid verification of the protein structure. Conversely, disagreement indicates a possible error or modification in the proposed structure and provides information concerning the location of the error or modification. Mass spectrometric peptide mapping has been used for the verification and elucidation of primary structures of proteins deduced from their DNA sequences and the identification of sites of variation in mutants and post-translationally modified proteins (1). Two mass spectrometric ionization methods, fast atom bombardment and ^{252}Cf plasma desorption, have been used for protein analysis by this procedure. Naylor et al (5) have shown that in fast atom bombardment mass spectrometric analysis of peptide mixtures, a

0006-291X/90 $1.50

selective desorption and ionization of hydrophobic peptides occurs together with a suppression of hydrophilic components. Such discrimination effects have been mitigated by partial fractionation (3) or chemical derivatization of the peptides in the mixture (5). By contrast, Neilsen and Roepstorff (6) observed that in positive ion ^{252}Cf plasma desorption analysis of tryptic digests, peptides carrying net positive charges are desorbed preferentially from nitrocellulose surfaces.

In this Communication, we describe the utilization of electrospray ionization mass spectrometry (8) for the peptide mapping of proteins. As an example, we present a detailed peptide mapping analysis of human apolipoprotein AI, a 28 kDa protein. The procedure involves complete or near complete digestion of the protein with trypsin and mass spectrometric determination of the molecular weights of the resulting peptide fragments without any chromatographic fractionation or chemical derivatization.

MATERIALS: Human apolipoprotein AI was provided by Drs. J. Breslow, E. Brinton, and Y. Ito (Rockefeller University). The sample was found (9) to be pure with an observed MW 28,078.1±0.8 Da in close agreement with that calculated (28078.6 Da) from the published sequence (10). Bovine trypsin, horse heart cytochrome C, and bovine heart cytochrome C were obtained from Sigma Chemical Co. (St. Louis, MO).

METHODS: Digestion With Trypsin: A 5 pmol/ul solution of trypsin was prepared by dissolving the enzyme in 0.06M NH_4HCO_3 solution (pH=8.7). The protein sample was then added to 20 ul of the enzyme solution with an enzyme substrate ratio of 1:25. The digestion was carried out for 12 hours at 37^0C. The resulting tryptic peptide fragment mixture was lyophilized and dissolved in a solution containing methanol, water and acetic acid with a ratio 47.5:47.5:5.0 v/v, to obtain a concentration of 20 pmol/ul of each component for mass spectrometric analysis.
Mass Spectrometric Analysis: The electrospray ionization mass spectrometer and the procedures adopted to acquire mass spectra have been described previously (9). Briefly, the sample solution was electrosprayed from a syringe needle in ambient air at a rate of 0.6 ul/min. The resulting ionized species were transported into the vacuum of the quadrupole mass spectrometer through a 203 mm long, heated metal capillary tube for mass-to-charge ratio (m/z) analysis. The spectra were acquired using a commercial data system Vector-1 (Teknivent, St. Louis, MO) on an IBM AT compatible computer. The data system records ion abundances at integer m/z values and, thus, produces data of limited accuracy. In order to obtain more accurate m/z values, the centroids of the peaks of interest were determined by scanning a narrow range of m/z values (typically 5-20) in the so called "calibration mode" (9).

RESULTS AND DISCUSSION: The electrospray ionization mass spectrum of a tryptic digest of human apolipoprotein AI is shown in Fig. 1. The spectrum,

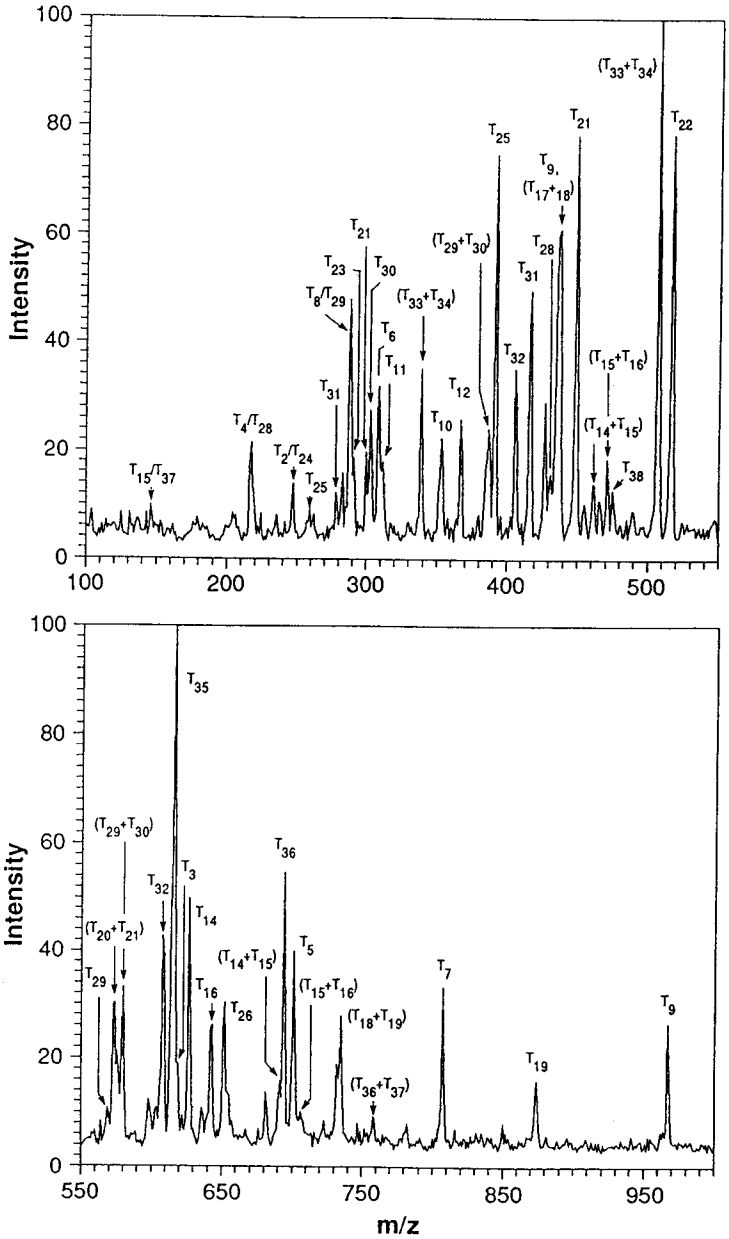

Fig. 1. Electrospray ionization mass spectrum between m/z 100 and 1100 of the unfractionated tryptic digestion products from human apolipoprotein AI. The spectrum is an average of five scans, each acquired in 88 secs. The electrospray of the digest solution (water : methanol : acetic acid = 47.5 : 47.5 : 5.0) was carried out at a rate of 0.6 ul/min. Total sample consumed = 88 pmol. T_i designates the ions of the tryptic peptide component i (Table 1).

shown between m/z 100 and 1100, is an average of five scans, each acquired in 88 secs. The solution of the peptide mixture (20 pmol/ul of each component), was electrosprayed at a rate of 0.6 ul/min. Thus 88 pmol of

sample was consumed in acquiring the spectrum. The spectrum shows the presence of a large number of ion peaks, each corresponding to a tryptic peptide. The labels, T_i, identify the tryptic peptides whose sequences are given in the second column of Table 1. Also given in Table 1 are the calculated molecular weights and the observed m/z values of the tryptic peptides.

In general, the tryptic peptides were observed as doubly charged ions, designated $(M+2H)^{2+}$ in Table 1, where M is the peptide molecule. Since the peptides obtained from tryptic digestion have a free amino-terminus and a basic amino acid residue (either arginine or lysine) at the carboxy-terminus, the attachment of two protons to the two basic sites is expected in electrospray ionization (8,9,11,12). One exception is the carboxy-terminal tryptic peptide of the protein that may not contain a basic residue at its carboxy-terminus. A second important exception concerns peptides containing a histidine residue (T_{21}, T_{25}, T_{26}, T_{31}, and T_{32}) which produce both doubly and triply protonated ions (Table 1). Peptide fragments resulting from incomplete tryptic digestion having an internal lysine or arginine (e.g. T_{14}+T_{15}, T_{29}+T_{30}, and T_{33}+T_{34}) also produced both doubly and triply charged ions. Small peptides consisting of 1 or 2 amino acid residues, generally produce only low intensity singly charged ions.

Ions resulting from all the tryptic fragments of apolipoprotein AI were observed (Fig. 1 and Table 1) with the exception of a few mono- and di-peptides and the fragment T_{18}. However, an intense doubly protonated ion resulting from incomplete tryptic cleavage between T_{17} and T_{18} (designated T_{17}+T_{18} in Table 1) confirms the presence of T_{18}. Incomplete digestion products of this type provide additional information on the structure of the protein. For example, in addition to the detection of ions corresponding to the individual fragments T_{14} and T_{15}, ions resulting from the incompletely cleaved peptide T_{14}+T_{15} were also observed.

The majority of ions were observed to have m/z values within $\pm$0.2 of the corresponding calculated values with the exception of the small, low intensity singly charged ions. The observed m/z values of these low intensity ions agree to within $\pm$0.8 of the calculated values. Partial overlap of ion peaks results from

Table 1. Sequences, calculated molecular weights[a], and observed m/z values (from Fig. 1) of tryptic peptides from human apolipoprotein AI

Fragment	Sequence	Calculated molecular weight	Observed m/z values $(M+H)^+$	$(M+2H)^{2+}$	$(M+3H)^{3+}$	Fragment	Sequence	Calculated molecular weight	Observed m/z values $(M+H)^+$	$(M+2H)^{2+}$	$(M+3H)^{3+}$
T_1 (1-10)	DEPPQSPWDR	1226.3		614.1		T_{25} (154-160)	AHVDALR	781.9		391.8	262.6
T_2 (11-12)	VK	245.3	246.8			T_{26} (160-171)	THLAPYSDELR	1302.4		651.5	435.7
T_3 (13-23)	DLATVYVDVLK	1235.4		618.7		T_{27} (172-173)	QR	303.3	d		
T_4 (24-27)	DSGR	433.4		217.0[b]		T_{28} (174-177)	LAAR	430.5		217.0[b]	
T_5 (28-40)	DYVSQFEGSALGK	1400.5		701.1		T_{29} (178-182)	LEALK	572.7		287.9	
T_6 (41-45)	QLNLK	614.7		308.9		T_{30} (183-188)	ENGGAR	603.6		303.2[d]	
T_7 (46-59)	LLDNWDSVTSTFSK	1612.7		807.1		T_{31} (189-195)	LAEYHAK	830.9		416.6	278.6
T_8 (60-61)	LR	287.4	287.9			T_{32} (196-206)	ATEHLSTLSEK	1215.3		608.5	406.3
T_9 (62-77)	EQLGPVTQEFWDNLEK	1933.1		967.4		T_{33} (207-208)	AK	217.3	219.0[b]		
T_{10} (78-83)	ETEGLR	703.7		352.9		T_{34} (209-215)	PALEDLR	813.91			
T_{11} (84-88)	QEMSK	621.7		312.1		T_{35} (216-226)	QGLLPVLESFK	1230.5		616.1	
T_{12} (89-94)	DLEEVK	731.8		367.1		T_{36} (227-238)	VSFLSALEEYTK	1386.6		694.2	
T_{13} (95-96)	AK	217.3	219.0[b]			T_{37} (239)	K	146.2		c	
T_{14} (97-106)	VQPYLDDFQK	1252.4		627.1		T_{38} (240-243)	LNTQ	474.5	475.4		
T_{15} (107)	K	146.2	c			$T_{14} + T_{15}$		1380.6		691.2	461.5
T_{16} (108-116)	WQEEMELYR	1283.4		642.6		$T_{15} + T_{16}$		1411.5		706.6	471.3
T_{17} (117-118)	QK	274.3	c			$T_{17} + T_{18}$		869.0		435.8	e
T_{18} (119-123)	VEPLR	612.7	c			$T_{20} + T_{21}$		1152.3		577.0	385.3
T_{19} (124-131)	AELQEGAR	872.9		437.3		$T_{29} + T_{30}$		1158.3		580.2	387.2
T_{20} (132-133)	QK	274.3	c			$T_{33} + T_{34}$		1013.2		507.0	338.9
T_{21} (134-140)	LHELQEK	896.0		449.2	300.0	$T_{36} + T_{37}$		1514.7		758.4	e
T_{22} (141-149)	LSPLGEEMR	1031.2		516.6							
T_{23} (150-151)	DR	289.3	291.2								
T_{24} (152-153)	AR	245.3	246.8								

a Average molecular weight is calculated from the natural isotopic abundances.
b Peaks corresponding to the $(M+2H)^{2+}$ ions from T_4 and T_{28} were not resolved.
c Protonated ions from these peptides were not observed. However, ions corresponding to larger incompletely cleaved peptides containing these residues were observed.
d The $(M+H)^+$ ion of peptide T_{27} could not be resolved from the $(M+2H)^{2+}$ ion of peptide T_{30}.
e The peaks were not resolved completely.

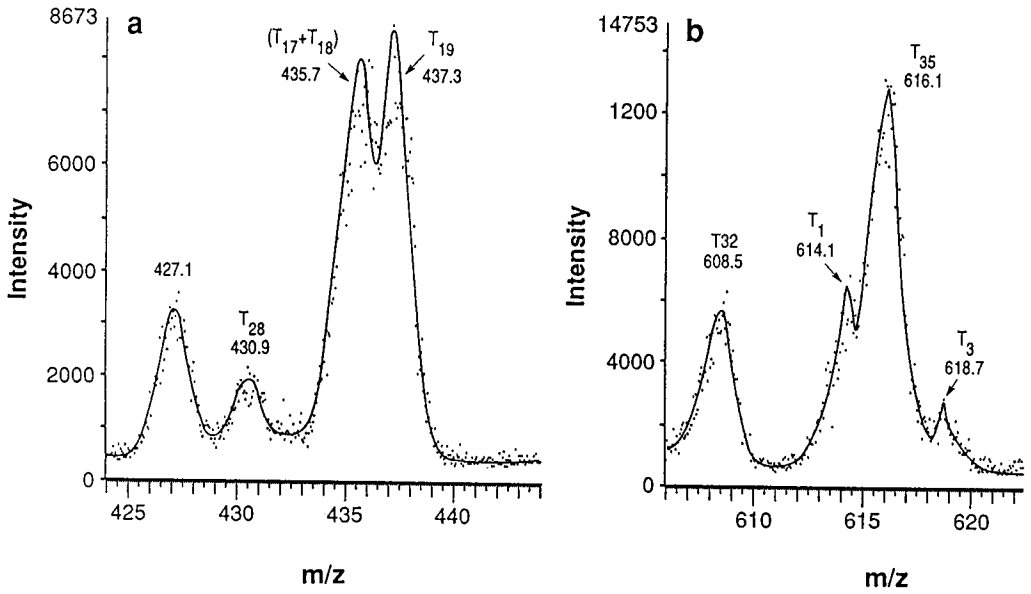

Fig. 2. Detailed mass spectra of two small regions of the mass spectrum shown in Fig. 1, where the close proximity of the m/z values resulted in unresolved peaks. These two spectra were obtained by scanning a narrow region in the "calibration mode" of the data system (9). Each spectrum was acquired in 30 secs.

ions originating from different peptides that are separated by only 1 or 2 m/z units. In such cases, a slow scan over a narrow region (10-20 m/z unit) in the "calibration mode" of the data acquisition system is obtained (9) and the centroids of the incompletely resolved peaks can be determined. For example, the congested regions near m/z 435 and 616 (Fig. 1) were redetermined by the latter procedure. The detailed mass spectra of these two regions are shown in Fig. 2. Although a complete separation of the ion peaks was not obtained, all the components were observed and their centroids determined (Table 1).

We have also utilized the described procedure for the analysis of several other proteins. In general, ions from all tryptic fragments were observed, with the exception of some small peptides with MW <400 Da. Thus, for example, the mass spectra obtained from the tryptic digest of horse heart cytochrome C and bovine heart cytochrome C clearly identify the 3 amino acid residues that differ between these two proteins (13). The discrimination effects observed in other mass spectrometric mapping techniques appear to be less serious in the present method. This procedure is simple, rapid, sensitive and can be utilized

without any prior fractionation using a relatively inexpensive single quadrupole mass analyzer.

ACKNOWLEDGMENTS: This work was supported in part by grants RR00862 and RR07065, Division of Research Resources, NIH and GM38274, National Institutes of general Sciences, NIH.

REFERENCES

1. Chowdhury, S.K. and Chait, B.T. (1989) in Annual Reports in Medicinal Chemistry - 24 (Allen, R.C., ed.), pp. 253-263, Academic press, New York.
2. Morris, H.R., Panico, M., and Taylor, W. (1983) Biochem. Biophys. Res. Commun. 259, 299-305.
3. Gibson, B.W. and Biemann, K. (1984) Proc. Natl. Acad. Sci. USA 81, 1956-1960.
4. Takato, T., Hitouji, T., Shimonishi, Y., Tanake, T., Inouye, S., and Inouye, M. (1984) J. Biol. Chem. 259, 6105-6109.
5. Naylor, S., Findeis, A.F., Gibson, B.W., and Williams, D.H. (1986) J. Am. Chem. Soc. 108, 6359-6363.
6. Neilsen, P.F. and Roepstorff, P. (1989) Biomed. Environ. Mass Spectrom. 18, 131-137.
7. Chait, B.T., Chaudhary, T., and Field, F.H. (1987) in Methods in Protein Sequence Analysis - 1986 (Walsh, K.A., ed.), pp 483-492, The Humana Press, NJ.
8. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989) Science 246, 64-71.
9. Chowdhury, S.K., Katta, V., and Chait, B.T. (1990) Rapid Commun. Mass Spectrom. vol 4, in press.
10. Protein Identification Resource (PIR), National Biomedical Research Foundation (NBRF) Release 19.0 (3187 December 1988).
11. Covey, T.R., Bonner, R.F., Shushan, B.I., and Henion, J.D. (1988) Rapid Commun. Mass Spectrom. 2, 249-256
12. Loo. J.A., Udseth, H.R., and Smith, R.D., (1989) Anal. Biochem. 179, 404-412.
13. Chowdhury, S.K., Katta, V., and Chait, B.T. (in preparation).