Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 6873–6877, September 1990
Chemistry

Rapid, sensitive analysis of protein mixtures by mass spectrometry

(sinapinic acid/laser desorption/apolipoproteins/milk)

RONALD C. BEAVIS AND BRIAN T. CHAIT

Department of Mass Spectrometry and Gas Phase Ion Chemistry, The Rockefeller University, New York, NY 10021

Communicated by F. W. McLafferty, June 15, 1990 (received for review May 22, 1990)

ABSTRACT **We have developed a method for determining the molecular masses of proteins in complex mixtures by mass spectrometry. The method has the capacity to examine the components of mixtures without using any chromatographic separation steps and will tolerate relatively large amounts of buffers and inorganic contaminants. It allows the simultaneous determination of protein molecular masses from 1 to 40 kDa with an accuracy of ±0.01% and above 40 kDa with reduced accuracy. The lower limit for practical detection of a protein is a concentration of ≈0.1 μM, and <1 μl of such a solution is consumed. The analysis is very fast: <15 min is necessary to perform the complete analysis, including sample preparation, introduction into the mass spectrometer, mass spectrum collection, and data reduction. The mass spectrum that is obtained does not require elaborate interpretation because there is no fragmentation of the ionized protein (or protein subunit) molecule. Therefore, there is a one-to-one correspondence between the peaks in the mass spectrum and the proteins present in the original mixture. The spectra assume the appearance of chromatograms, with the abscissa being mass-to-charge ratio rather than chromatographic retention time.**

Probably the most useful characterization technique currently available for identifying proteins in mixtures is SDS/PAGE. This method will separate a complex mixture of proteins into a series of bands on a polyacrylamide gel that can be visualized and studied by a variety of techniques. A characteristic of SDS/PAGE is that the position of a band on a gel can be correlated to the molecular mass of either a protein or a protein subunit (1, 2). This correlation of migration distance with molecular mass has been very useful in protein analysis, providing an estimate of the number of amino acid residues in an unknown molecule. Although the accuracy of the molecular mass measurement is commonly assumed to be 5–10%, certain classes of proteins [e.g., glycoproteins (3)] are known to migrate differently than the standard proteins used as mass calibration markers.

The introduction of new ionization techniques such as ^{252}Cf plasma desorption (4, 5), fast atom bombardment (6, 7), electrospray ionization (8–11), and matrix-assisted laser desorption (12–16) has raised the possibility of using mass spectrometry as an alternative technique for determining the molecular masses of proteins (or protein subunits) in complex mixtures with a much higher accuracy than SDS/PAGE. While mass spectrometry does not provide any separation of components in a preparative sense, it does measure the masses of isolated gas-phase protein ions directly by their behavior in electromagnetic fields. The isolation of individual protein molecules in the gas phase removes effects caused by complex solution-phase behavior of proteins, such as aggregation and changes in conformation, that can alter the migration of a band in SDS/PAGE. The application of mass spectrometry to the analysis of protein mixtures has, so far, been limited to electrospray ionization measurements of a mixture of purified bovine and porcine insulin (17) and a mixture of ubiquitin and thioredoxin (9).

This paper reports the development of a mass spectrometric method for determining the molecular masses of proteins in crude or semipurified mixtures. The type of mass spectrometric analysis used is based on the matrix-assisted laser desorption technique, originally described by Karas and Hillenkamp (12–15). This technique involves mixing a small amount (≈1 pmol) of a protein with a 10^4-fold molar excess of a "matrix" material in solution and then drying the solution to form a deposit on a metal substrate. This deposit is then loaded into a time-of-flight mass spectrometer and examined using a laser pulse (10^7 W/cm^2, λ = 266 nm, τ = 10 nsec). The matrix material used by Karas and Hillenkamp is nicotinic acid, which absorbs the laser radiation and volatilizes, taking the intact polypeptide chains with it into the gas phase. Intact, ionized proteins with molecular masses in excess of 100 kDa have been observed.

We have modified the method of Karas and Hillenkamp and have improved its mass resolution, mass accuracy, and reproducibility by the use of a simplified time-of-flight analyzer (18), new matrix materials (in particular 3,5-dimethoxy, 4-hydroxy-*trans*-cinnamic acid, sinapinic acid) (19), and a different laser wavelength (355 nm) (20). The use of the new matrix material, sinapinic acid, has proved to be beneficial in several important respects. Sinapinic acid has the property of being able to selectively ionize protein molecules in the presence of high concentrations of contaminating materials, such as lipids, carbohydrates, and salts. It is also relatively nonselective in its behavior towards proteins of radically different primary structures and modifications (such as phosphorylation and glycosylation). The general affinity of this matrix for proteins, coupled with its ability to tolerate high concentrations of ionic contaminants, allows the application of mass spectrometry to mixtures of a type that has been completely inaccessible to mass spectrometry previously, i.e., the protein components of crude biological extracts.

EXPERIMENTAL PROCEDURES

The linear time-of-flight mass spectrometer and laser desorption ion source were both constructed at Rockefeller University and have been described (18, 19). Briefly, the frequency tripled output of a Q-switched Lumonics HY400 neodymium/yttrium aluminium garnet laser (355 nm, 10-nsec output pulse) was focused by a lens (12-inch focal length) through a fused silica window onto a sample inside the mass spectrometer. The product ions formed by the laser were accelerated by a static electric potential of 30 kV. The ions then drifted down a 2-m tube maintained at a vacuum of 30 μPa and their arrival at the end of the tube was detected and recorded using a Lecroy TR8828D transient recorder. The transient records of up to 200 individual laser shots were summed together and the resulting histogram was plotted as a mass spectrum. Peak centroid determinations and data reduction were performed using a VAX workstation.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Protein samples were prepared for laser desorption analysis by a single standard procedure. The laser desorption matrix material (sinapinic acid) was dissolved in aqueous 30% (vol/vol) acetonitrile containing 0.1% (wt/wt) trifluoroacetic acid, to make a standard solution at 20°C (≈50 mM). A solution containing the protein sample of interest was then added to the matrix solution to give a final protein concentration of 0.1–10 μM. A small aliquot (0.5 μl) of this mixture was then applied to a flat metal probe tip (2-mm diameter) and dried at room temperature with forced air. The resulting deposit was washed in 4°C distilled water by immersing the tip for 10 sec. This washing step aided the removal of soluble ionic contaminants from the protein/matrix deposit, without removing the proteins or matrix. Once the sample was washed, it was inserted into the mass spectrometer and analyzed. The entire protocol, from the beginning of sample preparation to finished mass spectral analysis, took ≈15 min.

The protein standards used for Fig. 1 were all purchased from Sigma and were used without further purification. The high density lipoprotein fraction of human plasma for Fig. 2 and the apolipoprotein AI used for Fig. 4 were obtained as a gift from J. Breslow (Rockefeller University). The bovine milk used for Fig. 3*A* was pasteurized, homogenized commercial dairy milk obtained from a local supermarket. The human milk used for Fig. 3*B* was a gift and obtained from a subject 3 months postpartum.

The cyanogen bromide digests shown in Fig. 4 were obtained by two different procedures. The nearly complete digest shown in Fig. 4*A* was made using a 30 μM solution of human apolipoprotein AI dissolved in aqueous 0.1 M HCl. A molar excess of cyanogen bromide was added to 50 μl of the solution and the reaction was allowed to proceed in the dark for 2 hr. The reaction mixture was then vacuum-evaporated and redissolved in 100 μl of aqueous 0.1% trifluoroacetic acid. The digest in Fig. 4*B* was performed using a 30 μM solution of protein in aqueous 70% formic acid. After the addition of cyanogen bromide, the reaction vial was immediately placed into the vacuum evaporator and dried.

RESULTS AND DISCUSSION

The most general result of the use of the simple protocol described above is that the time-of-flight mass spectra obtained from mixtures of proteins allowed the molecular mass determination of each of the components of the mixture. The spectra obtained were easy to interpret because no fragmentation of proteins into smaller fragment ions was observed. The lack of fragmentation includes disulfide bonds, which were not observed to be broken during the desorption process. Therefore, there is a one-to-one correspondence between the protonated molecule ions present in the mass spectrum and the molecular species originally found in the mixture.

This relatively nonselective detection of proteins in a complex mixture is a feature of the use of cinnamic acid derivatives as matrices. It is our experience that the laser desorption matrix originally described by Karas and Hillenkamp (12–15), nicotinic acid, does not produce uniform responses from mixtures of proteins. The presence of some proteins in the original solution strongly suppresses the signals from other components. These suppression effects lead to spectra dominated by only a few of the species present. It is reasonable to assume that while the matrix/protein solution is drying, the matrix and the protein molecules must become intimately mixed for the matrix to have the desired effect on the protein's subsequent desorption. This assumption suggests that the relatively nonselective behavior of the cinnamic acid derivative matrices may be caused by a relatively nonselective interaction between the matrix molecules and those of a protein. Such an interaction may explain the apparent lack of segregation of the proteins and the matrix during the drying process.

The mass spectrum obtained from a mixture of five standard proteins is given in Fig. 1. The amount of each protein in the aliquot loaded onto the probe tip was 0.5 pmol. Each protein produced a singly charged, singly protonated signal ($[M+H]^+$) and a doubly charged, doubly protonated signal {$[M+2H]^{2+}$, m/z = 0.5 × (molecular mass +2)}. Each of these signals also had an additional satellite peak corresponding to a photochemically generated matrix-molecule adduct species ($[M+206]^+$) (19). From this spectrum, the molecular mass of each of the component proteins was determined with an accuracy of ±0.01%. Each molecular mass is actually determined twice in the spectrum because both a singly and a doubly charged species are observed.

A general feature of the mass spectra obtained by this method is the presence of intense signals corresponding to the individual polypeptide subunit chains of each protein. The major ion species in a mass spectrum are typically the singly protonated molecule ions of the subunit chains. Noncovalent complexes, even tightly bound ones such as the ribonuclease S protein/peptide pair (19), produce signals that correspond to the individual chains. There are indications that some proteins may retain some of their quaternary structure (15), but this finding does not appear to be general. The lack of quaternary structure may be caused by a denaturation of the native protein during the sample deposition and/or the laser desorption process. Any discussion of quaternary structure in this context should be made cautiously, however, because the idea of a quaternary structure held together by solvent-dependent forces may not be applicable to gas-phase ions.

Fig. 2 shows an example of a biologically generated mixture of proteins, the mass spectrum of the delipidated high density lipoprotein fraction of human plasma. Mass determination of the peaks in the spectrum identified the proteins in this fraction, allowing the rapid characterization of all of the major proteins and glycoproteins present. The majority of the intense peaks in Fig. 2 have been assigned to known components of the high density fraction (22) by their observed molecular mass, as indicated in the figure. The extent of glycosylation (such as the set of CIII proteins $CIII_1$ and

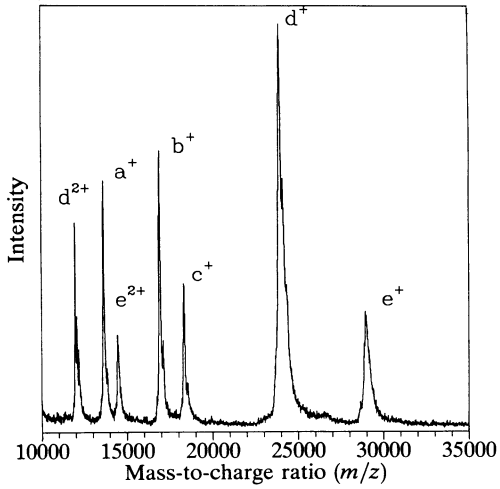

FIG. 1. Mass spectrum obtained from a mixture of five proteins. Approximately 0.5 pmol of each protein was present in the mixture added to the probe tip. The proteins were bovine pancreatic ribonuclease A (a), equine skeletal muscle myoglobin (b), bovine milk β-lactoglobulin (c), bovine pancreatic trypsinogen (d), and bovine erythrocyte carbonic anhydrase II (e). The peaks labeled as "x^+" and "x^{2+}" refer to the singly and doubly protonated protein ion species, respectively.

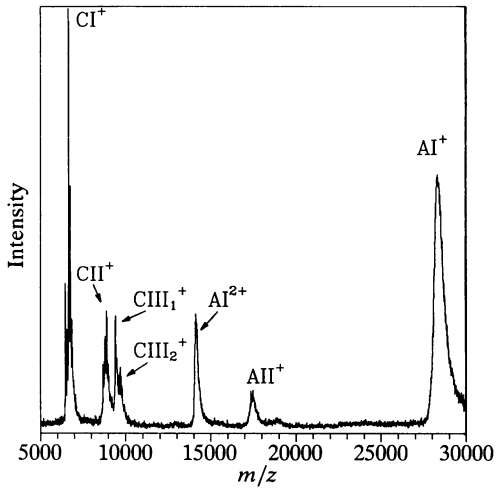

FIG. 2. Mass spectrum obtained from a delipidated human high density lipoprotein fraction. The protein peaks were assigned by their molecular masses according to the usual nomenclature for human apolipoproteins (21), with their charge and protonation states indicated as in Fig. 1.

$CIII_2$) can be readily assessed. Any mass shift or peak twinning of the known protein components would indicate the presence of genetic variants or protein modifications.

An example of the selectivity of this technique for the analysis of proteins in the face of ionic and nonionic components in a mixture is shown in Fig. 3, a comparison of the spectra obtained from bovine and human milk. The spectrum in Fig. 3*A* was obtained from pasteurized, homogenized bovine milk, as obtained from a commercial dairy with no purification. In both cases, 1 μl of the milk sample was diluted into 40 μl of the sinapinic acid solution, and 0.5 μl of the resulting solution was applied to the probe tip for analysis. The main proteins in the mixture were identified from their molecular masses and are labeled. For comparison, the mass spectrum obtained from raw human milk is shown in Fig. 3*B*.

To our knowledge, the type of protein analysis shown in Fig. 3—i.e., examination of a crude biological fluid with no purification—is without precedent in mass spectrometry. The capacity to perform such an analysis can be attributed to the selectivity of this particular method for molecules containing long polypeptide chains. Other desorption and ionization methods (e.g., fast atom bombardment and electrospray atmospheric pressure ionization) are strongly quenched when samples contain significant amounts of nonproteinaceous material, particularly ionic compounds such as salts. Salt concentrations that totally suppress the signals from these other methods (concentrations >1 mM) do not have any strong effect on the quality of the spectra obtained by the matrix-assisted laser desorption method described. A major component of this observed selectivity for proteins is probably caused by segregation of the proteins and contaminants during the solution drying/deposition process. The more (or less) soluble components of the mixture will precipitate from the drying solution at a different time and at a macroscopically different place than the protein/matrix aggregate. The relatively specific interaction of the protein and matrix molecules themselves may also produce some separation and is certainly necessary for the washing procedure described above to selectively remove salts from the dried deposit.

Another class of mixture that is of analytical importance in protein chemistry is the polypeptide mixture produced by cyanogen bromide digestion of a protein (cyanogen bromide selectively cleaves proteins at methionine residues). While cyanogen bromide digest products have been used to a

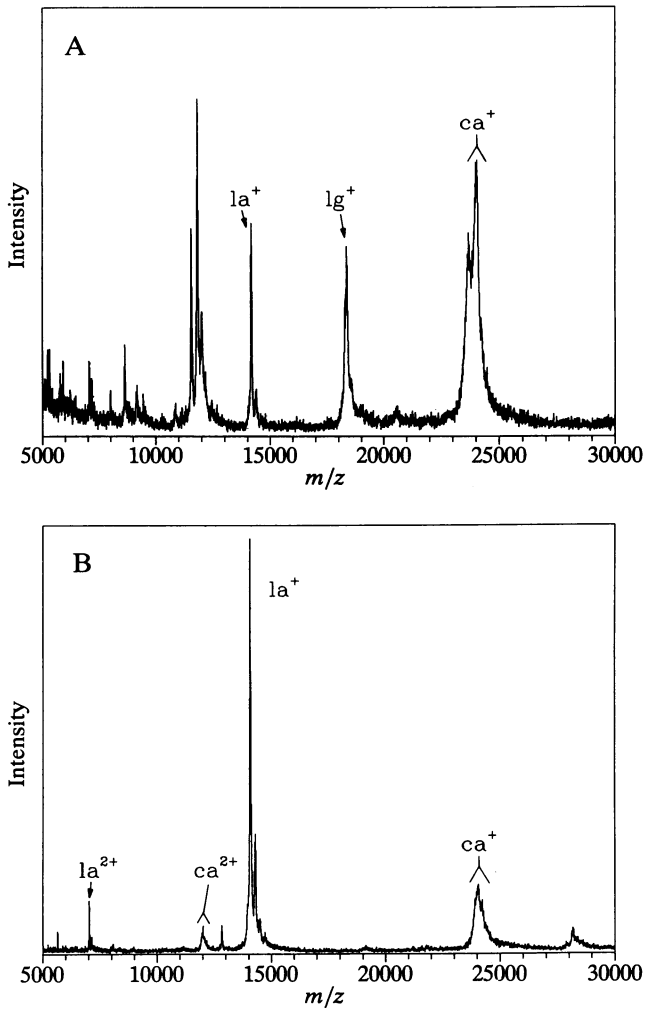

FIG. 3. Mass spectra obtained from commercial bovine milk (*A*) and human breast milk (*B*). The lactalbumin (la), lactoglobulin (lg), and casein (ca) peaks are labeled, with the protonation states of the ions indicated as in Fig. 1. The unlabeled low-mass ions were not identified.

limited extent in the sequencing of proteins by mass spectrometry (23), in general the polypeptides produced are too massive for fast atom bombardment/secondary ion mass spectrometry. The mixture of cleavage products from a cyanogen bromide digest is, however, well suited to matrix-assisted laser desorption analysis (24). Because of the low abundance of methionine residues, the fragments produced are relatively massive (1–20 kDa) and there is a sufficiently small number of products so that interpretation of spectra obtained from the unpurified digest mixture is not too difficult. Fig. 4*A* shows the nearly complete digestion of human apolipoprotein AI by cyanogen bromide. All of the expected scissions are observed, resulting in four main products. The molecular mass of these products can be determined from this spectrum to within 0.01%.

A useful alternative to complete cyanogen bromide digestion is to produce an incomplete cleavage of the protein, leaving a statistical distribution of product polypeptides. If the rates of the bond scission reactions are similar, then a "ladder" will be produced allowing the alignment of the fragments into the complete molecule. A mass spectrum that illustrates such a mixture is shown in Fig. 4*B*. This mixture was the result of adding cyanogen bromide to the protein solution (in 70% formic acid) and immediately placing the reaction vial into a vacuum evaporator. These results can be

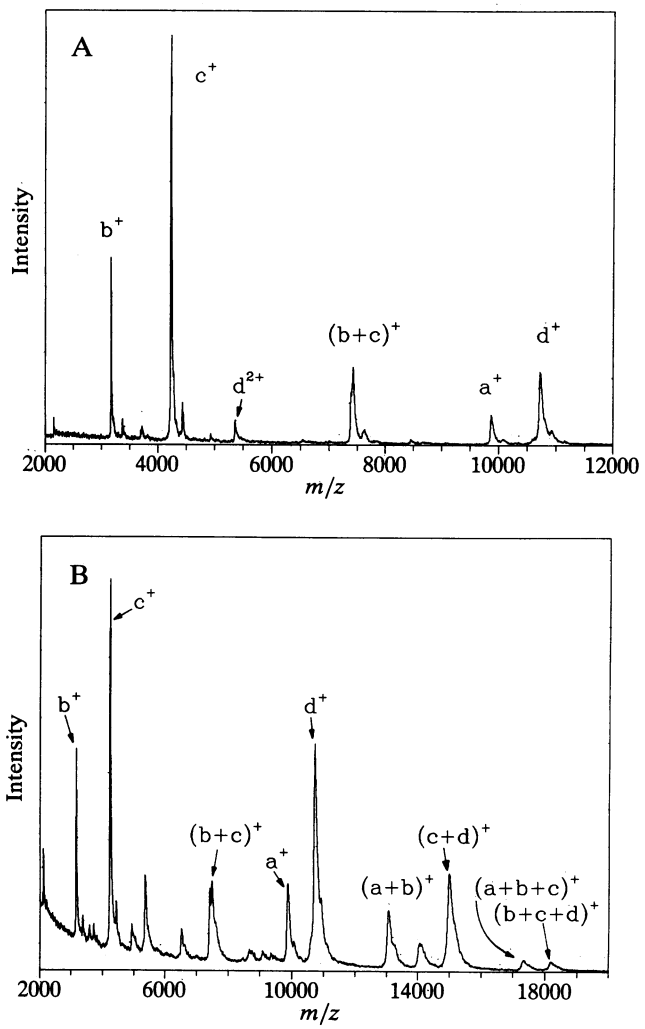

FIG. 4. Mass spectra of two different cyanogen bromide digests of human apolipoprotein AI: one spectrum was obtained from a 2-hr reaction, i.e., nearly complete (*A*); the other spectrum (*B*) was obtained from a brief reaction, showing the products of incomplete digestion (note the change in mass scale between *A* and *B*). For clarity, the doubly protonated, doubly charged species in *B* are not labeled. The peak labels correspond to the various possible peptides produced by scissions at the three methionines in the molecule as shown in Fig. 5. All of the signals except those containing the C-terminal peptide (peptide d) correspond to the homoserine lactone product.

easily interpreted to align the final products (Fig. 4*A*) into their proper order in the intact protein. The alignment scheme for the peptides in Fig. 4 is represented schematically in Fig. 5. The mass spectrum of such a ladder mixture serves as both a molecular mass constraint and an ordering device in conjunction with either classic or genetic sequencing methods.

The use of partial, rather than complete, digestion to characterize a molecule greatly increases the speed of the analysis, from the 24 hr conventionally used to obtain a complete reaction to <1 hr. The increased speed reduces the extent of formylation of the product species formed, which can be a significant problem in reactions carried out in aqueous formic acid for long periods of time.

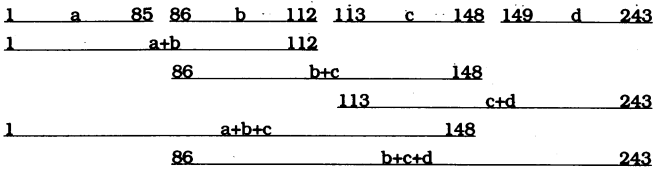

FIG. 5. An assignment of the peptides indicated in Fig. 4. The top line indicates the peptides formed by complete methionine bond scission, and the products of an incomplete reaction are shown below. Numbers indicate amino acid residue position in the polypeptide.

The discovery that complicated mixtures can be examined by laser desorption mass spectrometry has led to the application of the technique to many samples that were considered too crude for the application of other mass spectrometric techniques. The results of our experience with a variety of such samples have identified two types of contaminants that quench the matrix-assisted laser desorption process: (*i*) ionic detergents (e.g., SDS) and (*ii*) low-volatility solvents (e.g., dimethyl sulfoxide and glycerol). The first class, ionic detergents, must be completely removed from a sample before good mass spectra can be obtained. Presumably the strong interaction of these detergents with the individual protein molecules changes the solution-phase properties of the protein sufficiently to interfere with their incorporation into the matrix during the sample drying step. Simple extraction procedures, such as acetone precipitation, are not sufficient to remove those contaminants. Purification of a contaminated sample by reverse-phase HPLC or the single-phase extraction suggested by Henderson *et al.* (25) has proved useful for removing these detergents. The other class of contaminants, low-volatility solvents, if present in sufficiently high concentration will also interfere with the sample drying step by leaving a coating of unevaporated solvent on the surface of the sample. Methods of dealing with this problem that have proved useful, depending on the particular application, are dilution, dialysis, or lyophilization.

Commonly used biological buffering agents do not have to be removed from a protein solution prior to use. The types of buffers that have been used successfully are citrate/phosphate, glycine, Hepes, Tris, ammonium bicarbonate, and ammonium acetate. These buffers have been used at concentrations of 50–200 mM. Acidification agents such as trifluoroacetic acid and hydrochloric acid can also be used. Denaturation agents such as urea and guandinium hydrochloride can be tolerated at concentrations of <2 M.

The examples shown were selected from a variety of protein mixtures that have been examined, including egg white, human blood cell lysates, *Escherichia coli* cell lysates, crude seed extracts, enzyme digest mixtures, and commercial preparations of proteins containing many impurity proteins and peptides. Over 160 different proteins have been examined by this technique in our laboratory. This experience suggests that the application of laser desorption mass spectrometry to biological extracts will allow the use of gas-phase molecular masses to identify particular protein components in a manner analogous to the current application of SDS/PAGE, but with higher resolution and mass accuracy. The method can be applied to the rapid screening of biological fluids for the determination of variations in protein sequence and the presence of posttranslational modifications such as glycosylation with high (≈1 pmol) sensitivity.

This research was supported by grants from the National Institutes of Health (RR00862, GM38724, and BSRG S07 RR07065).

1. Weber, K. & Osborn, M. (1969) *J. Biol. Chem.* **224**, 4406–4412.
2. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
3. Segrest, J. P. & Jackson, R. L. (1972) *Methods Enzymol.* **28**, 54–63.
4. Sundqvist, B. U. R. & Macfarlane, R. D. (1985) *Mass Spectrom. Rev.* **4**, 421–460.
5. Lacey, M. P. & Keough, T. (1989) *Rapid Commun. Mass Spectrom.* **3**, 323–328.
6. Barber, M. & Green B. N. (1987) *Rapid Commun. Mass Spectrom.* **1**, 80–83.
7. Biemann, K., Martin, S. A., Scoble, H. A., Johnson, R. S.,

Papayanopoulos, I. A., Biller, J. E. & Costello, C. E. (1986) in *Mass Spectrometry in the Analysis of Large Molecules*, ed. McNeal, C. J. (Wiley, Chichester, U.K.), pp. 131–149.
8. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. (1989) *Science* **246**, 64–71.
9. Smith, R. D., Loo, J. A., Edmonds, C. G., Barinaga, C. J. & Udseth, H. R. (1990) *Anal. Chem.* **62**, 882–899.
10. Covey, T. R., Bonner, R. F., Shushan, B. I. & Henion, J. D. (1988) *Rapid Commun. Mass Spectrom.* **2**, 249–256.
11. Henry, K. D., Williams, E. R., Wang, B. H., McLafferty, F. W., Shabanowitz, J. & Hunt, D. F. (1990) *Proc. Natl. Acad. Sci. USA* **86**, 9075–9078.
12. Karas, M. & Hillenkamp, F. (1988) *Anal. Chem.* **60**, 2299–2301.
13. Karas, M., Bahr, U. & Hillenkamp, F. (1989) *Int. J. Mass Spectrom. Ion Processes* **92**, 231–242.
14. Spengler, B., Karas, M., Bahr, U. & Hillenkamp, F. (1989) *Ber. Bunsenges. Phys. Chem.* **93**, 396–402.
15. Karas, M., Bahr, U., Ingedoh, A. & Hillenkamp, F. (1989) *Angew. Chem. Int. Ed. Engl.* **28**, 760–761.
16. Salehpour, M., Perera, I., Kjellberg, J., Hedin, A., Islamian, M. A., Haakansson, P. & Sundqvist, B. U. R. (1989) *Rapid Commun. Mass Spectrom.* **3**, 259–263.
17. Loo, J. A., Udseth, H. R. & Smith, R. D. (1988) *Rapid Commun. Mass Spectrom.* **2**, 207–210.
18. Beavis, R. C. & Chait, B. T. (1989) *Rapid Commun. Mass Spectrom.* **3**, 233–237.
19. Beavis, R. C. & Chait, B. T. (1989) *Rapid Commun. Mass Spectrom.* **3**, 432–435.
20. Beavis, R. C. & Chait, B. T. (1989) *Rapid Commun. Mass Spectrom.* **3**, 436–439.
21. Gotto, A. M., Jr., Pownall, H. J. & Havel, R. J. (1986) *Methods Enzymol.* **128**, 3–41.
22. Brinton, E. A., Eisenberg, S. & Breslow, J. L. (1989) *J. Clin. Invest.* **84**, 262–269.
23. Robbin, P. W., Trimble, R. B., Wirth, D. F., Hering, C., Maley, F., Maley, G. F., Das, R., Gibson, B. W., Royal, N. & Biemann, K. (1984) *J. Biol. Chem.* **259**, 7577–7583.
24. Kratzin, H. D., Wiltfang, J., Karas, M., Neuhoff, V. & Hilschmann, N. (1989) *Anal. Biochem.* **183**, 1–8.
25. Henderson, L. E., Oroszlan, S. & Konigsberg, W. (1979) *Anal. Biochem.* **93**, 153–157.