Reprint Series
25 September 1992, Volume 257, pp. 1885–1894

SCIENCE

Weighing Naked Proteins: Practical, High-Accuracy Mass Measurement of Peptides and Proteins

Brian T. Chait and Stephen B. H. Kent

Weighing Naked Proteins: Practical, High-Accuracy Mass Measurement of Peptides and Proteins

Brian T. Chait and Stephen B. H. Kent

Two new technologies have made the study of proteins by mass spectrometry straightforward. Proteins with molecular masses of up to more than 100 kilodaltons can be analyzed at picomole sensitivities to give simple mass spectra corresponding to the intact molecule. This development has allowed unprecedented accuracy in the determination of the molecular weights of proteins. A number of "case studies" are used to present the revolutionary impact that these powerful new ways of looking at proteins are having on biological research.

Technologies that enable the accurate measurement of the molecular weights (MWs) (*1*) of biopolymers promise to facilitate greatly those biological studies that are dependent on the detailed definition of proteins. However, most of the techniques developed to date for the measurement of the masses of proteins have accuracies limited to no better than 5 to 10%. The most widely used of these techniques is sodium dodecylsulfate–polyacrylamide gel electrophoresis (SDS-PAGE) (*2*). SDS-PAGE has assumed a pivotal role in biological research because of the power of simple visualization of the total protein content of a sample, together with crude information on the relative MWs and approximate amounts of the proteins present. An SDS-PAGE analysis of the low-MW apolipoproteins from the high-density lipoprotein (HDL) fraction of human plasma (*3*) is shown in Fig. 1A, in which the different apolipoproteins are visualized as discrete stained bands on the electrophoretic gel (*4*). A comparable analysis is shown in Fig. 1B in the form of a mass spectrum of naked ionized apolipoprotein molecules (*5*). The spectrum was obtained from a similar HDL fraction from human plasma in which the measurement was made with one of a new generation of highly accurate mass measurement technologies called matrix-assisted laser desorption/ionization time-of-flight (TOF) mass spectrometry (MS) (*6*). Each peak in the mass spectrum corresponds to a distinct apolipoprotein. The position along the *x*-axis provides a measure of the mass of the apolipoprotein with an accuracy of ±0.01%.

As this example shows, innovations are occurring in MS with important implications for biological researchers. Since 1988, two new ways have emerged for getting proteins into the gas phase as intact molecular species bearing integral excess charges: matrix-assisted laser desorption MS (*6–8*), an example of which is shown above, and electrospray ionization MS (*9–12*). Past developments in MS with potential applications in biological research have had some success but frequently involved sophisticated techniques that depended on expensive and difficult-to-maintain instrumentation. This combination has meant that the methods have been restricted to dedicated MS laboratories and have not been widely available nor directly usable in the laboratory of the biological researcher.

In contrast, the new developments in protein ionization are readily applicable to the solution of a diverse range of problems and can be combined with relatively simple mass measurement devices. The laser desorption experiment is optimally combined with TOF mass measurement, whereas the electrospray method is optimally combined with a quadrupole mass filter. Both methods give mass accuracy of up to 1 part in 10,000 for proteins with MWs less than 30 to 40 kD and somewhat reduced mass accuracy for larger proteins. These new techniques are in many ways analogous to SDS-PAGE and reversed-phase high-performance liquid chromatography (HPLC), the two most commonly used existing analytical methods for the study of proteins. Both the amounts of proteins used in these standard analytical protein chemistry techniques and the form of the sample are directly compatible with either matrix-assisted laser desorption MS or electrospray MS. No special handling of samples is required, and the picomole amounts used are consistent with existing practice in the biological research laboratory.

This article focuses on the practical application of these two new MS techniques in biological research involving proteins (*13*). The principles of the matrix-assisted laser desorption and electrospray ionization methods are described, and the power and utility of accurate mass measurements of proteins and peptides in biological research are demonstrated by detailed examples.

B. T. Chait is at The Rockefeller University, New York, NY 10021. S. B. H. Kent is at The Scripps Research Institute, La Jolla, CA 92037.

Matrix-Assisted Laser Desorption MS

All mass spectrometers designed to analyze proteins consist of two essential components, the ion source and the mass analyzer. In the ion source, a seemingly unlikely phase transition is effected: Proteins introduced as solids or in solution are converted into intact, naked ionized molecules in the gas phase (*5*). Subsequently, in the mass analyzer, the mass-to-charge (m/z) ratios of the naked protein molecule ions are determined.

Recently, methodology has been developed whereby intact protein ions could be generated in large numbers by laser photon bombardment of protein-containing samples (*14, 15*). Karas and Hillenkamp (*14*) demonstrated the intense production of intact, naked ionized protein molecules when proteins dilutely imbedded in a solid matrix are bombarded with intense, short duration bursts of ultraviolet (UV) laser light. The solid matrix consists of low-MW organic molecules that strongly absorb the UV irradiation.

A simple and effective mass spectrometer in which this type of ion source is used is illustrated diagrammatically in Fig. 2A (*16*). Samples are prepared for mass analysis by adding protein or proteins to a concentrated aqueous solution containing a large molar excess of a matrix-forming material (*6*) such as 3,5-dimethoxy-4-hydroxycinnamic acid (sinapinic acid) (*17*). A small volume of this mixture containing ~1 pmol of the protein or proteins is dried on a sample probe and inserted into the mass spectrometer. In the mass spectrometer, the sample (consisting of a layer of matrix microcrystals containing isolated protein molecules) is bombarded with short duration (1 to 10 ns) pulses of UV wavelength (typically, 337 nm from a nitrogen laser or 355 nm from a frequency-tripled Nd:YAG laser). The interaction of the laser pulse with the sample causes a fraction of the matrix and trapped proteins to go into the gas phase and causes the protein to be ionized. A static electric field is imposed upon ions generated from the sample by application of a high potential (typically ±30 kV) to the sample probe with respect to a closely spaced accelerat-

ing electrode. Protein ions are thus accelerated through the orifice in the electrode and enter into a field-free flight-tube (typical length is 50 to 200 cm).

The masses of the protein ions can be simply determined by TOF analysis. Because all of the ions are accelerated through a fixed potential difference, the velocity of the ions is proportional to $(m/z)^{-1/2}$. As the ions pass through the field-free flight-tube, they separate into a series of spatially discrete individual ion packets, each traveling with a velocity characteristic of its m/z ratio. A detector positioned at the end of the field-free flight-tube produces a signal as each ion packet strikes it. A recording of the detector signal as a function of time yields a TOF spectrum. The difference between the start time, set by the occurrence of the laser pulse and common to all ions, and the arrival time of an individual ion at the detector is proportional to $(m/z)^{+1/2}$ and can be used to calculate the ion's m/z ratio. Such a calculation can be used to convert the x-axis of the spectrum (TOFs) into a m/z ratios axis (a conventional mass spectrum; see Fig. 1B). The TOF analyzer is highly efficient because all ions of different m/z ratios arising from a single laser shot are measured; they simply arrive at the ion detector at different times.

The typical features of a matrix-assisted laser desorption mass spectrum obtained from a protein is shown in Fig. 2B; a mass spectrometer similar to that shown in Fig. 2A was used (*18*). The simple spectrum, which was obtained from a mouse monoclonal immunoglobulin G (IgG) antibody, is dominated by peaks arising from singly and doubly (and to a lesser extent triply) protonated intact IgG. The singly and doubly protonated molecules have charge states of +1 and +2, respectively, yielding m/z ratios of (M + 1)/1 and (M + 2)/2, respectively (where M is the molecular mass of IgG). Simple algebraic manipulation of these two measured m/z ratios give two independent measures of the average molecular mass (*19*) of IgG, which in this experiment was determined to be 148,140 daltons.

Performance specifications of laser desorption TOF-MS for a high-quality commercial instrument include a resolution of 300 (*20*), a mass determination accuracy up to 1 part in 10^4, a sensitivity of better than 1 pmol, and a spectrum acquisition time of 1 min (*18*). The method appears almost universal for proteins that can be dissolved in appropriate solvents, such as a volume/volume ratio of 2:1 of 0.1% trifluoroacetic acid–acetonitrile or neat hexafluoroisopropanol (for proteins with hydrophobic character) (*21*). A selection of proteins that have been successfully analyzed in the laboratory of B.T.C. are given in Table 1. A special feature of the technique that makes it practical for solving biological problems is its ability to measure complex mixtures of peptides and proteins in the presence of large molar excesses of salts, buffers, lipids, and other species (*22*) (see Fig. 1B). There are practical limitations that need to be considered. These limitations include the occurrence of adduct artifacts (*16*) that limit the mass accuracy for masses greater than 30 to 40 kD, the requirement that both the protein and the matrix material be soluble in the solvent mixture used, and the poisoning effect on the mass spectra of traces of ionic detergent (such as SDS) or involatile liquid additives [such as glycerol and dimethyl sulfoxide (*22*)].

The physicochemical events leading to the transfer of proteins to the gas phase and their ionization in matrix-assisted laser desorption/ionization have not yet been fully elucidated. The matrix is believed to serve several functions, including absorption of energy from the laser light and the isolation of individual protein molecules within the large molar excess of solid matrix (*23, 24*). The protein-matrix mixture typically forms a microcrystalline layer spontaneously on drying the sample on the insertion probe tip. Upon irradiation with a short duration burst of laser light, one model for the mechanism assumes that the uppermost layers of matrix are induced to undergo a phase transition from the solid to the gas phase. The subsequent expansion of these matrix molecules into the vacuum drags the matrix-isolated protein molecules into the gas phase (*25*). During the transfer to the gas phase, the proteins undergo ionization through proton transfer reactions with the matrix by reaction processes that remain to be elucidated (*26, 27*).

Electrospray Ionization MS

The electrospray phenomenon is a process for producing naked, intact protein molecules in ionized form from a dilute solution of the protein by creating a spray of fine, highly charged droplets in the presence of a strong electric field. Recently, Fenn and co-workers found that electrospray is a particularly powerful means for producing peptide and protein ions suitable for study by MS (*28, 29*).

A diagrammatic representation of an electrospray ionization mass spectrometer is shown in Fig. 3A. It makes use of an atmospheric pressure interface to the mass spectrometer, wherein the sample solution is simply sprayed in the atmosphere toward an orifice or pipe leading into the vacuum region enclosing the mass analyzer. The protein mixture to be analyzed is in solution

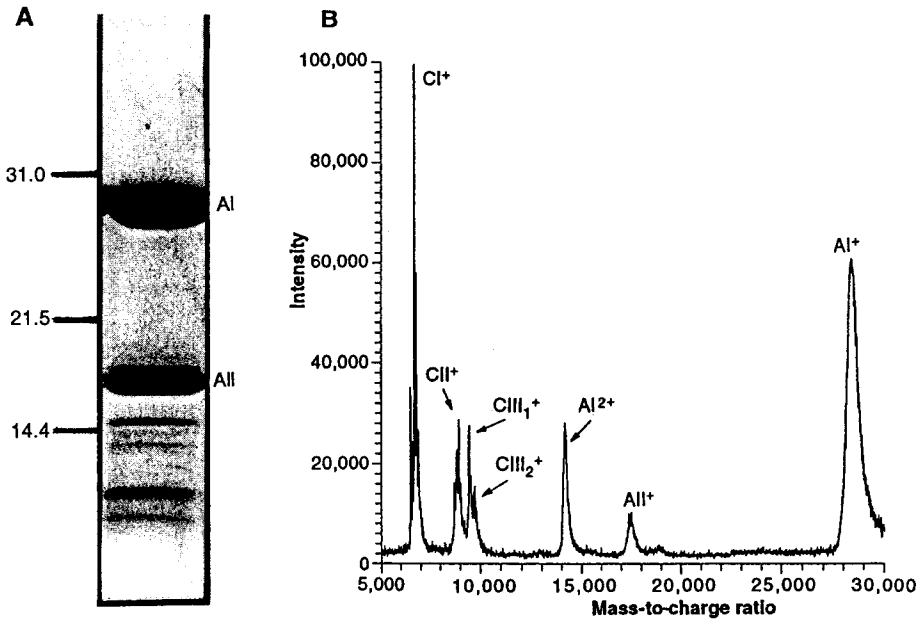
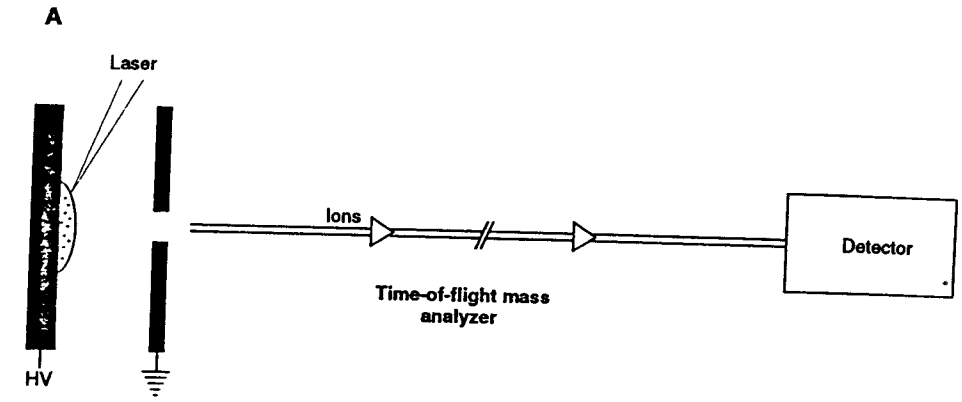


Fig. 1. (**A**) Sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) analysis of high-density lipoprotein (HDL) particles from human plasma, showing portion of the gel containing the most abundant apolipoproteins. The positions of three molecular weight standards are indicated at 31.0, 21.5, and 14.4 kD, respectively. [Adapted from (*4*) with permission © American Society for Biochemistry and Molecular Biology] (**B**) Matrix-assisted laser desorption mass spectrum (*6*) obtained from a delipidated human HDL fraction. The protein peaks were assigned by their molecular weights according to the usual nomenclature for human apolipoproteins (*3*). The "+" designates a singly protonated (charged) ion and "2+" a doubly protonated ion. [Adapted from (*22*) with permission, © National Academy of Science]



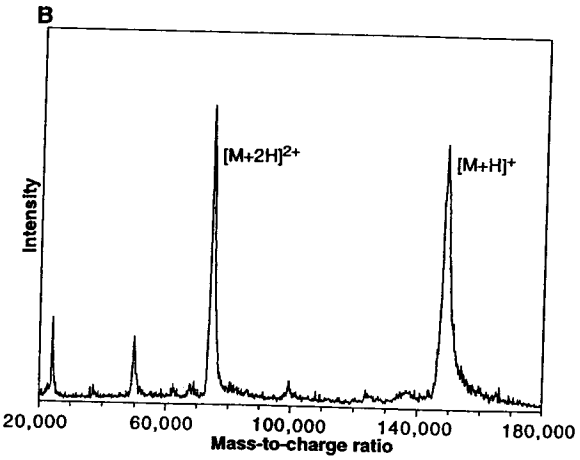
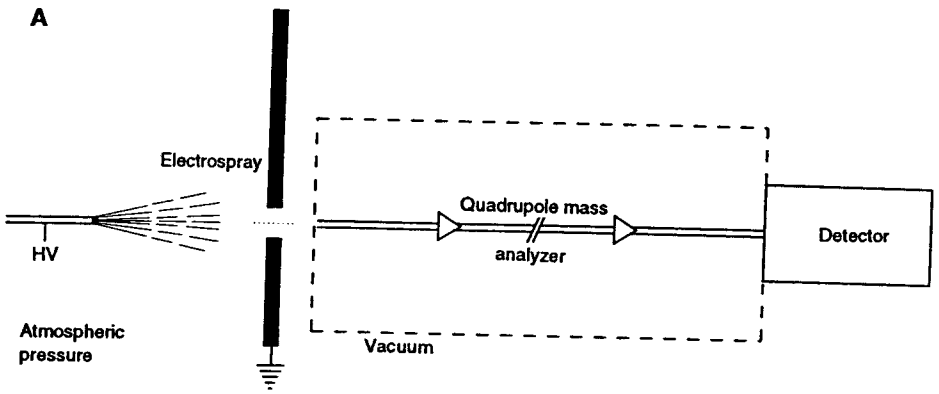


Fig. 2. (**A**) Matrix-assisted laser desorption mass spectrometer with a TOF *m/z* ratio analyzer; HV, high voltage. (**B**) Matrix-assisted laser desorption mass spectrum of monoclonal immunoglobulin G_1 (*Mus musculus*), measured molecular mass of 148,140 ± 150 daltons. $[M+nH]^{n+}$ designates the intact protein with *n* attached protons. The amount of protein applied to the sample probe for insertion into the mass spectrometer was 0.5 pmol. [Adapted from (*18*) with permission, © American Chemical Society]



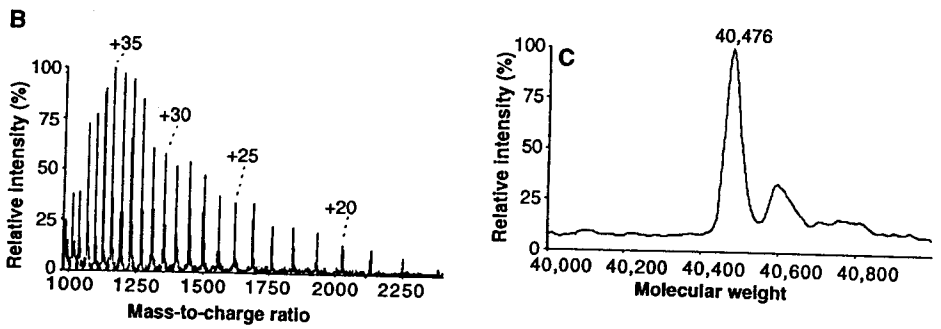


Fig. 3. (**A**) Electrospray ionization mass spectrometer with a quadrupole *m/z* ratio analyzer. (**B**) Electrospray ionization mass spectrum of tropomodulin (*43*). The numbers above the peaks designate the charge states of tropomodulin detected in the present experiment. (**C**) Reconstructed mass spectrum after an algorithm (*40*) was applied to convert the family of ion peaks arising from tropomodulin to a single peak corresponding to the neutral molecule. [MS by S.B.H.K.]

in water or in a water-organic solvent mixture, typically with a trace of acetic or trifluoroacetic acid added to promote ionization of the sample. This solution is sprayed from a region of strong electric field at the tip of a metal syringe needle maintained at approximately +5000 V (for positive-ion spectra). The fine droplets so formed carry excess positive charge and are attracted to the inlet, which is held at a lower potential. Through the application of directed flows of dry gas or heat or both to the droplets during their journey into the vacuum system, solvent contained in the droplets is caused to evaporate. Production of isolated multiply ionized proteins in the gas phase occurs from these shrinking droplets (*9–12, 29–31*). Large numbers of solvent molecules are frequently observed to remain attached to the ionized protein molecules that enter the vacuum system. These ionized protein molecules undergo collisions with residual neutral gas molecules in the entrance region of the vacuum system. By adjusting the magnitude of the electric field in this region, the collision energy can be controlled to completely desolvate the intact ionized protein molecules (*32*). The resulting naked proteins, each molecule carrying an excess positive charge related to its charge state in the original solution (*33–37*), are directed into a quadrupole filter (*38*) for *m/z* ratio analysis.

Several variations of the electrospray experiment exist, but all contain the essential elements described above. In one robust version, so-called "ion spray," a concentrically applied nebulizer gas at the capillary-needle tip is used to assist the formation of suitable fine droplets, a feature that facilitates direct coupling of reversed-phase HPLC with electrospray MS (*12*).

Mass accuracies of about 1 part in 10,000 and resolutions of about 1 in 1,000 are routinely obtained in electrospray-quadrupole MS. Molecules with MWs <60,000 can be analyzed in a straightforward fashion, although much larger proteins have been analyzed (*39*). Optimum sample concentrations are essentially the same as those used for analytical reversed-phase HPLC: ~0.1 to 0.5 μg/μl, and total sample consumption in an experiment is typically ~5 to 50 μl (that is, 1 to 10 μg). Picomole amounts of total sample are consumed during the experiment. The experiment is a simple, practical one and can be performed as easily as running an HPLC analysis on a protein sample.

Perhaps the most striking feature of the electrospray MS experiment is that the resulting mass spectra contain a family of charge states arising from a single protein (Fig. 3B) (*9, 11*). For many denatured proteins, the resulting charge states average

approximately one positive charge per 1,000 daltons. The mass filter measures the m/z ratio of an ion. A protein of MW 29,970 daltons with 30 additional protons would have a charge of +30, a mass of 30,000 daltons, and would appear in the mass spectrum at m/z = 30,000/30 = 1,000. Because of this multiple charging phenomenon, a modest m/z range quadrupole filter can be used to measure large proteins. Each member of the series of charge states for a particular protein molecule is related to its nearest neighbors by having one more or one less proton. Simple mathematical calculation allows the mass of the parent molecule to be determined from any two consecutive charge states of known m/z ratio (*9, 11*).

The complex mass spectrum resulting from the multiple charge state phenomenon (Fig. 3B), at first impression, appears to be a disadvantage of the electrospray experiment, and indeed congestion of the spectrum sometimes occurs for complex mixtures. However, the researcher soon realizes that the multiple charge states provide highly redundant (and therefore highly precise) information about the MW of a protein species present in a particular experiment. In addition, algorithms have been developed that convert the families of ion peaks arising from a single molecular species to a single peak corresponding to the neutral molecule (*40*). The resulting spectrum has the appearance of the more conventional plot of intensity versus mass (Fig. 3C).

Although many proteins have been successfully analyzed by electrospray ionization MS, certain classes of proteins have proved difficult to measure. These include certain insoluble membrane-bound proteins and proteins with very stable tertiary structure or low numbers of basic residues. Substantial progress is being made on developing techniques for improving the mass spectrometric response of these difficult-to-analyze proteins (*31, 41, 42*). Most buffer salts and detergents (neutral or charged) interfere with the simple electrospray experiment; samples in common buffers such as phosphate or containing sodium chloride would not yield useful spectra under standard conditions. The ideal solvent for a protein sample is water-organic solvent with a trace of acid present. Because this is the state in which proteins are recovered from a reversed-phase HPLC experiment, on-line HPLC–electrospray MS can be routinely performed and is extremely useful. HPLC-MS is also an effective way of removing interfering salts from a protein sample prior to analysis (for example, see Fig. 3B). Off-line analysis of protein-containing components collected from an analytical HPLC run is also routine and useful.

A typical application of the electrospray method for ionizing protein samples is illustrated in Fig. 3. Here, a tropomyosin-binding membrane protein, tropomodulin (*43*), was isolated from human erythrocyte membranes and was found to have a molecular mass of 40,476 ± 4 daltons. This result was consistent with the amino acid sequence of the protein deduced from complementary DNA (cDNA) sequencing of the corresponding gene (*44*) and demonstrated that the mature, functional protein had been processed after translation by removal of the amino-terminal Met residue, followed by acetylation of the new amino terminal (calculated MW of 40,481). The amino terminal of tropomodulin had been previously shown to be blocked to Edman degradation.

The charge state associated with a protein in the electrospray method is sensitive to the conformational state of the protein molecule (*33–37*). Thus, denatured proteins have more charges than native protein molecules. Electrospray MS has been used to study conformational transitions in proteins in solution (*33, 36, 41*). The results correlate with direct measurements in solution (such as circular dichroism and nuclear magnetic resonance). Recently, it has also been found that the electrospray ionization process is sufficiently mild that noncovalent complexes between protein and specific ligands can be observed (*34, 45–48*). This approach has been applied to the study of receptor-drug interactions (*45*) and to protein-protein interactions (*47*). Finally, electrospray MS has been used to determine the nature of either covalent or noncovalent intermediates formed in enzyme-catalyzed reactions (*46, 49*).

Correlations of Processed Proteins with Their Genes

Once the cDNA sequence of a gene has been determined, a simple, accurate measurement of the MW of the corresponding protein can provide a host of valuable information. If the measured mass of the protein agrees with that calculated from the gene sequence, it is likely that the deduced sequence is correct, the amino and carboxyl terminals of the mature protein have been correctly assigned, and the protein contains no posttranslationally modified amino acid residues. However, a difference between the measured and predicted MWs implies either an error in the cDNA deduced sequence or a posttranslational modification or processing of the protein.

ARPP-16. The power of accurate MW determination in this context is illustrated here by consideration of ARPP-16, a cAMP

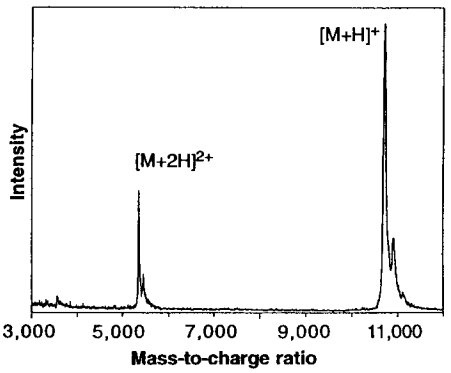


Fig. 4. Matrix-assisted laser desorption mass spectrum of the protein ARPP-16 (*50*). [Adapted from (*6*) with permission, © American Chemical Society]

Table 1. Examples of proteins successfully analyzed by laser desorption TOF mass analysis. Chosen from more than 200 examples analyzed by R. C. Beavis in the laboratory of B.T.C.

Insulins	Human casein
Mouse macrophage inducing proteins 1a, 1b, 2, and 3	Rat trypsin II
Synthetic transforming growth factor α	α-Chymotrypsin
Mouse ubiquitin	Bacteriorhodopsin (*Halobacterium halobium*)
Porcine proinsulin	Hen egg white ovomucoid
Betabellin	Human steroid dehydrogenase
Synthetic and recombinant HIV-I protease	Subtilisin Carlsberg (*Bacillus subtilis*)
Peptidyl-prolyl cis-trans isomerase (*Neisseria meningitidis*)	Bovine carbonic anhydrase II
Equine cytochrome c	V8 protease (*Staphylococcus aureus*)
SH-2 Src	Aldolase (*S. aureus*)
Streptavidin (*Streptomyces avidinii*)	Porcine pepsin
Human milk lysozyme	Yeast alcohol dehydrogenase
Bovine superoxide dismutase	Rabbit muscle aldolase
Horse skeletal apomyoglobin	Rabbit muscle actin
Phospholipase A_2	Human IgG Fab
Bovine β-lactoglobulin A	Trypanosome trypanathione reductase
Difolate reductase (*Escherichia coli*)	β-Amylase
Recombinant flavodoxin (*Azobacter vinelandii*)	Human corticosteroid binding hormone
Human interferon-α	Tropomyosin dimer
	Colicin Ia (*E. coli*)
	β-Galactosidase (*E. coli*)
	Human polyclonal IgG

(adenosine 3',5'-monophosphate)–regulated phosphoprotein from the soluble fraction of bovine caudate nuclei (*50*). Analysis by SDS-PAGE of the protein yielded an apparent MW of 16,000. The amino terminus was not amenable to amino acid sequence analysis because it was chemically blocked. Inspection of the cDNA sequence and the results of partial amino acid sequence analysis suggested a MW considerably lower than that inferred from the SDS-PAGE analysis (*50*). The matrix-assisted laser desorption mass spectrum obtained from the unphosphorylated form of ARPP-16 is shown in Fig. 4. The mass spectrum exhibits two intense peaks corresponding to the singly and doubly protonated intact protein. The average of the two measurements yields a molecular mass of 10,708.8 daltons. This measured mass is much lower than that determined by SDS-PAGE but close to that inferred from the cDNA sequence with an amino-terminal methionine (calculated MW = 10,665.2). The difference between the measured and the calculated masses is 43.6 daltons, a value consistent (within the uncertainty of the measurement) with the presence of an acetyl blocking group (mass difference of 42 daltons) at the amino terminus. Taken together with the cDNA sequence data and the partial amino acid sequence data, the simple mass measurement thus provides compelling confirmation of the predicted primary structure of ARPP-16 as well as detailed information concerning a subtle modification of the protein. It is emphasized that this detailed information was obtained rapidly (in less than 15 min) from a total quantity of 1 pmol of protein.

Sometimes, differences are observed between the measured and calculated MWs that are more difficult to interpret than in the example of ARPP-16 given above. In such cases, a useful strategy involves degradation of the protein by chemical or enzymatic means and measurement by matrix-assisted laser desorption mass spectrometry of the total mixture of peptide products so generated (*22*, *51*). Comparison of the accurately measured masses of the degradation products with those predicted from the cDNA sequence rapidly yields information on the sites and natures of modifications and errors.

Feline immunodeficiency virus (FIV) viral proteins. FIV is a mammalian retrovirus closely related to human immunodeficiency virus (HIV) (*52*) and shares basically similar epidemiology and course of infection in cats. This virus is studied as a model for the human disease AIDS (acquired immunodeficiency syndrome). The FIV life cycle is similar to that of HIV. Viral replication involves a series of key elements catalyzed by virus-encoded enzymes. One such element is the proteolytic processing of the very large polyprotein translation products of the FIV *gag* and *pol* genes, to give the smaller structural proteins and enzymes required for productive virus particle formation. In order to define these sites of proteolytic processing, a highly purified FIV virus preparation was inactivated, the particles disrupted, and the total mixture of released proteins subjected to direct analysis by reversed-phase HPLC (on-line)–electrospray MS (*53*). The resulting profile of total ion current versus elution position (time) is shown in Fig. 5A.

Such an experiment generates mass spectral data for individual protein components eluting from the HPLC column. The total ion current profile gives an indication of where components may have eluted. Mass spectra of individual proteins eluting from the HPLC can be examined by looking at the appropriate spectral records. At least seven viral proteins were detected in this single experiment, which represent all but two of the mature translation products of the *gag* and *pol* genes. The observed MWs were correlated with translation products of the viral genome (*52*) that would be expected if processing events analogous to those observed in related retroviruses occurred.

The mass spectrum of one component, of measured mass 14,735 ± 2 daltons eluting at 34 min, is shown in Fig. 5, B and C. This spectrum was interpreted as follows as corresponding to the mature, functional form of the FIV matrix protein. The matrix protein is encoded at the 5' end of the *gag* gene. Its terminus was inferred from the amino terminal of the neighboring *gag* protein, the capsid protein, experimentally determined by Edman degradation of SDS-PAGE–separated viral proteins. This contiguous stretch of 135 amino acids gives a calculated mass of

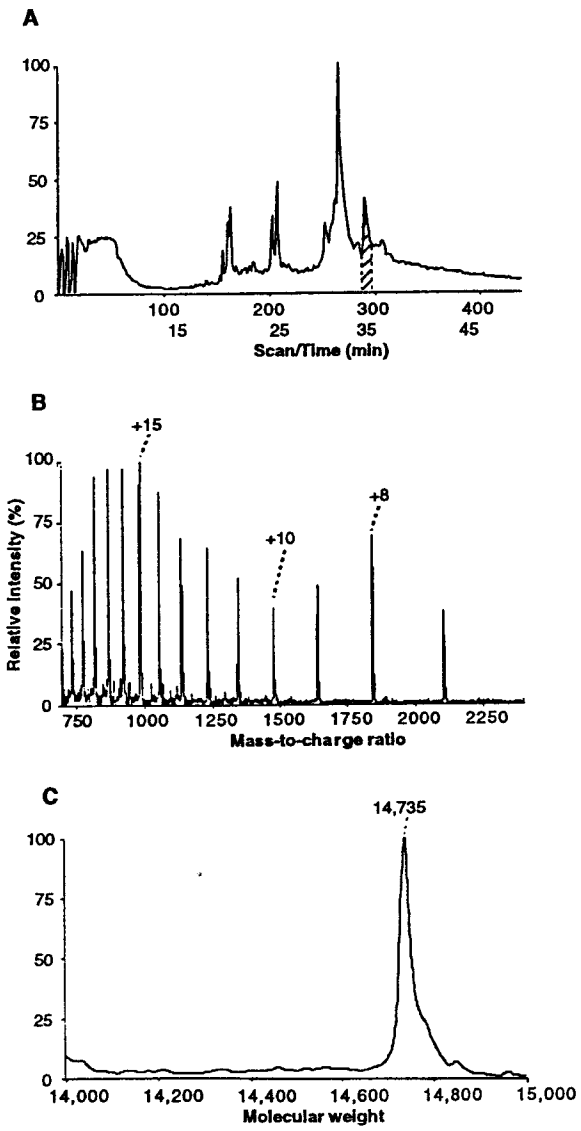


Fig. 5. On-line reversed-phase HPLC–electrospray MS of a total virus protein preparation for the feline immunodeficiency virus (FIV). (**A**) Plot of the total current of electrosprayed ions as a function of elution time of the various protein components. (**B**) Electrospray mass spectrum of the viral protein component eluting at ~34 min [indicated by the cross-hatched peak in (A)]. (**C**) Reconstructed mass spectrum of (B). Measured MW is 14,735 ± 2. Calculated MW for the mature FIV matrix protein is 14,738. [HPLC-MS by S.B.H.K.]

14,659 daltons, which is significantly different from that observed (14,735 daltons). The *gag* open reading frame has a Met at codon 1. In other retroviruses, the amino acid at position 1 was shown by standard protein chemistry techniques to have been removed, and the revealed amino terminal acylated with myristic acid (C_{14} fatty acid) (*54*). For the FIV matrix protein, correcting the calculated molecular mass for the removal of a Met residue (131 daltons) and addition of a myristic acid moiety (210 daltons) gave a calculated mass of 14,738 daltons for the mature FIV matrix protein, in good agreement with the molecular mass 14,735 ± 2 daltons found experimentally.

Glycoprotein Analysis

α_1-Acid glycoprotein. The elucidation of the carbohydrate portion of glycoproteins provides a formidable analytical challenge, especially when only small amounts of sample are available. Previously, proteins containing large proportions of carbohydrate have also proven quite refractory to MS analysis (*13*). Such glycoproteins, however, produce intense mass spectra by matrix-assisted laser desorption/ionization. The spectrum obtained from α_1-acid glycoprotein, a human blood plasma protein containing ~40% by weight of carbohydrate (*55*), is shown in Fig. 6A. From enzymatic analysis, it was known that the oligosaccharides are asparagine linked and of the complex form that contains a large number of neuraminic acid moieties. The spectrum is dominated by peaks corresponding to the singly and doubly protonated glycoprotein. These peaks are extremely broad, reflecting the large amount of heterogeneity in the carbohydrate portion of the protein. The resolution of the mass spectrometer used to obtain this spectrum [$M/\Delta M$ = 300 full-width at half-maximum (fwhm)] was insufficient to permit resolution of the individual glycoprotein components. However, the difference between the average mass of the molecule (36,800 daltons, deduced from the centroids of the broad peaks) and the mass of the peptide backbone (21,539 daltons, calculated from the known amino acid sequence) gives a measure of the average mass of carbohydrate attached to the protein (15,300 daltons). The width of the peaks (2,200 daltons fwhm) gives a measure of the amount of heterogeneity in the compound.

Treatment of the compound with glycosidases and subsequent remeasurement of the MW yields information concerning the number of the particular types of sugars that are cleaved by the enzymes. For example, Fig. 6B shows the mass spectrum of the sample of α_1-acid glycoprotein subsequent to digestion with neuraminidase. The peaks have shifted to lower mass (33,000 daltons) as a result of the cleavage of the neuraminic acid moieties, and the peak width has decreased. The shift of 3,800 daltons indicates the presence of an average of 13 neuraminic acid moieties (mass = 291 daltons) in the glycoprotein. The decrease in peak widths results from the removal of the neuraminic acid heterogeneity and provides a measure of this heterogeneity. Residual heterogeneity from the remaining carbohydrate is evident and can now be partially resolved (Fig. 6B). In principle, the analysis could be continued by further treatment of the sample with selected glycosidases. In this manner, a considerable amount of information concerning the make-up of picomole amounts of glycoproteins can be rapidly deduced.

Tissue plasminogen activator (TPA). This proteolytic enzyme has important therapeutic applications in the treatment of human coronary disease, and TPA has been produced by recombinant DNA methods in suitably engineered cell lines and marketed as a pharmaceutical. Recombinant human TPA produced in Chinese hamster ovary cells is a glycoprotein with a single polypeptide chain of 527 amino acids and a mass of 59,000 daltons (*56*). Glycosylation at two or three sites adds another 3,000 to 9,000 daltons to the molecular mass (*57*). For reproducible preparation of this protein and to lay the basis of understanding the molecular origins of activity, it was important to define both the sites of glycosylation and the nature of the carbohydrate moieties attached to the recombinant protein (*58*).

Reversed-phase HPLC peptide mapping of a tryptic digest of the protein was used to evaluate the glycosylation state of the recombinant TPA (Fig. 7). Both UV detection (at 214 nm) (Fig. 7A) and nebulizer gas-assisted electrospray MS (*12*) (Fig. 7B) were used to detect and characterize the eluted peptides. The effluent stream from the standard 4.6 by 150 mm C_{18} reversed-phase HPLC column (10 nmol of digest loaded) was split (20:1 ratio), with the large portion going to the UV detector and the smaller portion going to the mass spectrometer. Mass spectral data were acquired continuously and stored for subsequent interpretation. Comparison of the elution profiles in Fig. 7, A and B, shows that detection by the two methods gives similar but not identical results.

Accurate mass measurement leads directly to rapid identifications of the peptides present in a digest from a protein of known sequence. Even coeluting species are readily differentiated and identified by their accurately measured MWs.

The tryptic glycopeptides have a similar MS behavior, in terms of sensitivity of response and charge states, as the carbohydrate-free tryptic peptides. Contour "two-dimensional" (2-D) plots of m/z versus HPLC elution time proved to be ideal for the detection of families of glycopeptides containing the same peptide but heterogeneous carbohydrate moieties. The more heavily glycosylated peptides elute slightly earlier in the reversed-phase HPLC, and the signature of a glycopeptide family of molecular species is a characteristic diagonal in the 2-D plot (*58*).

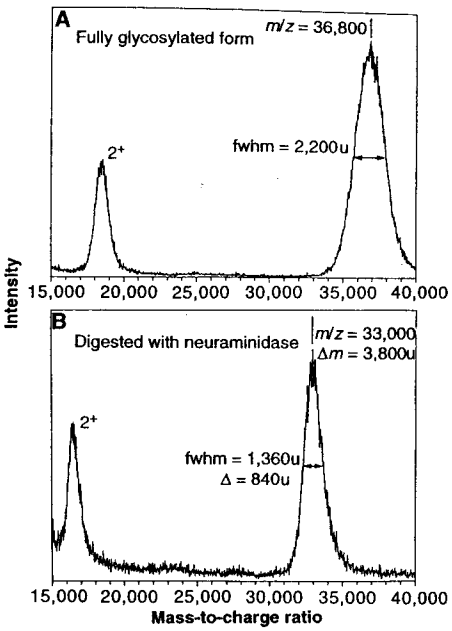


Fig. 6. Matrix-assisted laser desorption mass spectra of α_1 acid glycoprotein (**A**) prior to and (**B**) after treatment with neuraminidase. [Experiment was performed by R. C. Beavis in the laboratory of B.T.C.]

With the use of the tryptic digest–HPLC–electrospray MS approach, three sites of N-linked glycosylation in recombinant TPA were found and characterized. The data for one of these are illustrated in Fig. 7C. Not only is the site of glycosylation identified (peptide T45 with sequence Cys-Thr-Ser-Gln-His-Leu-Leu-Asn-Arg), but accurate mass measurement of the various glycosylated forms of the peptide leads to identification of the various carbohydrate compositions in Fig. 7C (*58*).

In addition to the utility in detecting and characterizing glycopeptides, HPLC-MS peptide mapping of a protein is ideal for evaluating the presence of modifications such as oxidation, phosphorylation (*59*), fatty acylation, and a variety of other modifications, including those labile to the conditions of the Edman degradation.

Phosphoprotein Analysis

Phosphoproteins play a central role in many intracellular processes, including signal transduction and the regulation of cell division. The site and extent of phosphoryla-

tion of key proteins are believed to play an important regulatory role in many intracellular signaling pathways.

The enzyme "cAMP-dependent protein kinase" is a complex of mass of 178 kD and is made up of two regulatory and two catalytic subunits. In the presence of cAMP, which binds to the regulatory subunits, the complex dissociates and releases the enzymatically active catalytic subunits. The 3-D molecular structure of the catalytic subunit of cAMP-dependent protein kinase, 350 amino acids with a molecular mass of 40,440 daltons, has been determined by x-ray crystallography (60).

The type α catalytic subunit of cAMP-dependent protein kinase from the mouse was cloned and expressed in *Escherichia coli* (61). The recombinant protein was isolated as a mixture of molecular species, all containing the same peptide chain but differing from each other in the degree of phosphorylation at specific residues. Three isoforms of the recombinant enzyme were prepared in highly purified form for structural analysis (61). Each was believed to be a homogeneous molecular species. The final stage of purification gave the protein samples in a high salt buffer. Therefore, the MWs of the three protein isoforms were determined by on-line reversed-phase HPLC–electrospray MS, which simultaneously evaluated the purity of the proteins and desalted them in a form suitable for electrospray MS analysis (Fig. 8).

The isoforms differed from one another by the mass of a single–PO_3H group (80 daltons). Based on the predicted sequence, the mature form of the recombinant polypeptide chain has a calculated mass of 40,440 daltons (62). Isoform I had a molecular mass of 40,759 daltons, 319 daltons above the calculated mass and therefore contained four phosphate groups (319/80). Isoform II had a mass of 40,678 daltons (that is, −81 daltons with respect to I and +238 daltons with respect to the parent polypeptide chain) and thus contained three phosphate groups. Isoform III had a mass of 40,600 daltons (that is, −78 daltons with respect to II, −159 daltons with respect to I, and +160 daltons with respect to the parent) and had two phosphate groups. These data confirmed that the isoforms were of high purity and were consistent with the fact that they were homogeneous molecular species. The differences in phosphate content were consistent with prior data (63), indicating a phosphate content of two, three, and four groups, respectively, for the three isoforms of the cloned enzyme.

As these results show, phosphoproteins can be analyzed in a straightforward fashion by electrospray MS. The accuracy of MW measurement is well within the limits for useful determination of the degree of phosphorylation of intact proteins of typical size. Combined with enzymatic digestion and HPLC-MS peptide mapping (compare with Fig. 7 above), this is a powerful technique for the structural characterization of phosphoproteins.

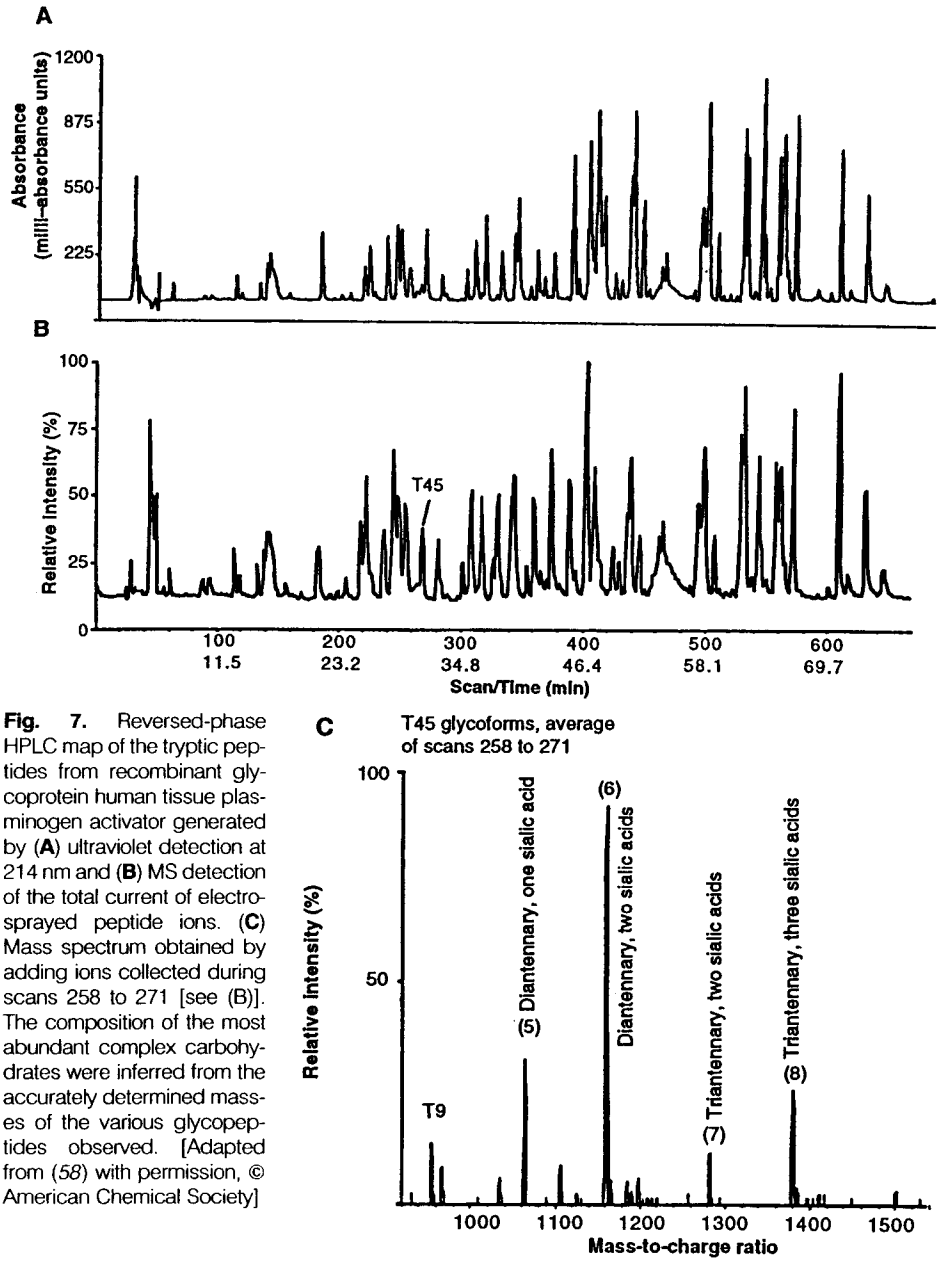


Fig. 7. Reversed-phase HPLC map of the tryptic peptides from recombinant glycoprotein human tissue plasminogen activator generated by (**A**) ultraviolet detection at 214 nm and (**B**) MS detection of the total current of electrosprayed peptide ions. (**C**) Mass spectrum obtained by adding ions collected during scans 258 to 271 [see (B)]. The composition of the most abundant complex carbohydrates were inferred from the accurately determined masses of the various glycopeptides observed. [Adapted from (*58*) with permission, © American Chemical Society]

Protein Sequence Determination

Determination of the amino acid sequence of a protein molecule plays a central role in much biological research. Typically, the biological researcher's first direct observation of a protein would be by 1-D or 2-D PAGE. A critical step in the study of many biologically important proteins is the determination of limited stretches of amino acid sequence data from 10 to 100 picomole amounts of the natural protein isolated from a biological source. Such limited sequence data are frequently the key information used to identify and clone the gene corresponding to the protein of interest. The nucleic acid sequence of the gene is then determined and translated to afford the complete amino acid sequence of the translation product. After cloning and expression of the identified gene, amino acid sequence data are used to confirm the structure of the protein produced. At the present time, amino acid sequence data are almost invariably generated by automated Edman degradation of a protein from the amino terminal, either of the intact protein or of peptides separated after proteolytic digestion of the protein. In some instances, tandem MS methods (13) are proving in-

creasingly useful, especially for proteins with blocked amino termini.

Protein ladder sequencing with one-step MS readout. Recently, we have described a new approach (*64*) to the determination of the amino acid sequence of picomole amounts of a protein that takes advantage of the ability of matrix-assisted laser desorption MS to accurately and rapidly measure protein mixtures. Chemistry is used to generate, in a controlled fashion, a family of sequence-defining fragments from a polypeptide chain. The family of sequence-defining fragments is read out with laser desorption MS to simultaneously generate the complete data set in a single operation as a protein sequencing ladder (Fig. 9). Mass differences between consecutive peaks define the identity of a particular amino acid, based on the distinctive mass of each genetically coded amino acid (except Leu/Ile), and the family of fragments defines the sequence of amino acids in the original peptide chain. One straightforward way of generating the sequence-defining set of fragments from a peptide or protein is to carry out the Edman degradation in the presence of a terminating agent. The protein ladder sequencing method relies on the capabilities of matrix-assisted laser desorption MS as a general way of measuring the MW of proteins with high accuracy. This sequencing method is as yet unproven for very small amounts of proteins of unknown sequence isolated from biological sources, but the potential exists for speeding up and simplifying protein sequencing. This development could lead to more widespread, routine use of protein sequence analysis in biological research.

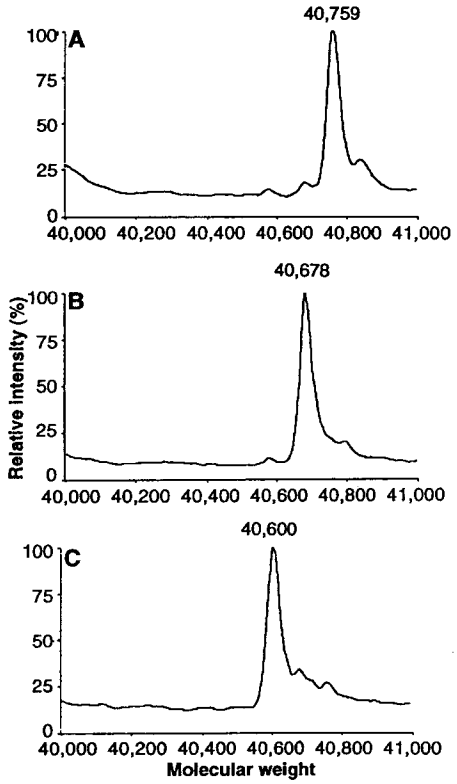


Fig. 8. Electrospray mass spectra of three isoforms of the recombinant phosphoenzyme α catalytic subunit of cAMP-dependent protein kinase. The data were acquired from three separate reversed-phase HPLC–electrospray MS experiments. A total of ~20 μg (500 pmol) of each protein was used. (**A**) Isoform I; measured mass of 40,759 daltons. (**B**) Isoform II; measured mass of 40,678 daltons (−81 daltons with respect to isoform I; that is, less one phosphate group). (**C**) Isoform III; measured mass of 40,600 daltons (−78 daltons; −159 daltons with respect to isoform I; that is, less two phosphate groups]. The mass spectra were reconstructed from raw data showing multiply charged molecules with a median charge of +50. [HPLC-MS experiments were run by S.B.H.K.]

Future Developments

Although remarkable progress has been made during the past few years, much work remains. We list below a selection of improvements and developments that are likely to further widen the impact of MS on biological research involving proteins.

For matrix-assisted laser desorption/ionization, there is a need to further improve the methods of sample preparation and ion production so that the ions of interest are produced devoid of interfering adduct ion species (*16*), with a minimum of metastable decay (*65*) and a controlled extent of ionization (*26*). Because the natural environment of many proteins is aqueous, it is desirable to develop conditions under which water (in liquid or solid form) can be effectively used as a matrix (*66*). A water matrix could eventually lead to the direct examination of biological specimens.

It is important that the resolution of the currently available TOF instrumentation for proteins be increased [for example, by the effective use of an electrostatic mirror (*67*)] to the resolution limit imposed by the envelope of the isotope distribution over the full mass range of interest. This development would give the optimum mass accuracy and would resolve compounds that have closely similar masses. Alternative means for achieving this goal include mass analysis by double focusing magnetic deflection (*68*), Fourier transform ion cyclotron resonance (*69*), and 3-D electric multipole ion trapping (*70*). These analyzers are also of interest because of the possibilities for manipulating selected and trapped polypeptide ions in interesting ways. Of particular interest is the possibility of dissecting ion-

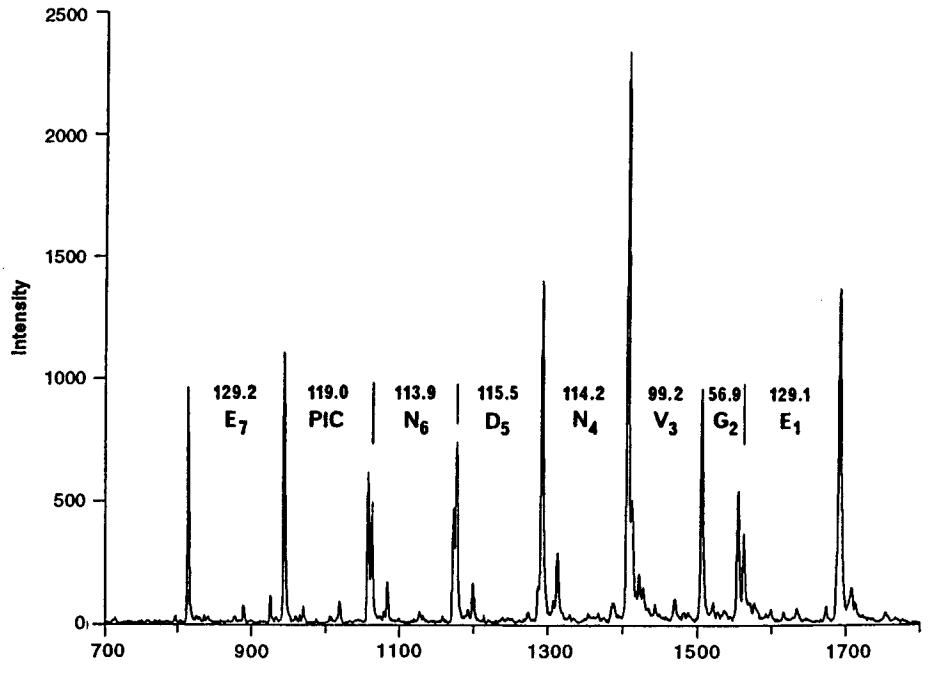


Fig. 9. Example of a protein sequencing ladder data set. The peptide [Glu¹]fibrinopeptide b [sequence EGVNDNEEGFFSAR (*71*), MW = 1570.6] was subjected to Edman degradation in the presence of a terminating agent [5% by mole phenylisocyanate (PIC)]. No intermediate separation and analysis of the low MW products released at intermediate cycles were performed. After seven cycles of Edman degradation, a portion of the total reaction mixture was read out in a single operation by laser desorption MS. The measured masses, identities, and sequence of the first seven amino acids of the peptide are marked on the mass spectrum. Note that Asp and Asn were readily distinguished. [Experiment was performed by R. Wang in the laboratory of B.T.C.]

ized polypeptides chemically in a highly controlled manner in the gas phase.

For the electrospray ionization of proteins, there is a need to render the ionization process more universally successful and uniform and to make the production of ions from proteins more tolerant of nonproteinaceous ionic constituents such as buffers and salts. It is of considerable interest to extend the m/z range of analyzers beyond that currently available with commercial quadrupole instrumentation for electrospray ionization. Such m/z range extension should facilitate the study of proteins that are not highly charged under solution conditions of interest (for example, conditions designed to maintain the native conformation of the protein) as well as the study of the association of proteins with other proteins, DNA, enzyme substrates, and nonprotein prosthetic groups.

Both matrix-assisted laser desorption and electrospray ionization MS are inherently highly sensitive techniques whose sensitivity, in practice, is frequently limited by sample handling prior to the MS measurement. There is thus a pressing need to devise improved methods for sample handling and introduction into the mass spectrometer of picomole to femtomole amounts of peptides and proteins, especially for samples present in low concentrations (10^{-7} to 10^{-9} M) in solution. The ability to measure such small quantities of sample constituents of interest would also require the reduction of the unwanted background and improvements in instrumental intensity dynamic range.

Further work on improving the coupling between traditional methods of protein separations including 1-D and 2-D electrophoresis would greatly facilitate the identification of proteins of interest and the elucidation of posttranslational modifications. If MS is to fulfill its potential in biological research, it is imperative that instruments be placed in the laboratories of biological workers. This would require that the instruments be made still easier to use and less expensive. We believe that, in the future, the new MS capabilities will change the way that biological research involving proteins is approached.

REFERENCES AND NOTES

1. The terms "mass" and "weight" are used interchangeably throughout this text.
2. Y. P. See and G. Jackowski, in *Protein Structure: A Practical Approach*, T. E. Creighton, Ed. (IRL Press, New York, 1990).
3. A. M. Gotto, Jr., H. J. Pownall, R. J. Havel, *Methods Enzymol.* **128**, 3 (1986).
4. S. S. Deeb *et al.*, *J. Biol. Chem.* **266**, 13654 (1991).
5. Here, a naked protein is defined as the isolated protein molecule, totally devoid of attached solvent molecules.
6. F. Hillenkamp, M. Karas, R. C. Beavis, B. T. Chait, *Anal. Chem.* **63**, 1193A (1991).
7. M. Karas, U. Bahr, U. Giessman, *Mass Spectrom. Rev.* **10**, 335 (1991).
8. B. Spengler and R. Kaufman, *Analysis* **20**, 91 (1992).
9. J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse, *Science* **246**, 64 (1989).
10. J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, *Mass Spectrom. Rev.* **9**, 37 (1990).
11. R. D. Smith *et al.*, *ibid.* **10**, 359 (1991).
12. T. R. Covey, R. F. Bonner, B. I. Shushan, J. D. Henion, *Rapid Commun. Mass Spectrom.* **2**, 249 (1988).
13. Not covered here is the application of previous MS techniques [J. A. McCloskey, vol. 193 of *Methods in Enzymology* (Academic Press, San Diego, 1990); D. M. Desiderio, Ed., *Mass Spectrometry of Peptides* (CRC Press, Boca Raton, FL, 1991); A. L. Burlingame, T. A. Baillie, D. H. Russell, *Anal. Chem.* **64**, 467R (1992)] such as fast-atom bombardment (FAB) ionization [M. Barber, R. S. Bordoli, R. S. Sedgwich, A. N. Tyler, *J. Chem. Soc. Chem. Commun.* **1981**, 325 (1981); M. Barber and B. N. Green, *Rapid Commun. Mass Spectrom.* **1**, 80 (1987)] and ^{252}Cf plasma desorption [R. D. Macfarlane and D. F. Torgerson, *Science* **191**, 920 (1976); R. J. Cotter, *Anal. Chem.* **60**, 781A (1988)], nor of methods such as tandem MS structural analysis of peptides [K. Biemann and H. A. Scoble, *Science* **237**, 992 (1987); K. Biemann, *Annu. Rev. Biochem.* **61**, 977 (1992); D. F. Hunt, J. R. Yates III, J. Shabanowitz, S. Winston, C. R. Hauer, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 6233 (1986)], despite the fact that these can be of great use in specific examples of biological research.
14. M. Karas and F. Hillenkamp, *Anal. Chem.* **60**, 2299 (1988).
15. K. Tanaka *et al.*, *Rapid Commun. Mass Spectrom.* **2**, 151 (1988).
16. R. C. Beavis and B. T. Chait, *ibid.* **3**, 233 (1989).
17. _____, *ibid.*, p. 432.
18. _____, *Anal. Chem.* **62**, 1836 (1990).
19. The average molecular weight of a protein with atomic composition $C_vH_wO_xN_yS_z$ is $(v \times 12.011) + (w \times 1.00794) + (x \times 15.9994) + (y \times 14.00674) + (z \times 32.066)$, where the atomic weights are calculated by using the natural abundances of the isotopes [J. R. DeLaeter, *J. Phys. Chem. Ref. Data* **17**, 1791 (1988)].
20. The resolution is defined as $(m/z)/\Delta(m/z)$, where m/z is the measured mass-to-charge ratio of the ion of interest and $\Delta(m/z)$ is the full-width at half-maximum height of the corresponding peak in the mass spectrum.
21. M. A. Baldwin *et al.*, in preparation.
22. R. C. Beavis and B. T. Chait, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6873 (1990).
23. K. Strupat, M. Karas, F. Hillenkamp, *Int. J. Mass Spectrom. Ion Processes* **111**, 89 (1991).
24. R. C. Beavis and J. N. Bridson, unpublished results.
25. R. C. Beavis and B. T. Chait, *Chem. Phys. Lett.* **181**, 479 (1991).
26. R. C. Beavis, T. Chaudhary, B. T. Chait, *Org. Mass Spectrom.* **27**, 156 (1992).
27. H. Ehring, M. Karas, F. Hillenkamp, *ibid.*, p. 472; M. E. Gimon *et al.*, *ibid.*, p. 827.
28. C. M. Whitehouse, R. N. Dreyer, M. Yamashita, J. B. Fenn, *Anal. Chem.* **57**, 675 (1985).
29. C. K. Meng, M. Mann, J. B. Fenn, *Z. Phys. D.* **10**, 361 (1988).
30. S. K. Chowdhury, V. Katta, B. T. Chait, *Rapid Commun. Mass Spectrom.* **4**, 81 (1990).
31. M. H. Allen and M. L. Vestal, *J. Am. Soc. Mass Spectrom.* **3**, 18 (1992).
32. J. A. Loo, H. R. Udseth, R. D. Smith, *Rapid Commun. Mass Spectrom.* **2**, 207 (1988); R. D. Smith *et al.*, *J. Am. Soc. Mass Spectrom.* **1**, 53 (1990).
33. S. K. Chowdhury, V. Katta, B. T. Chait, *J. Am. Chem. Soc.* **112**, 9012 (1990); V. Katta and B. T. Chait, *Rapid Commun. Mass Spectrom.* **5**, 214 (1991).
34. V. Katta and B. T. Chait, *J. Am. Chem. Soc.* **113**, 8534 (1991).
35. J. A. Loo *et al.*, *Rapid Commun. Mass Spectrom.* **5**, 101 (1991).
36. J. C. Y. LeBlanc *et al.*, *Org. Mass Spectrom.* **26**, 831 (1991).
37. R. Guevremont, K. W. M. Siu, J. C. Y. LeBlanc, S. S. Berman, *J. Am. Soc. Mass Spectrom.* **3**, 216 (1992).
38. P. H. Dawson, Ed., *Quadrupole Mass Spectrometry and Its Applications* (Elsevier, New York, 1976).
39. R. Feng and Y. Konishi, *Anal. Chem.*, in press.
40. M. Mann, C. K. Meng, J. B. Fenn, *ibid.* **61**, 1702 (1989); B. B. Reinhold and V. N. Reinhold, *J. Am. Soc. Mass Spectrom.* **3**, 207 (1992).
41. U. A. Mirza, S. L. Cohen, B. T. Chait, unpublished results.
42. A. M. Falick, P. Schindler, A. Van Dorsselaer, *Proceedings of the Fortieth ASMS Conference on Mass Spectrometry and Allied Topics* (American Society for Mass Spectrometry, Washington, DC, in press).
43. V. M. Fowler, *J. Cell Biol.* **111**, 471 (1990).
44. L. A. Sung *et al.*, *J. Biol. Chem.* **267**, 2616 (1992).
45. B. Ganem, Y. T. Li, J. D. Henion, *J. Am. Chem. Soc.* **113**, 6294 (1991).
46. _____, *ibid.*, p. 7818.
47. M. Baca and S. B. H. Kent, *ibid.* **114**, 3992 (1992).
48. A. K. Ganguly *et al.*, *ibid.*, p. 6559.
49. R. T. Alpin, J. E. Baldwin, C. J. Schofield, S. G. Waley, *FEBS Lett.* **277**, 212 (1990); D. S. Ashton *et al.*, *ibid.* **292**, 201 (1991); R. Menard *et al.*, *ibid.* **295**, 27 (1991).
50. A. Horiuchi *et al.*, *J. Biol. Chem.* **265**, 9576 (1990).
51. R. M. Caprioli, B. Whaley, K. K. Mock, J. S. Cottrell, *Techniques in Protein Chemistry II* (Academic Press, San Diego, 1991), p. 497; M. Schar, K. O. Bornsen, E. Gassman, *Rapid Commun. Mass Spectrom.* **5**, 319 (1991); R. S. Annan, H. J. Kochling, J. A. Hill, K. Biemann, *ibid.* **6**, 298 (1992).
52. R. L. Talbott *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5743 (1989); N. C. Pedersen, E. W. Ho, M. L. Brown, J. K. Yamamoto, *Science* **235**, 790 (1987).
53. J. Elder and S. B. H. Kent, unpublished data.
54. L. E. Henderson, H. C. Krutzsch, S. Oroszlan, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 339 (1983).
55. K. Schmid *et al.*, *Biochemistry* **12**, 2711 (1973); P. G. Board, I. M. Jones, A. K. Bentley, *Gene* **44**, 127 (1986).
56. R. C. Choloupek *et al.*, *J. Chromatogr.* **463**, 375 (1989); D. Pennica *et al.*, *Nature* **301**, 214 (1983).
57. D. Collen *et al.*, *J. Pharmacol. Exp. Ther.* **231**, 146 (1984).
58. V. Ling *et al.*, *Anal. Chem.* **63**, 2909 (1991).
59. J. Hou *et al.*, *Biochemistry*, in press; K. Palezewski *et al.*, in preparation.
60. D. R. Knighton *et al.*, *Science* **253**, 407 (1991).
61. F. Herberg and S. S. Taylor, in preparation.
62. J. C. Chrivia, M. D. Uhler, G. S. McKnight, *J. Biol. Chem.* **263**, 5739 (1988).
63. W. Yonemoto and S. S. Taylor, unpublished data.
64. B. T. Chait, R. Wang, R. C. Beavis, S. B. H. Kent, *Proceedings of the Fortieth ASMS Conference on Mass Spectrometry and Allied Topics* (ASMS, Washington, DC, in press).
65. R. C. Beavis and B. T. Chait, in *Methods and Mechanisms for Producing Ions from Large Molecules*, vol. 209 of the *NATO ASI Series B: Physics*, K. G. Standing and W. Ens, Eds. (Plenum, New York, 1991); B. Spengler, D. Kirsch, R. Kaufman, *Rapid Commun. Mass Spectrom* **5**, 198 (1991).
66. R. W. Nelson, R. M. Thomas, P. Williams, *Rapid Commun. Mass Spectrom.* **4**, 348 (1990).
67. G. P. Baptista *et al.*, *ibid.* **5**, 632 (1991).
68. J. A. Hill, R. S. Annan, K. Biemann, *ibid.*, p. 395; R. S. Annan, H. J. Kochling, J. A. Hill, K. Biemann, *ibid.* **6**, 298 (1992); C. N. McEwen and B. S. Larsen, *ibid.*, p. 173; C. L. Wilkens, J. A. Castoro, C. Koster, *Proceedings of the Fortieth ASMS Conference on Mass Spectrometry and Allied Topics* (ASMS, Washington, DC, in press).
69. K. D. Henry, J. P. Quinn, F. W. McLafferty, *J. Am. Chem. Soc.* **113**, 5447 (1991); J. A. Castoro, C. Koster, C. Wilkins, *Rapid Commun. Mass Spectrom.* **6**, 239 (1992); C. L. Wilkins, J. A. Castro, C. Koster, *Proceedings of the Fortieth ASMS Conference on Mass Spectrometry and Allied Topics* (ASMS, Washington, DC, in press).
70. K. A. Cox, J. D. Williams, R. G. Cooks, *Biol. Mass Spectrom.* **21**, 226 (1992); G. J. Van Berkel, G. L.

Glish, S. A. McLuckey, *Anal. Chem.* **62**, 1284 (1990); M. E. Bier, J. C. Schwartz, I. Jardine, G. Stafford, *Proceedings of the Fortieth ASMS Conference on Mass Spectrometry and Allied Topics* (ASMS, Washington, DC, in press).
71. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
72. This article was supported in part by grants from the National Institutes of Health (RR00862 and GM38274) and by funds from the Markey Foundation. We thank R. C. Beavis and R. Wang for their considerable assistance in obtaining experimental results and G. McMilleon for preparing the manuscript.