[22] Matrix-Assisted Laser Desorption Ionization Mass–Spectrometry of Proteins

By RONALD C. BEAVIS and BRIAN T. CHAIT

Introduction

In this chapter, we provide a practical guide to the application of matrix-assisted laser desorption/ionization–mass spectrometry (MALDI–MS) for the analysis of peptides and proteins. We describe in detail the best methods that are currently available for preparing samples for MALDI–MS, because good sample preparation is the key to successful mass analysis. We consider aspects of the method that are important for obtaining high-quality data. Finally, we describe a selection of strategies for studying proteins with this powerful new technique.

Pulses of laser light have been employed since as early as 1976[1] to produce intact gas phase peptide ions from solid samples. The resulting peptide ions could then be analyzed by mass spectrometry. These early investigations and subsequent measurements over the following decade produced useful mass spectra from only a few short peptides. In addition, the probability for obtaining a useful mass spectrum depended critically on the specific physical properties of the peptide (e.g., photoabsorption spectrum, volatility) under study. This situation changed dramatically with the development by Karas and Hillenkamp.[2] MALDI–MS provides the means to volatilize proteins readily and to make the conditions for volatilization largely independent of the specific physical properties of the protein. This effect is achieved in two steps. The first step involves preparing an appropriate sample by dilutely embedding proteins in a matrix of small organic molecules that strongly absorb ultraviolet wavelength laser light. The second step involves ablation of bulk portions of this solid sample

[1] M. A. Posthumus, P. G. Kistemaker, and H. L. C. Meuzelaar, *Anal. Chem.* **50,** 985 (1978).
[2] M. Karas and F. Hillenkamp, *Anal. Chem.* **60,** 2299 (1988).

METHODS IN ENZYMOLOGY, VOL. 270

by intense, short-duration pulses of the laser light. In the ablation step, molecular components of the solid are put into the gas phase and ionized, producing intact protein ions. The molecular masses of these protein ions are easily determined by time-of-flight mass analysis. The marvelous development by Karas and Hillenkamp has been steadily refined and has become a method of choice for characterizing peptides and proteins.[3–8]

MALDI–MS is versatile and effective for the analysis of peptides and proteins because of the special properties and capabilities of the technique. Some of these capabilities and properties are as follows.

Biological samples can be examined without extensive purification.
Common biochemical additives such as buffers, salts, glycerol, chelating agents, chaotropic agents, and certain detergents do not interfere with the analysis.
Most classes of proteins can be examined, provided that the protein can be dissolved in appropriate solvents.
Posttranslationally modified proteins can be measured.
Useful mass spectra can be obtained from complex mixtures of peptides and proteins.
Proteins with masses ranging to greater than 100 kDa can be analyzed.
The total amount of protein required for an analysis is usually in the range of 1–10 pmol.
Protein molecular masses can be determined with mass accuracies as high as 1 part in 10,000.
Complete analyses can be made in a matter of minutes.

It is instructive to compare the properties of MALDI–MS for the analysis of proteins by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS–PAGE), a method widely used in biological research. SDS–PAGE is a universal technique for separating and analyzing proteins because of the effectiveness of the detergent SDS for dissolving proteins and for converting them into entities that migrate on electrophoretic gels with relative velocities that depend on their size. By contrast, there are certain restrictions (see below) on the detergents and additives that can be used in the preparation of samples for MALDI–MS. However, the more complex and expensive instrumentation required by MALDI–MS is justified because MS provides much higher mass accuracy determination (typically three to four orders

[3] F. Hillenkamp, M. Karas, R. C. Beavis, and B. T. Chait, *Anal. Chem.* **63,** 1193A (1991).
[4] B. T. Chait and S. B. H. Kent, *Science* **257,** 1885 (1992).
[5] A. L. Burlingame, R. K. Boyd, and S. J. Gaskell, *Anal. Chem.* **66,** 634R (1994).
[6] R. Aebersold, *Curr. Opin. Biotechnol.* **4,** 412 (1993).
[7] R. Wang and B. T. Chait, *Curr. Opin. Biotechnol.* **5,** 77 (1994).
[8] J. T. Stults, *Curr. Opin. Struct. Biol.* **5,** 691 (1995).

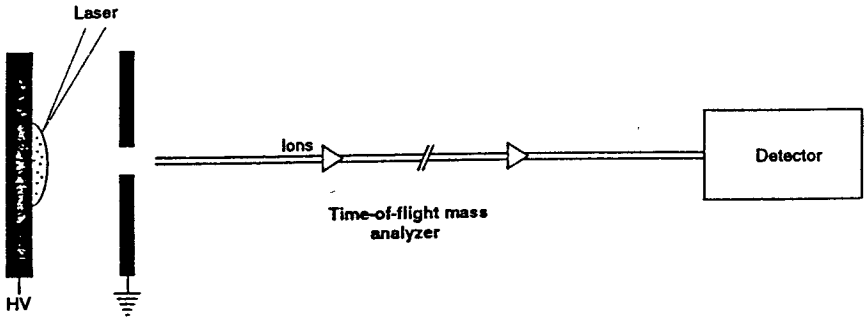


FIG. 1. Schematic diagram of a matrix-assisted laser desorption mass spectrometer with a linear time-of-flight mass analyzer.

of magnitude) and higher resolution than SDS–PAGE. MALDI–MS is also considerably faster than standard SDS–PAGE and can be used to analyze complex peptide mixtures generated by digestion of proteins.

The essential experimental setup for MALDI–MS is illustrated in Fig. 1. Short-duration (nanosecond) pulses of laser light are directed at matrix–protein sample mixtures (protein-doped matrix crystals; see below) that are inserted into the mass spectrometer on a mechanical probe. The laser light causes a portion of the matrix–protein sample to be volatilized and ionized. The resulting gas phase ions are accelerated to a fixed energy in an electrostatic field and directed into a field-free flight tube. After passing through the flight tube, the ions impact on an ion detector, whereupon the intervals (Δt_m) between the pulse of laser light and the ion impacts are measured. The masses of the ions passing through the flight tube can then be determined via the approximate relationship[9,10]

$$m = 2qV\Delta t_m^2/l^2 \quad (1)$$

where m is the mass of the ion, q is the charge on the ion, V is the potential through which the ion is accelerated, and l is the length of the flight tube.

MALDI–MS is a powerful and versatile new tool for the study of proteins and the required instrumentation is significantly simpler than conventional mass spectrometers. This relative simplicity makes the technique potentially accessible to many biologists. The extent to which MALDI–MS will impact on biological research will depend, to a large degree, both

[9] R. C. Beavis, "Time-of-Flight Mass Spectrometry," American Chemical Society Symposium Series 549 (R. Cotter, ed.), pp. 49–60. American Chemical Society, Washington, DC, 1994.
[10] H. Wolnick, *Anal. Instrument.* **16,** 15 (1987).

on the production of effective commercial instrumentation and on the development of methods for making MS easily accessible to biologists.

Preparing Samples

Sample preparation is crucial for matrix-assisted laser desorption. The surroundings of a protein molecule must be fashioned so that an intense light pulse can transfer the intact molecule into a vacuum. The best method for achieving this effect is to incorporate protein molecules into the crystals of a second material. The protein-containing crystal absorbs the light pulse and uses the absorbed energy to eject material from its surface. If the ejected volume contains protein molecules, they enter the vacuum along with the other ejected materials. Few compounds form crystals that incorporate proteins, eject them intact, and ionize them in the same process. The MALDI literature refers to any material with these properties as a *matrix.*

Several additional properties are important for a matrix compound to be analytically useful. Incorporating the protein into a growing matrix crystal implies that the protein and the matrix must be simultaneously in solution. Therefore, a matrix should dissolve and grow protein-doped crystals in commonly used protein solvent mixtures. The matrix crystal absorbs the light pulse, therefore the compound must contain a stable chromaphore. The protein and matrix must not react to form a stable product. From an instrumental point of view, the matrix crystals must remain in vacuum for extended periods of time (tens of minutes). Therefore, the sublimation rate of a matrix in vacuum should be as slow as possible.

The following sections introduce the matrix compounds commonly used, explain the requirements for protein samples, and give typical recipes for producing protein-doped matrix crystals. These sections do not exhaust all possible combinations of matrix compounds and recipes. Instead, these sections reflect widely used approaches to the problem. Modify a recipe if the requirements of a particular solvent differ from the description—however, keep in mind the general principles outlined below.

Selecting Matrices

Few compounds are good matrix materials for analytical protein MALDI. Many compounds form protein-doped structures that produce ions, but they are disqualified by other factors. The resulting mass spectra may have low signal intensities or poor mass resolution. Other materials work only for certain proteins or in a limited range of solvents. Most matrix compounds produce satellite signals called *adduct peaks* at slightly higher mass than the analyte molecule peaks. These peaks result from the photo-

TABLE I
PROPERTIES OF MALDI MATRICES

Matrix	Analytes[a]		Suggested solvent (water:organic)	Ionization[b]	Adduct[c]
	Peptides	Proteins			
Gentisic acid	+	+/−	9:1	+	M + 136
Sinapic acid	+/−	+	2:1	+	M + 206
3-Indoleacrylic acid	+	+	2:1	++	M + 185
4-HCCA	+	+	2:1	+++	—

[a] +, Matrix may be used for most peptides and proteins; +/−, matrix may (or may not) work.
[b] The more + signs, the more intense the signal, and the higher charge state of the most intense peak.
[c] Expected mass of the most intense satellite peak.

chemical breakdown of the matrix into more reactive species, which can add to the polypeptide. The best matrices have low-intensity photochemical adduct peaks.

Table I lists four compounds used for peptide and protein mass measurement and some information about relevant properties. These compounds can be used with 337-nm (nitrogen laser) or 354-nm [neodymium:yttrium/aluminum/garnet (3)] [Nd:YAG(3) laser] light. Table I gives an idea of the usefulness of a matrix. The utility of a specific matrix for a particular protein or peptide cannot be predicted in advance: trial and error may be necessary to establish a useful recipe. Table I also gives a guide to the most commonly used solvent mixtures. Besides the water and organic solvents (such as acetonitrile, methanol, propanol), the mixture may contain acids (e.g., trifluoroacetic acid), salts, lipids and some types of detergents (see below). An indication of the ion signal intensity and the average number of charges added to the protein in the ejection process are shown in Table I. α-Cyano-4-hydroxycinnamic acid (4-HCCA) does not produce strong photochemical adduct signals, although it seems to encourage copper attachment to some peptides, resulting in [M + Cu] peaks.

Gentisic acid (2,5-dihydroxybenzoic acid, molecular mass 154 Da, CAS 490-79-9) is a useful matrix for a wide variety of peptides and proteins.[11] Gentisic acid forms large dimorphic monoclinic crystals from water, water–alcohol, or water–acetonitrile mixtures. Many researchers use it as their first choice for analyzing peptide mixtures. The protein ions ejected from

[11] K. Strupat, M. Karas, and F. Hillenkamp, *Int. J. Mass Spectrom. Ion Processes* **111**, 89 (1991).

gentisic acid crystals appear to be relatively stable, making them ideal for reflectron mass spectrometers.[12,13]

Sinapic acid (or sinapinic acid, *trans*-3,5-dimethoxy-4-hydroxycinnamic acid, molecular mass 224 Da, CAS 530-59-6) is useful for a wide variety of peptides and proteins.[14] Sinapic acid forms small, thin monoclinic crystals from water–acetonitrile (or alcohol) mixtures. It has a strong affinity for proteins of all types and gives good results with mixtures of proteins. Small peptides (mass <3 kDa) may not produce strong signals.

trans-3-Indoleacrylic acid (molecular mass 203 Da, CAS 29953-71-7) has not found broad application for proteins and peptides,[15] although it has been used for industrial polymers. It produces intense signals and works well for complex mixtures of proteins and peptides.

4-HCCA (α-cyano-4-hydroxycinnamic acid, molecular mass 189 Da, CAS 28166-41-8) produces intense signals from peptides and proteins.[16] It forms polycrystalline clumps made up of twinned monoclinic prisms. Instruments with poor ion transmission efficiency rely on 4-HCCA to overcome their design problems. This matrix produces intense, multiply charged ions in the positive-ion spectra of proteins. The protein ions produced by 4-HCCA frequently undergo metastable decay in the mass spectrometer.[16a]

Preparing Protein Solutions for Analysis

The most important factor in incorporating a protein into matrix crystals is having sufficient protein dissolved in the crystal growing solution. For most proteins, a concentration of 1–10 μM in the crystal growing solution produces useful protein-doped crystals. Lower concentrations of protein may produce results. However, the use of low concentrations of proteins in small volumes means that care must be taken when handling the sample. Protein adsorption to the transferring pipettes, Eppendorf tube walls, etc., may reduce the actual protein concentration to negligible levels.

Protein samples that have been exposed to strong ionic detergents, such as sodium dodecyl sulfate, will not incorporate into matrix crystals. If strong

[12] W. Yu, J. E. Vath, M. C. Huberty, and S. A. Martin, *Anal. Chem.* **65,** 3015 (1993).
[13] R. Kaufmann, D. Kirsch, and B. Spengler, *Int. J. Mass Spectrom. Ion Processes* **131,** 355 (1994).
[14] R. C. Beavis and B. T. Chait, *Rapid Commun. Mass Spectrom.* **3,** 432 (1989).
[15] J. Z. Chou, M. J. Kreek, and B. T. Chait, *J. Am. Soc. Mass Spectrom.* **5,** 10 (1994).
[16] R. C. Beavis, T. Chaudhary, and B. T. Chait, *Org. Mass Spectrom.* **27,** 156 (1992).
[16a] M. Karas, U. Bahr, K. Strupat, F. Hillenkamp, A. Tsarbopoulos, and B. N. Pramanik, *Anal. Chem.* **67,** 675 (1995).

ionic detergents are a necessary part of an isolation procedure, they must be thoroughly removed. A method that successfully removes these detergents from small samples of proteins is a two-phase extraction.[17] Electroblotting from PAGE gels and chromatographic methods may be alternatives for detergent removal.[18,19] Acetone precipitation and dialysis usually do not remove enough detergent for MALDI sample production.

While ionic detergents must be avoided, most other protein solvents do not interfere significantly with sample production. Nonionic detergents (such as Triton X or octylglucoside) can be used[20] (see below for cautions). High concentrations of salts, buffers, urea, glycerol, and formic acid produce useful samples (see below for cautions). The mixture 1:2:3 (v/v) formic acid–2-propanol–water works well for many hydrophobic proteins and peptides.[21] Avoid buffer solutions containing sodium azide: its presence suppresses protein ion formation in the mass spectrometer.

Exposure of proteins to concentrated formic acid should be avoided, if possible, in experiments where accurate mass measurements are required. Formic acid reacts with amino groups (both N-terminal α-amino and lysine ε-amino groups), resulting in a formyl derivative of the protein. Multiple formylation leads to several peaks for each polypeptide, making mass determination ambiguous. If a procedure requires formic acid, exposure should be kept as short as possible. Low temperatures also slow the rate of formylation. Even with this caveat, experience suggests that aqueous 70% formic acid is the best general solvent for the cyanogen bromide peptide cleavage reaction. Dilute HCl (0.1 *N*) may also be used as a solvent for cyanogen bromide peptide cleavage; however, care must be taken to neutralize the pH of the solution before evaporating the solvent to dryness. (*Note:* Concentrated trifluoroacetic acid also reacts with free amino groups.)

The solvent chosen for the protein must be compatible with the matrix used to form the protein ion-emitting crystals. In most of the recipes that follow, an aliquot of a matrix solution is mixed with an aliquot of a protein solution to make the crystal-forming mother liquor. It is important that neither the matrix nor the protein precipitate when the two solutions mix. Particular care must be taken when the protein solvent does not

[17] L. E. Henderson, S. Oroszlan, and W. Konigsberg, *Anal. Biochem.* **93,** 153 (1979).
[18] W. J. Henzel, T. M. Billegi, J. T. Stults, S. C. Wong, C. Grimley, and C. Watanabe, *Proc. Natl. Acad. Sci. U.S.A.* **90,** 5011 (1993).
[19] W. Zhang, A. J. Czernik, T. Yungwirth, R. Aebersold, and B. T. Chait, *Protein Sci.* **3,** 677 (1994).
[20] O. Vorm, B. T. Chait, and P. Roepstorff, *in* "Proceedings of 41st ASMS Conference in Mass Spectrometry, San Francisco, 1993," p. 621a.
[21] S. L. Cohen and B. T. Chait, *Anal. Chem.* **68,** 31 (1996).

contain any organic solvent, which may lead to precipitation of the matrix on mixing.

Preparing Samples for Mass Spectrometry

Dried Droplet Method

The discovery of the dried droplet method allowed the application of laser desorption to proteins.[2] Drying a droplet of a protein–matrix solution remains the favorite method of most MALDI practitioners. The recipe for producing a dried droplet sample has several simple steps. First, make a saturated solution of the matrix material (see Table I for solvents) and mix in sufficient protein for a final concentration of 1–10 μM. This solution must be thoroughly mixed to ensure reproducible results. A convenient method is to place some saturated matrix solution (5–10 μl) in an Eppendorf tube and then add a smaller volume (1–2 μl) of a protein solution. Agitating the tube with a vortex mixer for a few seconds mixes the solution sufficiently. Second, place a droplet (0.5–2 μl) of the resulting mixture on the sample stage of the mass spectrometer. Dry the droplet at room temperature; blowing room temperature air over the droplet speeds drying. When the liquid has completely evaporated, the sample may be loaded into the mass spectrometer. There is no rush to load the sample. Dried droplets are quite stable: they can be kept in a drawer or in vacuum for days. The deposit may also be washed to etch the surface layer of the deposit crystals; the surface layer is the most heavily contaminated with involatile components of the original solution. Be careful when etching the crystals, as it is easy to wash them off the surface. We recommend thoroughly drying the sample (vacuum dried if possible) followed by a brief immersion in cold water (10–30 sec in 4° water). The excess water should be removed rapidly (e.g., by flicking the sample stage, or by suction).

Some authors suggest placing a drop of saturated matrix solution on the sample stage and then adding a similar volume of protein solution to form the final droplet. This method results in acceptable spectra for samples containing a single analyte. If the protein sample contains more than one protein or peptide component, we recommend thorough mixing of the two solutions in a tube before making the droplet. Prior, thorough mixing increases the reproducibility of the mass spectra obtained.

The simplicity of the dried droplet recipe surprises most people. This method gives good results for many types of protein samples. The recipe has some serious pitfalls that may be encountered.

The protein must be truly dissolved in the solvent. Making a slurry of a peptide powder and solvent is not sufficient.

Be careful of inadvertent changes in solvent composition. Such changes can easily occur when two solutions are mixed, resulting in the precipitation of either the protein or matrix (or both). A similar problem can result from the selective evaporation of organic solvents from aqueous solutions. The latter is a particular problem when small volumes of solution are stored in relatively large containers (e.g., 10 μl of solvent in a 1.5-ml Eppendorf tube).

Do not heat the droplet to speed drying. Changing the temperature of the solution alters matrix crystal formation and protein incorporation, usually in a bad way.

Use fresh matrix solutions whenever possible. Matrix solutions gradually decompose under normal laboratory conditions. Mix small volumes of solution as needed.

Do not use higher protein concentrations than recommended: the final protein concentration should be less than 10 μM. It is tempting to add more protein, in the hope of increasing signal intensity. This approach seldom works. Reducing the protein concentration (not increasing it) frequently increases the signal.

Do not use involatile solvents. The involatile solvents commonly used in protein chemistry are glycerol, polyethylene glycol, 2-mercaptoethanol, Triton X, dimethyl sulfoxide (DMSO), and dimethylformamide (DMF). These solvents interfere with matrix crystallization and coat any crystals that do form with a difficult-to-remove solvent layer. If you cannot avoid using these solvents (or if a droplet does not dry properly), try another method (see below).

Keep the concentration of nonprotein materials to a minimum. The dried droplet method tolerates the presence of salts and buffers well, but this tolerance has limits. Washing the crystals may help, although care must be taken not to wash the crystals off the substrate. If contaminants are suspected of suppressing the signal, try another method (see below).

The pH of the crystal-growing solution must be less than pH 4. The organic acid matrices described above become ions at pH > 4, completely changing their crystallization properties. Using aqueous 0.1% trifluoroacetic acid rather than pure water will usually solve any pH problems.

Do not leave crystals of undissolved matrix in the final protein–matrix solution. Intact matrix crystals act as crystal seeds, leading to inhomogeneous samples. Spinning the matrix solution in a tabletop centrifuge removes the large crystals that cause this problem.

Slow Crystallization (Growing Large Matrix Crystals)

The dried droplet method will fail for samples containing significant concentrations of involatile solvents or high concentrations of salts.[22] The problem seems to be that as the droplet dries, the contaminant concentrations become very high, interfering with matrix crystal growth and with protein incorporation. This concentration effect can be avoided by growing large, protein-doped matrix crystals under more controlled conditions. The technique described below has been shown to be effective for protein solutions containing low protein concentrations (<1 μM) and high concentrations of involatile solvents (e.g., 30% glycerol) or salts (4 M sodium chloride). In mixtures of peptides and proteins, this technique may discriminate in favor of the protein components.[21]

Large matrix crystals can be grown using the following procedure.[22] First, saturate the solvent mixture chosen for the mother liquor with the matrix at room temperature by placing 20–100 μl of the solvent mixture (containing the protein and other solutes) in a 1.5-ml Eppendorf tube with an excess of matrix. Swirl the mixture with a vortex mixer for at least 5 min and then centrifuge the mixture to remove large crystals. Carefully remove the supernatant with a Pasteur pipette and save it in a clean Eppendorf tube as the mother liquor. Seal the Eppendorf tube and heat the solution to approximately 40° to remove any small matrix crystals remaining in suspension. Remove the tube from the heater and allow it to come to room temperature before opening the tube. Place the opened tube into a holder on the rotating stage of a vortex mixer and set the mixer to its lowest speed. Then allow the solvent to evaporate for several hours while swirling the tube at low speed. This procedure results in a collection of protein-doped matrix crystals with long axes of about 0.5–1 mm (for sinapic acid) or small crystals (for 4HCCA).[21] Once the crystals have grown, remove the mother liquor from the Eppendorf tube with a Pasteur pipette. Wash the crystals by adding room temperature water and swirling the tube with a vortex mixer. Pellet the crystals by centrifugation and remove the solvent. Repeat the washing step several times. After the final removal of washing solvent, add a small amount of water, slurry the crystals into the water (taking care not to crush the crystals), and dump the contents onto a piece of filter paper or the sample stage.

There are many possible variations to this method.[21] The saturated matrix solution may be prepared first and cleared of excess crystals prior to the addition of the protein to the solution, taking care not to dilute the matrix solution. Crystallization may also be carried out by lowering the

[22] F. Xiang and R. C. Beavis, *Org. Mass Spectrom.* **28,** 1424 (1993).

temperature of the mother liquor by placing the open Eppendorf tube in a cold chamber or a refrigerator.

Crystals can be mounted on a sample holder in a number of different ways. For small crystals (<1 mm), transfer them to the sample holder while they are damp. The crystals usually adhere when as they dry. For larger crystals, it may be necessary to place a thin layer of polystyrene cement on the sample stage and adhere the crystals to the cement. Once they are mounted, the crystals can be examined using the same methods used for dried droplet samples.

Thin Polycrystalline Films

Preparing thin polycrystalline films involves producing a uniform layer of very small crystals on the sample stage of a mass spectrometer that are mechanically well adhered to the substrate.[23] It is a variant of the dried droplet method.[23] The crystals can be thoroughly washed without removing them from the surface. This effect is achieved by creating an activated layer of matrix on the surface of the substrate, which acts as an extended seeding site for growing matrix crystals. A droplet is then dried onto this activated surface, resulting in a thin film of protein-doped matrix crystals attached to the surface. The film grows rapidly, so it is not necessary to wait until the droplet is dry before washing the film, reducing effects caused by increasing contaminant concentrations as the droplet dries. The recipe given below results in reproducible polycrystalline films for most matrices, although it can be difficult to use gentistic acid because of its high solubility in water. The method has been shown to give useful signals from solvents containing as much as 50% (v/v) glycerol or 4 *M* urea.

First, prepare two solutions containing the matrix at room temperature, referred to below as solution A and solution B. Handle these solutions carefully: it is important to minimize transfer of undissolved crystalline solid from one step to the next. Prepare the solutions fresh each day.

Solution A contains the matrix alone. Prepare it using 0.3 ml of acetonitrile or any appropriate solvent mixture. Saturate the solvent with matrix in a 0.5-ml Eppendorf tube and swirl for at least 1 min with a Vortex mixer. Spin the tube in a benchtop centrifuge to deposit undissolved matrix particles on the bottom of the tube. Remove the supernatant to another tube.

Prepare solution B by saturating 0.3 ml of the selected solvent mixture with matrix and then remove excess matrix particles by centrifugation. Mix an aliquot of this matrix-saturated solution with a protein-containing

[23] F. Xiang and R. C. Beavis, *Rapid Commun. Mass Spectrom.* **8,** 199 (1994).

solution to produce a final protein concentration of approximately 1 μM. It is important to remember that the solvent produced by mixing the protein-containing solution and the matrix-containing solution must keep both the matrix and the protein in solution. Precipitating either solute leads to poor mass spectra.

To make the sample deposit, first place 0.5–1 μl of solution A on the sample stage of the mass spectrometer and allow it to dry. Press firmly on the surface with a clean Kimwipe and stroke the crystalline layer several times, crushing the crystals and rubbing them into the surface. Brush the crushed matrix to remove any loose matrix particles. The layer of crushed matrix serves as the "activated" seeding site for the drying droplet (see below). Alternatively, an activated layer can be made by electrospraying solution A onto the sample stage, being careful that the sprayed layer does not become wet during the spray process. Electrospraying is more difficult than simply crushing crystals, but the results are good and may be worth the trouble for particularly contaminated samples.

Apply 0.5–1 μl of solution B to the spot bearing the smeared matrix material. An opaque film forms over the surface of the substrate below the droplet within a few seconds, covering the metal. After about 1 min, the probe tip can be immersed in room temperature water to remove involatile solvents and other contaminants. It is not necessary to let the droplet dry before washing: the film does not wash off easily. Remove excess water and allow it to dry before loading into the mass spectrometer.

The aliquot of solution B used to create the film must be free of particulate matter. Any particles of matrix present in this solution result in rapid nucleation of matrix crystals throughout the solution, interfering with the formation of the film. If solution B is obviously cloudy, further centrifugation is necessary to clear the liquid.

Seeded Microcrystalline Films

The seeded microcrystalline film method[24] of sample preparation is a direct replacement for the dried droplet technique. It produces more uniform samples than the dried droplet technique and the crystals are better adhered to the surface of the sample stage. The uniformity of the deposit apparently improves the mass resolution and mass accuracy obtained in reflectron-type mass spectrometers. The well-adhered crystals are less likely to be lost during wash–etch cycles. This technique works well with α-cyano-4-hydroxycinnamic acid.

This technique involves first laying down a thin layer of small matrix crystals on the sample stage. A droplet containing the analyte is placed on

[24] O. Vorm, P. Roepstorff, and M. Mann, *Anal. Chem.* **66,** 3281 (1994).

top of the crystalline deposit. The solvent containing the analyte is carefully chosen so that it will redissolve part of the matrix layer. The resulting matrix–analyte solution dries, depositing the matrix on the seed sites provided by the undissolved portion of the original layer.

Two solutions must be prepared (solutions A and B). The solvent for solution A is acetone containing 1–2% water. Matrix is added to form a nearly saturated solution. A droplet (0.5 μl) of solution A is placed on the sample stage and allowed to spread out and dry. The deposit formed should be uniform. The volume of solution A used and the matrix concentration depend on the geometry of the sample stage: the reader is encouraged to experiment with different conditions to suit specific needs.[24]

Solution B contains the peptide/protein of interest at a concentration of approximately 1 μM in an aqueous solution that contains 20–30% organic solvent. Place approximately 0.5 μl of solution B onto the surface prepared from solution A and allow it to dry at room temperature. When the solution is completely dry, the deposit can be washed by placing a drop of pure water [or 0.1% aqueous trifluoroacetic acid (TFA)] onto the deposit for 2–10 sec and then removing excess water.

Measuring Molecular Masses

Mass spectrometers come in a wide variety of configurations. All mass spectrometers can be broken down into three main parts: an ion source, a mass analyzer, and an ion detector. Matrix-assisted laser desorption is carried out in the ion source, producing a current of protein ions whose mass can be measured using a mass analyzer and an ion detector.[3]

Ion Sources

Matrix-assisted laser desorption ion sources all consist of a sample stage (or probe) that is used to carry the protein-doped matrix crystals into the vacuum system of the mass spectrometer. The lasers that produce protein ion ablation are pulsed: intense laser light is emitted for a brief time (<10 nsec). All commercial ion sources use ultraviolet pulsed lasers, either a nitrogen laser (λ = 337 nm) or a Q-switched neodymium : yttrium aluminum garnet (Nd : YAG) laser, with its fundamental wavelength either tripled (λ = 354 nm) or quadrupled (λ = 266 nm). The light emitted by the laser is focused onto the protein-doped matrix crystals in the vacuum system to produce an illumination fluence of approximately 20 mJ/cm^2. Most of the commonly used matrices will absorb sufficient light to produce protein ions at all of these wavelengths. Longer wavelength lasers are currently favored

because they reduce the amount of absorption in the analyte protein molecule. Some results have been described using infrared lasers to desorb proteins,[25] but no commercial ion source uses this method.

The relationship between protein ion current and laser fluence is highly nonlinear. At low laser fluences (<10 mJ/cm^2), no protein ions are produced. If the laser fluence is increased, a point will be reached when protein ions suddenly appear ("threshold" fluence, ϕ_{th}). Increasing the laser fluence above the threshold value leads to a rapid increase in protein ion current.[26] It has been found empirically that mass analyzers and detectors work best when the laser fluence is near the threshold value.[27] Different matrices have different threshold fluences and fluence-to-protein ion-current curves [ϕ_{th}(α-cyano-4-hydroxycinnamic acid) < ϕ_{th}(sinapic acid) < ϕ_{th}(gentisic acid)]. For these reasons, the illumination fluence must be easily adjustable so that the operator can set a fluence appropriate for a particular sample.

Mass Analyzers and Detectors

Most currently available mass spectrometers with matrix-assisted laser desorption ion sources use time-of-flight mass analyzers and ion-to-electron conversion detectors. Laser desorption ion sources have also been used successfully with magnetic sector,[28] quadrupole ion trap,[29] and Fourier transform ion cyclotron resonance[30] mass analyzers.

Time-of-flight mass analyzers are easy to understand. Linear time-of-flight analyzers consist of a long, straight, empty flight tube with an ion source on one end and flat detector on the other end. A short-duration pulse of ions emerges from the ion source after acceleration to a common kinetic energy. Light ions move more quickly down the flight tube than heavy ions and therefore strike the detector first. If the signal coming out of the detector is plotted as a function of time, peaks corresponding to ions of different mass appear at different times. The mass-to-charge ratio

[25] K. Strupat, M. Karas, F. Hillenkamp, C. Eckerstkorn, and F. Lottspeich, *Anal. Chem.* **66,** 464 (1994).

[26] W. Ens, Y. Mao, F. Mayer, and K. G. Standing, *Rapid Commun. Mass Spectrom.* **5,** 117 (1991).

[27] R. C. Beavis and B. T. Chait, *Rapid Commun. Mass Spectrom.* **3,** 233 (1989); R. C. Beavis and B. T. Chait, *Anal. Chem.* **62,** 1836 (1990).

[28] J. A. Hill, R. S. Annan, and K. Biemann, *Rapid Commun. Mass Spectrom.* **5,** 395 (1991).

[29] V. M. Doroshenko, T. J. Cornish, and R. J. Cotter, *Rapid Commun. Mass Spectrom.* **6,** 153 (1992); K. Jonscher, G. Currie, A. L. McCormack, and J. R. Yates III, *Rapid Commun. Mass Spectrom.* **7,** 20 (1993); J. C. Schwartz and M. E. Bier, *Rapid Commun. Mass Spectrom.* **7,** 27 (1993); D. M. Chambers, D. E. Goeringer, S. A. McLuckey, and G. L. Glish, *Anal. Chem.* **65,** 14 (1993); J. Qin and B. T. Chait, *J. Am. Chem. Soc.* **117,** 5411 (1995); J. Qin, R. J. J. M. Steenvoorden, and B. T. Chait, *Anal. Chem.* (in press).

[30] See, e.g., Y. Li, R. T. McIver, Jr., and R. L. Hunter, *Anal. Chem.* **66,** 2097 (1994).

(m/z) corresponding to a particular peak can be calculated using simple equations and the time-to-intensity data can be converted into an m/z-to-intensity histogram.

Ions do not emerge from an MALDI ion source with exactly the same kinetic energy. This kinetic energy spread results in relatively low mass resolutions in linear time-of-flight analyzers. This effect can be partially corrected using an ion mirror to reflect the ions back toward a detector positioned near the ion source.[31] An analyzer with an ion mirror is referred to as a "reflectron." The mass resolution and mass accuracy of a reflectron analyzer can be superior to linear analyzers for $m/z < 10{,}000$. Above this mass-to-charge ratio, the performance of the two types of analyzers is comparable because large ions decay into smaller fragments in flight. This unimolecular decay broadens the signal in reflectrons but has little effect on linear analyzers. Similar improvements in resolution can be obtained using a pulsed extraction ion source[31a,31b] and a linear time-of-flight analyzer.

Ions are converted into signals using ion-to-electron conversion detectors. These detectors have a flat plate surface that is exposed to the incoming protein ions. The ions strike the surface (either a metal or metal oxide coating) and the collision results in the prompt emission of electrons and small negative ions. These electrons and ions are accelerated to another surface that produces electrons exclusively. These electrons are amplified using conventional electron multipliers (microchannel plates or discreet dynode multipliers). *Note:* Multipliers that are easily paralyzed by intense protein or matrix ion signals should not be used.[32]

Mass Calibration

After preparing a sample and obtaining a mass spectrum, one is confronted with the task of assigning masses to the peaks in the spectrum. Time-of-flight mass spectra are recorded as intensity versus time-of-flight tables. The time-of-flight data are converted into mass-to-charge ratio (m/z) data by using a simple equation:

$$(m/z)^{1/2} = at + b$$

The constants a and b can be calculated from the known geometry of an ion source and mass analyzer; however, in practice this calculation is rarely

[31] X. Tang, R. C. Beavis, W. Ens, F. Lafortune, B. Schueler, and K. G. Standing, *Int. J. Mass Spectrom. Ion Processes,* **85,** 43 (1988).

[31a] R. S. Brown and J. J. Lennon, *Anal. Chem.* **67,** 1998 (1995).

[31b] M. L. Vestal, P. Juhasz, and S. A. Martin, *Rapid Commun. Mass Spectrom.* **9,** 1044 (1995).

[32] R. C. Beavis and B. T. Chait, Methods and mechanisms for producing ions from large molecules. NATO ASI Series B. *in* "Physics" (K. G. Standing and W. Ens, eds.), Vol. 269, p. 227. Plenum Press, New York, 1991.

performed. A simpler and more accurate method of determining a and b is to select two peaks of known m and z and derive the constants from their measured flight times. These values of a and b can be used to calibrate other spectra obtained using the same instrument and voltage settings. Applying a set of calibration constants measured in a previous experiment is known as using an "external" calibration.

External calibrations are not always accurate. If experimental conditions change slightly, the use of an external calibration will result in mass assignments that are slightly incorrect, producing a systematic error in mass assignment. This problem can be corrected using known peaks in each spectrum to calculate a calibration for that particular data set, a procedure known as using an "internal" calibration. Internal calibrations are inherently more accurate than external calibrations; however, they require at least two peaks of accurately known mass in every spectrum, which can be difficult to arrange.

The most popular method for the internal calibration of a protein mass spectrum is to mix a known protein into the sample solution. The calibrant protein is added at a concentration such that it produces a signal with an intensity similar to the analyte of interest. Proteins produce ions with different charge states, usually $z = +1$ and $+2$, although there may be signals from ions with $z \gtrsim +3$. These charge states produce signals at predictable m/z values. For a protein of mass M, the series of m/z peaks produced are $z = 1$, $m/z = M + 1$; $z = 2$, $m/z = (M + 2)/2$; $z = 3$, $m/z = (M + 3)/3$, etc. Therefore a single protein calibrant usually produces at least two peaks that can be used to calculate the calibration constants. It may be necessary to try several concentrations of calibrant before a spectrum is obtained with both the calibrant and analyte proteins at similar intensities so that a good calibration can be obtained.

Calibrant proteins most commonly used include bovine insulin, horse muscle myoglobin, horse muscle cytochrome c, and bovine trypsinogen. These proteins are used because they normally occur in only one isoform and pure samples can be obtained from commercial suppliers at low cost. Also, they do not oxidize easily and commercial samples do not contain significant protease activity.

It must be pointed out that good mass assignments depend on good signals. If a signal is weak and noisy, the mass assignment will not be as accurate as for intense peaks with large signal-to-noise ratios. The signal-to-noise ratio can be improved by adding together digitized signal transients (spectra) from many individual laser shots taken on the same sample spot. A general rule of thumb is that digitized signal transients from 50 to 100 individual laser shots should be summed to obtain an acceptable signal-to-noise ratio.

Mass Spectrometric Strategies for Studying Proteins

Correlation of Processed Proteins with Their Genes by Molecular Mass Measurements

A simple, accurate molecular mass measurement can provide highly useful data concerning a protein sample. The resulting mass spectrum can give information concerning (1) the correctness of a hypothetical structure, (2) the purity of the protein preparation as well as data on the masses of any impurities that are present, (3) protein microheterogeneity, and (4) the presence and likely identities of posttranslational modifications.

Once the cDNA sequence of a gene has been determined, a simple, accurate measurement of the molecular mass 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 residues. However, a difference between the measured and predicted molecular masses implies either an error in the cDNA-deduced sequence or a posttranslational modification of the protein. The power of accurate molecular mass determination in this context is illustrated by consideration of ARPP-16, a cAMP-regulated phosphoprotein from the soluble fraction of bovine caudate nuclei.[33] Analysis by SDS–PAGE of the protein yielded an apparent molecular mass of 16,000 Da. The amino terminus was chemically blocked and could not be sequenced by Edman degradation. Inspection of the cDNA sequence suggested a molecular mass considerably lower than that inferred from the SDS–PAGE analysis.[33] The MALDI mass spectrum obtained from the unphosphorylated form of ARPP-16 is shown in Fig. 2. 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 10,709 Da. 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 molecular mass of 10,665 Da). The difference between the measured and calculated masses is 44 ± 2 Da, a value consistent with the presence of an acetyl-blocking group (calculated mass difference of 42 Da) at the amino terminus. Taken together with the cDNA sequence data and partial amino acid sequence data, the simple mass measurement provides compelling confirmation of

[33] A. Horiuchi, K. R. Williams, T. Kurihara, A. C. Nairn, and P. Greengard, *J. Biol. Chem.* **265,** 9576 (1990).

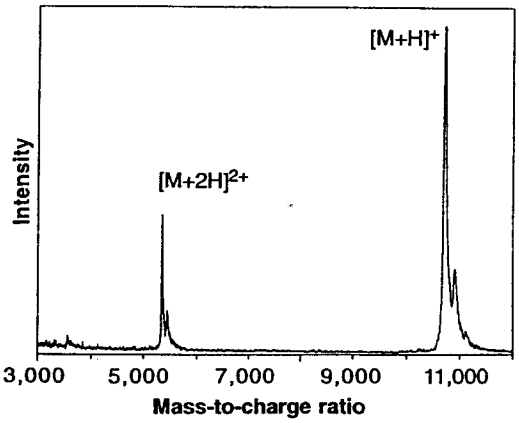


FIG. 2. Positive ion matrix-assisted laser desorption mass spectrum of the protein ARPP-16, obtained from a matrix of sinapic acid using the dried-droplet sample preparation method (see text). M designates the intact molecule. The small satellite peaks to the right of the main peaks arise from a photochemically produced adduct (see text). [Reproduced from B. T. Chait and S. B. H. Kent, *Science* **257,** 1885 (1992).]

the predicted primary structure of ARPP-16 as well as information concerning a subtle modification of the protein. This detailed information was obtained rapidly (in less than 15 min) from a total quantity of 1 pmol of protein.

Frequently, differences are observed between the measured and calculated molecular masses that are more difficult to interpret than in the example of ARPP-16 given above. The difficulty in defining the source of these mass differences arises from a number of possible sources. Some of these are as follows.

The mass spectrometric response may be too weak for accurate definition of the mass spectral peak centroids. A weak mass spectrometric response is normally the result of insufficient protein sample, failure to properly dissolve and manipulate the sample, or the presence of impurities.

The uncertainty in the mass determination accuracy increases as a function of the molecular mass of the protein. For proteins with molecular masses <30,000 Da, the mass accuracy is in the range of 0.01–0.1%. For proteins with molecular masses >30,000 Da, the mass accuracy is in the range of 0.03–0.3%.

The protein may have a high degree of heterogeneity and the individual components may not be resolved in the mass spectrum. This situation is commonly encountered for glycoproteins.

In all such cases, interpreting the source of mass differences can often be facilitated by an analytical strategy that involves degradation of the protein by chemical or enzymatic means and measurement by MALDI–MS of the total mixture of peptide products (see below).

Identification of Sites of Proteolytic Cleavage of Proteins

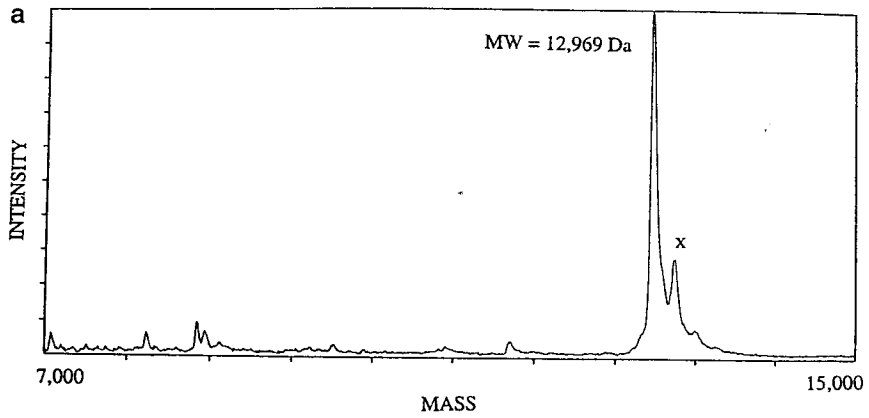
An accurate molecular mass determination provides a method of choice for the rapid and reliable identification of sites of proteolytic processing or degradation in proteins. If the protease responsible for the cleavage(s) under study is known and has high specificity, a simple molecular mass determination may be sufficient to yield an unambiguous identification of the portion of the protein produced. If, however, the activity of the protease is unknown or broad, it may be advantageous to define the amino terminus of the cleaved protein by Edman sequencing of a few residues. The carboxyl terminus (and hence the whole portion of the protein of interest) can then be defined without ambiguity by an accurate molecular mass measurement.

The principle of the MS analysis described above is illustrated by consideration of a preparation of streptavidin from *Streptomyces avidinii* that has undergone proteolytic cleavage.[34] The MALDI mass spectrum of this preparation (obtained from Boehringer Mannheim, Indianapolis, IN) is shown in Fig. 3a[35] together with the sequence of the intact precursor subunit (Fig. 3b). Edman degradation of the preparation defines the amino-terminal residues as EAGIT..., showing that the amino terminus begins at residue 14 (Fig. 3b). The molecular mass of the dominant component in the preparation was measured to be 12,969 ± 2 Da. With the knowledge that the amino terminus begins at residue 14, we can calculate (using the known sequence of streptavidin) the molecular masses of the modified protein with all possible C-terminal truncations. The molecule containing residues 14–136 has a calculated molecular mass of 12,971 Da, in good agreement with the measured value. Thus the accurate molecular mass determination has defined unambiguously the C terminus of this modified streptavidin preparation. The sequence of the modified streptavidin is underlined in Fig. 3b. An analogous strategy can be used to define the N terminus if information concerning the C terminus is available. Such a strategy is of particular use for N-terminally blocked proteins.

Assessing Homogeneity and Purity of Protein Preparations

It is often necessary to assess the homogeneity and purity of native, recombinant, or synthetic protein preparations. MALDI–MS provides one

[34] E. A. Bayer, H. BenHur, Y. Hiller, and M. Wilchek, *Biochem. J.* **258,** 369 (1989).
[35] B. T. Chait, *Structure* **2,** 465 (1994).



b

14

DPSKDSKAQVSAAEAGITGTWYNQLGSTFIVTAGADGALTGTYESAV

GNAESRYVLTGRYDSAPATDGSGTALGWTVAWKNNYRNAHSATTW

136

SGQYVGGAEARINTQWLLTSGTTEANAWKSRLVGHDTFTKVKPSAA

SIDAAKKAGVNNGNPLDAVQQ

FIG. 3. Mass spectrometry to identify proteolytic cleavage sites. (a) The positive ion MALDI–MS of proteolytically modified streptavidin from *Streptomyces avidinii,* using a sinapic acid matrix and the dried droplet sample preparation method. The peak labeled "x" is an artifact of the MALDI process (see text) and does not arise from an impurity in the sample. However, several small peaks observed in the mass spectrum below mass 12,000 Da do arise from low-abundance impurities in the sample. Interestingly, these impurities were observed to be absent from the mass spectrum of the protein after crystallization (S. E. Darst and B. T. Chait, unpublished data, 1995). (b) The sequence of streptavidin. The sequence of the proteolytically modified streptavidin is underlined. [Reproduced from B. T. Chait, *Structure* **2,** 465 (1994). Copyrighted by Current Science.]

of the most rapid, straightforward, and informative routes for such assessment, which is usually carried out by molecular mass determination followed by more detailed peptide mapping (if necessary). Examples include verification of recombinant[36,37] and synthetic proteins,[38,39] assessment of the

[36] T. M. Billegi and J. T. Stults, *Anal. Chem.* **65,** 1709 (1993).
[37] J. J. M. DeLlano, W. Jones, K. Schneider, B. T. Chait, J. M. Manning, G. Rodgers, L. J. Benjamin, and B. Weksler, *J. Biol. Chem.* **268,** 1 (1993).
[38] R. C. D. Milton, S. C. F. Milton, and S. B. H. Kent, *Science* **256,** 1445 (1992).
[39] B. T. Chait, R. Wang, R. C. Beavis, and S. B. H. Kent, *Science* **261,** 89 (1993).

integrity and purity of proteins used to grow crystals for X-ray diffraction measurements,[35,40] and comparison of different forms of a protein having distinct biological activities.

The utility of MALDI–MS for assessing the integrity and purity of a protein preparation is illustrated by examination of an artificial protein expressed in *Escherichia coli*.[41] A synthetic DNA was constructed by multimerizing oligonucleotide fragments encoding two copies of the repeated undecapeptide of the target protein **I**.

ASMTGGQQMGRDPMFKYSRDPMG-[AGAGAGAGPEG]$_{14}$-ARMHIRPGRYQLDPAANKARKEAELAAATAEQ

I

Amino acid analysis confirmed the composition of purified **I**, and Coomassie blue staining of SDS–PAGE gels revealed no contaminating proteins—the product was observed to migrate as a single tight band. However, the molecular mass reported by SDS–PAGE was found to be 43,000 Da, which is more than twice the expected molecular mass of 17,207 Da. To resolve the apparent molecular weight discrepancy, the protein was analyzed by MALDI–MS.

The mass spectrum of **I** is shown in Fig. 4A. Although the measured m/z (17,264 ± 2) shows the electrophoretically determined mass to be grossly in error, the observed value remains significantly different from the predicted m/z of 17,208. To determine the origin of the difference, the sequence of a 741-base pair DNA fragment containing the protein-coding region of the expression plasmid was determined. This analysis revealed C → T transitions in codons 96 and 101, causing two alanine → valine substitutions in the expressed protein. The calculated m/z of the protein with the altered repetitive sequence **II** is 17,264—the experimentally determined value.

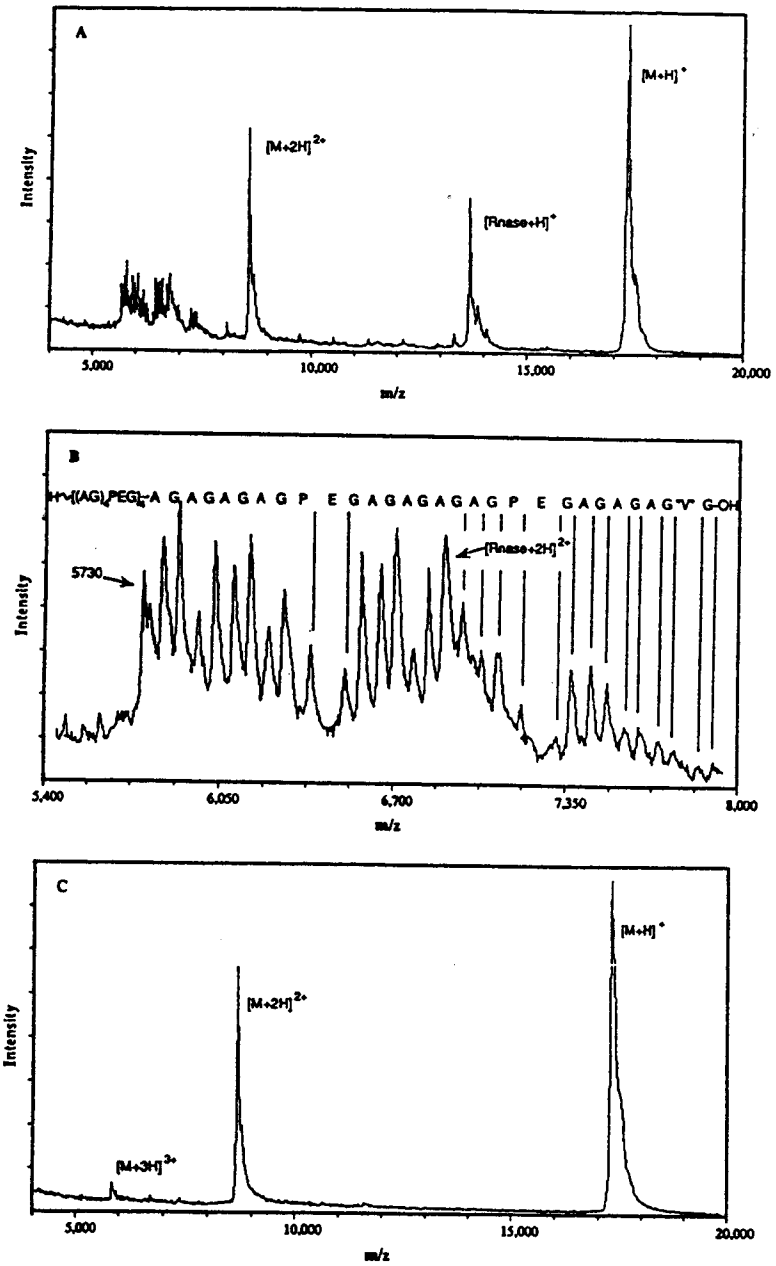
-[AGAGAGAGAGPEG]$_6$[AGAGAGVGPEG][VGAGAGAPEG][AGAGAGAGAGPEG]$_6$-

II

Signals arising from low molecular weight polypeptides are also visible in the spectrum shown in Fig. 4A. Analysis of these signals, expanded in Fig. 4B, shows that the peak at m/z 5730 corresponds to fragment **III**, which consists of the intact N-terminal sequence of **I** followed by four copies of the repeating undecapeptide. The calculated mass of the molecular ion derived from **III** is 5733, in agreement with the observed value.

[40] K. L. Clark, E. D. Halay, E. Lai, and S. K. Burley, *Nature* (*London*) **364,** 412 (1993); X. Xie, T. Kokubo, S. L. Cohen, U. A. Mirza, A. Hoffman, B. T. Chait, R. G. Roeder, Y. Nakatani, and S. K. Burley, *Nature* (in press).

[41] R. C. Beavis, B. T. Chait, H. S. Creel, M. J. Fournier, T. L. Mason, and D. A. Tirrell, *J. Am. Chem. Soc.* **114,** 7384 (1992).


A
[M+H]+
[M+2H]2+
[Rnase+H]+
Intensity
5,000
10,000
15,000
20,000
m/z
B
H-V-[(AG)2PEG]2-A G A G A G A G P E G A G A G A G A G P E G A G A G A G V G-OH
5730
[Rnase+2H]2+
Intensity
5,400
6,050
6,700
7,350
8,000
m/z
C
[M+H]+
[M+2H]2+
[M+3H]3+
Intensity
5,000
10,000
15,000
20,000
m/z

-ASMTGGQQMGRDPMFKYSRDPMG-[AGAGAGAGPEG]$_4$-
III

More striking, however, is the fact that each succeeding signal can be rationalized by addition of a single amino acid residue, proceeding in the N- to C-terminal direction along sequence **I**. Thus, one can read portions of the periodic sequence directly from the mass spectrum, including one of the substituted valines. Although the origin of these fragments has not yet been determined, they probably arise from the action of exo- and endopeptidases, either *in vivo* and *in vitro* or both. The fragments are easily removed by dialysis as shown in Fig. 4C, which shows only the singly, doubly, and triply ionized molecular ions of the intact protein.

Subsequent to the discovery of the contaminating fragments, the sample was rerun on a 25-cm 15% polyacrylamide gel, and the proteins were visualized by Coomassie blue staining. The fragments were not detected by this method, even when the sample was overloaded (up to 50 μg of protein per lane).

Defining Sites and Nature of Posttranslational Modifications by Peptide Mapping

To determine the modifications in a protein, its accurately measured molecular mass is compared with the molecular mass calculated from the cDNA sequence of the corresponding gene or the amino acid sequence of the protein. The difference between the measured and calculated molecular masses provides information concerning the possible nature of the modification(s). If more detailed information concerning the nature and site of a modification is required, it is usually necessary to subject the protein to further analysis, involving enzymatic or chemically induced degradation of the protein followed by further mass spectrometric measurement of the resulting peptide fragments. Partial definition of a posttranslational modification is illustrated by reference to the *a*-chain of the giant extracellular

FIG. 4. Analysis of an artificial protein by MALDI–MS. (A) Positive ion spectrum of the target protein **I** with bovine ribonuclease A as an internal calibrant. The dried-droplet sample preparation method was used with a sinapic acid matrix. (B) Spectrum expanded in the region of low molecular mass contaminants. The peaks progress starting with N-terminal fragment **III** (calculated m/z 5733, observed m/z 5730) with the sequential addition of amino acid residues through three repeats of the target repetitive sequence. The substituted valine is apparent in the seventh repeat. (C) Mass spectrum of the protein sample after removal of the contaminating fragments by dialysis. (Reprinted from R. C. Beavis *et al., J. Am. Chem. Soc.* **114,** 7384. Copyright 1992, American Chemical Society.)

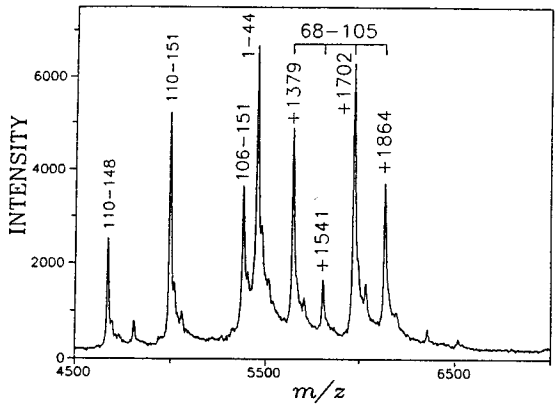


FIG. 5. Positive-ion MALDI–MS spectrum of the *a* chain of the giant extracellular hemoglobin from the earthworm *Lumbricus terrestris* after reduction with dithiothreitol and digestion with endoproteinase Lys-C.[43] The dried droplet sample preparation method was used with a matrix of α-cyano-4-hydroxycinnamic acid. Peptide 68–105 was found to be glycosylated.

hemoglobin of the earthworm, *Lumbricus terrestris*.[42] Measurement of the molecular mass of the *a*-chain revealed that it was heterogeneous, having three major components (a_1, 19,386 ± 15 Da; a_2, 19,221 ± 10 Da; and a_3, 18,901 ± 15 Da) with significantly higher molecular masses than that calculated from the known sequence (17,525 Da). The heterogeneity and the differences between these values suggest that the protein is glycosylated and that a_1 has one more hexose than a_2, which has two more hexoses than a_3. More detailed information concerning the site of glycosylation was obtained by reduction of the *a*-chain with dithiothreitol and digestion with Lys-C endoproteinase. The mass spectrum of the unfractionated digest (between *m/z* 4500 and 7000) is given in Fig. 5.[43] The various fragments can be readily assigned because the digesting enzyme is highly specific and because the interpretation of the mass spectrum is simplified by the dominance of singly charged ions. Comparison of the experimentally determined peptide fragment masses with the masses of the unmodified fragments calculated from the sequence reveals that the protein is glycosylated on the peptide spanning residues 68–105. This detailed analysis of proteolytic fragments reveals the presence of four components that differ from one another by integral numbers of hexoses (hexose residue mass of 162 Da).

[42] D. W. Ownby, H. Zhu, K. Schneider, R. C. Beavis, B. T. Chait, and A. F. Riggs, *J. Biol. Chem.* **268,** 13539 (1993).

[43] K. Schneider, D. W. Ownby, H. Zhu, R. C. Beavis, A. F. Riggs, and B. T. Chait, *in* "Proceedings of 41st ASMS Conference in Mass Spectrometry, San Francisco, 1993," p. 900a.

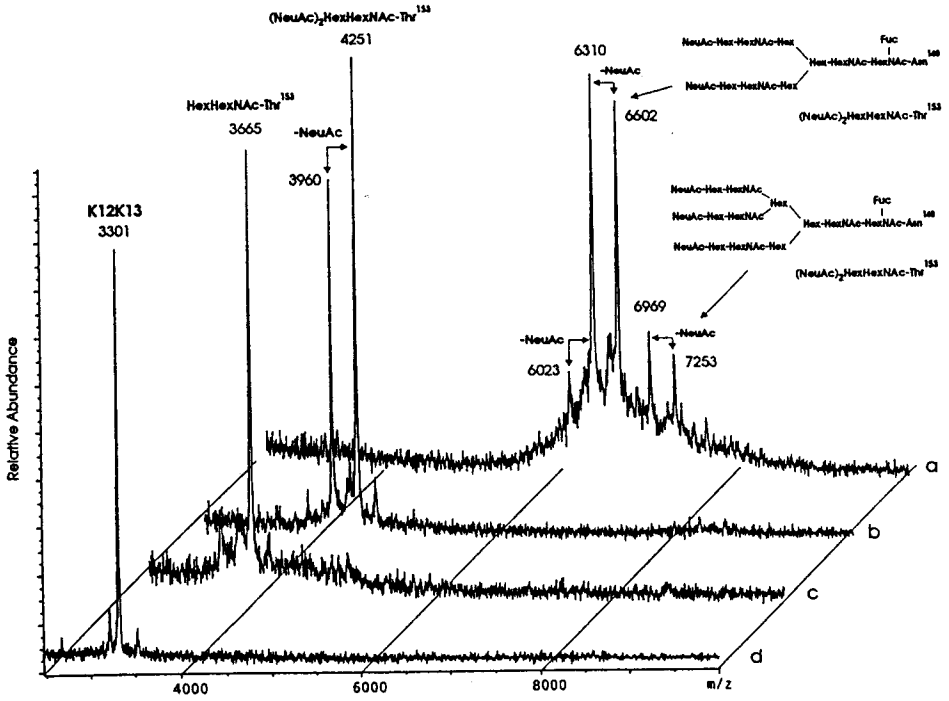


FIG. 6. Positive-ion MALDI–mass spectra of the glycopeptide encompassing residues 138–163 of recombinant human macrophage colony-stimulating factor containing both N- and O-linked oligosaccharides prior to and after treatment with a series of glycosidases. (a) The intact peptide, which was then sequentially treated with (b) PNGase F, (c) neuraminidase, and (d) *O*-glycanase. The major carbohydrate structures attached to the peptide are shown for each step. The extent of glycosylation and the proposed carbohydrate compositions may be determined on the basis of the observed mass differences from one step to the next and the known specificity of the glycosidase. (Reprinted with permission from M. C. Huberty *et al., Anal. Chem.* **65,** 2791. Copyright 1993, American Chemical Society.)

The difference between these masses and the sequence-derived mass may be accounted for by glycosylation involving 8–12 hexoses. Peptide 68–105 contains three possible sites for O-linked glycosylation. Precise definition of the site of glycosylation requires further analysis, e.g., proteolytic subdigestion or sequencing. Likewise, more detailed definition of the composition and structure of the attached carbohydrate requires further analysis. An example of an analysis that provides such compositional information is given below.

Information concerning the carbohydrate portion of a protein can be

obtained by mass determination of component glycopeptides prior to and after treatment with specific glycosidases. This approach is illustrated by analysis of a glycopeptide obtained by digestion of recombinant human macrophage colony-stimulating factor using *Achromobacter* protease I.[44] The MALDI–MS spectrum of the intact glycopeptide (containing both N- and O-linked oligosaccharides) encompassing residues 138–163 of the protein consists of a complex distribution of peaks in the range *m/z* 6000–8000 (Fig. 6a). The two most intense peaks, at *m/z* 6310 and 6602, differ by 292 Da, suggesting the presence of neuraminic acid (291 Da). Treatment of the glycopeptide with PNGase-F (*Flavobacterium meningosepticum*) produces two main products with *m/z* 3960 and 4251 (Fig. 6b), having masses 2351 Da lower than the major components in Fig. 6a. This observation demonstrates that the removed carbohydrate is asparagine linked and provides a measure of the masses of components of the complex carbohydrate. Treatment of the residual glycopeptide with neuraminidase (*Arthrobacter ureafaciens*) removes the neuraminic acid moieties to produce a homogenous product at *m/z* 3665 (Fig. 6c). Finally, treatment of this product with *O*-glycanase (*Streptococcus pneumoniae*) reduces the molecular mass by 364 Da to *m/z* 3301 (Fig. 6d), revealing the presence of a site of O-glycosylation. The molecular mass of the residual peptide is consistent with the calculated mass of the protonated peptide with a deamidated asparagine residue and no further modifications. Therefore, it appears that all of the modifications have been removed by the glycosidase treatments. Compositional analysis of the oligosaccharides can be deduced from these measured mass differences either *ab initio* or by reference to compositional databases.[44]

Peptide Maps of Proteins Eluted from One- and Two-Dimensional Gels

The high sensitivity of MALDI–MS permits the measurement of digestion products generated from proteins separated by one- and two-dimensional (2D) gel electrophoresis.[18,45,48] The combination of electrophoresis

[44] M. C. Huberty, J. E. Voth, W. Yu, and S. A. Martin, *Anal. Chem.* **65,** 2791 (1993).

[45] Y. K. Wang, P.-C. Liao, J. Allison, D. A. Gage, P. C. Andrews, D. M. Lubman, S. M. Hanash, and J. R. Strahler, *J. Biol. Chem.* **268,** 14269 (1993); W. Henzel, T. B. Billeci, J. T. Stults, S. C. Wong, C. Grimley, and C. Watanabe, *in* "Techniques in Protein Chemistry V" (J. W. Crabb, ed.), p. 3. Academic Press, San Diego, CA, 1994; S. Geromanos, P. Casteels, C. Elicone, M. Powell, and P. Tempst, *in* "Techniques in Protein Chemistry V" (J. W. Crabb, ed.), p. 143. Academic Press, San Diego, CA, 1994; S. D. Patterson and R. Aebersold, *Electrophoresis* **16,** 1791 (1995).

[46] Deleted in proof.

[47] Deleted in proof.

[48] W. Zhang, A. J. Czernik, T. Yungwirth, R. Aebersold, and B. T. Chait, *Protein Sci.* **3,** 677 (1994).

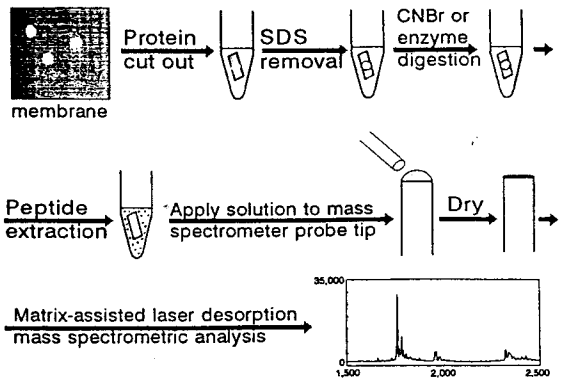


FIG. 7. Strategy employed for the mass spectrometric mapping of peptides generated from proteins electroblotted onto Immobilon CD membranes from two-dimensional electrophoretic gels. [Reprinted from W. Zhang *et al., Protein Sci.* **3,** 677 (1994).]

and mass spectrometry is an effective means for analyzing relatively insoluble proteins and is particularly powerful for assessing the differences between protein isoforms that exhibit, for example, different phosphorylation states. One variation of this approach[48] involves electroblotting of proteins separated by 2D electrophoresis onto a membrane with a cationic surface. The isolated proteins are subjected to chemical and/or enzymatic degradation directly on the membrane, and the resulting unfractionated mixture of peptide fragments is extracted from the membrane into a solution that is compatible with analysis by MALDI–MS. Accurate mass determination of these peptide fragments provide a facile means for detecting the presence of modifications and for correlating such modifications with the differential mobility of different isoforms of a given protein during 2D electrophoresis. The strategy employed for mapping peptides from proteins separated by 2D electrophoresis and electroblotted onto cationic membranes (Immobilon CD; Millipore, Bedford, MA) is outlined in Fig. 7.[48] Protein spots visualized by reverse staining of the blotting membrane are excised, washed, and subjected to chemical or enzymatic digestion. The resulting mixture of peptide fragments is extracted from the membrane, combined with the laser desorption matrix, and analyzed without fractionation.[48a,48b]

Protein Identification from Peptide Maps

A need exists for rapid, sensitive means of identifying proteins that have been isolated on the basis of their biological activity, response to

[48a] A. Sherchenko, M. Wilm, O. Vorm, and M. Mann, *Anal. Chem.* **68,** 850 (1996).
[48b] J. Qin, Doctoral Dissertation, The Rockefeller University, 1986.

stimuli, or specific association with other biomolecules. Such identifications are now frequently made using partial protein sequence information. MALDI–MS provides the basis for a new strategy for protein identification that is fast, sensitive, and accurate.[18,49] This strategy involves mass determination of peptide fragments generated from the protein of interest by an enzyme or chemical reagent with high specificity, followed by screening of these masses against an appropriate database of peptide fragments. The fragment database is calculated from a protein database using the known properties of the cleavage reagent. Proteins can be correctly identified with a relatively limited set of proteolytic peptides, and the approach has sufficient sensitivity for use with proteins separated by 2D electrophoresis (see above). This approach has been improved by using tandem mass spectrometry of the proteolytic peptides to introduce additional constraints for the database search.[51a,52]

Probing Protein–Protein Interactions

The use of MALDI–MS in conjunction with affinity-based biochemical techniques provides potentially powerful new means for probing protein–protein and protein–ligand interactions.[52] The concept is illustrated by reference to a method for rapid mapping of linear epitopes in proteins that are bound by monoclonal antibodies.[53,54] The method consists of three steps (Fig. 8).

The approach is demonstrated through the mapping of a binding epitope in the peptide melittin against a monoclonal antibody that was previously determined to bind to an epitope located at residues 20–26 of melittin.[55,56] Partial digestion of melittin by endoprotease Lys-C yielded four peptide

[49] M. Mann, P. Hojrup, and P. Roepstorff, *Biol. Mass Spectrom.* **22,** 338 (1993); D. J. C. Pappin, P. Hojrup, and A. J. Bleashy, *Curr. Biol.* **3,** 327 (1993); P. James, M. Quadroni, E. Carapoli, and G. Gonnet, *Biochem. Biophys. Res. Commun.* **195,** 58 (1993).
[50] Deleted in proof.
[51] Deleted in proof.
[51a] M. Mann and M. Wilm, *Anal. Chem.* **66,** 4390 (1994); J. R. Yates III, J. K. Eng, A. L. McCormack, and D. Schieltz, *Anal. Chem.* **67,** 1426 (1995); K. R. Clauser, S. C. Hall, D. M. Smith, J. W. Webb, L.-E. Andrews, H. M. Tran, L. B. Epstein, and A. L. Burlingame, *Proc. Natl. Acad. Sci. U.S.A.* **92,** 5072 (1995).
[52] T. W. Hutchens and T-T. Yip, *Rapid Commun. Mass Spectrom.* **7,** 576 (1993).
[53] D. Suckau, J. Kohl, G. Karwath, K. Schneider, M. Casaretto, D. Bitter-Suermann, and M. Przybylski, *Proc. Natl. Acad. Sci. U.S.A.* **87,** 9848 (1990).
[54] Y. Zhao and B. T. Chait, *Anal. Chem.* **66,** 3723 (1994); Y. Zhao, T. W. Muir, S. B. H. Kent, E. Tischer, J. M. Scardina, and B. T. Chait, *Proc. Natl. Acad. Sci. U.S.A.* (in press.)
[55] T. P. King, L. Kochoumian, and A. Joslyn, *J. Immunol.* **133,** 2668 (1984).
[56] P. F. Fehlner, R. H. Berg, J. P. Tam, and T. P. King, *J. Immunol.* **146,** 799 (1991).

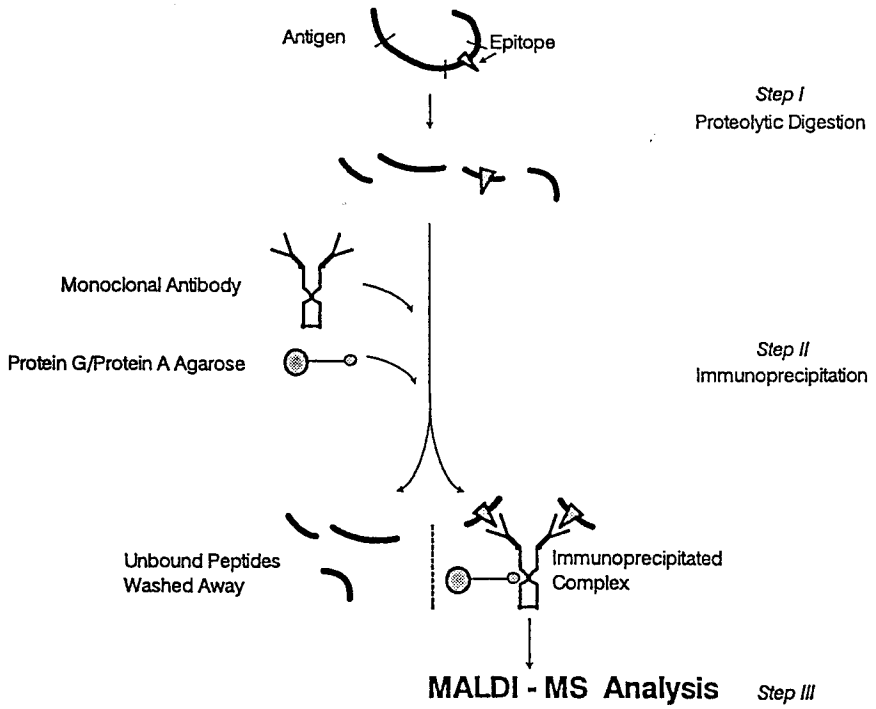


FIG. 8. Three-step strategy for linear epitope mapping. In the first step, an antigen–protein complex is digested by a proteolytic enzyme to produce an appropriate set of peptide fragments. In the second step, peptide fragments containing the linear epitope are selected and separated from the pool of peptide fragments by immunoprecipitation with the monoclonal antibody. In the final step, the immunoprecipitated peptides are identified by MALDI–MS. The method allows the rapid determination of antigenic sites without tedious peptide synthesis or protein mutagenesis. [Reprinted with permission from Y. Zhao and B. T. Chait, *Anal. Chem.* **66,** 3723 (1994).]

fragments that gave intense mass spectral peaks (Fig. 9A). The measured molecular masses of these peptides correspond to the molecular masses predicted for melittin fragments 8–23, 8–26, 1–23, and 1–26 (Fig. 9C). The other two peaks in the spectrum (designated i_1 and i_2) arise from unidentified impurities that were present in the melittin sample. After immunoprecipitation with the antibody, the two impurities and peptide fragments 1–23 and 8–23 were washed away, while fragments 1–26 and 8–26 were identified in the immunoprecipitation complex (Fig. 9B). To further define the antigenic site, an analogous experiment was carried out after chymotrypsin digestion of melittin. The experiment showed that only one of the three chymotryptic peptides was bound by the antibody (Fig. 9C). Inspection of

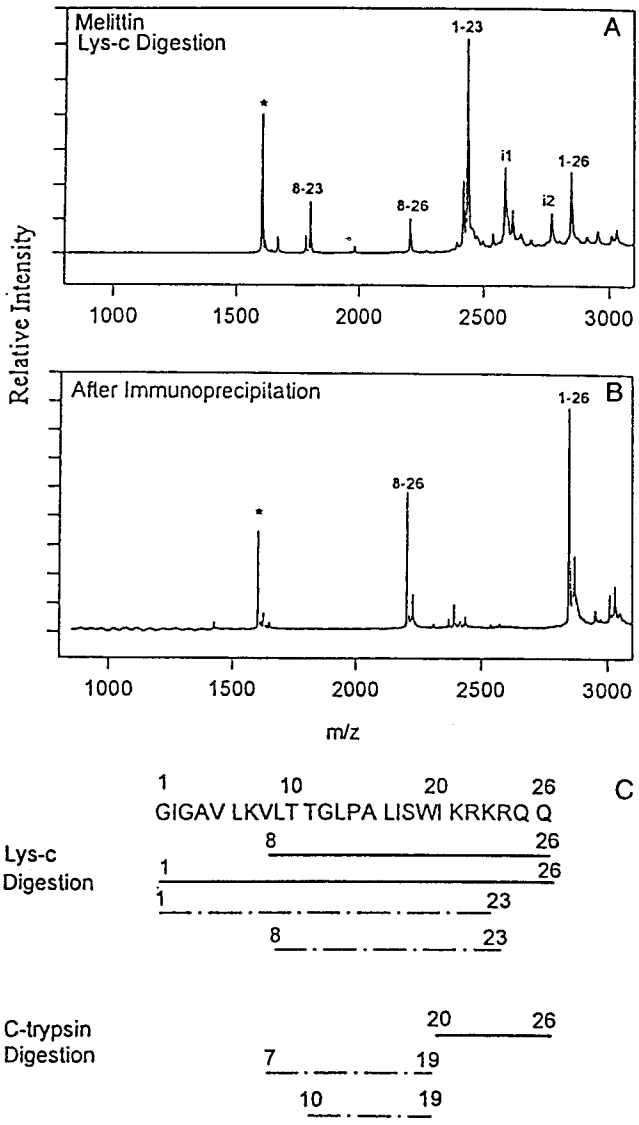


FIG. 9. Positive ion MALDI–MS spectra of (A) peptide fragments produced by endoprotease Lys-C digestion of melittin and (B) peptide fragments isolated by immunoprecipitation with the anti-melittin monoclonal antibody (see text). The peak labeled with an asterisk (*) corresponds to the internal calibrant dynorphin A 1–13, and peaks i_1 and i_2 correspond to unidentified impurities. (C) Amino acid sequence of melittin. The peptide fragments produced by endoprotease Lys-C and chymotrypsin digestion are indicated by lines. The solid lines represent those peptide fragments that were bound by the anti-melittin monoclonal antibody, and the dashed lines represent those that were not bound. [Reprinted with permission from Y. Zhao and B. T. Chait, *Anal. Chem.* **66**, 3723 (1994).]

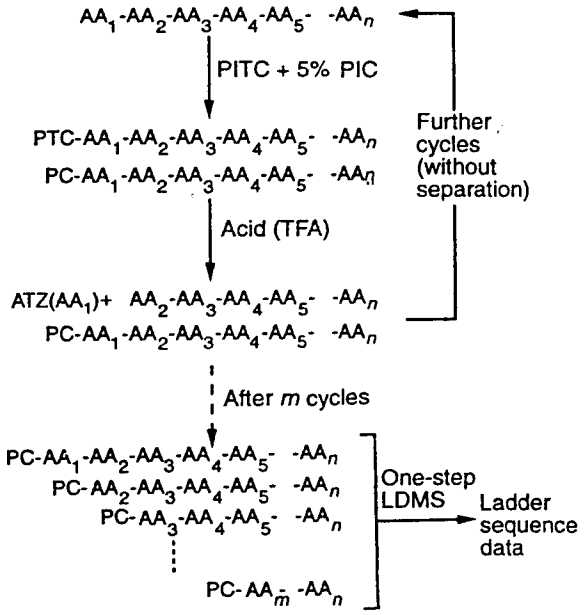


FIG. 10. Protein ladder sequencing principle exemplified by the generation of a set of sequence-determining fragments from an intact peptide chain with controlled ladder-generating chemistry. A stepwise degradation is carried out with a small amount of terminating agent present in the coupling step. In this case, 5% phenylisocyanate (PIC) was added to the phenylisothiocyanate (PITC). The phenylcarbamyl (PC) peptides formed are stable to the trifluoroacetic acid (TFA) used to cyclize and cleave the terminal amino acid (AA) from the phenylthiocarbamyl (PTC) peptide. Successive cycles of ladder-generating chemistry are performed without intermediate isolation or analysis of released amino acid derivatives. Finally, the mixture of PC peptides is read out in one step by MALDI–MS. [Reprinted with permission from B. T. Chait and S. B. H. Kent, *Science* **257,** 1885 (1992). © AAAS.]

the combined data (Fig. 9C) demonstrates that the region encompassing residues 20–26 is sufficient for binding to this monoclonal antibody.

Probing Protein Structure by Wet Chemistry in Combination with MALDI–MS

A powerful set of strategies for probing structural aspects of proteins uses the combination of appropriate wet chemical manipulation of the proteins followed by MALDI–MS measurements of the resulting chemical changes. Examples include strategies for amino acid sequencing; counting acid groups, amino groups, thiols, etc.; determining the solvent accessibility of particular amino acid side groups; chemical cross-linking for determining

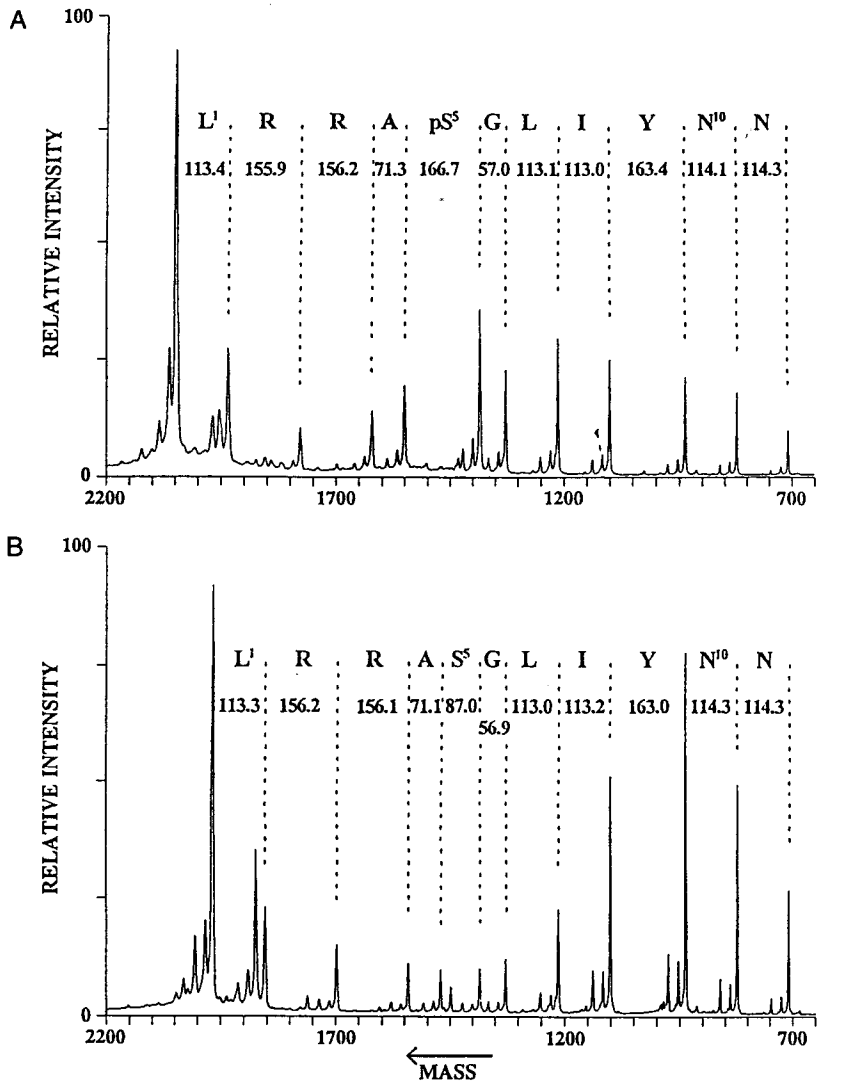


FIG. 11. Protein ladder sequencing of the 16-residue synthetic peptide: Leu-Arg-Arg-Ala-Ser(P_i)-Gly-Leu-Ile-Tyr-Asn-Asn-Pro-Leu-Met-Ala-Arg-amide. (A) Phosphorylated peptide. (B) Unphosphorylated peptide. Each peptide sample was subjected to 10 cycles of ladder-generating chemistry. Data defining the 11 N-terminal residues are shown. The Ser(P_i) residue is characterized by a mass difference of 166.7 Da observed in positive five (Ser, calculated residue mass 87.1 Da; Ser(P_i), calculated residue mass 167.1 Da). There is no evidence for loss of phosphate. [Reprinted with permission from B. T. Chait and S. B. H. Kent, *Science* **257,** 1885 (1992). © AAAS.]

the spatial relationship between proteins; and the footprinting of associated molecules on proteins. We illustrate the utility of such wet chemical manipulation in conjunction with MALDI–MS, using the technique of protein ladder sequencing.[39]

Protein ladder sequencing[39] combines multiple steps of wet degradation chemistry with a final, single-step mass spectrometric readout of the amino acid sequence. First, a sequence-defining concatenated set of peptide fragments, each differing from the next by a single amino acid residue, is chemically generated in a controlled fashion. Second, MALDI–MS is used to read out the complete fragment set in a single operation as a "protein sequencing ladder" data set.

A concatenated set of peptide fragments can be generated in a controlled fashion by carrying out rapid stepwise degradation in the presence of a small amount of terminating reagent, as indicated in Fig. 10. A small proportion of peptide chain blocked at the amino terminus is generated at each cycle. A predetermined number of cycles is performed without intermediate separation or analysis of the released amino acid derivatives. The resulting mixture is read out in a single operation by MALDI–MS. The mass spectrum contains protonated molecule ions corresponding to each terminated polypeptide species present. The mass differences between consecutive peaks each correspond to an amino acid residue, and their order of occurrence in the data set defines the sequence of amino acids in the original peptide chain.

An example of the ladder sequence analysis is shown in Fig. 11 for both phosphorylated and unphosphorylated forms of a 16-residue peptide containing a serine residue that can be phosphorylated by 3′,5′-cyclic AMP-dependent protein kinase. After 10 cycles of ladder-generating chemistry on each form of the peptide, the 2 separate sequence-defining fragment mixtures were each read out by MALDI–MS. The protein ladder-sequencing method directly identified and located a phosphoserine at position 5 in the peptide. Such direct localization and identification of posttranslational modifications are of great potential utility for biological research.

Acknowledgments

The present work was supported in part by grants from the National Institutes of Health (RR00862 and GM38274) and by the Skirball Institute of New York University.