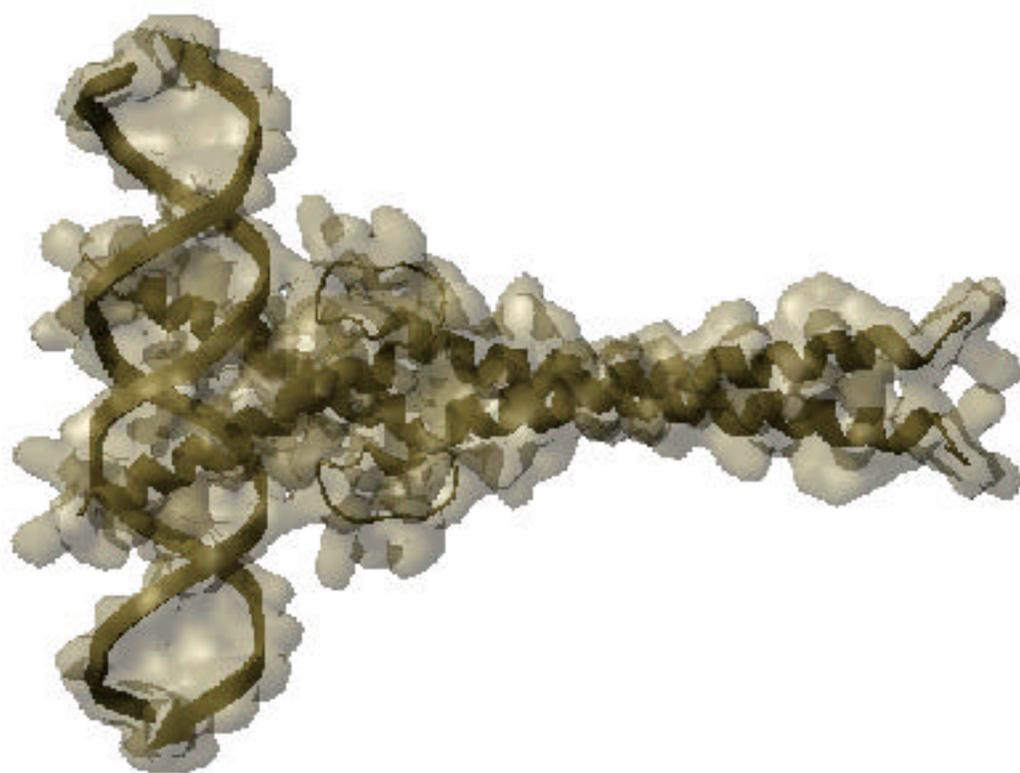


# Journal of Biomolecular Techniques

Volume 9, No. 3, September 1998



**ABRF** Online <http://www.abrf.org>

the official journal of the Association of Biomolecular Resource Facilities

Cover Illustration: Transcription factor Max/DNA complex. [see Ferre-D'Amare, et al., *Nature* 1993;363:38-45 and Kriwacki and Siuzdak, page 5 this issue]. The data file (1AN2.pdb) was obtained from the Brookhaven Protein Structure Database [Abola EE, et al., Protein Data Bank, In: Allen FH, Bergerhoff G, Sievers R, editors. Crystallographic databases-information content, software systems, scientific applications. Bonn/Cambridge/Chester: Data Commission of the International Union of Crystallography; 1987, p. 107-32; Bernstein FC, et al., The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 1977;112:535-42]. The image was produced using WebLab ViewerPro v3.0 (Molecular Simulations, Inc.).

## Combined Use Of Proteases And Mass Spectrometry In Structural Biology

Richard W. Kriwacki<sup>1</sup> and Gary Siuzdak<sup>2</sup>

<sup>1</sup>St. Jude Children's Research Hospital and <sup>2</sup>The Scripps Research Institute

Proteolysis and mass spectrometry methods have been extended to the analysis of higher-order protein structure. Proteases have long been used as probes of native structure, and this approach has been rejuvenated and used in concert with various mass spectrometry techniques. We discuss the application of protease as probes of native structure, delineate the mass spectrometry methods that are appropriate in these studies, and offer several innovative case studies to illustrate key concepts in the combined use of proteolysis and mass spectrometry in studies of biomolecular assemblies.

### Introduction

Biologic science is in the midst of an information and technical revolution because of the convergence of developments in many different areas, including genome and proteome sequencing, bioinformatics, and experimental and predictive structural biology. At the forefront of these developments are new analytic tools for meeting the demands of increased sequence data production. Prominent among these key technologies is biomolecular mass spectrometry (MS) because of the astounding sensitivity and accuracy of its protein mass determinations and the wealth of information about molecular composition inherent in a molecular mass. MS provides protein primary structure determination capabilities characterized by high accuracy and sensitivity and offers the potential for high throughput. New molecular and structural methods must be developed in order to remain astride the information revolution. The basis for these methods is the combined use of proteolytic digestion, mass analysis, and computer-based data analysis.

These methods can identify proteins, such as those from bands in acrylamide gels, using mostly automated protocols and subpicomole quantities. The target protein is completely digested using a sequence-specific protease such as trypsin (eg, within the gel matrix). The resulting fragments are extracted and prepared for mass analysis. MS analysis can yield information on fragment masses with accuracy approaching  $\pm 5$  ppm, or  $\pm 0.005$  Da for a 1000-Da peptide. The protease fragmentation pattern is compared with the patterns predicted for all proteins within a database, and matches are statistically evaluated. Because the occurrence of Arg and Lys residues in proteins is statistically high, trypsin cleavage (specific for Arg and Lys) usually produces a large number of fragments that have a reasonable prob-

ability of uniquely identifying the target protein. The success of this strategy relies on the existence of the protein sequence within the database, and with the sequences of whole genomes for several organisms complete (eg, *A. fulgidus*, *Bacillus subtilis*, *Caenorhabditis elegans*, *Escherichia coli*, *Saccharomyces cerevisiae*) and others well underway (eg, *Schizosaccharomyces pombe*, *Homo sapiens*), the likelihood for matches is reasonably high. Exact matches are readily identified, and homologous proteins are identified, albeit with lower statistical significance, placing a target protein within a particular family in the absence of an exact match.

Methods developed for primary sequence identification and elucidation using MS are particularly well suited for the analysis of higher-order, native protein structure. MS protocols used in primary structure analysis are directly transferable to the analysis of native structure, because they are used in the readout of information that is indirectly related to structure after the proteolysis reactions are performed. Analysis methods, however, must be modified to take into account the added spectral complexity resulting from incomplete proteolysis under limiting conditions. Proteases have long been used as probes of native structure, and this approach has been rejuvenated because powerful MS-based methods allow virtually complete identification of proteolysis reaction products.

We review the application of protease as probes of native structure and discuss the MS methods that are appropriate in these studies. We provide several innovative case studies that illustrate key concepts in the combined use of proteolysis and MS in studies of biomolecular assemblies.

---

Corresponding Author: Richard W. Kriwacki, Department of Structural Biology, St. Jude Children's Research Hospital, 332 N. Lauderdale, Memphis, TN 38105-2794, Tel: (901) 495-2917, Fax: (901) 495-3032, E-mail: richard.kriwacki@stjude.org

## Limited Proteolysis as a Probe of Higher-Order Structure

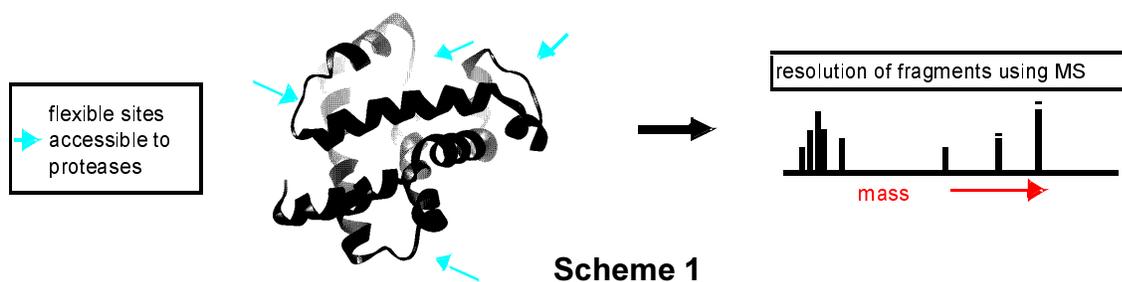
Figure 1 illustrates how proteases can be employed as probes of secondary, tertiary, and quaternary protein structure under protease-limited reaction conditions. The kinetic accessibility of a site within a protein to a protease depends on several factors, including the physical compatibility of local chemical structure with the enzyme active site (ie, sequence specificity), the accessibility of the site to the protease, and the flexibility of the site (1,2). In the analysis of protein primary structure, physical compatibility governs the selectivity of cleavage. In most cases, a sequence-specific protease is used, reducing the number of fragments that are produced, improving the likelihood for statistically significant matches between observed and predicted fragment masses, and reducing the opportunities for spurious matches. However, site accessibility and flexibility must be overcome in the analysis of primary structure so that all possible sites are cleaved and fragment maps are complete. Discrimination between cleavage sites on the basis of structure-dependent variations in accessibility and flexibility is the foundation for the analysis of native higher-order protein structure with proteases.

The arrows in Figure 1 mark potential cleavage sites within a hypothetical protein; these sites are surface exposed and located in flexible loop regions. The distribution of amino acids in a protein guides the choice of protease to be used as a structural probe. Ideally, sites should be evenly distributed throughout the sequence and have a reasonable likelihood of being accessible and flexible. Because amino acids with hydrophilic side chains are found in greater abundance on the surface of proteins at the solvent interface, proteases that cleave at hydrophilic sites are preferred in structure analysis. Trypsin and V8 pro-

tease, which cleave basic (K, R) and acidic sites (D, E), respectively, are good choices.

Table 1 lists proteases commonly used in protein structure analysis (3); more extensive lists are available elsewhere (4,5). The non-sequence-specific proteases such as subtilisin Carlsberg often are used as structural probes. This approach allows protein structure to be probed in a non-sequence-biased manner. The potential for generating large numbers of related fragments is high, complicating the analysis of mass spectral data and the identification of peptide fragments on the basis of mass. Despite this potential problem, sequence-specific and non-sequence-specific proteases can be effectively used in parallel for structural analysis.

Limited proteolysis has been used to probe the structures of individual proteins and those of multicomponent macromolecular assemblies. The simplest application involves the use of proteases in domain mapping, which is extensively used in structural biology. The term *domain* refers to minimal structural elements within proteins that often are associated with protein function (6). For technical reasons beyond the scope of this review, protein samples for x-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy ideally should possess highly ordered domain structures and be free of unstructured elements at the amino and carboxyl termini. Unstructured segments can spoil the formation of crystals, cause nonspecific aggregation, and give rise to resonance interference in NMR spectroscopy studies. Although not exclusively true, structured domains often are the functional elements within proteins, and determination of structure for these domains therefore provides insights into structure-function relationships. Because proteins often are composed of several individual domain building blocks connected by relatively unstructured linker domains, the



**Figure 1.** Schematic illustration of the use of proteolytic cleavage as a probe of protein structure. The arrows mark surface exposed and flexible sites that would be susceptible to proteolytic cleavage. If a sequence specific protease were used, the marked sites would also have to contain the protease recognition sequence to sustain cleavage. Mass analysis of all fragments together yield the cleavage ‘map’ that provides information on secondary, tertiary and, in multicomponent assemblies, quaternary structure.

# Methods & Reviews

**Table I.** Proteases commonly used in structural analysis [after Konigsberg (3), Carrey (4) and Coligan (5)].

<i>Protease</i>	<i>Specificity<sup>a</sup></i>	<i>pH optima</i>	<i>Inhibitors<sup>b</sup></i>
Chymotrypsin	P <sub>1</sub> = W, F, Y P' <sub>1</sub> = nonspecific	7.5-8.5 Ca <sup>2+</sup> -activated	DFP, PMSF, TPCK
Elastase	P <sub>1</sub> = A, V, I, L, G, S, T P' <sub>1</sub> = nonspecific	7.5-8.5	DFP, PMSF
Endoproteinase Asp-N	P <sub>1</sub> = nonspecific P' <sub>1</sub> = D	6.0-8.0	EDTA, 1,10-phenanthroline
Endoproteinase Glu-C (V8 protease)	P <sub>1</sub> = E (or E and D) P' <sub>1</sub> = nonspecific	7.8 (4.0)	DFP
Endoproteinase Lys-C	P <sub>1</sub> = K P' <sub>1</sub> = nonspecific	8.5	DFP, TLCK
Pepsin	P <sub>1</sub> = nonspecific P' <sub>1</sub> = nonspecific, but cannot be V, A, G	2.0-4.0	DFP, PMSF, TPCK
Proteinase K	P <sub>1</sub> = nonspecific P' <sub>1</sub> = nonspecific	7.5-12.0	PMSF, DFP
Subtilisin Carlsberg	P <sub>1</sub> = nonspecific P' <sub>1</sub> = nonspecific	7.0-8.0	PMSF, DFP
Thermolysin	P <sub>1</sub> = nonspecific P' <sub>1</sub> = L, F, I, V, M, A	7.0-9.0	EDTA
Trypsin	P <sub>1</sub> = K, R P' <sub>1</sub> = nonspecific <u>but cannot be P</u>	8.5	DFP, PMSF, TLCK

cleavage sites



<sup>a</sup> P<sub>1</sub> ——— P'<sub>1</sub> ; P<sub>1</sub> is on the amino-terminal side of the scissile bond and P'<sub>1</sub> on the carboxy-terminal side.

<sup>b</sup> DFP, diisopropyl fluorophosphate; EDTA, ethylenediaminetetraacetic acid; PMSF, phenylmethylsulfonyl fluoride; TLCK, tosyllysine chloromethyl ketone ;TPCK, tosylamido-2-phenylethyl chloromethyl ketone.

reduction of a multidomain protein to its individual domains through proteolysis provides a mechanism for functional characterization on a domain-by-domain basis.

Limited proteolysis was the original method for deletion analysis of protein structure-function relationships, but this approach has largely been replaced by DNA-based deletion analysis. Because proteases discriminate on the basis of structure, they still deserve a place as first-line tools in the analysis of protein structure, especially in combination with biomolecular MS (7). In domain mapping experiments, protease reaction kinetics must be controlled to yield only limited digestion (ie, single-hit kinetics), because extensive peptide bond scission reduces protein thermodynamic stability. Through local unfolding, multiple cleavages

expose secondary cleavage sites that provide information only indirectly related to native structure. Under limiting, or single-hit conditions, only the most kinetically accessible sites are cleaved, providing reliable information on three-dimensional protein structure.

In addition to probing static protein structure, protease sensitivity can be used to monitor changes in structure caused by, for example, ligand binding (1,2) and the addition of denaturants (8). Limited proteolysis can be used to probe the structure of multicomponent assemblies, including peptide-protein complexes, protein-protein complexes, and protein-DNA complexes. A common feature of these applications is that the protease is used to provide contrast between the associated and unassociated states of

the system. The formation of an interface between a protein and another macromolecule excludes solvent molecules and macromolecules such as proteases and protects otherwise accessible sites from protease cleavage.

## Mass Spectrometry Techniques Used In Protease Mapping Studies

MS has become an integral part of biologic research primarily because of the development of matrix-assisted laser desorption and ionization (MALDI) (9) and electrospray ionization (ESI) (10). MALDI and ESI have greatly advanced our ability to characterize large, thermally labile molecules by providing an efficient means of generating intact, gas-phase ions. These two techniques have been used to gain molecular weight information on biologic samples with unprecedented speed, accuracy, and sensitivity. Developments in instrumentation (11) coupled with newer sampling methods have enabled higher levels of sensitivity, increased mass range, and better mass accuracy and promoted an increasing number of MS-based applications in the study of covalent and noncovalent protein structure. Both approaches offer unique and complementary capabilities.

ESI and MALDI are fundamentally different ionization techniques, but they achieve essentially the same end result: nondestructive vaporization and ionization. With electrospray, ions are formed directly from solution (usually an aqueous or aqueous and organic solvent system) by creating a fine spray of highly charged droplets in the presence of a strong electric field. Vaporization of these charged droplets produces singly or multiply charged gaseous ions. ESI can be interfaced with liquid chromatography in the analysis of proteolytic digests directly from the solution phase or with tandem mass analyzers such as ion traps and triple quadrupoles to perform collision-induced dissociation experiments on peptides.

In nano-ESI (12), a variation of ESI used for protein identification, the spray needle is made very small (tip  $\sim 5 \mu\text{m}$ ) and is positioned close to the entrance of the mass analyzer, resulting in greatly increased efficiency. For instance, the flow rates for nano-ESI sources are on the order of tens of nanoliters per minute, and the total amount of sample consumed typically is less than femtomoles. Another advantage is that the droplets formed are smaller than in normal ESI, making nano-ESI more tolerant of salts and other impurities.

MALDI mass analysis generates gas-phase ions by laser vaporization of a solid matrix-analyte mixture in which the matrix (usually a small crystalline organic compound) acts as a receptacle for energy deposition. The relatively low number of charge states generated with MALDI, along with its high sensitivity and ability to simultaneously generate ions from multicomponent mixtures, makes it especially well suited for complex biomolecular samples such as pro-

teolytic digests. Moreover, MALDI-MS offers a reliable way of analyzing proteins and peptides.

ESI and MALDI-MS commonly use quadrupole and time-of-flight (TOF) mass analyzers, respectively. ESI with quadrupole mass analyzers typically has accuracy on the order of 0.01%, and ESI with the quadrupole ion trap mass analysis offers the additional advantage of allowing collision-induced dissociation experiments to be performed without having multiple analyzers. ESI quadrupole ion traps are being used extensively in the analysis of tryptic peptide digests for accurate protein identification.

MALDI with TOF analyzers constitutes one of the simplest mass analyzing devices and has accuracy between 0.1% and of 0.005%. These systems operate by accelerating a set of MALDI-generated ions with the same amount of energy down a flight tube. Because the ions theoretically have the same energy, the ions with different  $m/z$  values reach the detector at different times. Although the TOF analyzer has limited resolving power with MALDI (typically  $<2000$ ), the addition of the reflectron, which reduces the kinetic energy distribution of ions that reach the detector, has improved performance. TOF reflectron mass analyzers are capable of generating high-resolution and high-accuracy mass measurements (errors  $<50$  ppm). ESI and MALDI also are being coupled to the ultrahigh-resolution ( $>10^5$ ) Fourier transform mass analyzer with part-per-million ( $<10$  ppm) accuracy. Higher accuracy is proving to be valuable in protein identification and protein mass mapping.

## Analysis of Peptides and Proteins

The utility of ESI and MALDI for primary structural analysis lies in their ability to provide accurate molecular weight information on intact compounds, information that is extremely useful for protein identification. For example, an unknown protein often can be unambiguously identified by mass spectral analysis of its constituent peptides produced by chemical or enzymatic treatment of the sample. MALDI is especially well suited for such analyses, because complex mixtures of peptides are directly amenable to MALDI analysis. The molecular weights of individual peptides in a protein digest are easily determined by using a combination of liquid chromatography and ESI-MS.

Peptide and protein analysis can be facilitated by initiating fragmentation in the gas phase. Fragment ions generated inside ESI and MALDI mass spectrometers by collision-induced dissociation (CID) often yield information about the primary structure of a sample (13). Tandem mass analysis techniques such as ESI ion traps, Fourier transform MS (FTMS), or triple quadrupoles and MALDI-FTMS involve selecting an ion of interest with the mass analyzer and isolating it in a collision cell. Once in the collision cell, the selected ion undergoes collisions with an inert gas such as argon, creating fragments that can be mass analyzed to

provide information about their sequence. This multiple mass analysis approach is often referred to as tandem MS or MS<sup>2</sup>. Because the CID behavior of peptides is already well characterized, tandem MS with CID can be used to acquire direct sequence information on small peptides (<3 kDa).

Two important advantages of MALDI-MS are its sensitivity and ability to analyze complex polypeptide mixtures. These features also are being used to sequence biopolymers. The protein ladder sequencing technique originated by Chait et al. (14) allows stepwise removal of each amino acid in a peptide, a process in which each residue is chemically or proteolytically removed from the amino-terminal end to produce sequence-defining peptide fragments. Alternatively, amino acids can be enzymatically removed from the carboxyl terminus (15). A MALDI mass spectral readout enables generation of the resulting protein sequencing ladder. This method, which allows each amino acid to be identified from the mass difference between successive peaks, can provide sequence information on peptides of more than 30 residues. Sequence data has been obtained from larger proteins by enzymatic cleavage combined with protein ladder sequencing (15-17).

## Software Packages for Spectrographic Analysis of Proteolytic Digests

Several software packages have been developed to facilitate the analysis of MS data for proteins and have been especially useful in the analysis of multicomponent fragmentation patterns generated through proteolytic digestion. Two such software packages are described to illustrate some of the features that are available. These two packages are available free of charge to not-for-profit organizations over the World Wide Web.

The software package PAWS (Proteometrics, Rockefeller University, New York, NY, freeware; [www.proteometrics.com](http://www.proteometrics.com)) offers a user-friendly, intuitive interface that allows a wide variety of MS-related, protein-based operations to be performed. Protein sequences can be loaded from files or entered manually and the monoisotopic or average molecular mass calculated. The program allows specific amino acid sites to be chemically modified and the modified mass to be calculated. Theoretical cleavage reactions (enzymatic or chemical) can be performed and the resulting fragment masses calculated and presented in tabular and graphical form. Searching tools have been incorporated to aid the analysis of MS data from proteolytic or chemical cleavage reactions. A simple search tool identifies peptide sequences derived from a known protein sequence that match a particular target mass. Searches consider all possible peptide sequences or only sequences consistent with fragmentation resulting from a particular sequence-specific cleavage reaction. Lists of fragment masses can be entered manually or imported from files, and mul-

iple searches can be performed. Search results are presented in graphical and tabular forms. The PAWS package offers a powerful and easy-to-use set of tools to predict cleavage patterns and fragment masses and to search for and identify peptides derived from known protein sequences that match experimentally determined masses.

The Protein Prospector package (University of California at San Francisco Mass Spectrometry Facility, San Francisco, CA; Drs. Karl Clauser and Peter Baker; [prospector.ucsf.edu](http://prospector.ucsf.edu)) offers some of the same capabilities as PAWS, along with extensive database searching capabilities. The software, which can be used over the Internet or installed locally after appropriate licensing, is divided into several modules that perform specific functions. These include MS-Digest, which can be used to generate theoretical protein digests; MS-Product, which can predict fragmentation in the mass spectrometer caused by post-source decay and collision-induced dissociation; and MS-Comp, which can suggest amino acid compositions consistent with experimental MS data, including parent peptide mass and immonium ion fragmentation data. Database searching tools include MS-Fit, which compares an observed set of proteolytic fragment masses with those predicted for all proteins in a database; MS-Tag, which compares tandem MS peptide fragmentation patterns with predicted patterns; and MS-Edman, which compares short segments of a peptide sequence with segments in a protein database. The database searching tools in Protein Prospector have been developed to aid the identification of unknown proteins on the basis of MS data but are also well suited for the analysis of MS data from protease-based structure mapping experiments. The MS-Fit module can be used to match observed fragment masses against those predicted for the protein being studied. Instead of using the entire nonredundant protein database for comparison, the content of the database is limited to the protein under study. Because the protease cleavage reactions for mapping studies are performed under limiting conditions, the fragments that are produced may span one or more uncleaved protease sites, increasing the number of theoretical fragments that must be considered during the comparisons. However, MS-Fit accommodates this need with a settable parameter corresponding to the maximum number of missed cleavages. This capability is also included in MS-Digest, allowing tables of all possible peptide fragments to be generated.

## Analysis of Protein-DNA Interactions

The first application of limited proteolysis and MALDI mass analysis to the study of a multicomponent biomolecular assembly was published in 1995 by Chait et al. (18). This combined approach was used in structural analysis of the transcription factor Max when free in solution and when bound to an oligonucleotide containing its

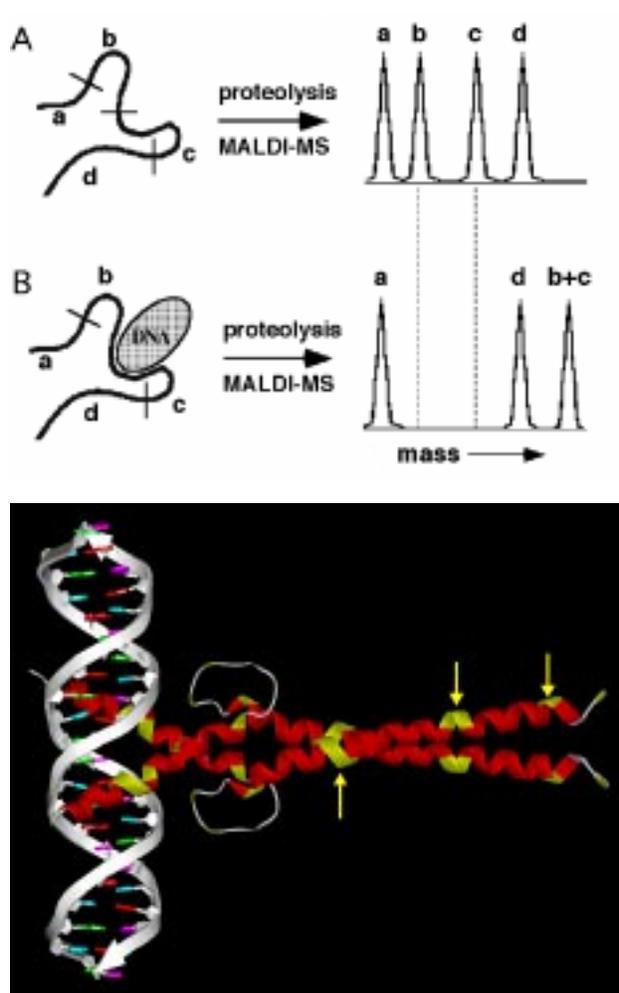
## Methods & Reviews

specific DNA binding site. Max is a member of the basic helix-loop-helix (bHLH) family of DNA-binding proteins and was the target of crystallographic studies. An extensive series of limited proteolysis experiments were conducted using free Max. The products of digestion reactions were analyzed using MALDI-TOF MS, demonstrating the suitability of this MS technique for analysis of multicomponent biomolecular samples in the identification of fragments and in their relative quantitation. The results showed that Max usually is very susceptible to proteolytic cleavage. However, Max is less susceptible to digestion by a variety of proteases at high ionic strengths, suggesting that salt stabilizes Max's structure. Because cleavage requires accessibility and flexibility, this result suggested that Max's structure is more highly ordered in the presence of higher salt concentrations, with loop regions in less flexible states. These results, indicating that Max may be relatively flexible in the absence of DNA, are consistent with the inability to crystallize Max in the free state.

Much more dramatic stabilization of Max was observed in the presence specific DNA. In this case, cleavage rates were reduced 100-fold, indicating significant stabilization in the presence of DNA. This stabilization stems from the protection of potential cleavage sites in formation of the Max-DNA interface and from the added thermodynamic stability imparted to Max by association with DNA. The cleavage pattern within the Max-DNA complex revealed that the bHLH domain is the minimal requirement for DNA binding and that the leucine-zipper domain is dispensable for this activity. The Max-DNA interaction sites were identified (Figure 2). These results provided valuable insights about Max-DNA binding structure-function relationships that guided the successful crystallization and structure determination of the Max-DNA complex (19).

### Analysis of Protein-Protein Interactions

The general approach outlined previously has been applied to the study of protein-protein interactions. The mapping of protein-protein complexes in situ, however, is complicated because peptide fragments are produced for all subunits within a complex. An MS-based approach to map protein-protein interactions that overcomes this complication was developed by Chait et al. (20), which maps the interaction of a protein growth factor with a monoclonal antibody. Basic fibroblast growth factor (bFGF) was digested with endoproteinase Asp-N, followed by immunoprecipitation of binding-competent peptides using a monoclonal IgG1 antibody (mAb). The multicomponent bFGF/mAb assemblies were subjected to MALDI-TOF MS analysis, allowing identification of the peptide segments of bFGF that constitute the binding epitope. Because the monoclonal antibody was raised against small bFGF-derived peptides that lack native structure, this method was successful in mapping the mAb-binding epitope. However, the method requires



**Figure 2.** Mapping a Protein/DNA interface using limited proteolysis and MALDI-MS [after Chait and co-workers(19)]. A) Schematic view of proteolysis of a protein alone followed by MALDI-MS analysis. Cleavage sites are identified on the basis of I) the observed fragment masses, II) the known primary sequence, and III) known protease sequence-specificity. B) Formation of the protein/DNA complex protects one site from cleavage, producing an altered protease cleavage 'map'. Comparison of the maps in A and B give information about amino acids at the protein/DNA interface. C) Experimental results for the Max/DNA complex shown in the context of the 3D crystal structure [Ferre-D'Amare, *et al.*(20)]. The sites marked in light grey are cleaved in the absence of DNA while only the sites marked by arrows are cleaved in the presence of DNA. The bHLH domain at the interface with DNA is protected.

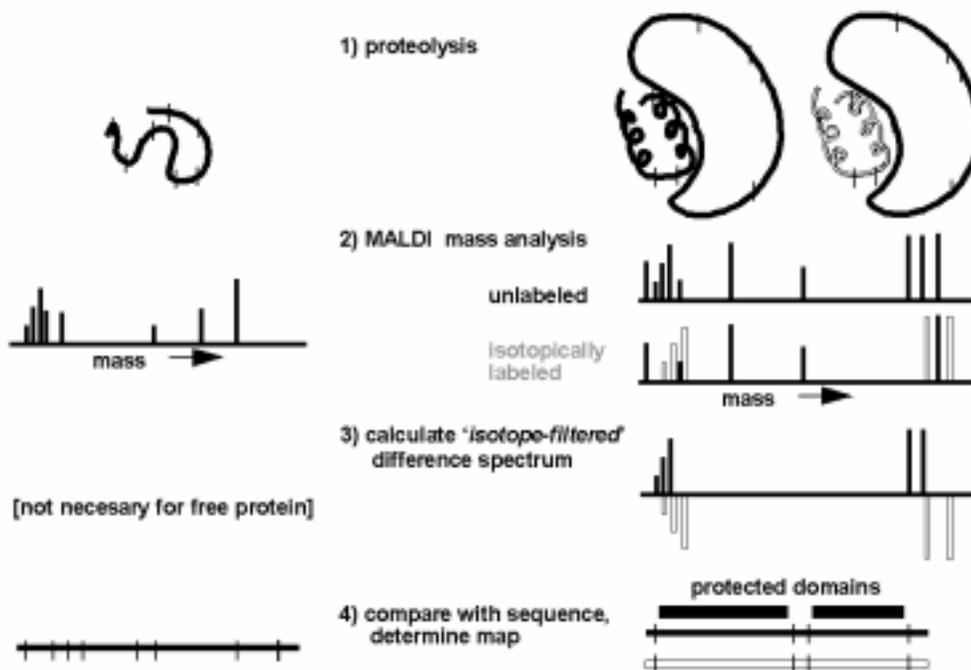
that a target protein sustain binding activity after proteolysis, and it would have only limited utility in the studies of proteins that require native structure for protein binding activity.

## Methods & Reviews

An alternative approach to mapping protein-protein interfaces that overcomes the complications described was demonstrated by Kriwacki and Siuzdak (8,21). The experimental scheme is illustrated in Figure 3. The method exploits the high mass accuracy, resolution, and sensitivity of MALDI-TOF MS, combined with the power of stable isotope labeling, and it offers access to isotope-filtered mass spectra of individual subunits within multiprotein assemblies. Figure 3 (panel 1), illustrates two concepts central to mapping interfaces within assemblies. First, proteolysis reactions are performed for one component before and after formation of a multiprotein assembly (panel 1, left and right, respectively). Second, proteolysis reactions for the complex are performed in duplicate, with one subunit prepared at natural isotopic abundance in one experiment and in an isotope-labeled form in a second (panel 1, right). Other proteins within the assembly are used at natural isotopic abundance in both experiments. Reaction products are analyzed using MALDI-TOF MS (panel 2). In panel 2 (right), a subset of peaks in the upper spectrum appear at shifted positions in the lower spectrum (black versus open bars); these shifts correspond to the mass differences between unlabeled and labeled fragments. The isotope-filtered mass spectrum is obtained by subtracting these two spectra (panel 3). After data analysis, regions within the target protein that are pro-

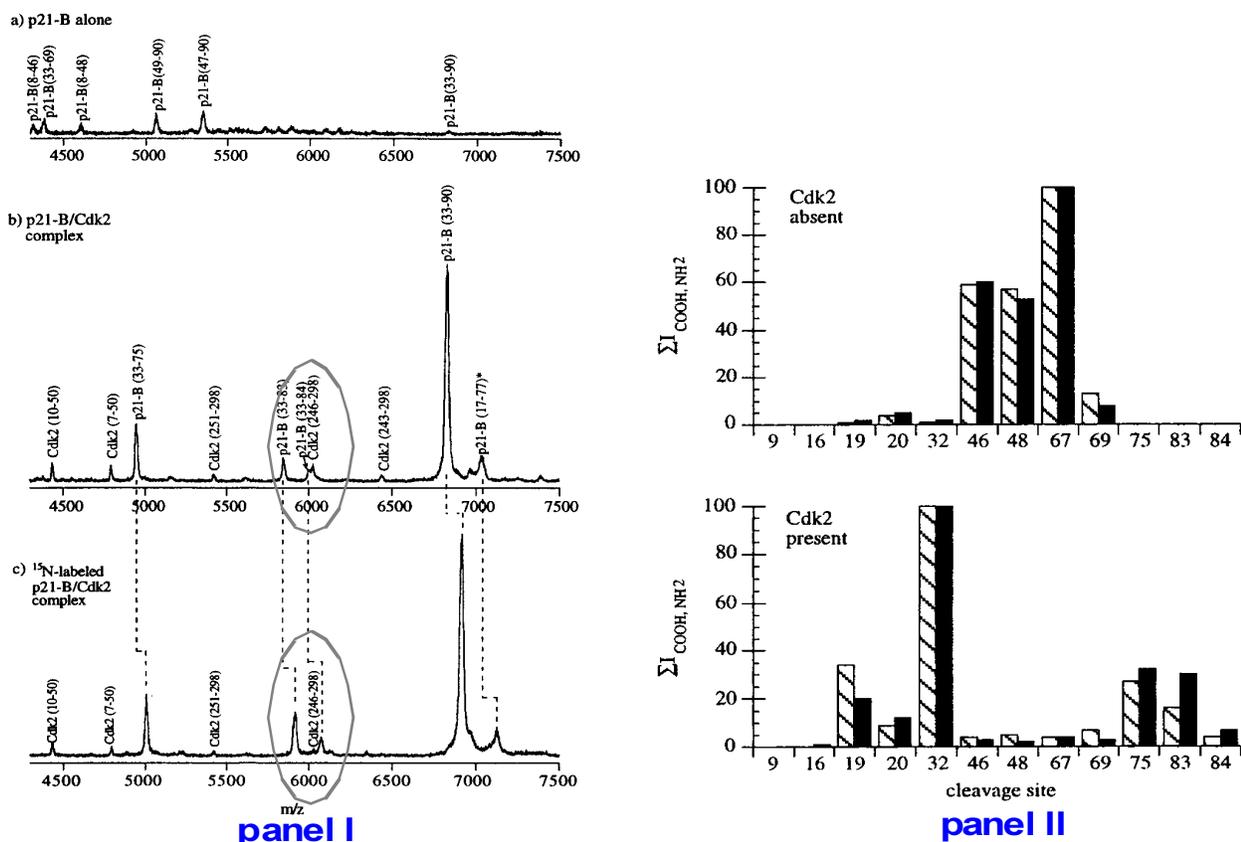
tected from proteolysis in the assembly are identified (panel 4, right).

This mapping strategy has been applied to the complex between p21, a cell cycle regulatory protein, and cyclin-dependent kinase 2 (Cdk2). Experimental data are shown in Figure 4 (left), with a histogram of the final results shown on the right. MALDI analysis of the tryptic fragments of p21-B (the kinase inhibitory domain of p21) in the presence and absence of Cdk2 revealed a segment of 24 amino acids in p21-B that is protected from trypsin cleavage, thereby identifying the segment as the Cdk2 binding site on p21-B. The top trace (Figure 4, panel 1, shows a region of the MALDI-TOF mass spectrum for p21 after proteolysis with trypsin in the absence of Cdk2. The middle and bottom traces show spectra after proteolysis of the p21/Cdk2 complex, with p21 at natural abundance (middle) and  $^{15}\text{N}$  labeling (bottom). Cdk2 is unlabeled in both cases. Analysis of the peaks in the top trace allowed identification of 28 trypsin sites in p21. Analysis of peaks in the lower panels showed that 4 of the 28 sites were protected from cleavage in the presence of Cdk2. The dotted lines in the lower traces mark peaks that exhibit an  $m/z$  shift from one spectrum to the other and reveal fragments derived from p21. The utility of the isotope labeling strategy is illustrated in the first panel of Figure 4, near  $m/z$  6000 (marked by ovals). In the



**Figure 3.** Schematic illustration of combined use of limited proteolysis, isotope labeling and MALDI-TOF mass analysis in protein-protein interface mapping. The left panel illustrates the analysis of p21 alone while the right panel illustrates the utility of isotope labeling one subunit within a multicomponent assembly in simplifying data analysis. At the top, right, regions within p21 (black lines and open lines) are protected from cleavage due to the formation of a complex with Cdk2 (kidney shape).

# Methods & Reviews



**Figure 4.** Interface mapping using limited proteolysis, isotope labeling and MALDI-TOF MS. Panel I, MALDI-TOF spectra for p21-B alone (top) and p21-B/Cdk2 complexes (middle and bottom panels). The dashed lines mark peaks that experience a mass increase in the lower spectrum with respect to the middle spectrum, allowing identification as p21-B-derived fragments. The ovals identify a region where the usefulness of isotope-labeling is particularly evident. Panel II, histograms showing trypsin accessibility versus position within p21-B amino acid sequence for p21-B alone (top) and p21-B/Cdk2 complex (bottom). The data represented by solid bars are derived from spectra for unlabeled samples and the hashed bars from spectra for  $^{15}\text{N}$ -labeled samples.

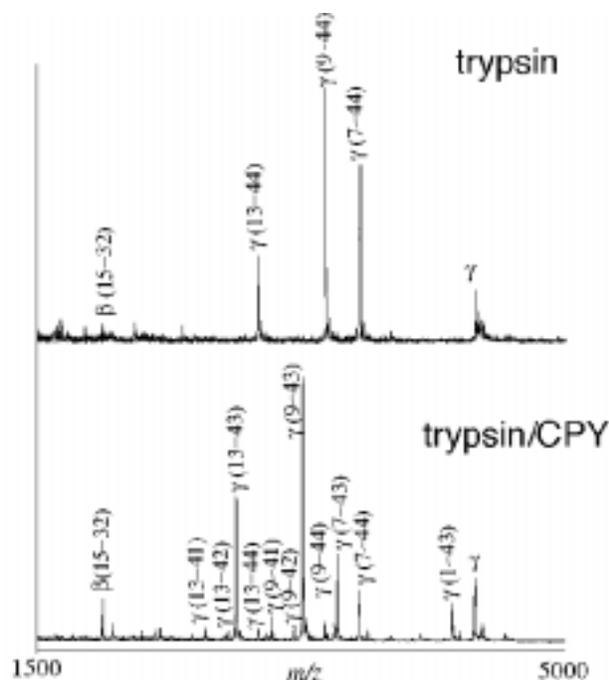
middle trace, two nearly coincident peaks appear (p21 residues 33-84, Cdk2 residues 246-298); in the bottom trace, one of these peaks shifts to a new position. This result unambiguously identifies the left-hand peak in the middle trace as originating from p21 and the other as originating from Cdk2.

The mass accuracy of MALDI-MS instruments ( $\pm 0.005\%$ ) allows reliable identification of most fragments from p21 and Cdk2 without resorting to isotope labeling. However, even this level of accuracy cannot allow identification in all cases because of the finite probability that fragments from the different subunits will have similar masses, as illustrated earlier. Figure 4 shows histograms of protease accessibility compared with position within the sequence for unlabeled and  $^{15}\text{N}$ -labeled p21 (hashed versus solid bars) in the absence and presence of Cdk2. The four protected sites are clearly revealed in the bottom panel.

Other methods for examining protein-protein interactions rely on the chemical cross-linking of protein complexes before MALDI analysis. In one case, multimeric proteins are subjected to MALDI analysis after reaction with a cross-linking agent such as glutaraldehyde to determine their stoichiometry (22). Glutaraldehyde is used to covalently link the protein subunits in solution. Subsequent mass analysis of the covalently linked complex by MALDI permits an accurate assessment of the oligomeric state of the protein. In principle, the cross-linked species can be subjected to limited or complete proteolysis to reveal the amino acid residues involved in intermolecular covalent bonds.

## Viral Studies

MS is useful for many higher-order structural studies, particularly the analysis of capsid quaternary protein structure (protein-protein interactions) of nonenveloped vi-



**Figure 5.** Digestion of Flock House viral capsid proteins performed directly on the virus. (Top) MALDI-MS data generated from the trypsin digest. (Bottom) Data generated from the carboxypeptidase Y digest of the trypsin digest (or endo/exo sequential digestion) of the viral capsid protein.  $\beta$  and  $\gamma$  denote the  $\beta$ -protein and  $\gamma$ -peptide segments of the  $\alpha$ -protein, respectively, as discussed in the text, while the numbers in parenthesis are the N-terminal and C-terminal amino acids of the proteolytic fragments.

ruses (23). Capsid protein subunits, which make up the protective outer shell of the virus, provide structural stability and play a major role in infectivity. Although such protein-protein interactions typically have been mapped through x-ray crystallography, protein mass mapping is gaining more recognition as an effective technique. Conventional protein mapping has been used for probing the primary structure (ie, amino acid sequence) of individual proteins by incorporating chromatography, gel electrophoresis, or both techniques. Although proteolytic cleavage can provide indirect information about the domain structure of proteins, the method has not been routinely applied to protein-protein complexes because of the limitations in resolving and identifying the multiple fragments produced with conventional methods. Protein mass mapping shows great promise because MS is well suited to the analysis of complex mixtures of biomolecules and viral proteins, offering high sensitivity, resolving power, and accuracy.

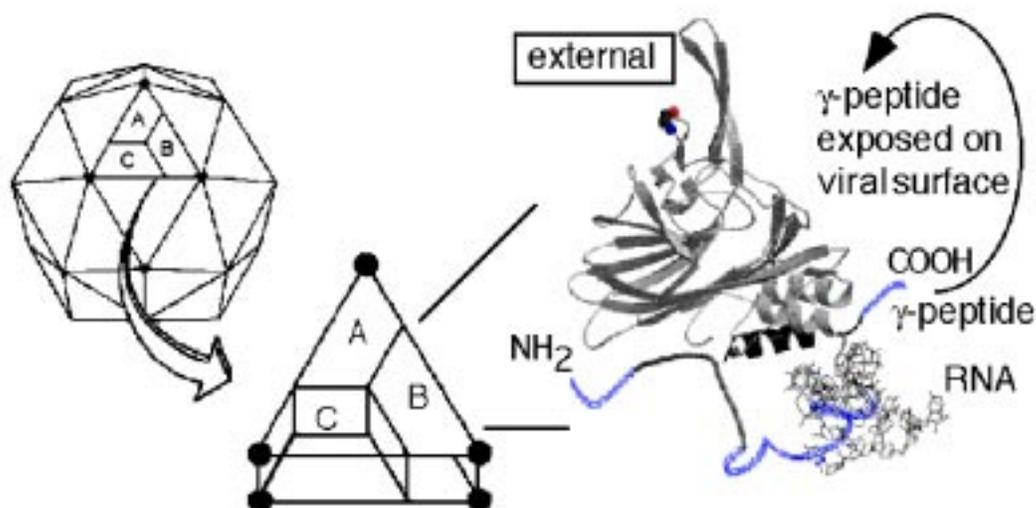
Limited proteolysis and MALDI-MS experiments (24) were performed on Flock House virus, a nonenveloped, icosahedral RNA animal virus with dimensions (~300 Å)

similar to those of rhinovirus and poliovirus. Its protein coat or capsid is composed of 180 copies of a single gene product,  $\alpha$ -protein, which is autocatalytically cleaved to peptides,  $\beta$ -protein and  $\gamma$ -peptide, during maturation. The autocatalytic cleavage products are easily observed through MALDI-MS. In using time-resolved proteolysis followed by MALDI-MS analysis, it was expected that the reactivities of virus particles to different proteases would reflect the surface-accessible regions of the viral capsid and offer a new way of mapping the viral surface. When these experiments were performed, cleavages on the surface-accessible regions were observed, but cleavages internal to the viral capsids (based on the crystal structure) were also generated. Observation of such cleavages was a surprising and initially perplexing result. In these studies, identification of the viral capsid protein fragments was facilitated by sequential digestion in which proteins were first digested by an endoprotease such as trypsin, followed by exposure to an exoprotease such as carboxypeptidase Y (Figure 5) (24). These results, along with the x-ray data, indicate that portions of the  $\beta$ -protein and  $\gamma$ -peptide are transiently exposed on the surface of the virus (Figure 6). These viral portions are implicated in RNA release and delivery.

Studies have been extended to the common cold virus (human rhinovirus), revealing previously undocumented viral structural dynamics and the inhibition of such dynamics by an antiviral agent (16). Results indicate that binding of the antiviral agent causes local conformational changes in the drug-binding pocket and stabilizes the entire viral capsid.

### Conclusions and Trends

The examples presented illustrate the utility of combining proteolysis and MS analysis in structural studies of proteins and multicomponent proteinaceous assemblies. The experimental techniques that these approaches rely on are simple, efficient, and readily accessible. Moreover, as biomolecular MS continues to expand as the technique of choice in bioanalytic studies, the necessary MS instrumentation is becoming a standard feature of biotechnology core facilities. Biologic scientists are in a position to obtain information on molecular interactions without the need for site-directed mutagenesis and free of the caveats associated with this technique. The MS component of the experiments still requires expert operation of increasingly complex instrumentation and will remain in the realm of support facilities. Because computer analysis of mapping data requires expertise and chemical knowledge of the underlying cleavage and modification reactions, it may be best provided as support services. These methods are within reach of the entire biologic community. The important task is to educate the biologic community about the utility of these structure mapping methods and to expand their implementation.



**Figure 6.** Crystal structure of Flock House virus shows that the  $\gamma$ -peptide and the N- and C-terminus of the  $\beta$ -protein are localized internal to the virus. Yet, proteolytic time-course experiments demonstrated that these domains are transiently exposed on the viral surface.

What are the future directions for MS-based proteolytic mapping methods? Although current MS instrumentation is limited by low resolution and accuracy at the upper mass limits, indirect mass analysis (ie, only the products of probing reactions are flown through the mass spectrometer) makes the “window of opportunity” quite broad. We have discussed applications to biomolecular assemblies that have and have not been supported by high-resolution structural data. In the early stages of structural studies, the MS-based probing methods are particularly well suited to provide rapid access to low-resolution maps that are used to guide high-resolution studies. However, this stage may be an end point in some investigations in which the identification of interacting residues is the desired information. As a complement to high-resolution structural information from x-ray crystallography or NMR spectroscopy, probing studies have already been shown to provide valuable and even startling insights into protein dynamics and structural rearrangements. These investigations mark the take-off point for studies that seek to quantitate the molecular kinetics and thermodynamics of the underlying dynamic phenomena. We posit that MS methods will assume a position similar to that of gel electrophoresis as a primary research tool in coming years because of its superior sensitivity, precision, accuracy, and throughput.

### References

1. Fontana A, Polverino de Laureto P, De Filippis V, Scaramella E, Zambonin. Probing the partly-folded states of proteins by limited proteolysis. *Fold Des* 1997;2:R17-26.
2. Fontana A, Zambonin M, Poleverino de Laureto P, De Filippis V, Clementi A, Scaramella E. Probing the conformational state of apomyoglobin by limited proteolysis. *J. Mol. Biol* 1991;266:223-30.
3. Konigsberg WH. Limited proteolysis of DNA polymerase as probe of functional domains. *Methods Enzymol* 1995;262:331-46.
4. Carrey EA. In: Creighton T, editor. *Protein Structure, a Practical Approach*. New York: IRL Press, 1989:117-44.
5. Coligan JE, Dunn BM, Ploegh HL, Speicher DW, Wingfield PT, editors. *Current Protocols in Protein Structure*, vol II. New York: John Wiley & Sons, 1997.
6. Kirschner K, Bisswanger H. Multifunctional Proteins. *Annu Rev Biochem* 1976;45:143-66.
7. Chait BT. Mass spectrometry—a useful tool for the protein X-ray crystallographer and NMR spectroscopist. *Structure* 1994;2:465-7.

## Methods & Reviews

8. Kriwacki RW, Wu J, Tennant L, Wright PE, Siuzdak G. Probing protein structure using biochemical and biophysical methods. Proteolysis, matrix-assisted laser desorption/ionization mass spectrometry, high-performance liquid chromatography and size-exclusion chromatography of p21/Waf1/Cip1/Sdi1. *J Chromatogr A* 1997;777:23-30.
9. Hillenkamp F, Karas M, Beavis RC, Chait BT. Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal Chem* 1991;63:A1193-1201.
10. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization—principles and practice. *Mass Spectr Rev* 1990;9:37-70.
11. Siuzdak G. *Mass Spectrometry for Biotechnology*. San Diego: Academic Press, 1996.
12. Wilm MS, Mann M. Electrospray and Taylor-cone theory, Dole's beam of macromolecules at last. *Int J Mass Spectr Ion Process* 1994;136:167-80.
13. Papayannopoulos IA. The interpretation of collision-induced dissociation tandem mass-spectra of peptides. *Mass Spectr Rev* 1995;14:49-73.
14. Chait BT, Wang R, Beavis C, Kent SB. Protein ladder sequencing. *Science* 1993;262:89-92.
15. Patterson DH, Tarr GE, Regnier FE, Martin SA. C-terminal ladder sequencing via matrix-assisted laser desorption mass spectrometry coupled with carboxypeptidase Y time-dependent and concentration-dependent digestions. *Anal Chem* 1995;67:3971-8.
16. Thiede B, Wittmann-Liebold B, Bienert M, Krause E. MALDI-MS for C-terminal sequence determination of peptides and proteins degraded by carboxypeptidase Y and P. *FEBS Lett* 1995;357:65-9.
17. Woods AS, Huang AYC, Cotter RJ, Pasternack GR, Pardoll DM, Jaffee EM. Simplified high-sensitivity sequencing of a major histocompatibility complex class I-associated immunoreactive peptide using matrix-assisted laser desorption/ionization mass spectrometry. *Anal Biochem* 1995;226:15-25.
18. Cohen SL, Ferre-D'Amare AR, Burley SK, Chait BT. Probing the solution structure of the DNA-binding protein Max by a combination of proteolysis and mass spectrometry. *Protein Sci* 1995;4:1088-99.
19. Ferre-D'Amare AR, Prendergast GC, Ziff EB, Burley SK. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* 1993;363:38-45.
20. Zhao Y, Muir TW, Kent SBH, Tischer E, Scardina JM, Chait BT. Mapping protein-protein interactions by affinity-directed mass spectrometry. *Proc Natl Acad Sci USA* 1996;93:4020-4.
21. Kriwacki RW, Wu J, Tennant L, Siuzdak G, Wright PE. Probing protein/protein interactions with mass spectrometry and isotopic labeling: analysis of the p21/Cdk2 complex. *J Am Chem Soc* 1996;118:5320-1.
22. Farmer TB, Caprioli RM. Mass discrimination in matrix-assisted laser desorption ionization time-of-flight mass spectrometry—a study using cross-linked oligomeric complexes. *J Mass Spectr* 1995;30:1245-54.
23. Siuzdak G. Probing viruses with mass spectrometry. *J Mass Spectr* 1998;33:203-11.
24. Bothner B, Dong XF, Bibbs L, Johnson JW, Siuzdak G. Evidence of viral capsid dynamics using limited proteolysis and mass spectrometry. *J Biol Chem* 1998;273:673-6.

