Mass spectrometry — a useful tool for the protein X-ray crystallographer and NMR spectroscopist

Even under ideal conditions, the determination of a protein structure at atomic resolution requires a considerable investment of time and effort. The determination becomes still more challenging when conditions are not ideal, as in cases where there are questions regarding the precise composition, integrity, or purity of the protein preparation. The development of new mass spectrometric methods for accurately determining the molecular weights of proteins permits a rapid check of the correctness of the accepted primary structure of proteins (either with or without covalent modifications, heavy atom derivatives, or isotopic substitutions), the high-resolution determination of the purity of protein preparations and the facile, reliable identification of protein domains prepared by proteolysis. These capabilities make mass spectrometry a useful tool for the protein X-ray crystallographer and NMR spectroscopist.

New mass spectrometric approaches to the analysis of proteins

The introduction of effective new methods for accurately measuring the molecular weights of individual peptides and protein molecules has vastly expanded the utility of mass spectrometry for the structural biologist. Two techniques, in particular — matrix-assisted laser desorption/ionization mass spectrometry, and electrospray ionization mass spectrometry — are beginning to have a significant impact on the speed and ease with which a wide variety of structural biology questions can be resolved [1]. These methods enable the rapid (data is obtained in minutes) and accurate (accuracies as high as 1 part in 10 000) analysis of picomole amounts of polypeptides with molecular weights up to and exceeding 100 kDa.

The special strengths of electrospray ionization mass spectrometry [2] include the ease with which it can be coupled directly to liquid chromatography, the high accuracy with which masses are measured and the resolving power of the technique, and the ability to analyze proteins directly from liquid solutions. Limitations include the deleterious effects of salts, buffers, and other biochemical additives on the quality of the mass spectra, and the finding that certain proteins do not readily yield useful mass spectra. Matrix-assisted laser desorption/ionization mass spectrometry [3], on the other hand, has contrasting strengths. These include immunity to additives widely used in biological research and

structure determinations (e.g., salts, buffers, glycerol), the ease with which complex mixtures of peptides and proteins can be directly measured, and the relatively wide applicability of the technique for the analysis of peptides and proteins. Limitations include the lower resolving power and mass determination accuracy, especially for large proteins. These strengths and limitations make the two techniques complementary. In both cases, the mass spectrometric analysis provides a measurement that can be likened to an SDS-PAGE analysis, but with much higher accuracy, resolution, and speed.

Characterizing proteins and assessing the integrity of protein preparations

Before a high-resolution structure determination is attempted, it is important that the primary structure of the protein be known in detail. Uncertainties concerning the primary structure may arise in several ways, depending on the source of the protein (natural, recombinant, or synthetic) and the available genetic and biochemical information. For the majority of proteins, information on the primary structure is deduced mainly from the cDNA sequence corresponding to the gene. If there are errors in this cDNA sequence, the deduced primary sequence of the protein will be in error. Alternatively, the mature protein, isolated from natural sources, may differ from the deduced sequence because of post-translational modifications, or unanticipated proteolytic processing or degradation. A rapid and reliable check of the correctness of the presumed primary structure can be made by a simple, accurate molecular weight determination of the protein. Any significant difference between the measured mass and that calculated for the putative structure implies an error in the sequence deduced from the cDNA, or a post-translational modification, or proteolytic processing or degradation of the protein [1,4,5]. It should be emphasized that although a matching molecular weight for the protein is a strong indicator of the correctness of the putative primary structure (and is certainly a necessary condition), it is not a sufficient condition for proving that the primary structure is correct. If more detailed verification of the primary structure is required (especially to elucidate the nature and site of a post-translational modification), it is usually necessary to subject the protein to further analysis, involving enzymatically or chemically induced degradation of the protein, fol-

lowed by conventional Edman sequencing or further mass spectrometric measurement of the resulting peptide fragments [5]. This examination of the peptide fragments can be most easily accomplished by analyzing the unfractionated peptide mixture using matrix-assisted laser desorption mass spectrometry [6] or by subjecting the mixture to liquid chromatography coupled to electrospray ionization mass spectrometry [7].

Proteins produced by recombinant techniques are subject to a number of special sources of error, including unanticipated mutations, modifications, termination and proteolytic degradation [5,8]. Again, a simple molecular mass measurement provides a fast and reliable check for correctness. At my own institution (The Rockefeller University), these simple molecular weight measurements have proved so informative and time-

saving that a mass spectrum is obtained on virtually every protein before it is subjected to X-ray crystallography or NMR spectroscopy. In addition, proteins produced with special amino acid residues (e.g. selenomethionine), with ^{13}C and/or ^{15}N enrichment, or with heavy atom derivatives can be analyzed to determine the stoichiometry of incorporation or enrichment. Mass spectrometry is also of particular utility for assessing the integrity of synthetic proteins and for detecting the occurrence of side reaction products [9].

Assessing the homogeneity of protein preparations

It is widely appreciated that the purity or homogeneity of a protein preparation is an important determinant of the success for obtaining crystals that diffract
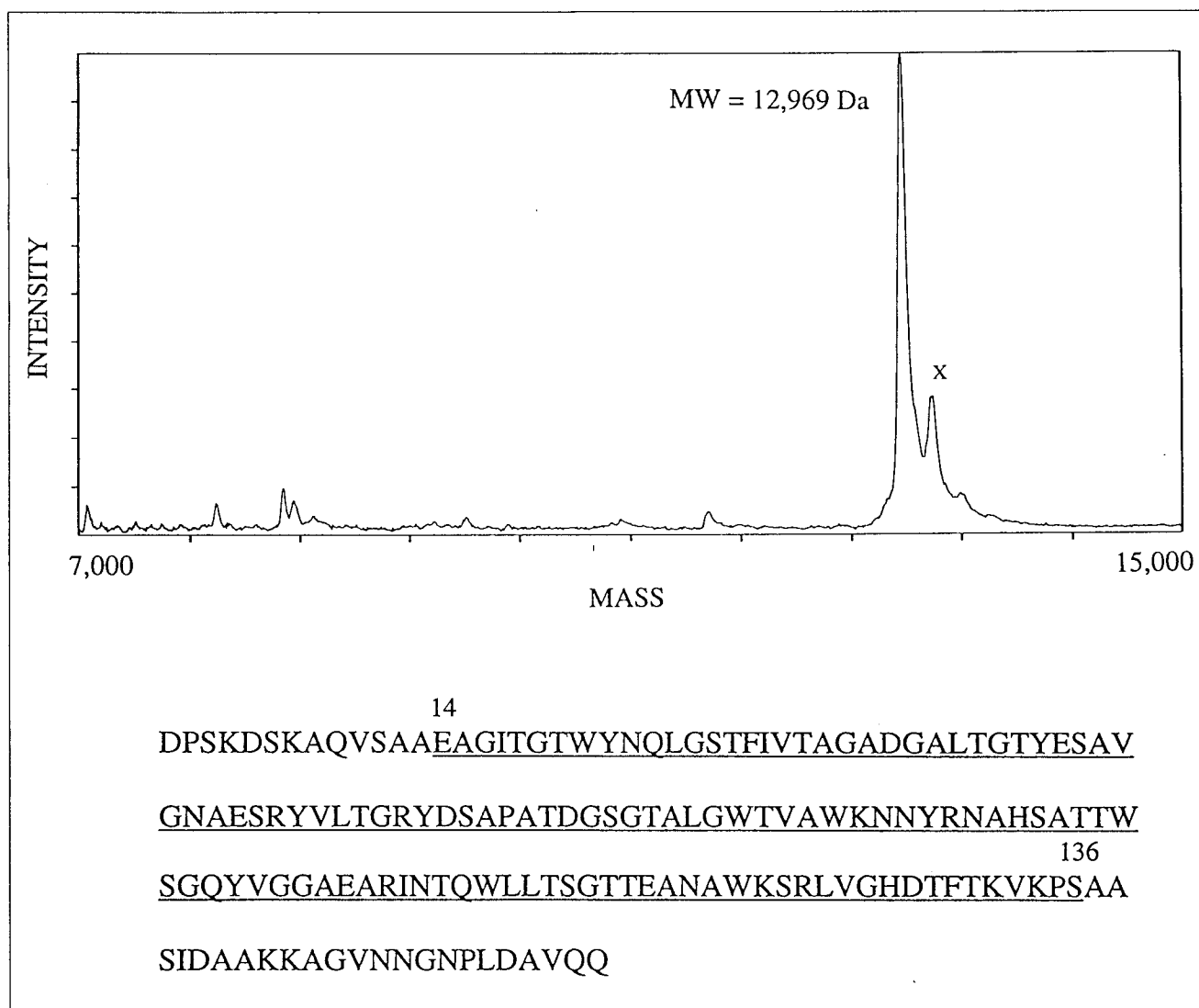

MW = 12,969 Da

X

7,000   15,000

MASS

14
DPSKDSKAQVSAA_EAGITGTWYNQLGSTFIVTAGADGALTGTYESAV_

GNAESRYVLTGRYDSAPATDGSGTALGWTVAWKNNYRNAHSATTW
                                                                    136
SGQYVGGAEARINTQWLLTSGTTEANAWKSRLVGHDTFTKVKPSAA

SIDAAKKAGVNNGNPLDAVQQ

Fig. 1. Mass spectrometry to identify proteolytic cleavage sites. **(a)** The matrix-assisted laser desorption mass spectrum of proteolytically modified streptavidin from *Streptomyces avidinii*. The peak labelled X is an artifact of the matrix-assisted laser desorption process [3] and does not arise from an impurity in the sample. On the other hand, several small peaks observed in the mass spectrum below mass 12000 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.(SE Darst and BT Chait, unpublished data). **(b)** Sequence of streptavidin. The sequence of the proteolytically modified streptavidin is underlined.

to high resolution. Less widely appreciated is the degree of subtlety of the effects of sample inhomogeneity. For example, a preparation of trypanathione reductase isolated from *Crithidia fasciculata*, which appeared completely homogenous by liquid chromatography, resisted repeated attempts to obtain diffraction-grade crystals. Analysis of the preparation by mass spectrometry revealed the presence of multiple forms of the protein, differing in molecular weight by only a few hundred daltons (5–6 amino acid residues). Once this microheterogeneity was detected, a higher resolution liquid chromatographic purification scheme was developed, which yielded pure protein that produced high quality crystals resulting in structure determination to 2.4 Å resolution [10]. In an even more dramatic example, our X-ray crystallographer colleagues have observed complete suppression of crystallization for protein preparations that are microheterogenous in a single amino-terminal arginine residue. Again, purification to homogeneity yielded a protein preparation that crystallized readily. These findings lead us to conclude that the presently described high-resolution mass spectrometric analysis provides a time-saving tool for assaying the homogeneity of protein preparations. Mass spectrometry is also of considerable value for detecting heterogeneity arising from variation in the carbohydrate portions of glycoproteins (with or without chemical or enzymatic modification), partial phosphorylation, oxidation, deamidation, heavy atom derivatives, *etc.*

Identification and definition of structural domains prepared by proteolytic cleavage of proteins

Proteolysis is widely used by structural biologists to reduce the size of proteins to manageable proportions or to isolate structurally rigid domains. Mass spectrometry proves a method of choice for the rapid and reliable identification of such proteolytically truncated proteins. If the protease (employed to cleave the protein) has high specificity, a simple molecular weight determination may be quite sufficient to yield an unambiguous, accurate identification of the portion of the protein produced. Mass spectrometry is vastly superior to SDS-PAGE for this purpose. If, however, the protease has broad specificity, it may be advantageous to define the amino terminus of the cleaved domain by Edman sequencing of a few residues. The carboxyl terminus (and hence the whole domain) can then be defined without ambiguity by an accurate molecular weight measurement.

To illustrate the principle of the mass spectrometric analysis described above, consider a preparation of streptavidin from *Streptomyces avidinii* that has undergone proteolytic digestion [11]. The matrix-assisted laser desorption mass spectrum of this preparation (obtained from Boehringer Mannheim, Indianapolis,

Indiana) is shown in Fig. 1 together with the sequence of the intact precursor subunit. Edman degradation of the preparation defines the amino-terminal residues as EAGIT…, showing that the amino terminus begins at residue 14. The molecular weight 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 weights of the modified protein with all possible carboxy terminal truncations. The molecule containing residues 14–136 has a calculated molecular weight of 12 971 Da, in good agreement with the measured value. Thus, the accurate molecular weight determination has defined unambiguously the carboxyl terminus of this modified streptavidin preparation. The sequence of the modified streptavidin is shown underlined in Fig. 1.

In summary, mass spectrometry provides one of the most rapid, straightforward and informative methods for assessing the identity, integrity, and purity of protein preparations, and is therefore likely to become an indispensable tool for the structural biologist.

References

1. Chait, B.T. & Kent, S.B.H. (1992). Weighing naked proteins. High-accuracy mass measurement of peptides and proteins. *Science* 257, 1885–1894.
2. Gross, M.L. (ed.) (1993). Special issues on electrospray ionization. *J. Am. Soc. Mass Spectrom.* 4, 523–684.
3. Hillenkamp, F., Karas, M., Beavis, R.C. & Chait, B.T. (1991). Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal. Chem.* 63, 1194A–1203A.
4. Aebersold, R. (1993). Mass spectrometry of proteins and peptides in biotechnology. *Curr. Opin. Biotechnol.* 4, 412–419.
5. Wang, R. & Chait, B.T. (1994). High-accuracy mass measurement as a tool for studying proteins. *Curr. Opin. Biotechnol.* 5, 77–84.
6. Beavis, R.C. & Chait, B.T. (1990). Rapid, sensitive analysis of protein mixtures by mass spectrometry. *Proc. Natl. Acad. Sci. USA* 87, 6873–6877.
7. Ling, V., *et al.*, & Shushan, B.I. (1991). Characterization of the tryptic map of recombinant DNA derived tissue plasminogen activator by high-performance liquid chromatography-electrospray ionization mass spectrometry. *Anal. Chem.* 63, 2909–2915.
8. Beavis, R.C., Chait, B.T., Creel, H.S., Fournier, M.J., Mason, T.L. & Tirrell D.A. (1992). Analysis of artificial proteins by matrix-assisted laser desorption mass spectrometry. *J. Am. Chem. Soc.* 114, 7384–7385.
9. Wlodawer, A., *et al.*, & Kent, S.B.H. (1989). Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* 245, 616–621.
10. Kuriyan, J., *et al.*, & Henderson, G.B. (1991). X-ray structure of trypanathione reductase from *Crithidia fasciculata* at 2.4 Å resolution. *Proc. Natl. Acad. Sci. USA* 88, 8769–8773.
11. Bayer, E.A., Ben-Hur, H., Hiller, Y. & Wilchek, M. (1989). Postsecretory modifications of streptavidin. *Biochem. J.* 258, 369–376.

Brian T Chait, Laboratory for Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, 1230 York Avenue, New York, NY 10021–6399, USA.