Isotope-Labeling Strategy for the Assignment of Protein Fragments Generated for Mass Spectrometry

Andrew Miranker,[†] Gary H. Kruppa,[‡] Carol V. Robinson,[†] Robin T. Aplin,[†] and Christopher M. Dobson*[,†]

Oxford Centre for Molecular Sciences
University of Oxford, South Parks Road
Oxford OX1 3QT, United Kingdom
Bruker Instruments Inc., Fortune Drive
Manning Park, Billerica, Massachusetts 01821

Received September 11, 1995

Soft ionization techniques in mass spectrometry provide a means of probing novel features of macromolecular structure and behavior.[1,2] Recent applications have included the characterization of both covalently and noncovalently linked systems in studies of reaction mechanisms[3] and the study of protein structure and dynamics using charge state distributions[4] and hydrogen deuterium exchange.[5−9] The latter has shown particular utility in the field of protein folding[10] where new light has been thrown on folding pathways,[5] the cooperativity of structural transitions,[11] and the interactions between partially folded states and molecular chaperones.[12] Of particular importance are techniques which allow determination of the positions along a polypeptide chain at which covalent modifications, such as hydrogen deuterium exchange, have taken place. A number of approaches are being developed and involve fragmentation of the polypeptide chain.[1,2] In this communication we describe initial studies of hen lysozyme, a protein whose sequence and structure are well chararacterized and on which extensive mechanistic studies of protein folding have been performed.[13,14] In particular we describe a heavy atom counting method as a powerful means for the identification and assignment of collision-induced dissociation (CID) fragments.

Hen lysozyme was ionized using an electrospray source from conditions (pH 3.8 formic acid in H_2O or 20% CH_3CN) under which the protein is in a native conformation.[5] The capillary skimmer potential was then increased until collisions with residual atmosphere were sufficiently energetic to result in extensive fragmentation (Figure 1) of all the parent charge states (predominantly +10 and +11).[15] The very high resolving power of some mass spectrometers, for example, the Fourier transform mass spectrometer (FTMS) used here, is such that fragments appear as multiplets primarily as a consequence of the 1.1% natural abundance of ^{13}C; e.g., the peaks at 605.4, 606.4, and 607.4 correspond to chemically identical species and represent populations having zero, one, or two ^{13}C atoms.[16] The resolution of individual natural abundance isotopes allows the charge state (and hence the mass) of the fragments to be measured directly. For example, the fragment with a monoisotopic *m/z* of 584.4 has isotopic peaks spaced 0.5 *m/z* apart (Figure 1, inset), indicating a charge of +2 and a mass of 1168.8 Da. The focus of this study is the group of 24 fragments found in the mass range 500−2000 *m/z*. Of the 24 fragments, 13 have a charge of +2 and 11 have a charge of +1.

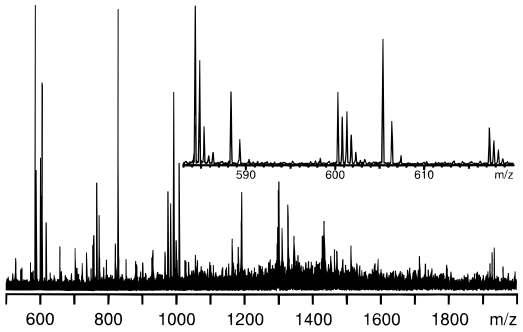

Figure 1. FTMS CID spectrum of hen lysozyme. (Inset) A 37 *m/z* region of the CID spectrum showing a distribution of natural abundance isotope peaks resulting from six fragment ions.

Polypeptide CID tends to be dominated by single breaks of the main chain.[17,18] The largest such fragment for lysozyme would be residues 1−5 as a consequence of a disulfide bond between residues 6 and 127. Most fragments produced in the present work have a *m/z* greater than that of residues 1−5, indicating the presence of internal sequence ions and/or fragmentation of a disulfide. An illustration of the magnitude of the assignment problem can be given for a typical fragment ion at 584.4. An exhaustive computer search[19] identifies 96 fragments whose calculated *m/z* is within 1000 ppm of this value and 30 within 50 ppm (Table 1A). Assignment on the basis of mass alone will be ambiguous for at least some fragments. Indeed, of the ∼400 000 calculated fragments of lysozyme which would be resolvable on a mass spectrometer with infinite resolution, only ∼300 will be resolved at 50 ppm resolution and ∼100 000 at 1 ppm. Various approaches have been devised to overcome problems of this type.[20−22] In this paper, we utilize atom counting to determine elements of the empirical formula

* Author to whom correspondence should be addressed.
[†] Oxford Centre for Molecular Sciences (OCMS).
[‡] Bruker Instruments, Inc.

(1) Mann, M.; Wilm, M. *Trends Biochem. Sci.* **1995**, *20*, 219−224.
(2) Siuzdak, G. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11290−11297.
(3) Just, I.; Selzer, J.; Wilm, M.; Voneichelstreiber, C.; Mann, M.; Aktories, K. *Nature* **1995**, *375*, 500−503.
(4) Chowdhury, S. K.; Katta, V.; Chait, B. T. *J. Am. Chem. Soc.* **1990**, *112*, 9012−9013.
(5) Miranker, A.; Robinson, C. V.; Radford, S. E.; Aplin, R. T.; Dobson, C. M. *Science* **1993**, *262*, 896−900.
(6) Wagner, D. S.; Melton, L. G.; Yan, Y.; Erickson, B. W.; Anderegg, R. J. *Protein Sci.* **1994**, *3*, 1305−1314.
(7) Zhao, Y. M.; Chait, B. T. *Anal. Chem.* **1994**, *66*, 3723−3726.
(8) Johnson, R. S.; Walsh, K. A. *Protein Sci.* **1994**, *3*, 2411−2418.
(9) Anderegg, R. J.; Wagner, D. S. *J. Am. Chem. Soc.* **1995**, *117*, 1374−1377.
(10) Miranker, A.; Robinson, C. V.; Radford, S. E.; Dobson, C. M. *FASEB J.* **1996**, *10*, 93−101.
(11) Hooke, S. D.; Eyles, S. J.; Miranker, A.; Radford, S. E.; Robinson, C. V.; Dobson, C. M. *J. Am. Chem. Soc.* **1994**, *117*, 7548−7549.
(12) Robinson, C. V.; Groβ, M.; Eyles, S. J.; Ewbank, J. J.; Mayhew, M.; Hartl, F. U.; Dobson, C. M.; Radford, S. E. *Nature* **1994**, *372*, 646−651.
(13) Dobson, C. M.; Evans, P. A.; Radford, S. E. *Trends Biochem. Sci.* **1994**, *19*, 31−37.
(14) Radford, S. E.; Dobson, C. M. *Philos. Trans. R. Soc. London B* **1995**, *348*, 17−25.
(15) Hen lysozyme was obtained from Sigma and dialyzed for 3 days against three changes of dilute HCl solution in H_2O at pH 3.8 to remove residual salts. All FTMS spectra were recorded on a Bruker Bio-APEX 47e spectrometer equipped with a 4.7 T magnet, external ion source, and electrospray source (Analytica of Branford, Branford, CT). Ions were accumulated in the FTMS trap for 300 ms, followed by chirp excitation and detection with a 240 kHz bandwidth. Parent multiply charged molecular ions of lysozyme were observed with a capillary skimmer potential of 60−100 V, while for CID spectra the capillary skimmer potential was increased to 160−250 V. Calibration was performed against a mixture of peptides.

(16) Buchanan, M. V.; Hettich, R. L. *Anal. Chem.* **1993**, *65*, 245A−259A.
(17) Yost, R. A.; Boyd, R. K. *Methods Enzymol.* **1990**, *193*, 131−153.
(18) Biemann, K.; Martin, S. A. *Mass Spectrom. Rev.* **1987**, *6*, 1−75.
(19) A library of fragments was computed from which to identify the closest possibilities for each of the observed ion masses. The program is implemented in Mathematica, can trace fragments through intact disulfides, and optionally allows disulfide cleavage. All fragments which include one or more cleavages of α-carbon−amide nitrogen, amide nitrogen−carbonyl carbon, or carbonyl carbon−α-carbon peptide backbone bonds are considered. In addition, all fragments in which one disulfide was cleaved at either the sulfur−sulfur or β-carbon−sulfur bond were also considered.
(20) Senko, M. W.; Beu, S. C.; Mclafferty, F. W. *Anal. Chem.* **1994**, *66*, 415−417.
(21) Tang, X. J.; Thibault, P.; Boyd, R. K. *Anal. Chem.* **1993**, *65*, 2824−2834.
(22) Sepetov, N. F.; Issakova, O. L.; Lebl, M.; Swiderek, K.; Stahl, D. C.; Lee, T. D. *Rapid Commun. Mass Spectrom.* **1993**, *7*, 58−62.
(23) Clore, G. M.; Gronenborn, A. M. *Methods Enzymol.* **1994**, *239*, 349−363 and others in this volume.

Table 1. (A) Assignment for the Fragment Ion at 584.4, Illustrating 9 of 30 Possibilities[25] Made on the Basis of *m/z* Alone (±50 ppm) and (B) the Nearest 3 Fragments with a Maximum of 2 Cleavages and the Correct Nitrogen and Carbon Atom Count

m/z calcd.	sequence	fragment
	(A)	
584.30	20−29	$(a_{29}y_{110}')^{2+}_{10}$
584.30	13−21	$(b_{21}x_{117}')^{2+}_{9}$
584.30	106−114	$(c_{114}y_{24})^{2+}_{9}$
584.31	9−19	$(b_{19}y_{121}'''')^{2+}_{11}$
584.31	1−7−127−129	$(a_7^6z_3^{127}'''')^{2+}_{10}$
584.31	13−22	$(a_{22}y_{117})^{2+}_{10}$
584.32	116−126	$(a_{126}z_{14})^{2+}_{11}$
584.33	55−63	$(a_{63}y_{75})^{2+}_{9}$
584.34	**120−129**	$(y_{10}s_{127}^0'')^{2+}_{10}$
	(B)	
569.29	106−114	$(b_{114}z_{24})^{2+}_{9}$
584.34	**120−129**	$(y_{10}s_{127}^0'')^{2+}_{10}$
594.29	14−23	$(a_{23}z_{116})^{2+}_{10}$

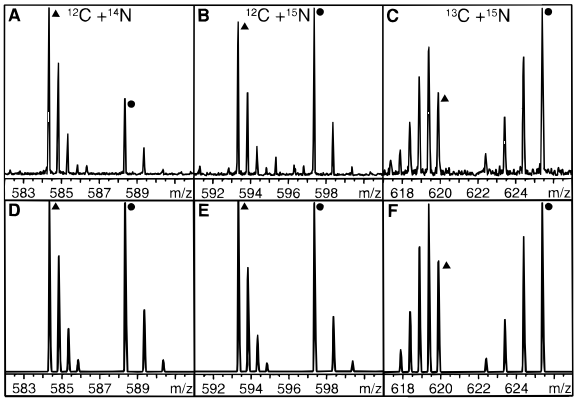

Figure 2. (A) Expansion of the CID spectrum of ^{14}N lysozyme from 582 to 592 *m/z*. (B) Expansion of CID spectrum of ^{15}N lysozyme from 591 to 601 *m/z*. (C) Expansion of CID spectrum of $^{15}N^{13}C$ lysozyme from 617 to 627 *m/z*. The offset in *m/z* from A to B is used to count the number of nitrogens in the fragments, while the offset from B to C is used to count the number of carbon atoms in the fragments. Note that the lightest peaks in A and B are compared with the most massive peaks in C. The isotopic labeling enables these two fragments to be assigned to residues 120−129 and 1−5 with molecular formula $C_{53}H_{88}N_{18}O_{12}$ and $C_{28}H_{46}N_9O_5$, respectively. The standard CID notation[30] for these fragments is $(y_{10}s_{127}^0'')^{2+}_{10}$ and b_5, respectively. Simulation of the distributions at natural abundance (D), 99.8% ^{15}N (E), and 99.8% ^{15}N, 97.3% ^{13}C (F) are shown. The primary structure of hen lysozyme is cross-linked by disulfide bonds between residues 6−127, 30−115, 64−80, and 76−94.

of fragment ions. The approach we use is to exploit isotopically labeled proteins now routinely produced for NMR studies of structure and dynamics.[23]

The pattern of fragmentation for isotopically labeled lysozymes (Figure 2B,C) was found to be closely similar to that of unlabeled lysozyme (Figure 2A) except that individual fragments were found to have shifted to greater *m/z*. Corresponding peaks in the spectra can therefore be used to count the number of nitrogen and carbon atoms in individual fragments. This procedure results in delineation of 18 nitrogen and 53 carbon atoms for the +2 charged fragment ion at 584.4 and 9 nitrogen and 28 carbon atoms for the +1 charged fragment ion at 588.4. For these fragments, only one possibility with the correct carbon and nitrogen count is found within 50 ppm of the experimental *m/z*; indeed, the next nearest *m/z* values for theoretical fragments in each case are 730 and 1700 ppm, allowing the assignments to be made with confidence. If we restrict the information from atom counting with a single reasonable assumption, for example, that the range of possible fragments includes a maximum of two main chain and/or disulfide bond cleavages, the next nearest theoretical fragment ion to that at 584.4 is 17 000 ppm distant (Table 1B),[24] and 20 of the 24 fragments studied here can be uniquely identified.[25]

The assignment of the CID spectra shows clearly the presence of internal fragments as well as fragmentation of the disulfide at 6−127. The unique assignment of fragment ions at 584.4, 600.3, 992.52, and 1008.5 to disulfide cleavage and the result that these ions are members of two sets of triplets characteristic of disulfide cleavage in which the fragment retains zero, one, or two sulfur atoms substantiates this result.[27]

Another important conclusion from analysis of the CID spectrum of native lysozyme compared to unstructured polypeptides is the presence of non-amide bond cleavage at specific sites.[17,18] Twenty-three of the twenty-four fragments studied here result from some combination of fragmentation at only 6% of the bonds along the polypeptide chain (between residues 5−7, 18−19, 29−30, 119−121, and 125−127). Furthermore, eight of the fragments have one bond break to the N-terminal side of either the 6−127 or the 30−115 disulfide. Of these, half result from fragmentation of the C_α−N bond with one assigned to a carbonyl carbon−C_α break.[19] This is consistent with fragmentation mechanisms in which the bond strength for both N−C bonds along the polypeptide chain is reduced upon protonation of the amide nitrogen.[18,26] These findings suggest that amide cleavage cannot be assumed perhaps because this cross-linked protein has sufficient structure in the gas phase for nonrandom localization and movement of labile charges.[28]

By exploitation of uniform isotopic labeling, the identification of many CID fragments has been found to be straightforward and should complement other approaches to assignment. We are presently extending these experiments to probe fragments partially labeled with deuterium following exchange experiments to obtain site-specific labeling information. While this strategy is complicated by issues dealing with molecular rearrangment,[28] the method has already been shown to be viable for helical peptides.[29]

Acknowledgment. The OCMS is supported by the UK BBSRC, EPSRC, and MRC. The research of C.M.D. is supported in part by an International Research Scholar award from the Howard Hughes Medical Institute. A.D.M. is a junior research fellow of Christ Church, Oxford, and is supported by NATO under a grant awarded in 1994. We thank D. Archer for labeled lysozyme and F. M. Richards for helpful discussion in the preparation of this manuscript.

Supporting Information Available: A table containing the experimental *m/z*, charge, number of carbon and nitrogen atoms, assignment by sequence, fragmentation notation,[30] and calculated *m/z* of the 24 fragments discussed in the text (1 page). See any current masthead page for ordering and Internet access instructions.

JA9531236

(24) On the basis of the assigned spectrum, an internal assessment of the mass accuracy of the instrument was found to be 8 ± 5 ppm.

(25) A peak assigned to an internal fragment in which both cleavages are at the peptide bond has identical mass and sequence to an assignment in which both cleavages are shifted by one bond toward the N or the C terminus of the polypeptide chain. These degenerate possibilities are ignored.

(26) McCormack, A. L.; Somogyi, A.; Dongre, A. R.; Wysocki, V. H. *Anal. Chem.* **1993**, *65*, 2859−2872.

(27) Bean, M. F.; Carr, S. A. *Anal. Biochem.* **1992**, *201*, 216−226.

(28) Johnson, R. S.; Krylov, D.; Walsh, K. A. *J. Mass Spectrom.* **1995**, *30*, 386−387.

(29) Anderegg, R. J.; Wagner, D. S.; Stevenson, C. L.; Borchardt, R. T. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 425−433.

(30) Roepstorff, P.; Fohlman, J. *Biomed. Mass Spectrom.* **1984**, *11*, 601. Nomenclature was extended to include disulfide breakage, s-. The subscript of s refers to the particular cysteine residue, while its superscript refers to the retained number of sulfur atoms. For fragments containing two peptides joined by an unbroken disulfide, superscripts to a-, b-, c- , x-, y-, and z- refer to the cysteine residue which forms the link. Charge is indicated for fragment ions whose charge is greater than +1. Primes indicate the number of hydrogen atoms added to the fragment and deviate from the established nomenclature in that they are always placed at the end of the fragment description. The possibility of charges localized on basic and/or acidic side chains precludes determination of whether charges are positioned at the cleavage site.