

Cooperative Elements in Protein Folding Monitored by Electrospray Ionization Mass Spectrometry

Shaun D. Hooke, Stephen J. Eyles, Andrew Miranker, Sheena E. Radford, Carol V. Robinson, and Christopher M. Dobson*

Oxford Centre for Molecular Sciences & New Chemistry Laboratory, University of Oxford South Parks Road, Oxford OX1 3QT, U.K.

Received December 22, 1994

Electrospray ionization mass spectrometry (ESI MS)^{1,2} can monitor the formation of persistent structure during protein folding through the increase in mass associated with the trapping of deuterons in hydrogen exchange experiments.^{3–5} Here, we demonstrate that differences in the cooperativity of folding events can be detected directly by mass analysis. The differences studied in this work result from sequence changes in variant lysozymes and are detected for pairs of proteins mixed and refolded simultaneously. The attributes of the technique include the rapidity of the method, the small quantities of protein required for analysis, and the certainty that, by monitoring mixtures of proteins simultaneously, each protein experiences identical folding conditions.

We have previously demonstrated that a partially folded state of lysozyme, protected only in one of its two domains (the α -domain), can be readily detected by this technique.³ Furthermore, the evolution of a state protected in both domains and the loss of the fully unprotected protein can also be measured. These data, together with results from pulse-labeling experiments monitored by NMR and a variety of stopped-flow optical techniques, have allowed us to propose a detailed model for the folding pathway of lysozyme which involves multiple pathways, distinct folding domains, and partially folded intermediate states.^{6,7} Thus, hen lysozyme represents an excellent model protein for studying the effects of variants on the cooperativity of folding events.

Figure 1 shows three charge states of the ESI mass spectrum of hen lysozyme, initially denatured and deuterated in its labile hydrogens, and then refolded and pulse labeled^{8,9} with protons 38 ms after the initiation of folding. The protein contains two isotopic states, one where ¹⁵N is in natural abundance (0.1%), referred to as ¹⁴N-hen lysozyme, and the other in which ¹⁵N was incorporated uniformly to >98%,¹⁰ referred to as ¹⁵N-hen lysozyme. The two isotopically distinct proteins can be readily resolved by ESI MS as a result of the mass difference arising from their differing ¹⁴N/¹⁵N content. Furthermore, additional peaks are observed as a consequence of the selective trapping of ²H in place of ¹H atoms into the proteins during the folding experiment.

(1) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* **1989**, *246*, 64.

(2) Smith, R. D.; Loo J.A.; Edmonds, C. G.; Barinaga, C. J.; Udseth, H. R. *Anal. Chem.* **1990**, *62*, 882.

(3) Miranker, A.; Robinson, C. V.; Radford, S. E.; Aplin, R. T.; Dobson, C. M. *Science* **1993**, *262*, 896.

(4) Robinson, C. V.; Groß, M.; Eyles, S. J.; Ewbank, J. J.; Mayhew, M.; Hartl, F. U.; Dobson, C. M.; Radford, S. E. *Nature* **1994**, *372*, 645.

(5) Wagner, D. S.; Melton, L. G.; Yan, Y.; Erickson, B. W.; Anderregg, R. J. *Protein Sci.* **1994**, *3*, 1305.

(6) Radford, S. E.; Dobson, C. M. *Phil. Trans. R. Soc. London, Ser. B*, in press.

(7) Dobson, C. M.; Evans, P. A.; Radford, S. E. *Trends Biochem. Sci.* **1994**, *19*, 31.

(8) Baldwin, R. L. *Curr. Opin. Struct. Biol.* **1993**, *3*, 84–91.

(9) Evans, P. A.; Radford, S. E. *Curr. Opin. Struct. Biol.* **1994**, *4*, 100–106.

(10) Roberts, I. N.; Mackenzie, D. A.; Jeenes, D. B.; Archer, D. B.; Radford, S. E.; Robinson, C. V.; Aplin, R. T.; Dobson, C. M. *Biotechnol. Lett.* **1992**, *14*, 897.

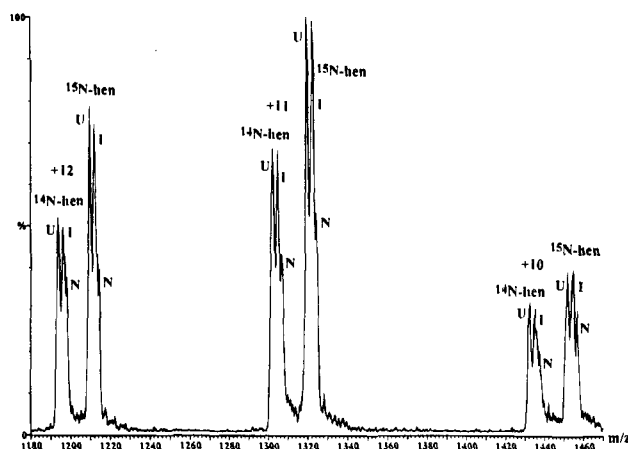


Figure 1. The ESI mass spectrum of approximately equimolar amounts of ¹⁴N-hen lysozyme and ¹⁵N-hen lysozyme pulse labeled 38 ms after initiation of folding. The measured masses for the ¹⁴N-hen lysozyme U, I, and N states are 14313.1 ± 1.1 , 14342.0 ± 0.5 , and 14368.03 ± 3.6 , respectively, and those for the ¹⁵N-hen lysozyme U, I, and N states are 14506.0 ± 0.2 , 14534.7 ± 1.3 , and 14560 ± 2.6 , respectively. All spectra were recorded on a VG BioQ quadrupole mass spectrometer equipped with an ABI dual syringe pump delivering 10 μ L/min of 20% acetonitrile and 80% water at pH 3.8. The samples for analysis were diluted into the same solution as the delivery solvent at a protein concentration of 25 pmol/ μ L. Samples were pulse labeled as described previously.¹² The spectrum is the sum of 10 scans and represents the raw data with minimal smoothing.

The cooperative nature of the folding process is immediately apparent since three distinct populations of molecules are observed for each of the charge states of both ¹⁴N-hen lysozyme and ¹⁵N-hen lysozyme.¹¹ The lowest molecular weight species in each charge state (U) correspond to molecules which have not refolded sufficiently to trap any deuterons prior to the labeling pulse. The species N correspond to those molecules which have folded to trap the number of deuterons found to be protected in the native state, resulting in a mass increase of 55 ± 2 Da over that of the unprotected state.³ The species labeled I represent the population of molecules with a mass intermediate between the masses of N and U resulting from the incorporation of 28 ± 1 deuterons at the time of the labeling pulse, in structure able to protect against exchange. Previous work, involving measurement of the deuterium occupancy at specific sites by NMR, has shown that the latter arises from the existence of a well-defined intermediate resulting from the folding of one of the two domains of the lysozyme structure (the α -domain) in the absence of persistent structure in the other (the β -domain).¹² ¹⁴N- and ¹⁵N-hen lysozymes clearly demonstrate identical folding behavior since the profile of peaks in the ESI mass spectrum was found to be identical in both charge state series not only at this time point but also throughout the time course of folding (3.5 ms to 2 s).

Comparison of the folding behavior of lysozymes from three different sources is shown in Figure 2. Each spectrum contains peaks from two proteins, one of which is always the hen lysozyme. The pairs of proteins have been refolded and pulse labeled simultaneously, and subsequently analyzed in the mixture by ESI MS. The proteins are from Lady Amhurst pheasant (LAP)¹³ (a), Japanese quail (JQ)¹⁴ (b), and human¹⁵

(11) The mass of each species was calculated from the three charge states +10, +11, and +12, shown in Figure 1. Masses quoted represent the centroid value which takes into account the asymmetry of the peak, arising from the natural isotopic distribution, together with the distribution of deuterons in the partially labeled species.

(12) Radford, S. E.; Dobson, C. M., & Evans, P. A. *Nature* **1992**, *358*, 302.

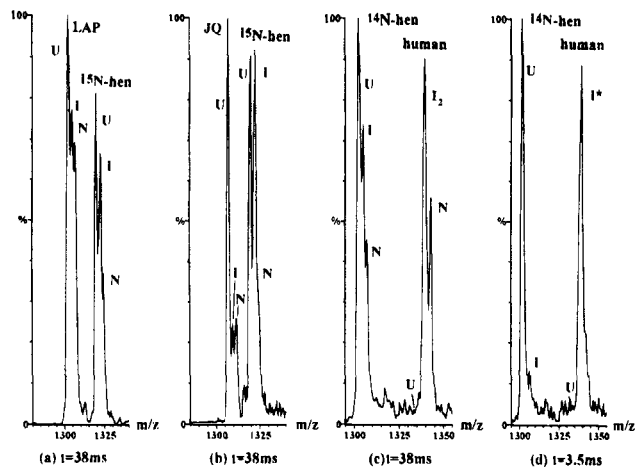


Figure 2. ESI mass spectra of (a) LAP and ^{15}N -hen lysozyme pulse labeled after 38 ms of refolding, (b) JQ and ^{15}N -hen lysozyme after 38 ms of refolding, and (c) human and ^{14}N -hen lysozyme after 38 ms of refolding and (d) after 3.5 ms of refolding. The +11 charge state is shown in each case, and spectra recorded as described above. The measured masses of the species are LAP U, I, and N $14\,312.8 \pm 0.4$, $14\,343.1 \pm 2$, and $14\,365.0 \pm 1.6$ Da, ^{15}N -hen lysozyme U, I, and N $14\,505.7 \pm 1.1$, $14\,535.2 \pm 1.1$, and $14\,559.1 \pm 1.1$ Da, JQ U, I, and N $14\,365.4 \pm 0.3$, $14\,393.7 \pm 0.5$, and $14\,418.1 \pm 0.6$ Da, ^{14}N -hen U, I, and N $14\,318.33 \pm 1.1$, $14\,345.5 \pm 1.7$, and $14\,367.8 \pm 1.1$ Da, human lysozyme I₂ and N $14\,729.8 \pm 1.1$ and $14\,761.9 \pm 1.8$ Da, and human I* $14\,721.9 \pm 2$ Da. Small differences in the population of I in the ^{15}N -hen lysozyme ($\pm 20\%$) arise from inherent errors in the pulse-labeling technique. For example, this method is exquisitely sensitive to small differences in temperature and pH between experiments. LAP and JQ lysozymes were prepared from egg white using repeated ion exchange chromatography, and human lysozyme was prepared as described previously.¹⁸

(c) sources, and the numbers of substitutions in the sequence of 129 residues from hen lysozyme are 6, 6, and 52, respectively. In experiments with LAP and JQ lysozymes, ^{15}N -hen lysozyme (14 498 Da) was used instead of the natural abundance ^{14}N -hen lysozyme as the control to enable complete mass resolution of the peaks from the two protein samples throughout the time course of folding.

The folding behavior of the four variant lysozymes is qualitatively similar since all demonstrate the formation of a well-defined intermediate (I) with protection of 28 ± 1 deuterons. The fact that the same number of amide deuterons is protected in the intermediate in each variant, despite their sequence differences, strongly suggests that the α -domain folds in a highly cooperative manner prior to the stabilized β -domain in each case. These data show, however, clear differences in the rate of formation of both the intermediate and the fully folded structures as demonstrated by their relative populations at this refolding time; for example, the cooperative formation of the α -domain in JQ lysozyme is considerably slower, relative to the β -domain, than that observed for the hen protein or indeed the LAP protein. It is interesting that all the amino acid substitutions in the JQ protein relative to the hen protein occur within the α -domain,¹⁶ while in the LAP protein substitutions occur within both domains.¹⁷ Given the high sensitivity in the

pulse-labeling technique to the precise experimental conditions, the presence of the hen protein in each of these experiments allows detailed comparisons to be made conclusively. Thus, even the small differences between the hen and LAP protein cannot be attributed to variations in experimental conditions since the proteins were folded in the same solution. This approach, of folding mixtures of proteins, is therefore of great value in comparative studies of protein folding.

The most substantive differences in the spectra occur in the comparison of the hen and the human lysozymes. After a folding time of 3.5 ms, Figure 2d, for human lysozyme a distinct population is observed with a mass 17 ± 1 Da larger than that of the mass measured for the unprotected protein in the control sample. After 38 ms, Figure 2c, the species with 17 ± 1 protected amides is no longer apparent, but a species which protects 28 ± 1 amides is observed, indicating that the formation of the α -domain is now complete. The detection of a well-defined species with fewer than 28 deuterons protected from exchange provides direct evidence for the formation of a smaller cooperative unit prior to the formation of the complete α -domain. NMR experiments have indicated that amide hydrogens in two of the α -helices (A and B), along with a short region of 3¹⁰ helix, are faster than the kinetics of protection of the remaining two helices (C and D)¹⁸ in the α -domain. The A and B helices together with the 3¹⁰ helix are packed together in the native structure of human lysozyme, suggesting that this unit is a well-defined subdomain of the native protein that is formed cooperatively during folding.

In hen lysozyme, stopped-flow optical experiments have indicated that a collapsed state containing extensive secondary structure forms within a few milliseconds of the initiation of refolding.¹² That high levels of hydrogen exchange protection occur only at later stages of folding has been associated, at least in part, with the need to reorganize incorrectly folded species prior to the formation of persistent structural units.^{6,7} The present results indicate that the time required for such reorganizational processes can differ significantly even for closely related proteins. Furthermore, the finding that for human lysozyme, virtually no unprotected protein is observed even after a folding time of 3.5 ms suggests that for this protein any misfolding events in the subdomain containing the A/B and 3¹⁰ helices are less probable or, if they occur, can be resolved much more readily than in the C/D helical region.

The changes in amino acid sequence between the four proteins examined here have little effect on the overall fold of the native states.¹⁹ It is intriguing, therefore, that such changes can have a profound effect on the formation of structure during folding. Examination of the sequence differences between natural variants having different folding properties can therefore give insight as to key residues affecting protein folding.¹⁸ The ability of the ESI MS method to probe differences rapidly and effectively with small quantities of protein suggests that it will be a particularly powerful technique in studies to elucidate the role of specific residues in protein-folding mechanisms.

Acknowledgment. The Oxford Centre for Molecular Sciences is supported by the UK BBSRC, EPSRC, and M.R.C. The research of C.M.D. is supported in part by an International Research Scholars award from the Howard Hughes Medical Institute. A.M. is supported by the North Atlantic Treaty Organisation under a grant awarded in 1994. S.E.R. is a 1983 Royal Society University Research Fellow. We acknowledge with thanks R. T. Aplin for helpful discussions, R. G. Procter for expert technical assistance, and D. Smith (Bristol University) for the gift of Japanese quail eggs.

JA944115C

(18) Hooke, S. D.; Radford, S. E.; Dobson, C. M. *Biochemistry* **1994**, *33*, 5867.

(19) McKenzie, A.; White, F. H. *Adv. Protein Chem.* **1991**, *41*, 173.

(13) Araki, T.; Kuramoto, M.; Torikata, T. *Agric. Biol. Chem.* **1990**, *54*, 2299.

(14) Houclusse, A.; Bentley, G. A.; Poljak, R. J.; Souchan, H.; Zhang, Z. Submitted.

(15) Artymuik, P. J.; Blake, C. C. F. *J. Mol. Biol.* **1981**, *152*, 737.

(16) JQ lysozyme has the following differences from the hen sequence F3Y, N19K, R21Q, G102V, N103H, and Q121N.

(17) LAP lysozyme has the following differences from the hen sequence: F3Y, H15L, Q41H, N77H, Q121N, and I124T.