Protein Folding Monitored at Individual Residues During a Two-Dimensional NMR Experiment

Jochen Balbach,* Vincent Forge,† Wai Shun Lau, Nico A. J. van Nuland, Keith Brew, Christopher M. Dobson‡

An approach is described to monitor directly at the level of individual residues the formation of structure during protein folding. A two-dimensional heteronuclear nuclear magnetic resonance (NMR) spectrum was recorded after the rapid initiation of the refolding of a protein labeled with nitrogen-15. The intensities and line shapes of the cross peaks in the spectrum reflected the kinetic time course of the folding events that occurred during the spectral accumulation. The method was used to demonstrate the cooperative nature of the acquisition of the native main chain fold of apo bovine α-lactalbumin. The general approach, however, should be applicable to the investigation of a wide range of chemical reactions.

The mechanism by which a protein folds to its native state after biosynthesis remains one of the central unresolved issues in structural biology (*1*). Studying the events occurring during protein folding is a major challenge to conventional structural methods such as NMR spectroscopy because of their rapidity and complexity (*2*). Most detailed structural information has therefore been sought indirectly—for example, through hydrogen-exchange pulse-labeling experiments (*3*)—or from the study of stable analogs of folding intermediates (*4*). These studies have provided information of considerable value, but it is essential to link such information to that from direct observations of the folding process (*3*, *4*). It has recently been shown that NMR spectroscopy can be applied in real time, after rapid mixing procedures, for such a purpose (*5*–*7*). However, to date, this approach has been limited to one-dimensional (1D) experiments and, hence, to low-resolution studies, because of the presumed limitations imposed by the relatively long accumulation times required for even the fastest 2D experiments (*8*). Two-dimensional procedures exploiting exchange phenomena have, however, been used to study fast chemical processes (*9*), including protein folding (*10*), occurring under equilibrium conditions. We have extended such procedures to nonequilibrium situations, demonstrating the power of modern heteronuclear NMR methodology to follow the development of structure at the level of individual residues in a protein. Moreover, the approach will provide an opportunity to apply high-resolution NMR spectroscopy to investigate reactions of other complex systems when rate processes are too rapid for conventional repetitive-accumulation procedures.

Instead of collecting a series of spectra at different stages of a reaction, the idea is to allow the reaction to occur during the accumulation of a single 2D experiment. The NMR spectrum, if recorded in an appropriate manner, will then contain the kinetic history of the events that take place during the reaction. Consider a ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) experiment (*11*). A pulse delay (t_1) is incremented during the course of the ^{1}H data accumulation to generate the indirect ^{15}N time domain. This effectively involves the collection of a series of independent data sets as a function of time. If a reaction occurs during this process, the magnitude of the signal from each species will be different at each of the t_1 increments, reflecting its appearance or disappearance with time. The signals from the different species in the frequency domain (F1) spectrum after double Fourier transformation will thus be convoluted by their kinetic profile in the reaction.

We used this approach to examine the folding of α-lactalbumin, a protein whose properties have contributed significantly to our understanding of a number of aspects of the folding process (*12*). Folding of the protein was initiated from a denatured state formed at pH 2.0 and 3°C. Under these conditions, the protein was in a molten globular state: The far-ultraviolet (UV) circular dichroism (CD) spectrum indicated extensive secondary structure, but the near-UV CD and fluorescence spectra indicated a lack of persistent tertiary interactions (*13*). The ^{1}H-^{15}N HSQC spectrum of this state (Fig. 1A) was poorly resolved, the majority of resonances being broad with little dispersion of the chemical shift; indeed, only a few resonances are detectable as individual peaks. This situation arises from exchange effects associated with slow fluctuations within the disordered state (*12*). The spectrum of the same sample recorded 80 min after the pH was increased to 7.0 in a rapid mix experiment (Fig. 1C) was, however, well resolved and corresponds to the spectrum of the fully native state, indicating that the protein had refolded completely.

A ^{1}H-^{15}N HSQC spectrum was initiated

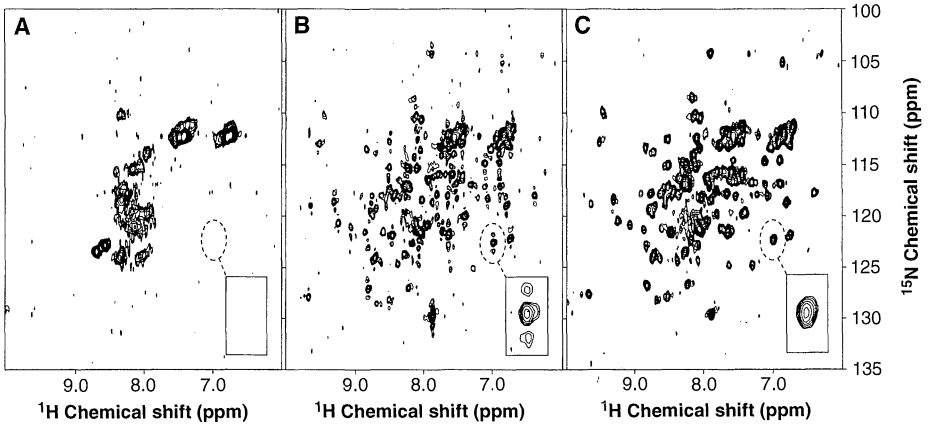

Fig. 1. ^{1}H-^{15}N HSQC spectra of bovine α-lactalbumin (BLA) at 3°C during different stages of the folding process: (**A**) HSQC of the denatured state at pH 2.0 (A state), (**B**) kinetic HSQC accumulated during folding, and (**C**) reference HSQC of the native state (N state). The spectra of the A and N states were recorded at equilibrium before and after the folding step (*26*). (**Insets**) Representative enlargements of the indicated region where the signal of Val^{92} resonates in the native state. The signal intensity in (B) is reduced compared to that in (C). As shown in the inset, additional components to the cross peaks are observed in the kinetic spectrum; these components have negative intensity.

J. Balbach, V. Forge, W. S. Lau, N. A. J. van Nuland, C. M. Dobson, Oxford Centre for Molecular Sciences, New Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QT, UK.
K. Brew, Department of Biochemistry and Molecular Biology, University of Miami School of Medicine, Miami, FL 33101, USA.

*Present address: Laboratorium für Biochemie, Universität Bayreuth, D-95440 Bayreuth, Germany.
†Present address: CNRS, Unité de Recherche Associée 2096, Section de Biophysique des Proteines et des Membranes, Department de Biologie Cellulaire et Moleculaire, Commissariat a L'Energie Atomique–Centre d'Etudes de Saclay, 91191 Gif-sur-Yvette, France.
‡To whom correspondence should be addressed.

immediately after the mix and accumulated for 31 min (Fig. 1B). Under these conditions, the refolding of the apo protein has a half-time of about 5 min; refolding was therefore complete within the accumulation time of the experiment. Examination of this spectrum shows resonances from two distinct species. One of these corresponds to the initial state, which under these conditions has been shown by stopped-flow CD, fluorescence spectroscopy, and 1D ^{1}H NMR to correspond closely to the low-pH denatured state from which refolding was initiated (*6*, *14*). Several broad resonances corresponding to this state are visible in the 2D spectrum, but the majority of the peaks have chemical shifts identical to those in the fully folded state. The intensities of these peaks are, however, less than those in the spectrum of the fully refolded protein, and the line shapes of the individual cross peaks show negative components in the ^{15}N dimension (Fig. 2). Both of these observations turn out to be key factors in the kinetic analysis.

A series of simulations was carried out for two simple kinetic schemes (Fig. 3). Both the intensities and line shapes of the resonances of different species in a reaction are strongly dependent on the reaction profile and reflect the explicit time course of the reaction. In a simple two-component system (A→B) (Fig. 3, left), the disappearance with time of species A, present at the beginning of the accumulation, will result in the faster decay of its magnetization in t_1 than would be the case for a nonreacting system, resulting in line broadening of its associated resonances after Fourier transformation. By contrast, the magnetization of the product species B will develop only during the course of the reaction, resulting in a gradual increase of its magnetization in t_1; this gives rise to narrower lines and negative components in the wings of the resulting peaks. For a reaction generating a transient species (A→B→C), more complex effects will occur (Fig. 3, right), and the intensity and line shape of the resonance of species B will reflect the history of the appearance and disappearance of the reaction intermediate (*15*).

These simulations provide a basis for interpretation of the data in Fig. 1B. The low intensity in the spectrum of the resonances of the initial denatured state can be attributed to the broadening associated with the disappearance of this state during the folding reaction, whereas the complex line shapes of the native state result from its gradual appearance during the spectral accumulation. The lack of additional peaks indicates that there is no significant population of a well-defined intermediate state. The line shapes of the well-resolved resonances provide us with an opportunity to monitor the kinetics of the development of the native character at the level of individual residues in the main chain of the protein. To achieve this monitoring, we simulated cross sections such as those in Fig. 2 using a model of the protein folding process in which the native structure at each site in the protein is assumed to develop by simple exponential kinetics (*16*). It was possible to fit experimental data well using such a procedure (Fig. 4 and Table 1).

The time constants for the formation of the native state measured by fluorescence and 1D NMR have been shown to be identical (*6*). These methods detect close packing of side chains in the native structure. The 2D experiment, by contrast, enables us to monitor the formation of native-like structure in the polypeptide main chain. The mean time constant for the amides given in Table 1 is 400 ± 110 s, which is

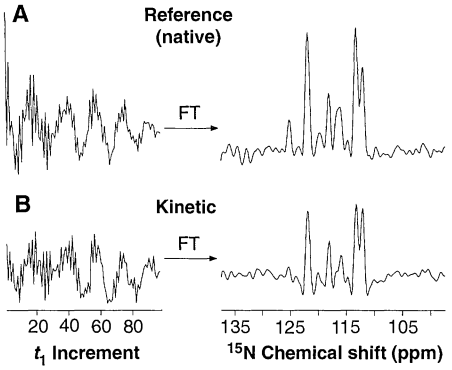

Fig. 2. Cross sections from (**A**) the reference HSQC of the native state and (**B**) the kinetic HSQC recorded during refolding at a ^{1}H chemical shift of 7.6 ppm. (A) Sections from the ^{1}H-^{15}N HSQC data set for the native-state contour plots in Fig. 1C. The time-domain spectrum (left) shows the magnetization as a function of the 96 t_1 increment steps. Each increment required 19.4 s to collect, resulting in a total accumulation time of 31 min. The corresponding frequency-domain spectrum (right) shows the different ^{15}N resonances correlated with ^{1}H resonances at this frequency. (B) Equivalent sections from the kinetic ^{1}H-^{15}N HSQC data set accumulated after the initiation of folding, shown as a contour plot in Fig. 1B.

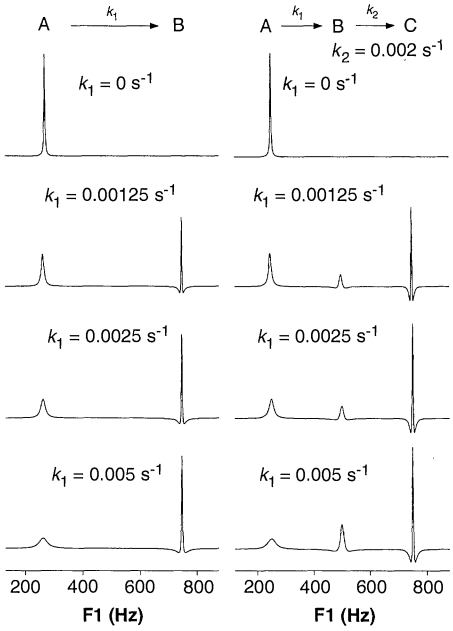

Fig. 3. Simulated HSQC line shapes for simple kinetic schemes. Left panel: line shapes resulting from a one-sided transition A → B for various rate constants k_1 in the indirect t_1 dimension of a 2D data set (*27*). The intensity of the resonance of spin A at frequency ν_A is additionally modulated in the t_1 dimension by $\exp(-tk_1)$, and that of spin B at ν_B is modulated by $[1 - \exp(-tk_1)]$, where t is the time after initiation of the reaction, assuming that at the beginning of the reaction only A is present. Right panel: An analogous line shape analysis for a transition involving an intermediate state B, emerging from state A with rate constant k_1 and evolving to state C with rate constant k_2. In these simulations, k_1 is varied while k_2 is kept fixed. The frequencies of the lines are 250, 500, and 750 Hz, respectively, and the natural linewidth of each is 14 Hz.

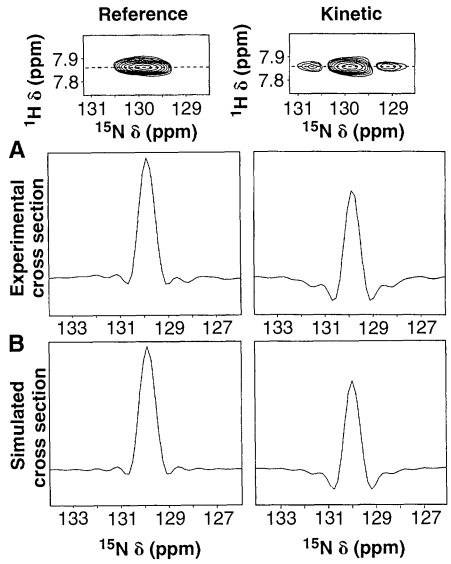

Fig. 4. (**A**) Experimental and (**B**) simulated cross sections of Leu[123] centered at the ^{1}H chemical shift (δ) indicated by the dotted line in the reference (native) (left) and kinetic (right) HSQC spectra (top). A simulated FID was generated for each signal using the linewidth, chemical shift, and amplitude of the experimental cross section of the reference signal (A, left). After Fourier transformation this resulted in a simulated cross section for the native state (B, left). This FID was then multiplied by a series of single exponential functions to simulate kinetic cross sections corresponding to different time constants for the folding reaction. The best fit (B, right), corresponding to a time constant τ (the inverse of the rate constant k) of 410 s, was determined by minimizing the difference from the experimental cross section (A, right).

consistent within experimental error with the value of 470 ± 15 s found for aromatic side chains from fluorescence measurements under similar experimental conditions (*17*). Moreover, the time constants for the individual secondary structure elements are all within one standard deviation of the mean, with the exception of that for the β-sheet region between residues 40 and 50. This region may achieve a native-like structure in its main chain before other regions, but the small number of probes and their low signal-to-noise ratios suggest that such a conclusion is premature. The data indicate that, despite the early existence of a native-like far-UV CD (*13*) spectrum and a degree of hydrogen exchange protection (*14*), the native-like environment of the main chain in at least the majority of the protein structure emerges cooperatively and in concert with that of the side chains. The absence of resolved resonances corresponding to a well-defined intermediate in the spectrum indicates that before this final step in folding, the main-chain environment remains heterogeneous. It is particularly interesting in this regard that extensive line broadening has been reported for species formed in the early stages of the kinetic unfolding of ribonuclease A (*18*) and dihydrofolate reductase (*7*) and has been attributed to dynamic processes resulting from the loss of the tight packing characteristic of the fully native state.

A wide variety of multidimensional NMR experiments, in addition to the ^{1}H-^{15}N HSQC experiment used here, are based on the same principle of incremented delays (*19*); these methods could potentially provide additional information about the events occurring during the folding reaction. The general approach requires slow folding of the protein system under study; rates faster than ~0.05 s^{-1} will not give detectable effects for a quantitative analysis, but improved pulse sequences can be explored to reduce this limit substantially. Here, removal of a structural metal ion was used to achieve refolding rates appropriate for detection. Other strategies include the generation of slow-folding mutants (*20*) and the addition of denaturants to the refolding medium (*21*). One of the particularly important features of the present approach is that it allows specific kinetic schemes to be tested, and potentially distinguished, by explicit simulation. Moreover, with a knowledge of the spectra of the different components, the kinetic profiles can in principle be determined directly by deconvolution of the kinetic spectrum in terms of these components. This technique therefore provides a powerful means of kinetic analysis for a folding reaction, or indeed for chemical reactions in general.

Table 1. Time constants of refolding for different elements of secondary structure in BLA. The average time constants and the standard deviations are calculated over the number of residues studied in this work.

Structure element	Time constant (s)	No. of residues
A-B Loop	400	2
B-Helix	405 ± 174	4
β-Sheet	300 ± 71	3
B-C Loop	475 ± 115	4
C-Helix	386 ± 103	7
C-D Loop	400	1
C-Terminal loop	431 ± 41	4

REFERENCES AND NOTES

1. C. R. Matthews, *Annu. Rev. Biochem.* **62**, 653 (1993); K. A. Dill and A. R. Fersht, Eds., *Curr. Opin. Struct. Biol.* **6**, 1 (1996).
2. P. A. Evans and S. E. Radford, *Curr. Opin. Struct. Biol.* **4**, 100 (1994); A. D. Miranker and C. M. Dobson, *ibid.* **6**, 31 (1996); M. Engelhard and P. A. Evans, *Folding Des.* **1**, R31 (1996).
3. S. W. Englander and L. Mayne, *Annu. Rev. Biophys. Biomol. Struct.* **21**, 243 (1992); R. L. Baldwin, *Curr. Opin. Struct. Biol.* **3**, 84 (1993).
4. P. S. Kim and R. L. Baldwin, *Annu. Rev. Biochem.* **59**, 63 (1990); C. M. Dobson, *Curr. Opin. Struct. Biol.* **2**, 6 (1992); A. L. Fink, *Annu. Rev. Biophys. Biomol. Struct.* **24**, 495 (1995); O. B. Ptitsyn, *Curr. Opin. Struct. Biol.* **5**, 74 (1995).
5. C. Frieden, S. D. Hoeltzli, I. J. Ropson, *Protein Sci.* **2**, 2007 (1993).
6. J. Balbach *et al.*, *Nature Struct. Biol.* **2**, 865 (1995).
7. S. D. Hoeltzli and C. Frieden, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 9318 (1995).
8. Repetitive accumulation of fast 2D experiments has been used to study the slow association of a trimeric peptide [X. Liu, D. L. Siegel, P. Fan, B. Brodsky, J. Baum, *Biochemistry* **35**, 4306 (1996)].
9. R. R. Ernst, G. Bodenhausen, A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions* (Oxford Univ. Press, Oxford, 1987), chap. 9; A. Kumar, G. Wagner, R. R. Ernst, K. Wüthrich, *J. Am. Chem. Soc.* **103**, 3654 (1981); H. B. Olsen, H. Gesmar, J. J. Led, *ibid.* **115**, 1456 (1993).
10. C. M. Dobson and P. A. Evans, *Biochemistry* **23**, 4267 (1984); G. S. Huang and T. G. Oas, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6878 (1995).
11. The ^{1}H-^{15}N HSQC experiment [G. Bodenhausen and D. Ruben, *Chem. Phys. Lett.* **69**, 185 (1980)] involves detection of ^{15}N resonances indirectly by means of their correlation through scalar coupling with ^{1}H spins. The ^{1}H magnetization is detected during an acquisition time t_2 as a conventional free induction decay (FID), allowing ^{1}H chemical shifts to be resolved after Fourier transformation. A series of ^{1}H FIDs is collected after incrementation of a time t_1 in which the ^{15}N magnetization is allowed to evolve. A second Fourier transformation converts the resulting t_1 modulation into the frequency domain.
12. K. Kuwajima, *Proteins Struct. Funct. Genet.* **6**, 87 (1989); J. Baum, C. M. Dobson, P. A. Evans, C. Hanley, *Biochemistry* **28**, 7 (1989); A. T. Alexandrescu *et al.*, *ibid.* **32**, 5681 (1993); S. Sugai and M. Ikeguchi, *Adv. Biophys.* **30**, 37 (1994); L. C. Wu, Z. Peng, P. S. Kim, *Nature Struct. Biol.* **2**, 281 (1995); K. Kuwajima, *FASEB J.* **10**, 102 (1996).
13. K. Kuwajima, Y. Hiraoka, M. Ikeguchi, S. Sugai, *Biochemistry* **24**, 894 (1985); M. Ikeguchi, K. Kuwajima, M. Mitani, S. Sugai, *ibid.* **25**, 6965 (1986).
14. M. Arai and K. Kuwajima, *Folding Des.* **1**, 275 (1996); J. Balbach, V. Forge, W. S. Lau, K. Brew, C. M. Dobson, unpublished data.
15. The time to acquire a FID in the indirect (t_1) dimension of a 2D experiment depends on the number of scans used to record each FID in the direct (t_2) dimension. It can be adjusted to be, in general, optimal for the time course of the particular system under study, subject to the limits imposed by the sensitivity and resolution requirements of the system. In addition, intermediates will only be detectable if their resonances are resolved and distinct from those of the other states and if the rate constants for their formation and disappearance result in a significant population of a given species.
16. Provided that the different components within such a kinetic scheme react sufficiently slowly relative to the acquisition time of the direct ^{1}H dimension (t_2), and that the number of molecules that interconvert in this time is small compared to the total number, they can be treated as independent species (*22*); because the acquisition time in the present experiment is 213 ms, this independence clearly applies to the denatured and native states considered here.
17. The total intrinsic fluorescence intensity between 340 and 360 nm was measured with a Hitachi F4010 luminescence spectrometer after manual mixing. The excitation wavelength was 278 nm.
18. T. Kiefhaber, A. M. Labhardt, R. L. Baldwin, *Nature* **375**, 513 (1995).
19. A. Bax, *Curr. Opin. Struct. Biol.* **4**, 738 (1994).
20. M. Mücke and F. X. Schmid, *Biochemistry* **33**, 14608 (1994); L. Serrano, *Curr. Opin. Struct. Biol.* **4**, 107 (1994).
21. B.-I. Chen, W. A. Baase, J. A. Schellman, *Biochemistry* **28**, 691 (1989).
22. R. O. Kühne, T. Schaffhauser, A. Wokaun, R. R. Ernst, *J. Magn. Reson.* **35**, 39 (1979).
23. The recombinant protein used here differs from wild-type BLA by the presence of an initiator Met on the N terminus and by having Val substituted for Met^{90} [J. A. Grobler, M. Wang, A. C. W. Pike, K. Brew, *J. Biol. Chem.* **269**, 5106 (1994)].
24. L. E. Kay, P. Kiefer, T. Saarinen, *J. Am. Chem. Soc.* **114**, 10663 (1992); J. Schleucher *et al.*, *J. Biomol. NMR* **4**, 301 (1994).
25. D. J. States, R. A. Haberkorn, D. J. Ruben, *J. Magn. Reson.* **48**, 286 (1982).
26. We initiated the refolding of BLA (*23*) by changing the pH to 7.0 by injecting 50 μl of a solution of 333 mM tris-HCl at pH 8.8 and 20 mM EDTA (to sequester Ca^{2+} ions) into the NMR tube containing 450 μl of the 1.4 mM protein solution at pH 2 in a 93:7 mixture of H_2O and D_2O. The first scan for the HSQC during the folding event was started 5 s after the injection to ensure that mixing was complete. All three data sets were recorded with the gradient-enhanced method (*24*), using identical acquisition parameters and the States method for phase sensitivity in F1 (*25*). A minimal phase cycling of 2 was used to keep contributions of changes of the interconverting species in corresponding scans below the signal-to-noise level. We used 512 and 48 complex points of the recorded spectrum in F2 and F1, respectively, and spectral widths were 9615 Hz (F2) and 2404 Hz (F1). For all spectra, 90°- (F2) and 60°- (F1) shifted squared sine-bell window functions were applied. Spectra presented are averages over three experiments. All NMR spectra were recorded with a home-built NMR spectrometer operating at 600.2 MHz and were processed with the use of FELIX (Biosym Technologies, San Diego, CA).
27. G. Bodenhausen and R. R. Ernst, *J. Am. Chem. Soc.* **104**, 1304 (1982).

28. We thank S. L. Winder and P. J. Hore for help in developing the rapid mixing device. J.B., V.F., and N.A.J.v.N. were supported by the Human Capital and Mobility Programme of the European Community. K.B. is supported by a grant from NIH. The Oxford Centre for Molecular Sciences is supported by the Engineering and Physical Sciences Research Council, Biotechnology and Biological Sciences Research Council, and Medical Research Council (UK). C.M.D. is supported in part by an International Research Scholars award from the Howard Hughes Medical Institute.

17 June 1996; accepted 13 September 1996