Time Resolved Collapse of a Folding Protein Observed with Small Angle X-Ray Scattering

L. Pollack,[1,*] M. W. Tate,[1] A. C. Finnefrock,[1] C. Kalidas,[2] S. Trotter,[2] N. C. Darnton,[3] L. Lurio,[4]
R. H. Austin,[3] C. A. Batt,[2] S. M. Gruner,[1] and S. G. J. Mochrie[4]

[1]*Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853*
[2]*Nanobiotechnology Center, Cornell University, Ithaca, New York 14853*
[3]*Department of Physics, Princeton University, Princeton, New Jersey 08544*
[4]*Department of Physics, MIT, Cambridge, Massachusetts 02139*
(Received 10 September 2000)

High-intensity, "pink" beam from an undulator was used in conjunction with microfabricated rapid-fluid mixing devices to monitor the early events in protein folding with time resolved small angle x-ray scattering. This Letter describes recent work on the protein bovine β-lactoglobulin where collapse from an expanded to a compact set of states was directly observed on the millisecond time scale. The role of chain collapse, one of the initial stages of protein folding, is not currently understood. The characterization of transient, compact states is vital in assessing the validity of theories and models of the folding process.

DOI: 10.1103/PhysRevLett.86.4962 PACS numbers: 87.14.Ee, 87.15.He

The question of how a protein changes shape ("folds") from a quasilinear state to assume its native, folded form is of both fundamental and practical interest [1]. Ideally, a fully unfolded peptide chain (denatured protein) forms an expanded random coil. In contrast, a folded protein is compact, with a density approaching that of an amino acid crystal and with well-defined internal (secondary and tertiary) structures. Chain collapse and secondary structure formation are included in the critical initial steps of protein folding that convert the former states to the latter. In some cases, a non-native compact state (known as compact denatured/molten globule/intermediate [2]) is rapidly formed. Folding to the native state occurs on a longer time scale.

The relevance of rapidly formed, compact denatured states, specifically an understanding of their role in protein folding, is a topic of current controversy and interest [3–5]. Chain collapse can occur rapidly, typically much faster than the millisecond dead times of conventional mixing apparatuses, requiring experimental studies of folding with submillisecond time resolution to probe these states. Access to this regime has been achieved only very recently, with the advent of new technologies [6]. Measurements of fast rate constants as a function of temperature [6,7] or solvent conditions [4,5] can be used to determine the thermodynamic nature of collapse for example, to differentiate between the limiting cases of a cooperative or continuous transition. For both cytochrome-*c* [2,4] and the B1 domain of protein G [5] detailed analyses of folding rate constants for the initial, rapid collapse present clear evidence for a free energy barrier separating expanded from compact states. These compact states are therefore "productive intermediates," relevant to folding [5].

These recent experiments highlight the need to acquire experimental data that will test theories/simulations of heteropolymer collapse, including lattice models of protein folding. For detailed comparison, experiments must be able to probe the structure of transiently occurring states as well as their kinetics. Small angle x-ray scattering (SAXS) can be used to assess both the size and compactness of a protein in solution [8]. Other experimental techniques, used previously, measure these parameters less directly. Several groups have been using SAXS to study protein folding [9–13]. Time resolved SAXS (TRSAXS), with millisecond or faster resolution can provide the requisite information about transiently occurring states, including partially folded states [14]. This paper reports significant advances in TRSAXS experiments using microfabricated fluid mixers in conjunction with small, brilliant x-ray beams available at the advanced photon source (APS).

We have studied the initial stages of the folding of β-lactoglobulin (BLG), a 162 amino acid, β-barrel protein found in bovine milk. From previous SAXS experiments of refolding from an expanded, urea-denatured state, it is known that a compact state forms within the 30 ms dead time of the stopped-flow mixers employed [13]. Very recently, rapid mixing experiments using both Trp fluorescence and hydrogen exchange labeling indicate the presence of many kinetic phases in BLG folding [15]. In these experiments, signatures that are consistent with collapse are observed within 2 ms of the initiation of folding.

Microfabricated fluid mixers, previously used to trigger protein folding by submillisecond scale pH jumps [9], have been modified for use with chemical denaturants. In these experiments BLG folding is triggered by a rapid decrease in urea concentration. A schematic of the mixer is shown in Fig. 1.

In a separate set of experiments, the width of the focused stream was determined by multiphoton imaging of fluorescein dye flowing in the inlet channel, focused by water flowing in the side channels. Urea concentration as a function of position was obtained by scaling the diffusion coefficient of fluorescein to that of urea [18]. An average urea concentration was determined at each location along the outlet channel due to the nonhomogeneous flow profile as a function of depth. The urea concentration drops from 6 M to below 3 M in just under 2 ms at the center of the stream.

 0031-9007/01/86(21)/4962(4)$15.00

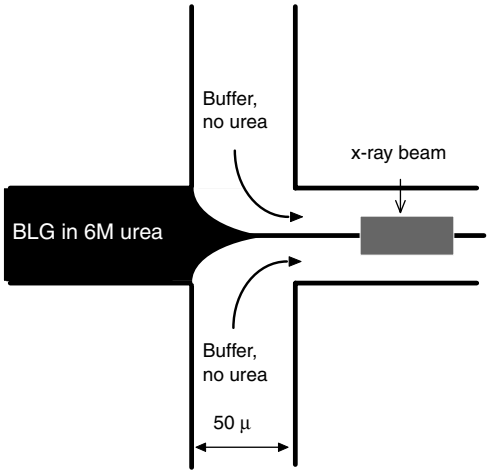

FIG. 1. A schematic of the rapid fluid mixer. To run the experiment, 1 mM unfolded protein in dilute phosphoric acid and 6 M urea (pH 2.8) flows in the inlet channel. Dilute phosphoric acid (pH 2.8) with no urea flows in the side channels, focusing the protein into a stream that is, on average, 5 μm wide. Small molecules such as urea diffuse rapidly out of the central stream, triggering the folding reaction. The pH was maintained below 3 at all times to prevent dimerization of the BLG (Refs. [13,16]). [The BLG used for these experiments was expressed and purified from the yeast Pichia pastoris (Ref. [17]).] The sample temperature was maintained at 27 °C throughout the experiment. The device was fabricated as described in Ref. [9]. For this work, the channel width has been reduced by a factor of 2, to 50 μm, while the channel depth remained 390 μm. The higher aspect ratio results in a more uniform stream as a function of distance from the top and bottom sealed surfaces which in turn results in more uniform diffusion times. One possible position of the x-ray beam is shown.

Flow speeds were also determined at all positions within the cell using fluorescence correlation spectroscopy [19].

The SAXS experiments were performed at IMM-CAT (APS) where pink beam (full width at half maximum = 2.6%) was produced by reflecting the first harmonic of APS undulator A (7.65 keV at a gap of 18 mm) off of a Si mirror [20]. The beam was collimated with a pair of crossed slits located 14 cm upstream of the sample and was set to 10 μm in the vertical direction, to illuminate only the protein jet, and 40 μm in the horizontal direction, to achieve the requisite time resolution of 240 μs per data point (at our average flow speed of 16.6 cm/s). The flux through the slits was 2.5×10^{11} x rays/sec. Exposure times were 40 s; typically two or more images were averaged. Pink beam can be used only with a flowing sample because of the potential for radiation damage.

Scattered x rays were detected using a custom-built CCD detector [21] located 40 cm downstream of the sample. At this distance, although the detector allows access to $0.04 < q < 0.5$ Å^{-1}, the interesting data fall below $q = 0.2$ Å^{-1}. A rapid decrease in signal strength at larger q results in an increase in noise ($q = 4\pi \sin\theta/\lambda$ where 2θ is the scattering angle and $\lambda = 1.62$ Å is the x-ray wavelength at the peak of the undulator spectrum). We followed a rigorous procedure for acquiring data, taking a background image with only buffer flowing before and after each set of protein images. We accepted the data point only when the two background images were superimposable. At low q the spatial average of the scattered intensity follows a simple form [8]: $\langle I(q)\rangle = I_o \exp(-R_g^2 q^2/3)$. In this Guinier approximation, a straight line is fit through a plot of $\log_e(I)$ vs q^2. The slope of this curve provides a measure of the size of the protein, expressed as its radius of gyration, R_g; its y-intercept, I_o, yields information about the aggregation state of the protein. In some cases refolding proteins form aggregates that can be mistaken for intermediates [22]. If dimerization occurs the measured value of I_o will increase by a factor of 2 [8].

Kratky plots (Iq^2 vs q) can be used to show the power law dependence of the high q scatter. A peak in this plot indicates that the scatter falls off more rapidly than q^{-2}, typical of a compact, uniform object. Expanded objects have scattering that falls off as q^{-1} [8], hence the product Iq^2 continues to increase over a broad range of q. Thus the appearance of a peak in the Kratky plot is useful for assessing collapse of the molecule and contains more information than the R_g, a single number obtained by examination of only the small fraction of data in the Guinier regime [8]. The main result of this work, Kratky plots of BLG flowing inside the device, is shown in Fig. 2. In this series of graphs the product Iq^2 is plotted as a function of q for six different positions inside the device. The top frame shows protein in the inlet channel in the initial state (6 M urea, pH 2.8). The increase at high q indicates that this denatured state is expanded. The remaining frames show data taken at different locations along the outlet channel: (b) 0.8 ms before the urea concentration drops below 3 M (nearly identical to top trace, but with less S/N), (c) 1.3 ms after [urea] < 3 M, (d) 3.4 ms after, (e) 5.9 ms after, and (f) 8.0 ms after. A peak develops between frames (c) and (d) indicating some compaction on a time scale of scale of 2 ms. By scaling the data of (b) to that of (e) and (f) in a range of q where we expect most of the contribution to come from the expanded state, we estimate that by 6 ms, approximately 20% of the protein has become compact, and by 8 ms, 30% of the protein has become compact. The high q data taken at a second, higher protein concentration (2.5 mM, not shown) reproduces this trend. The lowest frame (g) shows data from a much larger (static) native sample of protein (data not acquired inside the mixer). In a control experiment, where the solution surrounding the protein was the same as that flowing in the side channels, no change in the state of the protein was detected.

Figure 3 shows the results of a Guinier analysis applied to the low q region of the same data. Typically, the Guinier approximation is valid for $R_g q < 1.3$ [23], as a result of limited access to data at very low q, we extended our fitting range to $q = 0.06$ Å^{-1}. The left frame shows a typical Guinier fit; the right frame shows R_g as a function of time. Within our experimental uncertainty, there is no

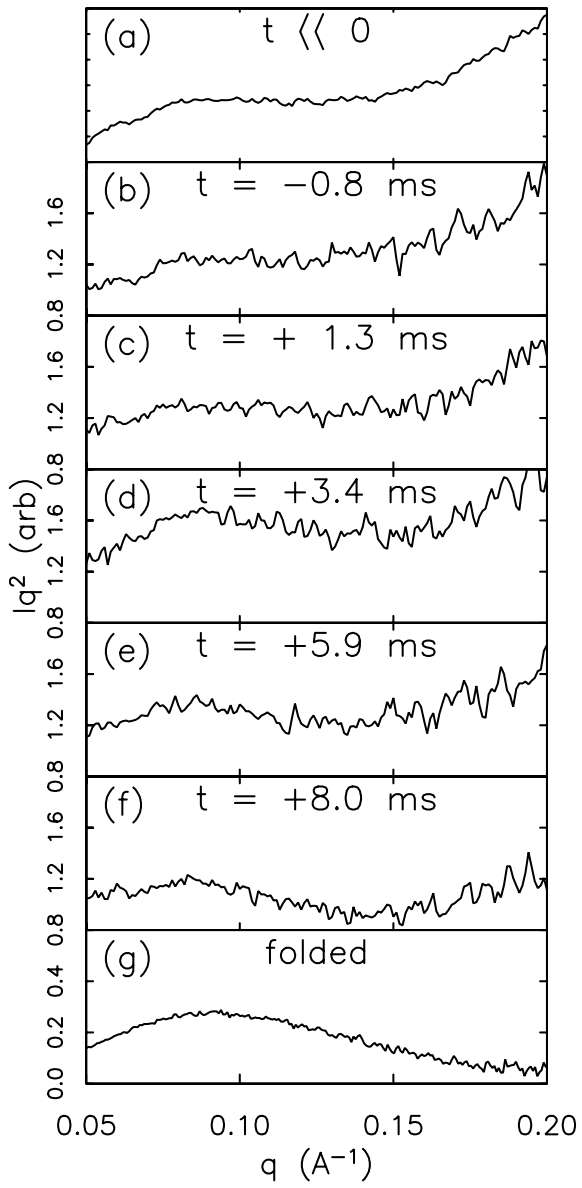

FIG. 2. Kratky plots showing collapse of the protein as a function of time, where $t = 0$ is defined to be the point at which the urea concentration is 3 M (native conditions). Collapse occurs between frames (c) and (d).

change in I_o as a function of time, indicating that the aggregation state of the protein does not change appreciably through the course of this measurement. Initially, at the high concentration (6 M) of chemical denaturant, we expect the protein to be monomeric. We therefore assume the persistence of this state throughout our measurement. We note that a small (less than 10%) decrease in I_o has been observed in static measurements on BLG between the urea unfolded and native states, reflecting changes in the hydration shell [10].

The R_g data of Fig. 3b are interpreted as follows. The initial protein concentration is 1 mM (18 mg/ml). At protein concentrations above 0.5 mM, static measurements reflect interparticle interference which artificially depresses

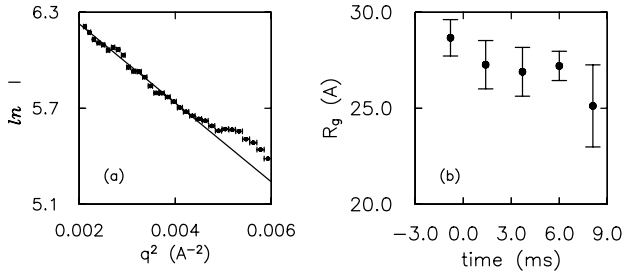

FIG. 3. The results of the Guinier analysis for the data of Fig. 2 [frames (b) through (f)] are shown. (a) shows raw data with a typical Guinier fit used to determine R_g. The size of the point indicates the error in the y direction. These data correspond to $t = 6$ ms. (b) shows the radius of gyration as a function of time. The small decrease is consistent with compaction.

the measured value of R_g relative to the value at infinite dilution. However, in our flow cell the protein concentration drops as a function of time (distance along channel) due to diffusion. The final protein concentration is more than a factor of 3 lower than the initial, below the concentration where strong interparticle interference effects have been observed. Therefore, since the effect of decreasing protein concentration should result in an apparent increase in R_g, the decrease that we observe is indicative of compaction. The trend from Fig. 3b suggests a slight decrease in R_g, and supports the main result of this work, the Kratky plots shown in Fig. 2. An R_g value of 16.4 Å is expected for monomeric, native BLG [24].

Experiments such as the one described in this Letter have the potential to distinguish between two very different folding mechanisms. In the first, a protein passes through a series of cooperative (all or nothing) transitions on its way to assuming its folded conformation. In the second, the protein passes through a series of partially folded states. One model of these states, discussed in Ref. [14], is of a compact core surrounded by random coil loops. The SAXS patterns corresponding to these two cases are distinguishable, but require a consistency of experimental conditions just beyond what we have achieved in this experiment. As a result of the small size of the beam, shifts of even a few microns in beam position can have a marked effect on the signal. The key to distinguishing between these cases lies in obtaining good statistics for the mid-q to high-q data.

In interpreting this experiment we make the assumption that the protein folds to completion under our experimental conditions. If we take the view that the first step in BLG folding is a cooperative transition between expanded and compact states, our conclusion is that collapse in this system is very slow: more than 70% of the protein remains in an expanded (denatured) state even 8 ms after conditions are favorable for refolding. Previous work [13,15] shows that subsequent folding to the native state requires seconds. These time scales are long: approximately 3 orders of magnitude longer than that measured for cytochrome-c. However, there are a number of explanations for the longer

times involved in BLG folding. The first possibility is that proteins that are entirely β sheet fold more slowly than those with all-α structure, due to nonlocal interactions in β-sheet formation [25]. A second possibility is the presence of five cysteine residues that form two disulfide bonds in the folded proteins. One of the disulfide bonds is known to form early in the folding process [26]. Models of folding predict that under some circumstances the presence of disulfide bonds may slow folding kinetics [27]. Third, the presence of numerous proline residues in BLG may be expected to increase the overall folding time [28]. The size of the protein does not appear to be a factor for small proteins [29].

Under the partially folded states model, the data of Fig. 2 can be interpreted as representing the scatter from a state that consists of a series of compact and looplike regions. In fact, the R_g of such a state might be large, consistent with the data of Fig. 3b. An intriguing possibility is that multiple paths to the native state exist and only a small fraction of the macromolecules reach a collapsed state on this time scale. With planned improvements to both the mixers and beam line, our immediate goal is to establish and maintain the constancy of conditions required to distinguish between these two scenarios.

Despite the numerous theoretical models addressing protein/heteropolymer folding, there is still a need for experimental data to test these predictions. TRSAXS experiments of the type described in this Letter provide a means to quantitatively test these models. Ultimately, no one theory is likely to explain the rich diversity of protein folding behavior. Rather, experiments and theory may discover a range of universality classes, distinguished by different mechanisms and folding parameters. TRSAXS provides this much-needed characterization of transiently occurring states. With these experiments we clearly demonstrate that we can time-resolve some of the initial stages of protein folding. The addition of information from other techniques on comparable time scales [15] will allow us to refine this picture, and is essential to unraveling the earliest events in protein folding.

We acknowledge useful discussions with W. Eaton, and the valuable assistance of J. Korlach, R. Williams, W. Zipfel, M. Novak, H. Gibson, and G. Toombes. This work was supported by the NSF under Grant No. DMR-9701453, through the STC program of the NSF under agreement No. ECS-986771 and by the DOE under No. DE-FG02-97ER62443. This work was performed at IMM-CAT using the beam line ID-8. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. W-31-109-Eng-38. These experiments also made use of the Cornell Nanofabrication Facility, supported by the NSF, Cornell University and industrial affiliates; and the Developmental Resource for Biophysical Imaging Opto-Electronics at Cornell, an NIH-NCRR facility.

*Permanent address: School of Applied and Engineering Physics, Cornell University, Ithaca, NY 14853.

[1] A. Fersht, *Structure and Mechanism in Protein Science* (W. H. Freeman and Co., New York, 1999).
[2] S. J. Hagen and W. A. Eaton, J. Mol. Biol. **00**, 1 (2000).
[3] K. A. Dill *et al.,* Protein Sci. **4**, 561 (1995); S. W. Englander, Annu. Rev. Biophys. Biomol. Struct. **29**, 213 (2000); A. R. Fersht, Proc. Natl. Acad. Sci. U.S.A. **92**, 10 869 (1995); T. R. Sosnick *et al.,* Proc. Natl. Acad. Sci. U.S.A. **94**, 8545 (1997).
[4] M. C. R. Shastry and H. Roder, Nature Struct. Biol. **5**, 385 (1998).
[5] S-H. Park, M. C. R. Shastry, and H. Roder, Nature Struct. Biol. **6**, 943 (1999).
[6] W. A. Eaton *et al.,* Annu. Rev. Biophys. Biomol. Struct. **29**, 327 (2000); H. Roder and M. C. R. Shastry, Curr. Opin. Struct. Biol. **9**, 620 (1999).
[7] J. Sabelko, J. Irvin, and M. Gruebele, Proc. Natl. Acad. Sci. U.S.A. **96**, 6031 (1999).
[8] O. Glatter and O. Kratky, *Small Angle X-Ray Scattering* (Academic, London, 1982).
[9] L. Pollack *et al.,* Proc. Natl. Acad. Sci. U.S.A. **96**, 10 115 (1999).
[10] G. V. Semisotnov *et al.,* J. Mol. Biol. **262**, 559 (1996).
[11] K. W. Plaxco *et al.,* Nature Struct. Biol. **6**, 554 (1999).
[12] L. Chen *et al.,* J. Mol. Biol. **276**, 225 (1998); D. Eliezer *et al.,* Science 270, 487 (1995).
[13] M. Arai *et al.,* J. Mol. Biol. **275**, 149 (1998).
[14] S. Doniach *et al.,* J. Mol. Biol. **254**, 960 (1995).
[15] K. Kuwata *et al.,* Nature Struct. Biol. **8**, 151 (2001).
[16] R. Townend and S. N. Timasheff, J. Am. Chem. Soc. **79**, 3613 (1957).
[17] T. R. Kim *et al.,* Protein Eng. **10**, 1339 (1997).
[18] W. Zipfel (unpublished); L. J. Gosting and D. F. Akeley, J. Am. Chem. Soc. **74**, 2058 (1952).
[19] L. Pollack *et al.* (to be published).
[20] A. R. Sandy *et al.,* J. Synchrotron Radiat. **6**, 1174 (1999).
[21] M. W. Tate *et al.,* J. Appl. Crystallogr. **28**, 196 (1995).
[22] M. Silow and M. Oliveberg, Proc. Natl. Acad. Sci. U.S.A. **94**, 6084 (1997).
[23] M. Kataoka and Y. Goto, Fold. Des. **1**, R107 (1996).
[24] G. Baldini *et al.,* Macromolecules **32**, 6128 (1999).
[25] S. E. Jackson, Fold. Des. **3**, R81 (1998).
[26] B. Forge *et al.* (to be published).
[27] C. J. Camacho and D. Thirumalai, Proteins Struct. Funct. Genet. **22**, 27 (1995).
[28] P. S. Kim and R. L. Baldwin, Annu. Rev. Biochem. **51**, 459 (1982).
[29] K. W. Plaxco, K. T. Simons, and D. Baker, J. Mol. Biol. **277**, 985 (1998).