

letters

Chain collapse can occur concomitantly with the rate-limiting step in protein folding

Kevin W. Plaxco^{1,2}, Ian S. Millett³, Daniel J. Segel³, Sebastian Doniach³ and David Baker¹

¹Department of Biochemistry, University of Washington, Seattle, Washington 98195, USA. ²Present address: Department of Chemistry and Interdepartmental Program in Biochemistry and Molecular Biology, University of California, Santa Barbara, Santa Barbara, California 93106, USA. ³Department of Applied Physics, Stanford University, Stanford, California 92343, USA.

We have directly characterized the extent of chain collapse early in the folding of protein L using time-resolved small angle X-ray scattering. We find that, immediately after the initiation of refolding, the protein exhibits dimensions indistinguishable from those observed under highly denaturing, equilibrium conditions and that this expanded initial state collapses with the same rate as that of the overall folding reaction. The observation that chain compaction need not significantly precede the rate-limiting step of folding demonstrates that rapid chain collapse is not an obligatory feature of protein folding reactions.

While the limited ability of water to solvate denatured states is widely thought to bring about the rapid, hydrophobically driven collapse of unfolded proteins, there has been much debate regarding the roles such collapsed species might play in the folding process^{1–10}. Here we report the results of time-resolved small angle X-ray scattering (SAXS) experiments^{11–13} aimed at directly monitoring the dimensions of the 62-residue IgG binding domain of protein L (protein L) during its refolding. Our studies indicate that, immediately after the initiation of refolding, protein L exhibits a radius of gyration (R_g) indistinguishable from that observed under strongly denaturing, equilibrium conditions. From this highly extended, initial species, the native state is formed by an apparently two-state process. These data provide direct evidence that collapsed species need not be significantly populated prior to the rate-limiting step in folding.

The equilibrium unfolding of protein L in guanidine hydrochloride (GuHCl) solutions is well described as a two-state process¹⁴. The results of static SAXS experiments (Fig. 1) are fully consistent with this model, producing values of ΔG_u (5.5 ± 0.8 kcal mol⁻¹) and m_{eq} (2.0 ± 0.3 kcal mol⁻¹ M⁻¹) within experimental error of previously reported values derived from far-UV circular dichroism measurements¹⁴. Static SAXS measurements indicate that the dimensions of native protein L are insensitive to GuHCl concentration and provide an estimate of 16.2 ± 0.2 Å for the native state R_g . The absorption of X-rays by the denaturant precludes direct measurements of R_g at 6 M GuHCl. Measurements collected at denaturant concentrations up to 5 M indicate, however, that to within experimental error the R_g of unfolded protein L remains a constant 26.0 ± 0.3 Å at 4 M GuHCl and above. The R_g observed for native protein L is consistent with the crystal structure (15.7 Å for a molecule lacking an extended, eight-residue histidine-tag), and that observed for the denatured state is consistent

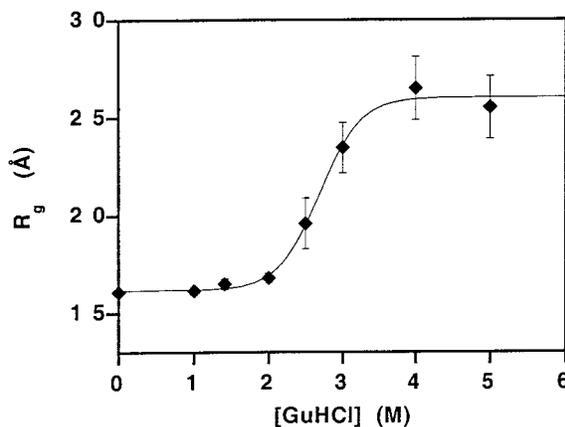


Fig. 1 The equilibrium unfolding of protein L is accurately described as a two-state process in which only the denatured and native states are populated. Equilibrium R_g values are well fitted ($r^2 = 0.995$) by a two-state, linear free energy model: $\Delta G = -\Delta G_u + m[\text{GuHCl}]$, where $\Delta G_u = 5.5 \pm 0.8$ kcal mol⁻¹ and $m = 2.0 \pm 0.3$ kcal mol⁻¹ M⁻¹. The two-state model (equation 1) provides estimates of the R_g of native (16.2 ± 0.2 Å) and fully denatured (26.0 ± 0.3 Å) states of protein L and provides no indication that the dimensions of the denatured state are altered at higher denaturant concentrations.

with those reported for other small, single-domain proteins^{15,16}.

Time-resolved SAXS experiments (Fig. 2) demonstrate that unfolded protein L does not undergo a rapid collapse upon transfer to conditions under which the native state is favored (1.4 M GuHCl, $\Delta G_u \sim 3$ kcal mol⁻¹). The R_g measured immedi-

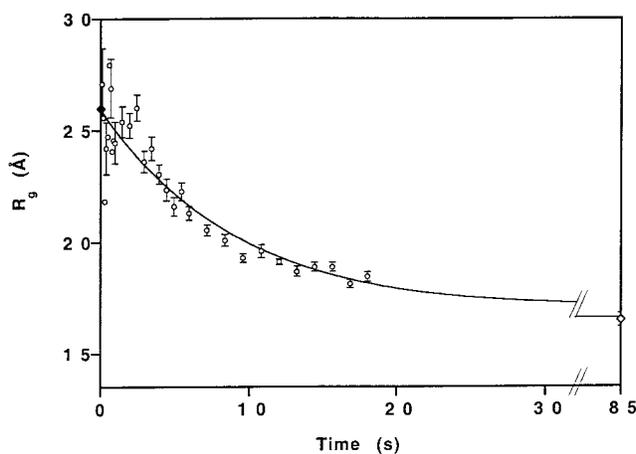


Fig. 2 Time-resolved SAXS indicates that the refolding of protein L lacks a burst-phase collapse event under the conditions employed. Within 100 ms of the initiation of refolding, the observed R_g (27.1 ± 1.6 Å) is indistinguishable from that of the equilibrium denatured state (filled diamond). Refolding from this initial species is well fitted (equation 2; $r^2 = 0.90$; mean amplitude and serial correlation of the residual, 3% and $r^2 = 0.06$, respectively) as a single-exponential process with a time constant, 8 ± 3 s, similar to the previously reported time constant for the recovery of native fluorescence (~ 4 s for a construct lacking the histidine-tag¹⁶). This two-state model extrapolates to within experimental error of the equilibrium R_g of protein L, predicting an initial R_g (25.9 ± 0.4 Å) again indistinguishable from that observed under highly denaturing conditions. This initial, fully unfolded species folds to apparently native protein L (fitted: 16.6 ± 1.9 Å; observed 16.5 ± 0.3 Å at 1.41 M GuHCl), and protein of native R_g (16.7 ± 0.4 Å) is recovered within the 85 s dead time of manual-mix refolding experiments (open diamond; note that this data point was omitted from the fit) For clarity, error bars are provided on only every third data point before 0.6 s.

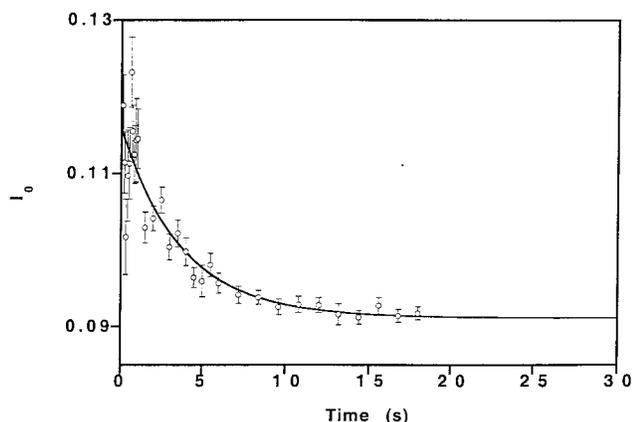


Fig. 3 A decrease in forward scattering (I_0) is observed during folding. This decrease, which has a decay time ($3.8 \pm 1.0 \text{ s}^{-1}$) similar to the rate of formation of species of native R_g (Fig. 2), probably reflects a decrease in the size of the hydration shell as the protein folds¹³. There is no indication of any significant aggregation during refolding (see text).

ately after the initiation of refolding (data bin 0–100 ms), $27.1 \pm 1.6 \text{ \AA}$, is indistinguishable from that observed under highly denaturing, equilibrium conditions. A fit of a simple, single exponential model to the time-resolved SAXS data predicts an unfolded R_g of $25.9 \pm 0.4 \text{ \AA}$, providing further evidence of the expanded character of the initial species. Data collection was terminated after 18 s in order to preclude radiation-induced protein damage. Under the conditions employed, however, protein of native R_g is recovered within the 85 s dead time of manual-mixing experiments (Fig. 2). Consistent with this, the two-state fit predicts that protein of native R_g ($16.6 \pm 1.9 \text{ \AA}$) is recovered from the initially extended material with a time constant similar to previously reported values derived using time-resolved fluorescence measurements¹⁷ (Fig. 2). These results demonstrate that, be they authentic folding intermediates or merely reflective of denatured state compaction under refolding conditions, collapsed species need not be significantly populated during the folding of protein L.

The large initial R_g does not arise due to the transient formation of oligomers producing an R_g coincidentally matching that of the monomeric, unfolded protein at equilibrium. A decrease in forward scattering (I_0 , X-ray scattering intensity extrapolated to the forward direction) is observed that occurs with the same rate as the overall folding process (Fig. 3). While such a change in I_0 is consistent with the formation of a small amount of multimer¹¹, it most likely arises from the increased hydration shell associated with an unfolded protein; for example, a 10% change in I_0 was observed at early time points in the refolding of an already largely collapsed lysozyme folding intermediate¹³. If we neglect for the moment this potentially significant hydration effect, the observed change in I_0 would reflect a dimer concentration of at most 11%. This dimer concentration would lead to an overestimate of the dimensions of the monomeric, unfolded species, with the true $R_g \geq 85\%$ of the observed R_g . Since hydration shell changes must account for much or all of the observed change in I_0 , the actual monomer R_g would be well above this lower limit. Thus even if limited, transient oligomerization were occurring, rapid compaction of the monomeric protein can be definitively discounted.

The apparent GuHCl concentration independence of the R_g of protein L under strongly denaturing conditions suggests that at or above 4 M GuHCl the denaturant is a sufficiently 'good' solvent that any increases in R_g at higher denaturant concentrations are within the experimental error limits of SAXS^{18–20}. Similar results have been reported for a mutant version of protein G (ref. 15), which, while lacking obvious homology, is similar in size and topology to protein L. It seems likely, however, that the conditions employed in our refolding experiments (1.4 M GuHCl; $\Delta G_u \sim 3 \text{ kcal mol}^{-1}$) represent a 'poor' solvent and thus one might predict significant chain compaction^{19,20}. The failure of the chain to collapse indicates that the free energy increases (rather than decreasing) as R_g is reduced below 26 \AA ; presumably the entropic cost of ordering the chain outweighs the hydrophobic interactions favoring compaction. Surmounting this free energy barrier may be an important component of the rate-limiting step in protein folding (Fig. 4).

In contrast to other methods²¹, time-resolved SAXS allow us to monitor directly the evolution of compact species during protein folding. Using this technique, we have demonstrated that under at least some conditions protein L does not fold by searching primarily through the set of collapsed conformations; rather, the vast majority of the folding time is spent in conformations as expanded as the denatured state. Folding presumably occurs when rare, short-lived fluctuations lead from this expanded state to form productive, high-

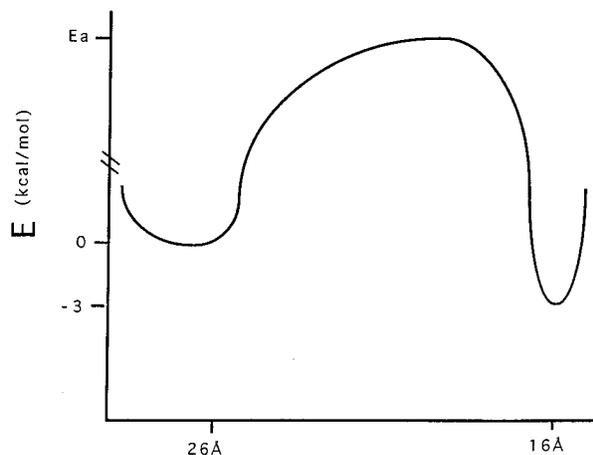


Fig. 4 A model of the folding barrier of protein L can be constrained using a number of recent experimental results. The concomitance of collapse and the rate-limiting step indicates that no stable, partially folded species arise before the kinetic barrier. The effects of co-solvent on the protein's folding kinetics suggest that the rate-limiting step occurs in an ensemble that sequesters approximately 3/4 of the solvent-excluded surface area of the native state^{14,24}; assuming spherical species, this would place the barrier at an R_g of $\sim 19 \text{ \AA}$. Studies of the sequence dependence of folding kinetics indicate that the rate-limiting step involves structure formation at one of the protein's two β -hairpins²⁵, and studies of the viscosity dependence of folding suggest the transition state region is fairly broad and that the rate-limiting step involves motions displacing a considerable amount of solvent²⁴. An enthalpic barrier of 10 kcal mol^{-1} is predicted from fits of the temperature dependence of folding rates to a modified Arrhenius model¹⁶, although -4 kcal mol^{-1} of this apparent barrier would arise due to the temperature dependence of the viscosity of water (W.A. Eaton, pers. comm.). Because of uncertainties regarding the magnitude of the pre-exponential factor, the entropic contribution to the free energy barrier, and thus the true barrier height, cannot be ascertained. Experimental studies of the relationship between native-state topology and folding rates²⁶, however, suggest that chain entropy also plays a significant role.

letters

energy collapsed species. Irrespective of the exact details of the folding process, the direct observation of the concomitance of collapse and folding for protein L under any conditions rules out those theories of folding that postulate an obligatory role for well-populated collapsed species.

Methods

Histidine-tagged IgG binding domain of protein L was purified as described¹⁴ and used without cleavage of the His-tag. All measurements were conducted with protein in 50 mM sodium phosphate pH 7.0 (phosphate) containing GuHCl as appropriate.

SAXS experiments were conducted at Beam Line 4-2 of the Stanford Synchrotron Radiation Laboratory. Static SAXS experiments were conducted at 10 mg ml⁻¹ protein L (R_g is independent of protein concentration over the range 2–10 mg ml⁻¹; data not shown). Lyophilized protein L was freshly dissolved in GuHCl/phosphate and centrifuged at ~10,000 g prior to equilibration for 10 min on ice. The sample cuvette was held at 5 ± 1 °C. R_g values were determined using the Guinier approximation²², with S^2 fitting ranges of 0.000018–0.0001 Å⁻¹. Equilibrium R_g values were fit to a linear free energy relationship using the equation:

$$R_g = [R_n^2 + (R_d^2 - R_n^2) \frac{e^{-(\Delta G_u/RT + m[\text{GuHCl}]/RT)}}{1 + e^{-(\Delta G_u/RT + m[\text{GuHCl}]/RT)}}]^{1/2} \quad (1)$$

where R_n and R_d are the R_g of native and denatured protein L, ΔG_u is the relative free energy of unfolding of protein L, and m measures the denaturant sensitivity of ΔG_u . The points in Fig. 1 represent the average of two to three independent measurements. Confidence limits reported for fitted parameters and error bars represent estimated one- σ confidence intervals.

Kinetic SAXS experiments were conducted as described^{13,23}. Refolding was initiated by rapid, stopped-flow dilution of 36 mg ml⁻¹ protein L in 6.0 M GuHCl with phosphate at a 1:3.25 ratio. The temperature was held fixed at 2.7 ± 0.5 °C, as measured by a thermocouple in the cell outlet. Stopped-flow data were collected for a total of 18 s in time-bins of 0.1 s (0–1 s), 0.5 s (1.5–6 s) or 1.2 s (7.2–18 s) duration. The Guinier plots, representing data averaged over 200 experiments, were analyzed as described above. Error bars represent estimated one- σ fitting errors. The kinetic SAXS R_g data were non linear least squares fitted to a single exponential model:

$$R_g = [R_n^2 e^{-t/\tau} + R_d^2 (1 - e^{-t/\tau})]^{1/2} \quad (2)$$

where τ is the folding time constant and the other parameters are as defined above. Kinetic I_0 values were also fit to a single exponential model. The improved fit to multi exponential models was not statistically significantly for either data set (data not shown).

Radiation damage, a potential problem for longer exposure

times at the X-ray fluxes required for time-resolved experiments, would be readily apparent in SAXS data through a large increase in forward scattering intensity. Fig. 3 and control experiments with native protein L (data not shown) indicate that significant radiation damage does not occur until >30 s of exposure, well after the termination of data collection at 18 s. A single time point at >18 s was collected by placing a manually mixed sample into the beam 85 s after the initiation of folding (Fig. 2). This time point was not included in the fit to equation 2 in order to avoid unduly biasing the estimate of the native R_g derived from stopped-flow results.

Acknowledgments

We thank H. Tsuruta for aid with and development of the stopped-flow apparatus used in this study and H. S. Chan, K. Fiebig, W. Parson, D. Shortle, D. Teller and P. Wolynes for critical readings of the manuscript. Data were collected at Beam Line 4-2 of the Stanford Synchrotron Radiation Laboratory (SSRL). The US Department of Energy and the National Institutes of Health support SSRL. A NIH FIRST award and Packard Foundation and NSF young investigator awards to D.B. supported portions of this research.

Correspondence should be addressed to D.B. email: dabaker@u.washington.edu

Received 14 December, 1998; accepted 17 March, 1999.

- Jackson, S.E. & Fersht, A.R. *Biochemistry* **30**, 10428–10435 (1991).
- Dill, K.A. et al. *Protein Sci.*, **4**, 561–602 (1995).
- Gutin, A.M., Abkevich, V.I. & Shakhnovich, E.I. *Biochemistry* **34**, 3066–3076 (1995).
- Baldwin, R.L. *Folding & Design* **1**, R1–8 (1996).
- Sosnick, T.R., Mayne, L. & Englander, S.W. *Proteins* **24**, 413–426 (1996).
- Sosnick, T.R., Shtilerman, M.D., Mayne, L. & Englander, S.W. *Proc. Natl. Acad. Sci. USA* **94**, 8545–8550 (1997).
- Roder, H. & Colón, W. *Curr. Opin. Struct. Biol.* **7**, 15–28 (1997).
- Guijarro, J.I., Morton, C.J., Plaxco, K.W., Campbell, I.D. & Dobson, C.M. *J. Mol. Biol.* **275**, 657–667 (1998).
- Thirumalai, D. *J. De Physique I* **5**, 1457–1467 (1995).
- Nymeyer, H., Garcia, A.E. & Onuchic, J.N. *Proc. Natl. Acad. Sci. USA* **95**, 5921–5928 (1998).
- Eliezer, D. et al. *Biophys. J.* **65**, 912–917 (1993).
- Kataoka, M. & Goto, Y. *Folding & Design* **1**, R107–114 (1996).
- Chen, L., Wildegger, G., Kiefhaber, T., Hodgson, K.O. & Doniach, S.A. *J. Mol. Biol.* **276**, 225–237 (1998).
- Scalley, M.L. et al. *Biochemistry* **36**, 3373–3382 (1997).
- Smith, C.K. et al. *Protein Sci.* **5**, 2009–2019 (1996).
- Segel, D.J., Fink, A.L., Hodgson, K.O. & Doniach, S. *Biochemistry* **37**, 12443–12451 (1998).
- Scalley, M.L. & Baker, D. *Proc. Natl. Acad. Sci. USA* **94**, 10636–10640 (1997).
- Nozaki, Y. & Tanford, C. *J. Biol. Chem.* **245**, 1648–1652 (1970).
- Alonso, D.O. & Dill, K.A. *Biochemistry* **30**, 5974–5985 (1991).
- Dill, K.A. & Shortle, D. *Annu. Rev. Biochem.* **60**, 795–825 (1991).
- Plaxco, K.W. & Dobson, C.M. *Curr. Opin. Struct. Biol.* **6**, 630–636 (1996).
- Glatter, O. & Kratky, O. *Small angle X-ray scattering*. (Academic Press, London; 1982).
- Tsuruta, H. et al. *Rev. Sci. Instr.* **60**, 2356–2358 (1989).
- Plaxco, K.W. & Baker, D. *Proc. Natl. Acad. Sci. USA* **95**, 13591–13596 (1998).
- Gu, H.D., Kim, D. & Baker, D. *J. Mol. Biol.* **274**, 588–596 (1997).
- Plaxco, K.W., Simons, K.T. & Baker, D. *J. Mol. Biol.* **277**, 985–994 (1998).