

Early Collapse is not an Obligate Step in Protein Folding

Jaby Jacob^{1,2}, Bryan Krantz¹, Robin S. Dothager¹, P. Thiyagarajan²
and Tobin R. Sosnick^{1,3*}

¹Department of Biochemistry
and Molecular Biology, The
University of Chicago, Chicago
IL 60637, USA

²Intense Pulsed Neutron Source
Argonne National Laboratory
Argonne, IL 60439, USA

³Institute for Biophysical
Dynamics, The University of
Chicago, Chicago, IL 60637
USA

The dimensions and secondary structure content of two proteins which fold in a two-state manner are measured within milliseconds of denaturant dilution using synchrotron-based, stopped-flow small-angle X-ray scattering and far-UV circular dichroism spectroscopy. Even upon a jump to strongly native conditions, neither ubiquitin nor common-type acylphosphatase contract prior to the major folding event. Circular dichroism and fluorescence indicate that negligible amounts of secondary and tertiary structures form in the burst phase. Thus, for these two denatured states, collapse and secondary structure formation are not energetically downhill processes even under aqueous, low-denaturant conditions. In addition, water appears to be as good a solvent as that with high concentrations of denaturant, when considering the over-all dimensions of the denatured state. However, the removal of denaturant does subtly alter the distribution of backbone dihedral ϕ, ψ angles, most likely resulting in a shift from the polyproline II region to the helical region of the Ramachandran map. We consider the thermodynamic origins of these behaviors along with implications for folding mechanisms and computer simulations thereof.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: protein folding; small-angle X-ray scattering; circular dichroism; kinetics; polyproline II

*Corresponding author

Introduction

The nature of the earliest events in protein folding remains a topic of intense interest. One view is that upon transfer of a denatured protein from a good solvent condition (i.e., high temperature, high denaturant concentration) to a poor solvent condition (i.e., low temperature, low denaturant

concentration), the polypeptide chain collapses from a random coil conformation to a stable compact state.^{1–21} In addition, protein folding simulations often tend to exhibit early collapse and secondary structure formation.^{22–28} Furthermore, intra-chain contact time-scales can be nanoseconds,^{21,29–31} although overall chain relaxation could be slower.³² Hence, one of the earliest events in protein folding arguably is an early collapse phase.³³

However, over 30 small proteins fold *via* a two-state mechanism without the accumulation of early intermediates.^{34,35} By definition for two-state proteins, only the native and unfolded states are populated during the folding process. In this ideal two-state situation, all native properties are recovered in the same kinetic step including collapse, secondary and tertiary structure formation. Two-state mechanisms are often supported by the “chevron” analysis of the denaturant dependence of folding rates.³⁶ If the equilibrium values for free energy, ΔG , and surface burial, m° -value, are equivalent to those calculated from kinetic measurements alone, then all the energy change

Present address: B. Krantz, Department of Microbiology and Molecular Genetics, Harvard Medical School, 200 Longwood Ave., Boston, MA 02115, USA.

Abbreviations used: csp, cold-shock protein; ctAcP, common-type acylphosphatase; HEWL, hen egg-white lysozyme; HX, hydrogen exchange; CD, circular dichroism; $D_{\text{physiological}}$, physiological denatured state under low-denaturant conditions; FRET, fluorescence resonance energy transfer; GdmCl, guanidinium chloride; I_σ , scattering intensity at zero angle; PPII, polyproline II; $P(r)$, pair distance distribution function; R_g , radius of gyration; SAXS, small-angle X-ray scattering; Ub, ubiquitin.

E-mail address of the corresponding author: trsosnic@midway.uchicago.edu

and surface burial in the entire two-state reaction can be accounted for in the observed kinetic phase. When these criteria are satisfied, stable species with significant free energy or surface area burial may not accumulate prior to the rate-limiting step. Therefore, a separate early collapse phase which buries surface area is inconsistent with the highly cooperative folding behavior observed for many small proteins.

Yet, sub-millisecond “burst-phase” circular dichroism (CD), fluorescence and fluorescence resonance energy transfer (FRET) signals have been observed for two-state proteins.^{9,20,21,35,37–43} We have proposed that these signal changes often represent the response of the unfolded state to the new, poorer solvent condition rather than the formation of a distinct intermediate.^{35,44–46} It is unclear how much chain contraction is associated with this relaxation process.

Small angle X-ray scattering (SAXS) is an ideal tool to measure the dimensions of macromolecules in solution.⁴⁷ The technique is particularly useful for the characterization of unfolded states which are not readily studied by high-resolution techniques.^{48–51} Furthermore, recent advances in time-resolved SAXS methods at synchrotron sources have made it possible to measure the dimensions of the refolding protein within milliseconds of the initiation of refolding.^{4,5,8,11,12,14,16,43,52,53}

Direct evidence for early collapse to a compact conformation has been obtained from SAXS measurements, albeit on multi-state proteins^{5,10–15} that contain disulfide bonds¹⁶ or prosthetic groups.^{8,52} A compact intermediate was observed for ubiquitin (Ub) refolding in the presence of 45% ethylene glycol at -20°C .⁴³ For the two-state model system, Protein L, however, SAXS measurements indicated that no collapse occurs in the burst phase; although this result was at a single, elevated denaturant concentration, 1.5 M guanidinium chloride (GdmCl).⁵⁴

To more extensively investigate the possibility of early collapse and secondary structure formation as well as the nature of the unfolded state under native conditions, we have performed time-resolved SAXS and CD_{222 nm} measurements on two proteins that are known to fold in a two-state manner, mammalian ubiquitin containing a F45W substitution (Ub, 76 aa)^{35,55} and common-type acylphosphatase (ctAcP, 98 aa).^{56,57} Upon dilution to less than 1 M denaturant, we find that both proteins fail to exhibit any appreciable chain contraction on the sub-millisecond time-scale. Whereas refolding ctAcP has no burst CD signal, Ub’s burst signal is slightly greater than its native value. However, the thermally denatured state shares this property. Hence, the burst signals for both proteins represent a solvent dependent response of the unfolded state, rather than the formation of specific intermediate. Burst CD signals as well as unfolded sloping baselines in general, likely are due to shifts from polyproline II (PPII) to helical backbone conformations in the denatured

ensemble. In total, these results demonstrate that upon transfer to aqueous conditions, only minimal local changes in the polypeptide structure occur, and an early collapse phase is not an obligatory step of protein folding.

Results

Time-resolved SAXS

We examine the compaction of Ub and ctAcP at the earliest stages of folding using SAXS at 20°C (Figures 1 and 2). The purpose of these measurements is to identify whether the polypeptide undergoes an early collapse phase. Rate constants are not determined with this method, because they can be readily obtained using either CD or fluorescence measurements. Hence, we measure the dimensions of the protein using SAXS only at a single early time point at each denaturant concentration during continuous-flow, where the refolding solution is aged for ~ 2.5 milliseconds. This strategy reduces the number of sample-intensive measurements (~ 15 mg protein/point) and permits us to examine the burst phase over an extensive denaturant range.

Equilibrium melting measurements reveal a cooperative unfolding transition at high denaturant concentration. This transition is commensurate with that obtained from far-UV CD (Figure 3). Native Ub⁵⁸ and ctAcP⁵⁹ have radii of gyration, $R_g = 13.9 \text{ \AA}$ and 14.6 \AA , which agree with their known values, 12.9 \AA and 13.9 \AA , respectively. When the proteins unfold, their R_g values increase to $\sim 26 \text{ \AA}$ and $\sim 31 \text{ \AA}$, respectively. The behavior of the R_g values as a function of denaturant concentration is consistent with a two-state folding model with the free energy being linearly dependent upon denaturant concentration according to $\Delta G([\text{denaturant}]) = \Delta G_{\text{Equil}}^{\text{H}_2\text{O}} + m^{\circ}[\text{denaturant}]$ (Table 1).

Time-resolved measurements indicate that right after dilution of denaturant, neither protein undergoes a measurable reduction in the R_g (Figure 1). This result holds down to the lowest concentrations we have investigated (0.65 M GdmCl for Ub and 0.88 M urea for ctAcP). Generally, the population of folded molecules is negligible in these continuous-flow measurements, because the folding time is much longer than the dead-time. For ctAcP, $k_f < 1 \text{ s}^{-1}$, the native population never exceeds 0.5%.

The fastest folding rate of Ub is 120 s^{-1} at 0.65 M Gdm. At this condition, $\sim 25\%$ of the molecules are estimated to have refolded within the dead-time. This fractional native population should result in a 2.5 \AA reduction in the observed R_g , which is slightly greater than the $\sim 1 \text{ \AA}$ value that we observe (i.e., we observe a 10%, rather than a 25% native fraction). SAXS measurements inherently are sensitive to this amount of native molecules, as seen in the equilibrium measurements

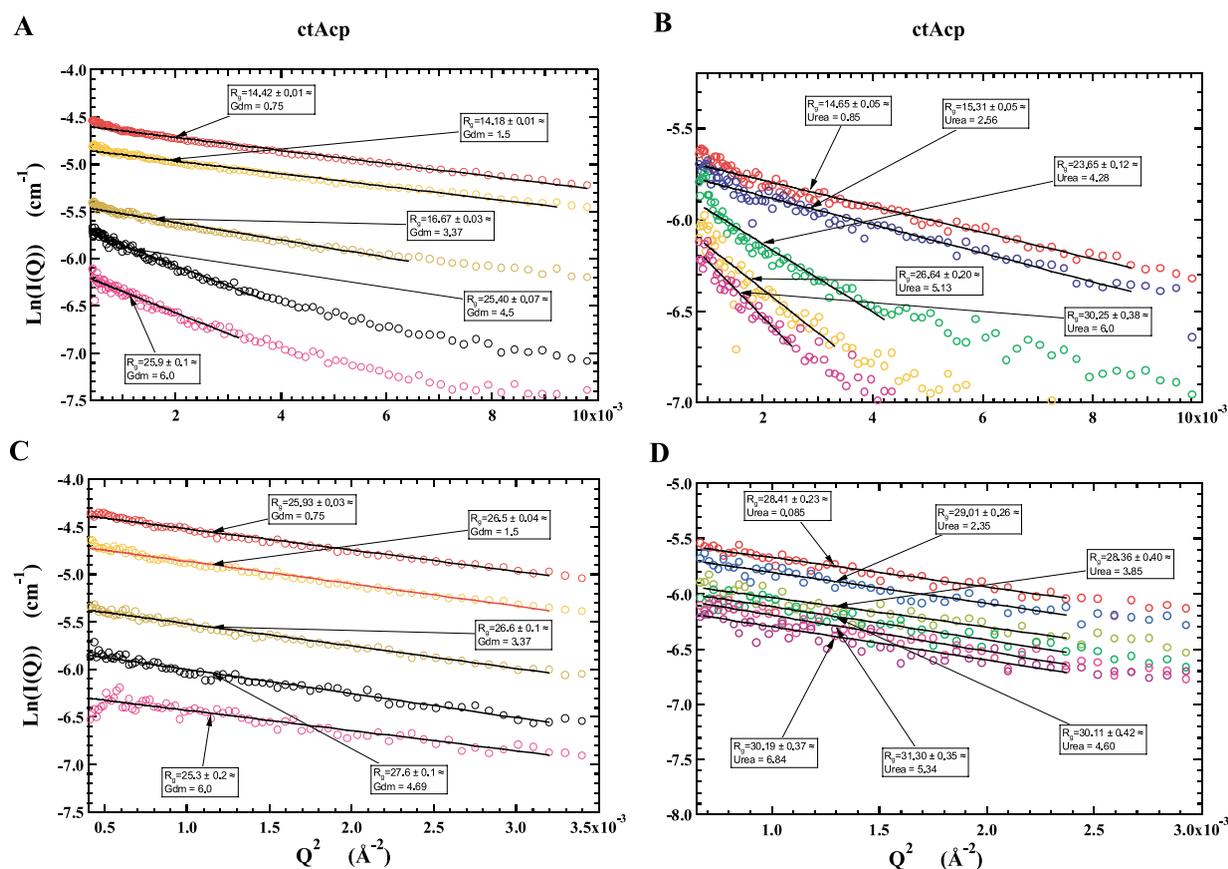


Figure 1. Equilibrium and time-resolved SAXS on Ub and ctAcP. Representative Guinier plots for equilibrium (upper panels) and burst-phase data (lower panels) obtained for Ub (left side) and CtAcP (right side). R_g values were obtained from a Guinier analysis performed using data range $Q_{\max} \leq 1.3/R_g$. Measurements were carried out at 20 °C in 50 mM Hepes, pH 7.0 (Ub) or 50 mM sodium acetate buffer, pH 5.5 (ctAcP). Unfolded samples were in 6 M GdmCl (Ub) or 8 M urea (ctAcP).

(Figure 3A). The sample has mixed with denaturant according to the scattering intensity (Figure 2A). Potentially, part of the 1.5 Å discrepancy, which is slightly larger than the repro-

ducibility of the time-resolved measurements, is due to an over-estimation of the dead-time, which depends upon the location of the X-ray beam on the capillary.

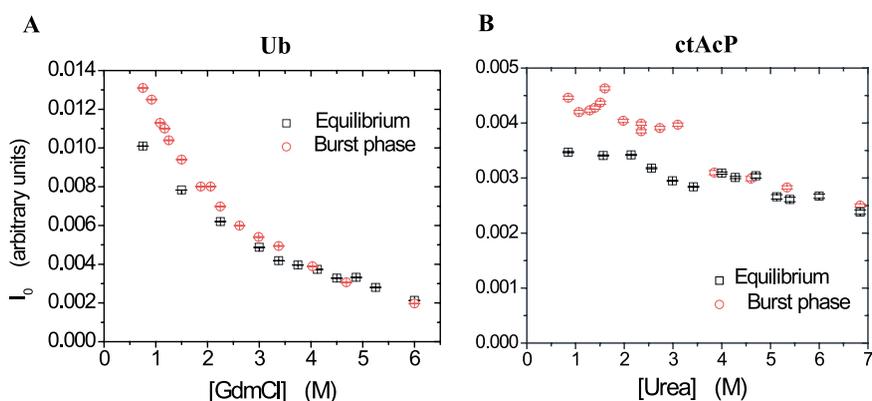


Figure 2. X-ray scattering intensity and oligomeric state. Intensity values extrapolated to zero angle, I_0 , for (A) Ub and (B) ctAcP. I_0 increases with oligomeric state and the electron contrast between protein and solvent. The equilibrium values for the folded proteins are for monomers as the observed R_g agrees with that calculated using the high-resolution structures. The similarity between the equilibrium and kinetic I_0 values, therefore, implies that during the kinetic study the proteins remained as monomers. The small discrepancy in the values at the lower denaturant concentrations may be due to the mild difference in electron contrast between a folded protein and the more solvated, unfolded chain present in the burst phase. The decrease in I_0 at higher denaturant concentration for both proteins is due to the reduction in the electron contrast between the protein and the increasingly electron dense solvent.

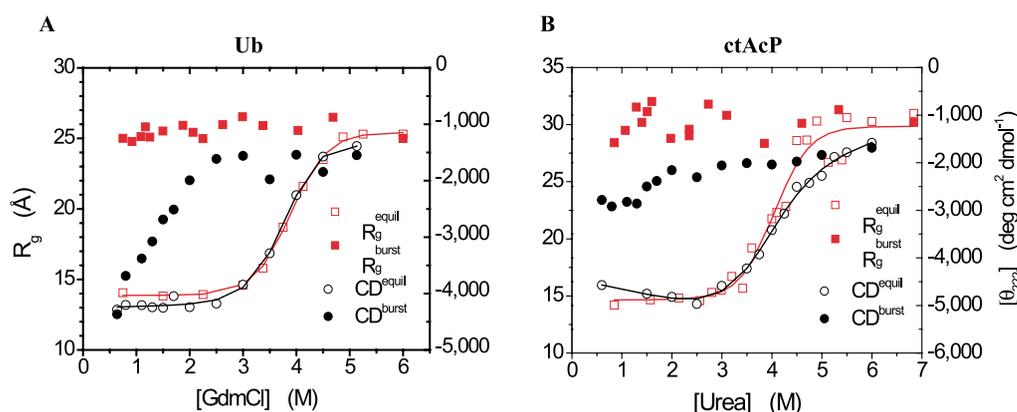


Figure 3. Denaturant dependence of equilibrium and burst phase R_g and CD values. R_g values are obtained from the representative Guinier analysis shown in Figure 1 while $CD_{222\text{ nm}}$ values were obtained from data such as shown in Figure 5.

Global shape changes can be analyzed qualitatively by means of a Kratky plot (Figure 4). In this plot of $Q^2I(Q)$ versus Q^2 , the scattering profile of a globular particle has a characteristic peak, whereas the profile for extended particles either increases at high Q , if it is an ideal random coil, or plateaus, if there is some degree of local structure in an otherwise extended chain. For both proteins, the Kratky plots obtained in the equilibrium measurements exhibit a cooperative unfolding transition with increasing denaturant going from a compact globular conformation to an extended coil. This transition is the same as observed with the R_g values. The plots for the folded proteins tend to increase at high Q , indicative of an open structure.

The burst-phase Kratky plots at all final denaturant concentrations do not exhibit the peak associated with a compact conformation. However, a small down turn or plateau occurs at $Q > 0.13\text{ \AA}^{-1}$, which is not observed in the equilibrium data of the unfolded state at high denaturant concentration. This downturn corresponds to changes in the local structure. These local changes, however, do not result in a net change in the R_g .

Shape changes are also seen in the pair-distribution function $P(r)$ (Figure 4C and D). This function is the probability distribution of vector lengths within the particle. The folded proteins have a compact conformation, while the burst-phase and equilibrium denatured states are expanded to a similar degree for both proteins.

Scattering intensity and oligomeric state

The oligomeric state of the protein is determined from the scattering intensity at zero angle, I_0 . For a monodisperse system, the I_0 value is related to the protein's molecular weight (MW), concentration (C , in mg/ml), and the electron density (ρ) difference, or contrast, between the protein and solvent:⁴⁷

$$I_0 \propto (\rho_{\text{protein}} - \rho_{\text{solvent}})^2 \text{MW} \times C \quad (1)$$

The I_0 value can be obtained from the y -intercept of the Guinier plot.

The I_0 values for both the equilibrium and kinetic measurements increase at lower denaturant concentration (Figure 2). This increase is due to the larger contrast between the protein and solvent, which is a result of decreasing electron density of the solvent with decreasing denaturant concentration. The higher electron density of GdmCl, used in the Ub studies, compared to density of urea, used in the ctAcP studies, leads to a larger change in I_0 as a function of denaturant concentration for Ub.

The equilibrium R_g values for native Ub and ctAcP agree with their known structures, indicating that they are monomeric in solution. The similarity between the equilibrium and kinetic I_0 values therefore indicates that the refolding proteins are monomeric as well. At the lowest denaturant concentrations, the equilibrium and kinetic I_0 values do slightly diverge. However, this comparison is between a folded protein and an

Table 1. Thermodynamic parameters from SAXS and $CD_{222\text{ nm}}$ measurements

| | SAXS | | $CD_{222\text{ nm}}$ | |
|-------|---|-------------|---|-------------|
| | m° (kcal mol ⁻¹ M ⁻¹) | C_m (M) | m° (kcal mol ⁻¹ M ⁻¹) | C_m (M) |
| Ub | 1.74 ± 0.09 | 3.99 ± 0.02 | 1.63 ± 0.32 | 3.84 ± 0.17 |
| ctAcP | 1.69 ± 0.36 | 4.16 ± 0.09 | 1.19 ± 0.25 | 3.80 ± 0.16 |

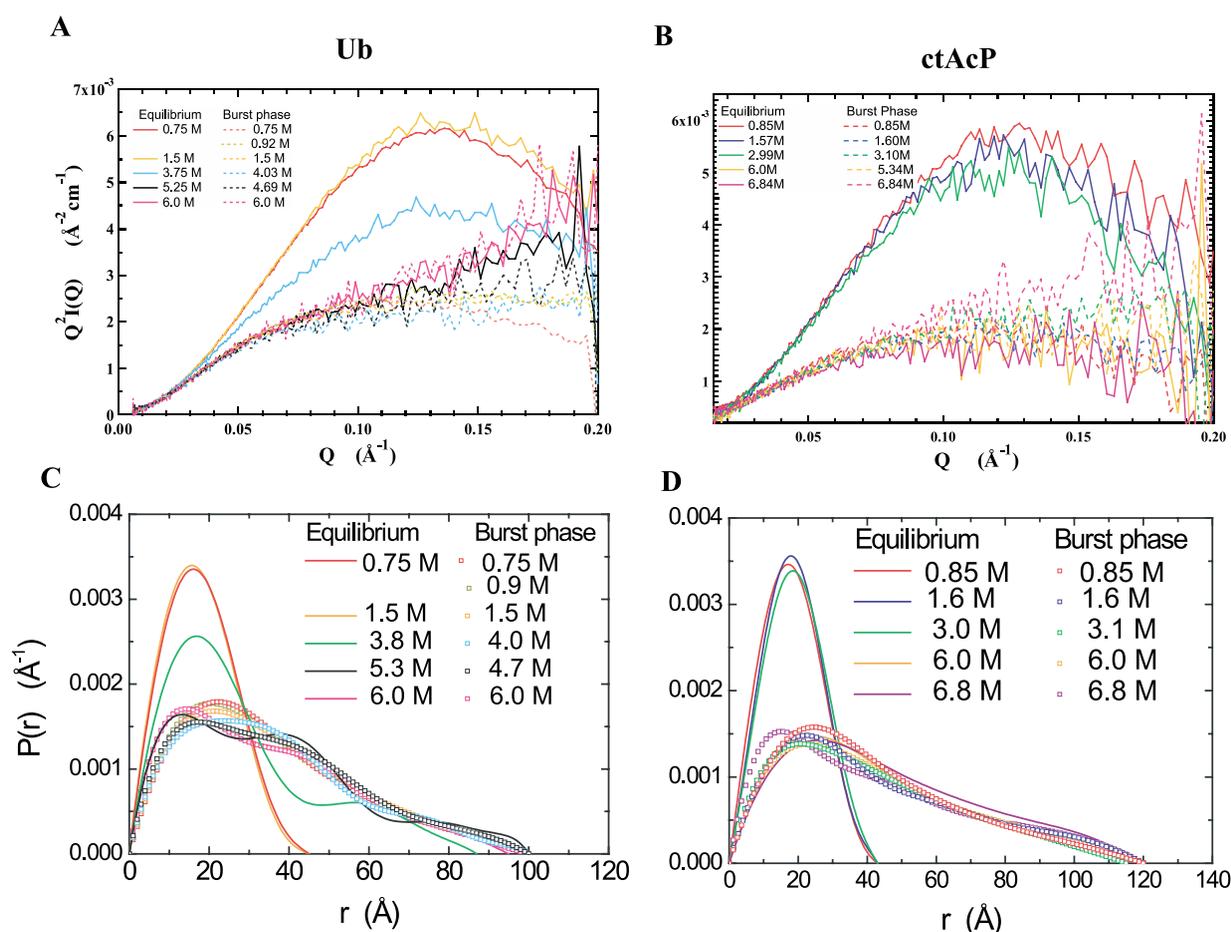


Figure 4. Globularity of Ub and ctAcP. Kratky and $P(r)$ plots for (A,C) Ub and (B,D) ctAcP after normalizing the data using I_0 values. Kratky plots serve to distinguish between compact (folded) conformations for which a peak is observed in the data, and extended (unfolded) conformations for which there is either an increase in intensity at high Q for a random coil conformation or a plateau if there is some local structure. For clarity, only a few representative plots are shown. The $P(r)$ plots similarly indicate that the folded states are compact, while both the burst phase and denatured states are similarly extended particles.

extended chain, which has different electron density contrast as there is a larger hydration shell associated with the extended chain.^{5,54,60,61} This effect accounts for the difference, although a minor amount of aggregation cannot be ruled out.

CD and fluorescence monitored folding

Prior fluorescence-monitored studies have demonstrated that Ub and ctAcP satisfy the two-state, chevron criteria.^{35,55–57} The equilibrium values for the free energy and surface burial are identical with the values obtained according to $\Delta G = \Delta G_f^\ddagger - \Delta G_u^\ddagger$ and $m^o = m_u + m_f$. This result establishes that no significant surface is buried and no significant free energy is expended prior to the major folding event.

Here, we follow the folding of Ub and ctAcP using $CD_{222 \text{ nm}}$ and fluorescence, which are measured simultaneously. Far-UV CD monitors the secondary structure content while fluorescence senses a change in the environment of the tryptophan and tyrosine residues in each protein.

For both proteins, the equilibrium CD denaturant titrations exhibit the same cooperative unfolding transition observed by SAXS (Figure 3 and Table 1). For Ub, the entire time-course of refolding is monitored (Figure 5) and the equilibrium CD values are obtained after folding has gone to completion. Due to the slow folding rates for ctAcP, pre-equilibrated samples are used to obtain the equilibrium CD values.

In the kinetic studies, ctAcP exhibits a mild burst-phase $CD_{222 \text{ nm}}$ signal (Figure 3B), while the fluorescence data lack a burst phase altogether (data not shown). The burst CD amplitude increases as the urea concentration is reduced, forming a near-linear extension of the unfolded protein's base-line.

For Ub, as GdmCl concentrations are reduced below 2.5 M, the burst $CD_{222 \text{ nm}}$ signal is larger than what would be expected from a linear extrapolation (Figure 3A). At 0.5 M GdmCl, the burst $CD_{222 \text{ nm}}$ signal is slightly larger than that of the native state. Nevertheless, the folding rate retains its denaturant dependence and the folding arm of

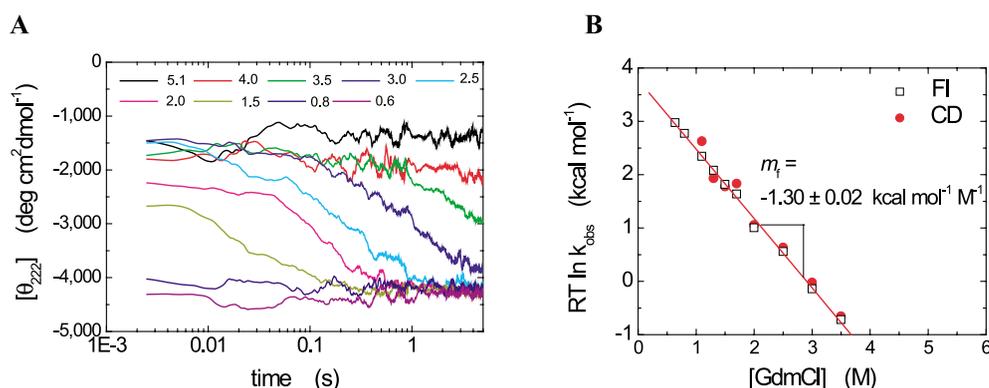


Figure 5. Denaturant dependence of Ub folding monitoring CD. A, Refolding traces at the GdmCl concentrations indicated. B, Rate constants are obtained from both the CD_{222 nm} and fluorescence traces, measured simultaneously. The continuous line is a fit to the higher accuracy fluorescence rates assuming a linear dependence of the free energy of folding. The fitted slope, or m_f value, is within the statistical precision of our previous measurements, and represents the amount of surface buried going from the denatured state to transition. Rates could not be obtained from CD_{222 nm} measurements at the two lowest denaturant concentrations as almost the entire CD signal is lost in the burst phase.

the chevron remains linear (Figure 5B). Hence, no surface burial occurs in the burst phase. In addition, the slope of the folding arm, the m_f -value, is $1.30(\pm 0.02)$ kcal mol⁻¹ M⁻¹, which agrees with our previous value of $1.35(\pm 0.02)$ kcal mol⁻¹ M⁻¹.⁵⁵

In contrast to our demonstration of two-state folding behavior for Ub, Roder and co-workers observed chevron roll-over and a fluorescent burst phase at 25 °C, but not at 8 °C.^{62,63} Accordingly, they proposed that an early folding intermediate does populate at the higher temperature. We were unable to observe either a roll-over or a burst phase with our stopped-flow apparatus,^{35,55} which has a one-millisecond, rather than a three-millisecond dead-time. In addition, we used both continuous-flow and double-jump protocols, which minimized the contribution of slow folding phases due to proline mis-isomerization and aggregation. Their data taken on the slower apparatus, however, contained significant amounts of the slower phases, which were included in the fitting process (see Figure 3D of Krantz *et al.*³⁵). This issue was most pronounced when the fast phase was nearly complete in the instrumental dead-time. As a result, the true folding rate was underestimated at the lowest denaturant concentrations at 25 °C, which inadvertently created a roll-over and a burst phase.

A time-resolved SAXS study conducted in 45% ethylene glycol at -20 °C observed the rapid formation of a compact, highly helical intermediate.⁴³ As this species did not populate under conditions used here, it presumably formed due to the presence of the concentrated cosolvent and the low temperature.

In summary, for the Ub burst phase, the polypeptide: (i) does not collapse, (ii) does not have a distinct fluorescence signal (data not shown), (iii) does not show any increase in surface burial, and (iv) does not form hydrogen bonds.⁶⁴ The absence of these changes requires some alternative

explanation for the origin of the burst-phase CD_{222 nm} signal. In addition, it is unclear why the CD_{222 nm} burst phase differs from previous work.⁶⁴ Therefore, additional thermal denaturation experiments were conducted to confirm our observations.

Thermal denaturation of Ub

In order to compare the CD_{222 nm} signal in the chemically denatured state to the thermally denatured state, we measured the far-UV CD spectra at higher temperatures (Figure 6). Changes in the CD spectra indicate that, at 95 °C, the protein has thermally denatured. In particular, the CD_{200 nm} signal goes from positive to negative, indicating an increase in the amount of coil or PII structure, defined as the extended conformation with dihedral angles $\Phi = -75^\circ$ and $\Psi = 150^\circ$.

The thermally denatured state has a stronger (more negative) CD_{222 nm} signal than the native state. This unusual increase upon thermal unfolding may be due in part to native Ub's low CD_{222 nm} value which results⁶⁵ from its high PPII content (32%, 24% and 34% for β , PPII, and α dihedral angles, respectively). The increase in CD_{222 nm} upon thermal denaturation probably is due to some residues with native dihedral angles in the PPII basin adopting helical ϕ , ψ angles in the denatured state (see below).

In summary, both the burst phase at 0.5 M GdmCl and the thermally denatured protein have nearly the same CD_{222 nm} signal. In addition, the absolute magnitude of the burst phase change is small ($\Delta CD \sim 3000$ deg cm² dmol⁻¹). These results, in combination with the lack of a change in the R_g , fluorescence signal, surface burial and hydrogen bond content between the burst phase and U_{6 M Gdm} argue that the burst-phase CD signal represents the adjustment of the denatured state to the new solvent condition, $D_{\text{physiological}}$, rather

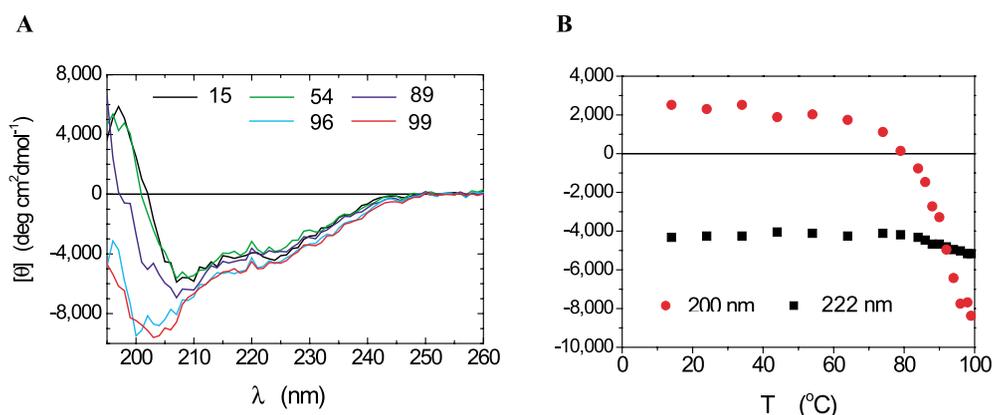


Figure 6. Temperature melting of Ub at pH 5.5. A, CD spectra of Ub at different temperatures (in °C). B, Unfolding of Ub monitored by CD at 200 nm (red circles) and 222 nm (black squares). The unfolding transition is clearly indicated by the behavior of $\text{CD}_{200 \text{ nm}}$ while the magnitude of $\text{CD}_{222 \text{ nm}}$ is higher in the unfolded state compared to folded state. This unusual increase upon thermal denaturation is likely due to a few residues on average shifting from the PPII basin to the helical basin of the Ramachandran plot. The lower pH (5.5 *versus* 7.0) was used to decrease the melting temperature.

than the formation of a meaningfully structured intermediate.

Discussion

Does an unfolded protein chain collapse rapidly and form secondary structure upon transfer to aqueous conditions? Is water a poor solvent for polypeptides? These issues are crucial to our understanding of the early steps in protein folding. We find that the dimensions of the unfolded state of Ub and ctAcP do not change upon transfer to strongly native conditions prior to the major folding event that leads to the native state. Furthermore, $\text{CD}_{222 \text{ nm}}$ data indicate that secondary structure does not form to an appreciable degree in the burst phase. These results demonstrate that the adjustment of the denatured state to the poorer, aqueous solvent condition can be a relatively minor process, and early collapse is not an obligatory step in protein folding.

In retrospect, this result may have been anticipated as both these proteins fold with only the unfolded and folded states being significantly populated.^{35,55–57} More importantly, both proteins satisfy the chevron criterion $m^0 = m_u + m_f$, which mandates that no denaturant sensitive surface is buried prior to the (sole) kinetic barrier. Hence, the denatured state at low denaturant has the same amount of denaturant sensitive surface area buried as it does at high denaturant concentration.

Nevertheless, we would have expected some contraction upon the transfer to low denaturant concentrations because the number of long-range hydrophobic interactions should have increased, even if only mildly, in the poorer solvent condition. As neither contraction nor surface burial occurs in the burst phase, we infer that long-range interactions do not increase appreciably in the denatured state under aqueous conditions. In

addition, the R_g of the denatured state under these conditions agrees with the self-avoiding random coil scaling behavior observed for many fully denatured proteins.⁴⁸ Therefore, the denatured states of Ub and ctAcP are dominated by local interactions and steric constraints at both low and high denaturant concentrations.

Collapse behavior in other proteins

Other SAXS measurements also found that the dimensions of denatured states do not change upon transfer to aqueous conditions. Protein L was measured in 1.5 M GdmCl in a time-resolved mode,⁵⁴ both hen egg-white lysozyme (HEWL)⁶⁶ and ribonuclease A (J.J., unpublished results) were measured in equilibrium mode under aqueous conditions using unfolded analogues created by reduction of each of their four disulfide bonds. In addition, once a protein has unfolded, its R_g does not increase with additional denaturant.^{48,49,67} Hence, the dimensions of unfolded state often remain unchanged even upon transfer to low denaturant, physiological condition. For these denatured states, water does not drive a generic collapse. Only upon folding to the native state does collapse occur.

However, this result is not universal and exceptions are found with proteins which often contain disulfide bonds or prosthetic groups.^{4,5,8,11,12,14,16,17,48,51,52,68} Cytochrome *c* provides a particularly interesting case as this protein has a very linear chevron folding arm.^{35,40} Even so, the denatured state responds to a change in solvent condition,^{35,40,45,46} with the R_g decreasing from ~ 30 Å in 3 M GdmCl to ~ 21 Å in water,^{8,52} consistent with the increase in the FRET signal.^{35,40,45,46}

For another chevron satisfying protein,⁶⁹ csp from *Thermotoga maritima*, a FRET-determined distance decreased by $\sim 25\%$ for the denatured

protein upon transfer to 0.5 M GdmCl.^{9,19} Although the large size of the two chromophores (four or six aromatic rings) opens up the possibility of their association, the denatured state lacking them also contracted, as assayed by tryptophan quenching.²¹ In addition, in the refolding of the csp from *Bacillus caldolyticus*, up to 70% of a FRET signal is quenched in the burst phase.²⁰ Nevertheless, Magg & Schmid conclude that in spite of this quenching, the burst-phase species cannot be a partially folded intermediate. Potentially, some of the change in the FRET signal is due to a different local environment resulting from a shift in Φ, Ψ conformers. Nevertheless, SAXS measurements, which directly monitor chain dimensions, would be extremely helpful in quantifying the amount of contraction that can occur in these otherwise two-state systems which give FRET burst signals.

Thermodynamic considerations

For over 30 small proteins, stable species do not accumulate on the unfolded side of the sole kinetic barrier.^{34,35,70} As intra-chain contact times are sub-microseconds, polymer-style collapse could be on this time-scale. As this time-scale is three to eight orders of magnitude faster than the overall folding times, the unfolded chain has ample time to collapse before it traverses the folding barrier. However, the chain does not do so for these highly cooperative proteins. Therefore, the absence of stable collapsed species prior to acquisition of the native state must be due to thermodynamic rather than kinetic constraints.

Although hydrophobic interactions are enhanced upon the shift to aqueous conditions, collapse still is inhibited by the loss of conformational entropy, both backbone and side-chain, and the desolvation of main-chain polar groups. As there is no increase in hydrophobic burial in the denatured state ($m^0 = m_u + m_f$), we conclude that hydrophobic association must be difficult to achieve without a costly desolvation of the polar backbone.

Previous results further support this reasoning. Kinetic ²H/H amide isotope effect experiments demonstrate that surface burial occurs commensurately with native hydrogen bond formation along the folding pathway.^{35,70} Stable hydrogen bonded secondary structure cannot be divorced from surface burial, thereby precluding early collapse without a significant amount of hydrogen bonding. Consistently, burst-phase hydrogen bond formation has not been detected by hydrogen exchange experiments on two-state proteins.^{41,64,71}

From a theoretical vantage point, stable backbone desolvation is very demanding as it can only occur in a context that satisfies hydrogen bonding requirements. The burial of each unsatisfied hydrogen bond partner costs several kcal mol⁻¹.⁷² Given that the net stability of a protein often is less than 10 kcal mol⁻¹, even a small number of failures results in an unstable collapsed species. Therefore,

a generic, hydrophobic collapse is unlikely. Due to the nearly universal trend for two-state folding in small proteins, these considerations are likely to be general, although exceptions can occur for functional reasons.¹⁷

Folding mechanisms

Main-chain burial requirements inhibit non-native chain condensation strongly, and in conjunction with entropic considerations, the early steps in the conformational search for the transition state are uphill in free energy. Stable species are difficult to form, and present only on the native side of the barrier. By the time the chain has achieved a stable conformation, it has such a significant proportion of surface burial and hydrogen bond formation that it will have already traversed the kinetic barrier.

Based upon these and other considerations, we proposed that formation of a stable collapsed species is the intrinsic rate-limiting step in protein folding.^{45,73,74} The limiting step itself is the formation of a transition state structure with a sufficient number of long-range contacts which adequately define the native topology.⁷⁵ At this point, the nucleation event, a considerable amount of configuration entropy is lost and folding can proceed rapidly downhill.

Implications to computer simulations

Even though folding studies with computer simulations probe single molecule trajectories, their average behavior should be consistent with our ensemble measurements (Figure 7). For a two-state system, the protein will jump back and forth between the unfolded and folded states.^{9,19,76} Transiently populated intermediates manifest themselves as brief stops between the two levels.⁷⁷ Hence, for simulations starting from a realistic unfolded state, the chain should, on an average, retain the properties of the unfolded state prior to a sharp transition leading to the native state. A realistic unfolded state would be a self-avoiding chain with the appropriate residue-dependent Ramachandran basin distributions, rather than, for example, a fully extended state that rapidly contracts.

Before the major folding transition, the chain can briefly sample collapsed conformations. However, the time-averaged population of such conformations should be much less than that of the unfolded state in order to agree with the experimental results. A greater population fraction would represent the formation of a stable collapsed intermediate before the rate-limiting step. We do not observe these collapsed species in burst-phase SAXS measurements. Even a low accumulation level of a collapsed species with significant surface burial would be detectable in an ensemble experiment. In addition to having a smaller R_{gr} , the intermediate would have buried surface that

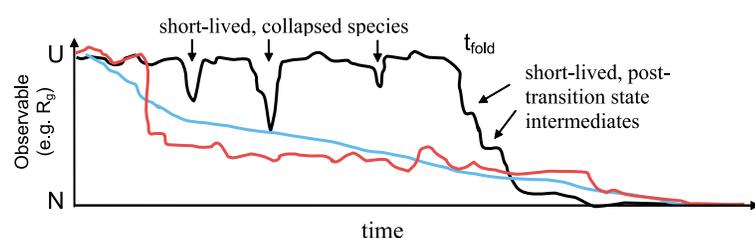


Figure 7. Two-state folding behavior and the lack of an early collapse phase places constraints on computer simulations. The black illustrative trace is consistent with ensemble behavior, as the observable probe has either the value of the unfolded state or native state for the majority of the trajectory. Only unstable, short-lived

collapsed species exist prior to major folding transition at t_{fold} . Likewise, post-transition state intermediates are short-lived compared to average folding time, obtained from multiple trajectories. The blue and red trajectories are inconsistent with ensemble two-state folding behavior and the present SAXS data, as collapsed intermediates populate, and the kinetic data will not satisfy the chevron criteria, $m^0 = m_u + m_f$.

would reduce the denaturant dependence of folding rate and produce a “roll-over” in the chevron folding arm. Observed folding arms often are linear to better than 10% of the m_f -value. Therefore, the fraction of time that the chain in a folding simulation is allowed to sample a non-native, collapsed conformation is severely limited.

Furthermore, all native-like properties should be acquired in the major transition, or shortly thereafter. Otherwise, the average of multiple simulations would be inconsistent with the observation that all the native properties are acquired in the same kinetic step.

Molecular origin of burst phase signals

Ub and ctAcP’s denatured states exhibit no contraction upon the shift to aqueous conditions, but the amount of short-range structure increases. For ctAcP, the change in CD is fully accounted for by a sloping base-line of the unfolded state. All the stopped-flow CD data on kinetically two-state proteins we are aware of share this property^{41,42,78} although the base line sometimes has mild curvature.^{44,46,79}

For Ub, the burst-phase CD has pronounced increase starting at 2.5 M GdmCl. Both the burst phase at 0.5 M GdmCl and the thermally denatured protein have larger CD signal than the native state. The difference may be due to the unusually low native $CD_{222\text{ nm}}$ value resulting from its high (24%) PPII content.⁶⁵ For both Ub and cytochrome *c*,⁴⁶ the thermally denatured state is a better approximation of $D_{\text{physiological}}$ than is the unfolded state at high denaturant concentration.

In summary, for all two-state systems considered, burst-phase CD signals simply represent the response of the unfolded state to the change in solvent condition. The origin of these signals is unclear as they often are not associated with stable hydrogen bonded structure.^{35,46,64,80} The spectroscopic signals could be due to transient, itinerant helix formation running throughout the chain. We propose that the residual CD signals could also result from a shift in the average backbone dihedral angle distribution from PPII to helical conformers. This increase in the number of peptide groups with ϕ , ψ angles in the helical basin of the

Ramachandran plot does not necessarily involve any organized helical structure. Shifts to the β -basin can be considered and would produce the same qualitative result, but this basin is not significantly populated in the denatured state.^{81–85}

PPII has a positive $CD_{222\text{ nm}}$ signal, and is the dominant conformation in the denatured state.^{86–94} PPII has a more exposed backbone and increased solvent–protein interactions compared to the helical conformation.^{95–99} Additionally, denaturants enhance PPII conformations (N. Kallenbach, personal communication).

Small shifts in basin populations even for short amino acid stretches are sufficient to quantitatively account for the denaturant dependence of the burst-phase CD. Even one to three consecutive helical residues have $-10,000\text{ mdeg cm}^2\text{ dmol}^{-1}$ to $-20,000\text{ mdeg cm}^2\text{ dmol}^{-1}$, respectively^{88,100} (R. Woody, personal communication). A subtle 20% shift from the helical basin to the PPII basin, using values of $0\text{ mdeg cm}^2\text{ dmol}^{-1}$ and $-15,000\text{ mdeg cm}^2\text{ dmol}^{-1}$ for $[\theta]_{\text{PPII}}$ and $[\theta]_{\text{helix}}$ respectively, would decrease the $CD_{222\text{ nm}}$ signal by $3000\text{ mdeg cm}^2\text{ dmol}^{-1}$. (This decrease is what we observe in Ub’s burst phase.)

In spite of the increase in helical basin population, the R_g does not change. However, this same result is found even when extensive amounts of isolated helix are formed in methanol^{101,102} and trifluoroethanol solutions.⁶⁶

Conclusions

Upon a jump to native conditions, the denatured chain remains extended for both Ub and ctAcP. Hence, water is a sufficiently good solvent that generic intra-chain contacts do not outweigh protein–solvent interactions to produce an early collapse phase. The local backbone changes that do occur are likely the result of a minor redistribution from PPII to helical conformers. These results place severe constraints on folding mechanisms and computer simulations of cooperatively folding proteins. The origins of these observations are due, in part, to thermodynamic issues including backbone desolvation, backbone and side-chain entropy, and hydrogen

bonding requirements. However, it is unclear whether strong protein–solvent interactions and lack of collapse are exclusive to protein sequences which fold in a two-state manner. Potentially, these properties are applicable to generic polypeptides, suggesting that chains of α -amino acid residues were chosen in part for having these properties, which would tend to reduce misfolding and aggregation.

Materials and Methods

Expression and purification

Proteins were expressed and purified as described^{35,55} in multiple batches to produce gram-scale quantities for SAXS experiments.

Small-angle X-ray scattering

SAXS experiments were carried out at the SAXS instrument on the BioCAT ID-18 beam-line of Argonne National Laboratory's Advanced Photon Source. Data were collected using a single chip (8.6 cm \times 4.9 cm) CCD area detector and exposure times were approximately 0.4 seconds for each measurement. Typically, four measurements were recorded at each concentration and these were averaged to obtain a single data point. Sample to detector distance was \sim 2 m and energy of X-ray radiation was set to 12 keV.

Experiments used a Biologic SFM 400 stopped-flow apparatus⁵⁵ with a 1.0 mm diameter cylindrical quartz capillary of 0.01–0.02 mm wall thickness. In order to minimize the dead-time in kinetic measurements, SFM 400 was mounted at the beam-line so that the X-ray beam passed through the lowest part of the capillary tube minimizing the distance between the mixer and the point where the X-rays pass through the capillary tube. Pre-equilibrated samples at the appropriate denaturant concentrations were used to obtain equilibrium data. A constant flow (flow speed 0.5 ml s⁻¹) was maintained to avoid radiation damage. Final protein concentration was \sim 1.5 mg ml⁻¹. In acquiring the kinetic data, the unfolded protein (concentration \sim 12 mg ml⁻¹) was diluted eightfold to appropriate denaturant concentration and data were acquired at a constant flow speed of 4 ml s⁻¹, corresponding to refolding dead-time of \sim 2.5 milliseconds. The buffer background scattering was obtained from an otherwise identical configuration.

Measurements were carried out at 20 °C in 50 mM Hepes, pH 7.0 (Ub) or 50 mM sodium acetate buffer, pH 5.5 (ctAcP). Unfolded samples were in 6 M GdmHCl (Ub) or 8 M urea (ctAcP).

CD and fluorescence measurements

Far-UV circular dichroism spectroscopy was performed with a Jasco Model 715 CD spectropolarimeter. Equilibrium CD spectra were recorded every 1 nm at 2-nm resolution in a 1 cm pathlength cell. The kinetic measurements used the Biologic SFM 4 stopped-flow apparatus interfaced with the Jasco spectropolarimeter and a 0.8 mm path length cuvette. The resolution was set to 2 nm and between ten and 40 traces were averaged at each condition. Simultaneous acquisition of the fluorescence signal was achieved using a second photo-

multiplier tube mounted at 90° to the incident beam and using a 300–400 nm band pass filter.

Time-dependent change in CD was placed on an absolute scale using the identical protein solution in the standard spectrophotometer sample holder. FI values were normalized by measurements on the folded and the unfolded protein. The signal from the buffer was subtracted at every denaturant concentration, and burst-phase amplitudes shown in Figure 3 were extrapolated back to zero time. The data in the burst phase are acquired by averaging 0.15–0.2 seconds of data in the continuous flow phase. Even though the dead volume (volume between mixer and the point of observation) is higher compared to that in SAXS measurements, a higher flow-speed of 6 ml s⁻¹ was used in these measurements leading to approximately the same dead-time of 2.5 milliseconds.

Identical buffer conditions were maintained in the CD and fluorescence measurements as for the SAXS measurements, except for lower protein concentrations. Final protein concentration was 92 μ M and 69 μ M for Ub and ctAcP, respectively. Approximately 40 shots were averaged at each condition. To reduce buffer and sample absorbance, temperature melting of Ub was carried out in 5 mM sodium acetate at a protein concentration of 19 μ M and pH 5.5. The lower pH was used to lower the melting temperature.

Data analysis

SAXS reports on the average structure of particles in solution. X-rays are scattered by the sample and the scattering profile is measured at very low angles θ .¹⁰³ Data are presented as the scattering intensity per solid angle, $I(Q)$, where the scattering vector Q , is defined as $Q = 4\pi \sin \theta / \lambda$, λ is the X-ray wavelength, and θ is the half scattering angle. At low $Q \leq 1.3/R_g$, the dimensions of a particle can be determined from the width of the inner part of the scattering profile, which can be approximated as a Gaussian, $I(Q) = I_0 e^{-Q^2 R_g^2/3}$. The R_g is the root-mean-square of the distances of all regions to the center-of-mass of the particle. Typically, the R_g is obtained from the slope of the low-angle region of a Guinier Plot of $\ln I(Q)$ versus Q^2 .

To measure the R_g value accurately, scattering data must be taken across angles where the Gaussian approximation holds. Our reliable data starts at $Q = 0.02 \text{ \AA}^{-1}$, or $QR_g = 0.6$ for the largest size ($R_g = 30\text{--}32 \text{ \AA}$ for unfolded ctAcP). This low Q_{\min} value, along with the linearity in the Guinier plot and the expected behavior of I_0 , demonstrate that the R_g can be reliably obtained from these data.

More precise structural parameters were derived from a $P(r)$ analysis using the entire scattering profile. The $P(r)$ function has a maximum at the most probable distance in the object (e.g., slightly larger than the radius for a sphere) and goes to zero at the maximum dimension, d_{\max} , of the object (e.g. the diameter). $P(r)$ functions were calculated according to:

$$P(r) = \frac{1}{2\pi^2} \int_{Q_{\min}}^{Q_{\max}} I(Q) Q r \sin(Qr) dQ \quad (2)$$

using the indirect Fourier inversion algorithms developed by Svergun[†].

[†] GNOM small-angle scattering data processing by means of the regularization technique. <http://www.srs.dl.ac.uk/ncd/computing/manual/gnom.html>

The equilibrium R_g values were fitted to a linear free energy relationship:

$$R_g = \sqrt{R_d^2 + \frac{R_n^2 - R_d^2}{1 + e^{-(mC_0 - m[\text{den}])/RT}}} \quad (3)$$

where R_n and R_d are the R_g of native and denatured protein, respectively, m is the equilibrium m -value and C_0 is the midpoint of the collapse transition.⁴⁹ The equilibrium CD data were fitted to standard two-state transitions with sloping linear baselines:

$$\begin{aligned} \text{CD} = & \text{CD}_u + S_u[\text{den}] \\ & + \frac{\text{CD}_f + S_f[\text{den}] - \text{CD}_u - S_u[\text{den}]}{1 + e^{-(mC_m - m[\text{den}])/RT}} \end{aligned} \quad (4)$$

where CD_f and CD_u represent the CD signals of the folded and unfolded states in water, respectively, S_f and S_u represent the denaturant dependencies of these values, and C_m , the midpoint of the transition.

The kinetic data were analyzed using the "chevron analysis" of the denaturant dependence of folding rate constants¹⁰⁴ where the standard free energy of folding, ΔG° , along with the standard activation free energy for folding ΔG_f^\ddagger , and unfolding, ΔG_u^\ddagger , are linearly dependent on denaturant concentration:

$$\Delta G^\circ([\text{den}]) = \Delta G_{\text{Equiv}}^{\text{H}_2\text{O}} + m^o[\text{den}] = -RT \ln(K_u) \quad (5a)$$

$$\Delta G_f^\ddagger([\text{den}]) = -RT \ln(k_f^{\text{H}_2\text{O}}) + m_f[\text{den}] + \text{Constant} \quad (5b)$$

$$\Delta G_u^\ddagger([\text{den}]) = -RT \ln(k_u^{\text{H}_2\text{O}}) - m_u[\text{den}] + \text{Constant} \quad (5c)$$

where R is the universal gas constant, T is the temperature, and the dependence on denaturant concentration, the m -values, report the degree of surface area burial during the folding process. When equilibrium and kinetic folding reactions are effectively two-state and are limited by the same activation barrier, the equilibrium values for the standard free energy and surface burial can be calculated from the kinetic measurements according to $\Delta G^\circ = \Delta G_f^\ddagger - \Delta G_u^\ddagger$ and $m^o = m_u + m_f$. Parameters were fit using non-linear least-squares algorithms implemented in the Microcal Origin software package.

Acknowledgements

We thank K. Plaxco, S. W. Englander, N. Kallenbach, R. Woody, A. Garcia, G. Rose, A. Kentsis, R. Pappu, D. Raleigh, and members of our group for discussions, unpublished results and preprints, and Dr Elena Kondrashkina for expert assistance at the beam-line. This work is supported by grants from the NIH (T.R.S.), The US Department of Energy, BES-Materials Science, under contract W-31-109-ENG-38 (P.T.), The Univ. of Chicago-Argonne National Laboratory Collaborative Seed Grant Program (T.R.S. & P.T.), and Packard Foundation Interdisciplinary Science Program (T.R.S., P.T., D. Lynn, S. Meredith and R. S. Berry). Use of the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science, under contract No. W-31-109-ENG-38. BioCAT is a National

Institutes of Health-supported Research Center RR-08630.

References

- Hagen, S. J. & Eaton, W. A. (2000). Two-state expansion and collapse of a polypeptide. *J. Mol. Biol.* **297**, 781–789.
- Agashe, V. R., Shastry, M. C. & Udgaonkar, J. B. (1995). Initial hydrophobic collapse in the folding of barstar. *Nature*, **377**, 754–757.
- Ballew, R. M., Sabelko, J. & Gruebele, M. (1996). Direct observation of fast protein folding: the initial collapse of apomyoglobin. *Proc. Natl Acad. Sci. USA*, **93**, 5759–5764.
- Pollack, L., Tate, M., Finnefrock, A., Kalidas, C., Trotter, S., Darnton, N. *et al.* (2001). Time resolved collapse of a folding protein observed with small angle X-ray scattering. *Phys. Rev. Letters*, **86**, 4962–4965.
- Chen, L., Wildegger, G., Kiefhaber, T., Hodgson, K. O. & Doniach, S. (1998). Kinetics of lysozyme refolding: structural characterization of a non-specifically collapsed state using time-resolved X-ray scattering. *J. Mol. Biol.* **276**, 225–237.
- Parker, M. J., Dempsey, C. E., Lorch, M. & Clarke, A. R. (1997). Acquisition of native beta-strand topology during the rapid collapse phase of protein folding. *Biochemistry*, **36**, 13396–13405.
- Nath, U. & Udgaonkar, J. B. (1997). Folding of tryptophan mutants of barstar: evidence for an initial hydrophobic collapse on the folding pathway. *Biochemistry*, **36**, 8602–8610.
- Akiyama, S., Takahashi, S., Kimura, T., Ishimori, K., Morishima, I., Nishikawa, Y. & Fujisawa, T. (2002). Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle X-ray scattering. *Proc. Natl Acad. Sci. USA*, **99**, 1329–1334.
- Lipman, E. A., Schuler, B., Bakajin, O. & Eaton, W. A. (2003). Single-molecule measurement of protein folding kinetics. *Science*, **301**, 1233–1235.
- Segel, D. J., Eliezer, D., Uversky, V., Fink, A. L., Hodgson, K. O. & Doniach, S. (1999). Transient dimer in the refolding kinetics of cytochrome c characterized by small-angle X-ray scattering. *Biochemistry*, **38**, 15352–15359.
- Segel, D. J., Bachmann, A., Hofrichter, J., Hodgson, K. O., Doniach, S. & Kiefhaber, T. (1999). Characterization of transient intermediates in lysozyme folding with time-resolved small-angle X-ray scattering. *J. Mol. Biol.* **288**, 489–499.
- Eliezer, D., Chiba, K., Tsuruta, H., Doniach, S., Hodgson, K. O. & Kihara, H. (1993). Evidence of an associative intermediate on the myoglobin refolding pathway. *Biophys. J.* **65**, 912–917.
- Doniach, S., Basile, J., Garel, T. & Orland, H. (1995). Partially folded states of proteins: characterization by X-ray scattering. *J. Mol. Biol.* **254**, 960–967.
- Chen, L., Hodgson, K. O. & Doniach, S. (1996). A lysozyme folding intermediate revealed by solution X-ray scattering. *J. Mol. Biol.* **261**, 658–671.
- Arai, S. & Hirai, M. (1999). Reversibility and hierarchy of thermal transition of hen egg-white lysozyme studied by small-angle X-ray scattering. *Biophys. J.* **76**, 2192–2197.
- Arai, M., Ito, K., Inobe, T., Nakao, M., Maki, K., Kamagata, K. *et al.* (2002). Fast compaction of alpha-lactalbumin during folding studied by

- stopped-flow X-ray scattering. *J. Mol. Biol.* **321**, 121–132.
17. Choy, W. Y., Mulder, F. A., Crowhurst, K. A., Muhandiram, D. R., Millett, I. S., Doniach, S. *et al.* (2002). Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques. *J. Mol. Biol.* **316**, 101–112.
 18. Dill, K. A. & Shortle, D. (1991). Denatured states of proteins. *Annu. Rev. Biochem.* **60**, 795–825.
 19. Schuler, B., Lipman, E. A. & Eaton, W. A. (2002). Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature*, **419**, 743–747.
 20. Magg, C. & Schmid, F. X. (2004). Rapid collapse precedes the fast two-state folding of the cold shock protein. *J. Mol. Biol.* **335**, 1309–1323.
 21. Buscaglia, M., Schuler, B., Lapidus, L. J., Eaton, W. A. & Hofrichter, J. (2003). Kinetics of intramolecular contact formation in a denatured protein. *J. Mol. Biol.* **332**, 9–12.
 22. Zagrovic, B., Snow, C. D., Khaliq, S., Shirts, M. R. & Pande, V. S. (2002). Native-like mean structure in the unfolded ensemble of small proteins. *J. Mol. Biol.* **323**, 153–164.
 23. Shimada, J. & Shakhnovich, E. I. (2002). The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation. *Proc. Natl Acad. Sci. USA*, **99**, 11175–11180.
 24. Duan, Y. & Kollman, P. A. (1998). Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, **282**, 740–744.
 25. Shen, M. Y. & Freed, K. F. (2002). All-atom fast protein folding simulations: the villin headpiece. *Proteins: Struct. Funct. Genet.* **49**, 439–445.
 26. Alonso, D. O. V. & Daggett, V. (1998). Molecular dynamics simulations of hydrophobic collapse of ubiquitin. *Protein Sci.* **7**, 860–874.
 27. Linhananta, A., Zhou, H. Y. & Zhou, Y. Q. (2002). The dual role of a loop with low loop contact distance in folding and domain swapping. *Protein Sci.* **11**, 1695–1701.
 28. Mayor, U., Guydosh, N. R., Johnson, C. M., Grossmann, J. G., Sato, S., Jas, G. S. *et al.* (2003). The complete folding pathway of a protein from nanoseconds to microseconds. *Nature*, **421**, 863–867.
 29. Lapidus, L. J., Eaton, W. A. & Hofrichter, J. (2000). Measuring the rate of intramolecular contact formation in polypeptides. *Proc. Natl Acad. Sci. USA*, **97**, 7220–7225.
 30. Krieger, F., Fierz, B., Bieri, O., Drewello, M. & Kiefhaber, T. (2003). Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding. *J. Mol. Biol.* **332**, 265–274.
 31. Chang, I. J., Lee, J. C., Winkler, J. R. & Gray, H. B. (2003). Bioinorganic Chemistry Special Feature: the protein-folding speed limit: intrachain diffusion times set by electron-transfer rates in denatured Ru(NH₃)₅(His-33)-Zn-cytochrome c. *Proc. Natl Acad. Sci. USA*, **100**, 3838–3840.
 32. Qiu, L., Zachariah, C. & Hagen, S. J. (2003). Fast chain contraction during protein folding: “foldability” and collapse dynamics. *Phys. Rev. Letters*, **90**, 168103.
 33. Sadqi, M., Lapidus, L. J. & Munoz, V. (2003). How fast is protein hydrophobic collapse? *Proc. Natl Acad. Sci. USA*, **100**, 12117–12122.
 34. Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold. Des.* **3**, R81–R91.
 35. Krantz, B. A., Mayne, L., Rumbley, J., Englander, S. W. & Sosnick, T. R. (2002). Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding. *J. Mol. Biol.* **324**, 359–371.
 36. Jackson, S. E. & Fersht, A. R. (1991). Folding of chymotrypsin inhibitor 2.1. Evidence for a two-state transition. *Biochemistry*, **30**, 10428–10435.
 37. Elöve, G. A., Chaffotte, A. F., Roder, H. & Goldberg, M. E. (1992). Early steps in cytochrome c folding probed by time-resolved circular dichroism and fluorescence spectroscopy. *Biochemistry*, **31**, 6876–6883.
 38. Roder, H. & Colon, W. (1997). Kinetic role of early intermediates in protein folding. *Curr. Opin. Struct. Biol.* **7**, 15–28.
 39. Shastry, M. C. & Roder, H. (1998). Evidence for barrier-limited protein folding kinetics on the microsecond time scale. *Nature Struct. Biol.* **5**, 385–392.
 40. Chan, C. K., Hu, Y., Takahashi, S., Rousseau, D. L., Eaton, W. A. & Hofrichter, J. (1997). Submillisecond protein folding kinetics studied by ultrarapid mixing. *Proc. Natl Acad. Sci. USA*, **94**, 1779–1784.
 41. Scalley, M. L., Yi, Q., Gu, H., McCormack, A., Yates, J. R., 3rd & Baker, D. (1997). Kinetics of folding of the IgG binding domain of peptostreptococcal protein L. *Biochemistry*, **36**, 3373–3382.
 42. Sato, S., Luisi, D. L. & Raleigh, D. P. (2000). pH jump studies of the folding of the multidomain ribosomal protein L9: the structural organization of the N-terminal domain does not affect the anomalously slow folding of the C-terminal domain. *Biochemistry*, **39**, 4955–4962.
 43. Qin, Z., Ervin, J., Larios, E., Gruebele, M. & Kihara, H. (2002). Formation of a compact structured ensemble without fluorescence signature early during ubiquitin folding. *J. Phys. Chem. B*, **106**, 13040–13046.
 44. Qi, P. X., Sosnick, T. R. & Englander, S. W. (1998). The burst phase in ribonuclease A folding and solvent dependence of the unfolded state. *Nature Struct. Biol.* **5**, 882–884.
 45. Sosnick, T. R., Mayne, L. & Englander, S. W. (1996). Molecular collapse: the rate-limiting step in two-state cytochrome c folding. *Proteins: Struct. Funct. Genet.* **24**, 413–426.
 46. Sosnick, T. R., Shtilerman, M. D., Mayne, L. & Englander, S. W. (1997). Ultrafast signals in protein folding and the polypeptide contracted state. *Proc. Natl Acad. Sci. USA*, **94**, 8545–8550.
 47. Svergun, D. I. & Koch, M. H. J. (2003). Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.* **66**, 1735–1782.
 48. Millet, I. S., Doniach, S. & Plaxco, K. W. (2002). Toward a taxonomy of the denatured state: small angle scattering studies of unfolded proteins. *Advan. Protein Chem.* **62**, 241–262.
 49. Millet, I. S., Townsley, L. E., Chiti, F., Doniach, S. & Plaxco, K. W. (2002). Equilibrium collapse and the kinetic “foldability” of proteins. *Biochemistry*, **41**, 321–325.
 50. Sosnick, T. R. & Trewthella, J. (1992). Denatured states of ribonuclease A have compact dimensions and residual secondary structure. *Biochemistry*, **31**, 8329–8335.
 51. Flanagan, J. M., Kataoka, M., Shortle, D. &

- Engelman, D. M. (1992). Truncated staphylococcal nuclease is compact but disordered. *Proc. Natl Acad. Sci. USA*, **89**, 748–752.
52. Pollack, L., Tate, M. W., Darnton, N. C., Knight, J. B., Gruner, S. M., Eaton, W. A. & Austin, R. H. (1999). Compactness of the denatured state of a fast-folding protein measured by submillisecond small-angle X-ray scattering. *Proc. Natl Acad. Sci. USA*, **96**, 10115–10117.
53. Semisotnov, G. V., Kihara, H., Kotova, N. V., Kimura, K., Amemiya, Y., Wakabayashi, K. *et al.* (1996). Protein globularization during folding. A study by synchrotron small-angle X-ray scattering. *J. Mol. Biol.* **262**, 559–574.
54. Plaxco, K. W., Millett, I. S., Segel, D. J., Doniach, S. & Baker, D. (1999). Chain collapse can occur concomitantly with the rate-limiting step in protein folding. *Nature Struct. Biol.* **6**, 554–556.
55. Krantz, B. A. & Sosnick, T. R. (2000). Distinguishing between two-state and three-state models for ubiquitin folding. *Biochemistry*, **39**, 11696–11701.
56. Taddei, N., Chiti, F., Paoli, P., Fiaschi, T., Bucciantini, M., Stefani, M. *et al.* (1999). Thermodynamics and kinetics of folding of common-type acylphosphatase: comparison to the highly homologous muscle isoenzyme. *Biochemistry*, **38**, 2135–2142.
57. Krantz, B. A., Srivastava, A. K., Nauli, S., Baker, D., Sauer, R. T. & Sosnick, T. R. (2002). Understanding protein hydrogen bond formation with kinetic H/D amide isotope effects. *Nature Struct. Biol.* **9**, 458–463.
58. Vijay-Kumar, S., Bugg, C. E., Wilkinson, K. D., Vierstra, R. D., Hatfield, P. M. & Cook, W. J. (1987). Comparison of the three-dimensional structures of human, yeast, and oat ubiquitin. *J. Biol. Chem.* **262**, 6396–6399.
59. Thunnissen, M. M., Taddei, N., Liguri, G., Ramponi, G. & Nordlund, P. (1997). Crystal structure of common type acylphosphatase from bovine testis. *Structure*, **5**, 69–79.
60. Svergun, D. I., Richard, S., Koch, M. H., Sayers, Z., Kuprin, S. & Zaccai, G. (1998). Protein hydration in solution: experimental observation by X-ray and neutron scattering. *Proc. Natl Acad. Sci. USA*, **95**, 2267–2272.
61. Merzel, F. & Smith, J. C. (2002). Is the first hydration shell of lysozyme of higher density than bulk water? *Proc. Natl Acad. Sci. USA*, **99**, 5378–5383.
62. Khorasanizadeh, S., Peters, I. D. & Roder, H. (1996). Evidence for a 3-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nature Struct. Biol.* **3**, 193–205.
63. Khorasanizadeh, S., Peters, I. D., Butt, T. R. & Roder, H. (1993). Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry*, **32**, 7054–7063.
64. Gladwin, S. T. & Evans, P. A. (1996). Structure of very early protein folding intermediates: new insights through a variant of hydrogen exchange labelling. *Fold. Des.* **1**, 407–417.
65. Sreerama, N. & Woody, R. W. (2003). Structural composition of betaI- and betaII-proteins. *Protein Sci.* **12**, 384–388.
66. Hoshino, M., Hagihara, Y., Hamada, D., Kataoka, M. & Goto, Y. (1997). Trifluoroethanol-induced conformational transition of hen egg-white lysozyme studied by small-angle X-ray scattering. *FEBS Letters*, **416**, 72–76.
67. Noppert, A., Gast, K., Zirwer, D. & Damaschun, G. (1998). Initial hydrophobic collapse is not necessary for folding RNase A. *Fold. Des.* **3**, 213–221.
68. Flanagan, J. M., Kataoka, M., Fujisawa, T. & Engelman, D. M. (1993). Mutations can cause large changes in the conformation of a denatured protein. *Biochemistry*, **32**, 10359–10370.
69. Perl, D., Welker, C., Schindler, T., Schroder, K., Marahiel, M. A., Jaenicke, R. & Schmid, F. X. (1998). Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins. *Nature Struct. Biol.* **5**, 229–235.
70. Krantz, B. A., Moran, L. B., Kentsis, A. & Sosnick, T. R. (2000). D/H amide kinetic isotope effects reveal when hydrogen bonds form during protein folding. *Nature Struct. Biol.* **7**, 62–71.
71. Sosnick, T. R., Mayne, L., Hiller, R. & Englander, S. W. (1994). The barriers in protein folding. *Nature Struct. Biol.* **1**, 149–156.
72. Honig, B. & Yang, A. S. (1995). Free energy balance in protein folding. *Advan. Protein Chem.* **46**, 27–58.
73. Sosnick, T. R., Mayne, L., Hiller, R. & Englander, S. W. (1995). *Peptide and Protein Folding Workshop*, Philadelphia, PA.
74. Sosnick, T. R. & Englander, S. W. (1996). What limits protein folding. In *Dynamics and the Problem of Recognition in Biological Macromolecules* (Jardetzky, O. & Lefevre, J-F., eds), vol. 288, pp. 65–71, Plenum Press, New York.
75. Plaxco, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985–994.
76. Deniz, A. A., Laurence, T. A., Belligere, G. S., Dahan, M., Martin, A. B., Chemla, D. S. *et al.* (2000). Single-molecule protein folding: diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc. Natl Acad. Sci. USA*, **97**, 5179–5184.
77. Tan, E., Wilson, T. J., Nahas, M. K., Clegg, R. M., Lilley, D. M. & Ha, T. (2003). A four-way junction accelerates hairpin ribozyme folding *via* a discrete intermediate. *Proc. Natl Acad. Sci. USA*, **100**, 9308–9313.
78. Zitzewitz, J. A., Bilsel, O., Luo, J., Jones, B. E. & Matthews, C. R. (1995). Probing the folding mechanism of a leucine zipper peptide by stopped-flow circular dichroism spectroscopy. *Biochemistry*, **34**, 12812–12819.
79. Takei, J., Chu, R. A. & Bai, Y. (2000). Absence of stable intermediates on the folding pathway of barnase. *Proc. Natl Acad. Sci. USA*, **97**, 10796–10801.
80. Robertson, A. D. & Baldwin, R. L. (1991). Hydrogen exchange in thermally denatured ribonuclease A. *Biochemistry*, **30**, 9907–9914.
81. Smith, L. J., Bolin, K. A., Schwalbe, H., MacArthur, M. W., Thornton, J. M. & Dobson, C. M. (1996). Analysis of main chain torsion angles in proteins: prediction of NMR coupling constants for native and random coil conformations. *J. Mol. Biol.* **255**, 494–506.
82. Serrano, L. (1995). Comparison between the phi distribution of the amino acids in the protein database and NMR data indicates that amino acids have various phi propensities in the random coil conformation. *J. Mol. Biol.* **254**, 322–333.
83. Penkett, C. J., Redfield, C., Dodd, I., Hubbard, J., McBay, D. L., Mossakowska, D. E. *et al.* (1997). NMR analysis of main-chain conformational

- preferences in an unfolded fibronectin-binding protein. *J. Mol. Biol.* **274**, 152–159.
84. Swindells, M. B., MacArthur, M. W. & Thornton, J. M. (1995). Intrinsic phi, psi propensities of amino acids, derived from the coil regions of known structures. *Nature Struct. Biol.* **2**, 596–603.
 85. Avbelj, F. & Baldwin, R. L. (2003). Role of backbone solvation and electrostatics in generating preferred peptide backbone conformations: distributions of phi. *Proc. Natl Acad. Sci. USA*, **100**, 5742–5747.
 86. Shi, Z., Olson, C. A., Rose, G. D., Baldwin, R. L. & Kallenbach, N. R. (2002). Polyproline II structure in a sequence of seven alanine residues. *Proc. Natl Acad. Sci. USA*, **99**, 9190–9195.
 87. Shi, Z., Woody, R. W. & Kallenbach, N. R. (2002). Is polyproline II a major backbone conformation in unfolded proteins? *Advan. Protein Chem.* **62**, 163–240.
 88. Eker, F., Griebenow, K. & Schweitzer-Stenner, R. (2003). Stable conformations of tripeptides in aqueous solution studied by UV circular dichroism spectroscopy. *J. Am. Chem. Soc.* **125**, 8178–8185.
 89. Pappu, R. V. & Rose, G. D. (2002). A simple model for polyproline II structure in unfolded states of alanine-based peptides. *Protein Sci.* **10**, 2437–2455.
 90. Creamer, T. P. & Campbell, M. N. (2002). Determinants of the polyproline II helix from modeling studies. *Advan. Protein Chem.* **62**, 263–282.
 91. Dukor, R. K. & Keiderling, T. A. (1991). Reassessment of the random coil conformation: vibrational CD study of proline oligopeptides and related polypeptides. *Biopolymers*, **31**, 1747–1761.
 92. Krimm, S. & Tiffany, M. L. (1974). The circular dichroism spectrum and structure of unordered polypeptides and proteins. *Isr. J. Chem.* **12**, 189–200.
 93. Wilson, G., Hecht, L. & Barron, L. D. (1996). Residual structure in unfolded proteins revealed by Raman optical activity. *Biochemistry*, **35**, 12518–12525.
 94. Keiderling, T. A. & Xu, Q. (2002). Unfolded peptides and proteins studied with infrared absorption and vibrational circular dichroism spectra. *Advan. Protein Chem.* **62**, 111–161.
 95. Adzhubei, A. A. & Sternberg, M. J. (1994). Conservation of polyproline II helices in homologous proteins: implications for structure prediction by model building. *Protein Sci.* **3**, 2395–2410.
 96. Creamer, T. P. (1998). Left-handed polyproline II helix formation is (very) locally driven. *Proteins: Struct. Funct. Genet.* **33**, 218–226.
 97. Mezei, M., Fleming, P. J., Srinivasan, R. & Rose, G. D. (2004). Polyproline II helix is the preferred conformation for unfolded polyalanine in water. *Proteins: Struct. Funct. Genet.* In the press.
 98. Kentsis, A., Mezei, M., Gindin, T. & Osman, R. (2004). Unfolded state of polyalanine is a segmented polyproline II helix. *Proteins: Struct. Funct. Genet.* In the press.
 99. Garcia, A. E. (2004). Characterization of non-alpha helical conformations in Ala peptides. *Polymers*, **45**, 669–676.
 100. Chin, D. H., Woody, R. W., Rohl, C. A. & Baldwin, R. L. (2002). Circular dichroism spectra of short, fixed-nucleus alanine helices. *Proc. Natl Acad. Sci. USA*, **99**, 15416–15421.
 101. Kamatari, Y. O., Konno, T., Kataoka, M. & Akasaka, K. (1998). The methanol-induced transition and the expanded helical conformation in hen lysozyme. *Protein Sci.* **7**, 681–688.
 102. Kamatari, Y. O., Konno, T., Kataoka, M. & Akasaka, K. (1996). The methanol-induced globular and expanded denatured states of cytochrome c: a study by CD fluorescence, NMR and small-angle X-ray scattering. *J. Mol. Biol.* **259**, 512–523.
 103. Glatter, O. & Kratky, O. (1982). *Small Angle X-ray Scattering*, Academic Press, London.
 104. Matthews, C. R. (1987). Effects of point mutations on the folding of globular proteins. *Methods Enzymol.* **154**, 498–511.

Edited by C. R. Matthews

(Received 11 December 2003; received in revised form 23 February 2004; accepted 24 February 2004)