Early Collapse is not an Obligate Step in Protein Folding

Jaby Jacob¹,², Bryan Krantz¹, Robin S. Dothager¹, P. Thiyagarajan² and Tobin R. Sosnick¹,³*

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 overall dimensions of the denatured state. However, the removal of denaturant does subtly alter the distribution of backbone dihedral ψ,ϕ 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

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, m²-value, are equivalent to those calculated from kinetic measurements alone, then all the energy change
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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{47} The technique is particularly useful for the characterization of unfolded states which are not readily studied by high-resolution techniques.\textsuperscript{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.\textsuperscript{4,5,11,12,14,16,43,52,53}

Direct evidence for early collapse to a compact conformation has been obtained from SAXS measurements, albeit on multi-state proteins,\textsuperscript{5,10–15} that contain disulfide bonds\textsuperscript{16} or prosthetic groups.\textsuperscript{6,52} A compact intermediate was observed for ubiquitin (Ub) refolding in the presence of 45% ethylene glycol at \(-20^\circ\text{C}\).\textsuperscript{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).\textsuperscript{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\textsubscript{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)\textsuperscript{35,55} and common-type acylphosphatase (ctAcP, 98 aa).\textsuperscript{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 (PPI) 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 Ub\textsuperscript{58} and ctAcP\textsuperscript{59} have radii of gyration, $R_g = 13.9\ \text{Å}$ and $14.6\ \text{Å}$, which agree with their known values, $12.9\ \text{Å}$ and $13.9\ \text{Å}$, respectively. When the proteins unfold, their $R_g$ values increase to $~26\ \text{Å}$ and $~31\ \text{Å}$, 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 $\Delta G([\text{denaturant}]) = \Delta G^{\text{eq}} + m^* [\text{denaturant}]$ (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, $k_i < 1\ \text{s}^{-1}$, the native population never exceeds 0.5%.

The fastest folding rate of Ub is $120\ \text{s}^{-1}$ at 0.65 M Gdm. At this condition, ~25% of the molecules are estimated to have refolded within the dead-time. This fractional native population should result in a 2.5 Å reduction in the observed $R_g$, which is slightly greater than the ~1 Å 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.
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 reproducibility 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 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 \( Q_{\text{max}} \approx 1.3/R_g \). Measurements were carried out at 20°C in 50 mM Hapes, 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 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 3A.**
Global shape changes can be analyzed qualitatively by means of a Kratky plot (Figure 4). In this plot of $Q^2 I(Q)$ 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 are peaked while the denatured 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 downturn or plateau occurs at $Q < 0.13 \, \text{Å}^{-1}$, 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 ($\rho$) difference, or contrast, between the protein and solvent:\(^{47}\)

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

(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

<table>
<thead>
<tr>
<th></th>
<th>$m^\circ$ (kcal mol$^{-1}$ M$^{-1}$)</th>
<th>$C_m$ (M)</th>
<th>CD$_{222,\text{nm}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ub</td>
<td>1.74 ± 0.09</td>
<td>3.99 ± 0.02</td>
<td>1.63 ± 0.32</td>
</tr>
<tr>
<td>ctAcP</td>
<td>1.69 ± 0.36</td>
<td>4.16 ± 0.09</td>
<td>1.19 ± 0.25</td>
</tr>
</tbody>
</table>
extended chain, which has different electron density contrast as there is a larger hydration shell associated with the extended chain. 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. The equilibrium values for the free energy and surface burial are identical with the values obtained according to

\[
DG = DG^\ddagger_f - DG^\ddagger_u \quad \text{and} \quad m^e = m^u + m^f.
\]

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 at 222 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 at 222 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 signal is larger than what would be expected from a linear extrapolation (Figure 3A). At 0.5 M GdmCl, the burst CD signal is slightly larger than that of the native state. Nevertheless, the folding rate retains its denaturant dependence and the folding arm of

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 I₀ 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.
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 \( 1.30(\pm 0.02) \text{ kcal mol}^{-1} \text{ M}^{-1} \), which agrees with our previous value of \( 1.35(\pm 0.02) \text{ kcal mol}^{-1} \text{ M}^{-1} \).\(^{55}\)

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\(^\circ\)C, but not at 8\(^\circ\)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\(^\circ\)C, which inadvertently created a roll-over and a burst phase.

A time-resolved SAXS study conducted in 45% ethylene glycol at \(-20\)\(^\circ\)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\(^\circ\)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 PPII 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 results\(^{65}\) from its high PPII content (32%, 24% and 34% for \( \beta \), PPII, and \( \alpha \) 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 \( f \), \( c \) 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 (\( \Delta CD \sim 3000 \text{ deg cm}^2 \text{ dmol}^{-1} \)). 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 Ub\( _{6 \text{ M Gdm}} \) argue that the burst-phase CD signal represents the adjustment of the denatured state to the new solvent condition, \( D_{\text{physiological}} \) rather

---

**Figure 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.
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, CD$_{222}$ 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. More importantly, both proteins satisfy the chevron criterion $m^p = m_u + m_f$, 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.

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, both hen egg-white lysozyme (HEWL) 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. 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. Cytochrome c provides a particularly interesting case as this protein has a very linear chevron folding arm. Even so, the denatured state responds to a change in solvent conditions, with the $R_g$ decreasing from $\sim 30$ A in 3 M GdmCl to $\sim 21$ A in water, consistent with the increase in the FRET signal.

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

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 CD$_{200}$ nm while the magnitude of CD$_{222}$ 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.
protein upon transfer to 0.5 M GdmCl. 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. In addition, in the refolding of the csp from *Bacillus caldolyticus*, up to 70% of a FRET signal is quenched in the burst phase. 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 $\Phi, \Psi$ 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. 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 ($m^0 = m_g + m_l$), 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 $^2$H/H amide isotope effect experiments demonstrate that surface burial occurs commensurately with native hydrogen bond formation along the folding pathway. Stable hydrogen bonded secondary structure cannot be divorced from the unfolded state, 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.

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$^{-1}$. 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.

**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. 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. 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. Transiently populated intermediates manifest themselves as brief stops between the two levels. 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 $R_g$ the intermediate would have buried surface that
Lack of Early Collapse in Protein Folding

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 $m_t$ 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, although the base line sometimes has mild curvature.

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 value resulting from its high (24%) PPII content. For both Ub and cytochrome $c$, the thermally denatured state is a better approximation of $D_{physiological}$ 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. 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 $\phi$, $\psi$ angles in the helical basin of the Ramachandran plot does not necessarily involve any organized helical structure. Shifts to the $\beta$-basin can be considered and would produce the same qualitative result, but this basin is not significantly populated in the denatured state.

PPII has a positive CD signal, and is the dominant conformation in the denatured state.

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 $-10,000$ mdeg cm$^2$ dmol$^{-1}$ to $-20,000$ mdeg cm$^2$ dmol$^{-1}$, respectively (R. Woody, personal communication). A subtle 20% shift from the helical basin to the PPII basin, using values of $0$ mdeg cm$^2$ dmol$^{-1}$ and $-15,000$ mdeg cm$^2$ dmol$^{-1}$ for $[\theta]_{222}$PPII and $[\theta]_{222}$helix, respectively, would decrease the CD signal by $3000$ mdeg cm$^2$ 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, trifluoroethanol solutions.

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 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 × 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 ~2 m and energy of X-ray radiation was set to 12 keV.

Experiments used a Biologic SFM 400 stopped-flow apparatus 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 ~1.5 mg ml⁻¹. In acquiring the kinetic data, the unfolded protein (concentration ~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 ~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. Fl 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 μM and 69 μM 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 μM 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,109 Data are presented as the scattering intensity per solid angle, I(Q), where the scattering vector Q, is defined as Q = 4π sin θ/λ, λ is the X-ray wavelength, and θ is the half scattering angle. At low Q ≪ 1.3/Rg 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) = I₀ e⁻²π²Q²Rg². The Rg is the root-mean-square of the distances of all regions to the center-of-mass of the particle. Typically, the Rg is obtained from the slope of the low-angle region of a Guinier Plot of ln(I(Q)) versus Q².

To measure the Rg value accurately, scattering data must be taken across angles where the Gaussian approximation holds. Our reliable data starts at Q = 0.02 Å⁻¹, or QRg = 0.6 for the largest size (Rg = 30–32 Å for unfolded ctAcP). This low Qmin value, along with the linearity in the Guinier plot and the expected behavior of I₀, demonstrate that the Rg 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, dmax, of the object (e.g. the diameter). P(r) functions were calculated according to:

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

using the indirect Fourier inversion algorithms developed by Svergun.†

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

$$R_e = \sqrt{R_d^2 + \frac{R_u^2 - R_d^2}{1 + e^{(-m\Delta G_{eq}[den])/RT}}}$$  \hspace{1cm} (3)

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

$$CD = CD_u + S_u[den] + \frac{CD_d + S_d[den] - CD_u - S_u[den]}{1 + e^{(-m\Delta G_{eq}[den])/RT}}$$  \hspace{1cm} (4)

where $CD_d$ and $CD_u$ represent the CD signals of the folded and unfolded states in water, respectively, $S_d$ 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 constants \(^{10}\) where the standard free energy of folding, $\Delta G_f^m$, along with the standard activation free energy for folding $\Delta G_{a,f}^m$, and unfolding, $\Delta G_{u,f}^m$, are linearly dependent on denaturant concentration:

$$\Delta G_f^m([den]) = \Delta G_{eq}^{H_2O} + m\Delta G_{eq}[den] = -RT \ln(K_m)$$  \hspace{1cm} (5a)

$$\Delta G_{a,f}^m([den]) = -RT \ln(k_{f}^{H_2O}) + m\Delta G_{eq}[den] + \text{Constant}$$  \hspace{1cm} (5b)

$$\Delta G_{u,f}^m([den]) = -RT \ln(k_{u}^{H_2O}) - m\Delta G_{eq}[den] + \text{Constant}$$  \hspace{1cm} (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 kinetic and protein 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 $\Delta G_f^m = \Delta G_{a,f}^m - \Delta G_{u,f}^m$ and $m^2 = m_u + m_t$. Parameters were fit using non-linear least-squares algorithms implemented in the Microcal Origin software package.

Acknowledgements

We thank K. Plaxco, S. W. Engleander, 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


51. Flanagan, J. M., Kataoka, M., Shortle, D. &


Edited by C. R. Matthews

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