

From snapshot to movie: ϕ analysis of protein folding transition states taken one step further

Tomas Ternström*, Ugo Mayor**[‡], Mikael Akke[§], and Mikael Oliveberg*[¶]

[†]Department of Biochemistry, Umeå University, S-901 87 Umeå, Sweden; and Departments of ^{*}Biochemistry and [§]Physical Chemistry 2, Lund University, P.O. Box 124, S-221 00 Lund, Sweden

Communicated by Peter G. Wolynes, University of Illinois at Urbana–Champaign, Urbana, IL, October 25, 1999 (received for review June 30, 1999)

Kinetic anomalies in protein folding can result from changes of the kinetic ground states (D, I, and N), changes of the protein folding transition state, or both. The 102-residue protein U1A has a symmetrically curved chevron plot which seems to result mainly from changes of the transition state. At low concentrations of denaturant the transition state occurs early in the folding reaction, whereas at high denaturant concentration it moves close to the native structure. In this study we use this movement to follow continuously the formation and growth of U1A's folding nucleus by ϕ analysis. Although U1A's transition state structure is generally delocalized and displays a typical nucleation–condensation pattern, we can still resolve a sequence of folding events. However, these events are sufficiently coupled to start almost simultaneously throughout the transition state structure.

In the quest to understand how proteins fold, an increasing number of studies have been focused on structural characterization of the transition state ensemble—the kinetic bottleneck for folding (1–3). The results provide fascinating snapshots of the high-energy conformational states that have greatly improved our view of how proteins adopt their structure. Interestingly, it emerges that small two-state proteins do not fold in a strictly hierarchical manner but establish their interactions more or less simultaneously within a diffuse and global nucleus (4). An irresistible challenge now is to refine the details of this process. How does the nucleus form and consolidate? What gives rise to different nucleation patterns? Is there a single obligate nucleus (5), or could multiple nuclei act in concert (6)?

As the transition state (\ddagger) never accumulates, details about its structure have been inferred indirectly from the folding rate and protein engineering (7). By systematically truncating side chains while probing the effects on the kinetics, it is possible to map out detailed interaction patterns in \ddagger . In effect, mutations that destabilize \ddagger are said to target contacts in its structure. The strength of the contacts is measured by the phi value (ϕ^\ddagger), which normalizes the stability loss of \ddagger to that of the native protein. $\phi^\ddagger = 1$ indicates that the site of mutation is fully structured in the transition state, whereas $\phi^\ddagger = 0$ indicates that the site is unfolded. The first two-state protein to be characterized by this method was CI2 (8). The results revealed a new type of folding behavior, apparently inconsistent with earlier hierarchical models developed from studies of larger multistate proteins (9). CI2's \ddagger shows a global distribution of fractional ϕ^\ddagger s diffusely centered on three residues with high ϕ^\ddagger s in the hydrophobic core (8). The result suggests a large and delocalized nucleus where secondary and tertiary structure condenses concomitantly around a leading density in its center. The behavior was denoted nucleation condensation (4). Similar, delocalized nuclei of fractional ϕ^\ddagger s have later been observed for Arc repressor (10), CheY (11), and λ repressor (12). But there is some variation in the nucleation patterns. With the homologous and all- β structure SH3 domains from src (13) and α -spectrin (14), it is evident that one end of the protein forms earlier than the rest, giving rise to “polarized” \ddagger structures with distinguishable interfaces between ordered and disordered regions.

In this study we explore a different strategy to investigate further the delocalized nucleation process: we scan the folding events with a moving transition state by tuning the protein stability. This procedure allows us to follow continuously the development and consolidation of the folding nucleus of the two-state protein U1A (15). Despite U1A's delocalized nucleation pattern, a clear but somewhat overlapping sequence of folding events is resolved. The first ordered interactions appear between strands 2 and 3 and helix 1, forming part of the hydrophobic core. Subsequently, strand 1 wraps around the still-expanded half-core and, finally, the structure is closed up by the condensation of strand 4 and helix 2. The latter event could be interpreted either as a protrusion of the main nucleus or as a secondary nucleation event. Notably, the folding events are so tightly coupled that they start almost simultaneously throughout the protein, thus causing the transition state structure to appear overall diffuse.

Materials and Methods

Materials. Mutagenesis was done by standard procedures using the Quick-Change kit (Stratagene). Expression and purification of protein was done as in ref. 15. Buffer was 50 mM Mes at pH 6.3 (19 mM acid and 31 mM salt). Guanidinium chloride (GdmCl) was ultrapure from GIBCO/Life Technologies. All stopped-flow analysis was done at 298 K on a SX 18 MV instrument from Applied Photophysics (Surrey, U.K.), and curve fitting was done with the KALEIDAGRAPH software (Abelbeck Software, Reading, PA). For experimental procedures see *Results*.

NMR. The ¹⁵N-labeled samples contained 1.0–1.2 mM protein dissolved in 600 μ l of 5 mM sodium acetate buffer (90% H₂O/10% D₂O) at pH 4.8/50 mM NaCl/80 μ M 2,2-dimethyl-5-silapentane-5-sulfonate (DSS). Experiments were done on 500- and 600-MHz Varian UNITY Inova spectrometers at ¹H Larmor frequencies of 499.78 and 599.89 MHz, respectively. A three-dimensional ¹⁵N total correlation spectroscopy–heteronuclear single quantum correlation (TOCSY–HSQC) experiment (16) using DIPSI-rc mixing (17, 18) was acquired at 600 MHz with spectral widths of 6601 Hz in ω_1 and ω_3 , and 1860 Hz in ω_2 , sampled over 128, 1024, and 48 complex points, respectively. ¹⁵N–¹H HSQC (19) spectra were acquired by using sensitivity-enhanced pulsed field gradient coherence selection (20–22) and GARP-1 (23) ¹⁵N decoupling during acquisition. The sample was dissolved in ice-cold D₂O and immediately placed at 303 K in the spectrometer, which had been tuned and calibrated with a sample of identical composition. Processing and analysis of the spectra were by FELIX97 (MSI, San Diego).

Abbreviation: GdmCl, guanidinium chloride.

[‡]Present address: Cambridge Centre for Protein Engineering, Hills Road, Cambridge CB2 2QH, U.K.

[¶]To whom reprint requests should be addressed. E-mail: mikael.oliveberg@chem.umu.se.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

^{15}N - ^1H HSQC spectra were processed with exponential and cosine apodization functions in ω_2 and ω_1 , respectively. Final size of the matrix was typically 2048×256 real points after zero-filling and Fourier transformation. Crosspeak intensities were evaluated as peak heights and decays were evaluated by standard procedures (24).

Results

Curved Chevron Plots and Transition State Movements. The gross appearance of the transition state (\ddagger) is often inferred from mass action—i.e., from the GdmCl dependence of the activation process for refolding ($\text{D} \rightleftharpoons \ddagger$) and unfolding ($\text{N} \rightleftharpoons \ddagger$). The GdmCl response is measured by the kinetic m values

$$m_u = \partial \log k_u / \partial [\text{GdmCl}] \text{ and } m_f = \partial \log k_f / \partial [\text{GdmCl}], \quad [1]$$

where k_u and k_f are the unfolding and refolding rate constants, respectively. Often, the m values are constants, implying a proportionality between m values and changes in the protein's solvent-accessible surface area (25). This proportionality is also evident from the good agreement between m values and changes in heat capacity (26, 27). For two-state folding where $K_{\text{D-N}} = k_u/k_f$ and consequently

$$m_u - m_f = m_{\text{D-N}} = \partial \log K_{\text{D-N}} / \partial [\text{GdmCl}]. \quad [2]$$

It is thus possible to assign the position of \ddagger in relation to D and N according to

$$\beta^\ddagger = -m_f/m_{\text{D-N}} = 1 - m_u/m_{\text{D-N}}, \quad [3]$$

where β^\ddagger is a normalized measure of how the stability of \ddagger changes upon addition of denaturant (25, 28). A more structural interpretation of β^\ddagger is the solvent exposure of \ddagger relative to that of N. Consistently, β^\ddagger may also be inferred from transient heat capacity changes (27) and transient pK_a shifts (29).

Experimentally, folding kinetics are evaluated from so-called chevron plots ($\log k$ vs. $[\text{GdmCl}]$) (Fig. 1). For two-state proteins these are characteristically V-shaped with fixed values of m_f , m_u , and β^\ddagger (3, 30). For U1A, however, the chevron plot is symmetrically curved, showing that the m values vary with $[\text{GdmCl}]$; m_f grows larger while m_u decreases (Fig. 1). Notably, the increase in m_f is precisely coupled to the decrease in m_u , maintaining a constant value of $m_{\text{D-N}}$ (Eq. 2). This suggests that U1A displays two-state kinetics at all $[\text{GdmCl}]$, and that the curvatures result from changes in β^\ddagger . That is, the curvature results from changes of U1A's transition state; at high concentrations of GdmCl, \ddagger moves closer to N, resulting in a smaller structural change upon activation $\text{N} \rightleftharpoons \ddagger$ and thus a lower value of m_u (31). A measure of the \ddagger movement is given by Eq. 3, and β^\ddagger ranges from ~ 0.2 at 0.45 M GdmCl to ~ 0.9 at 8 M GdmCl (Fig. 1). Analogous movements have been observed for barnase (32), Arc repressor (10), and CI2 (33) and have recently been induced by mutations in CI2 and S6 (34). The behavior is also suggested by theory and simulations in which the kinetic bottleneck changes location in the folding funnel upon changing of the stability (35).

An alternative explanation for curved unfolding limbs would be that strong GdmCl solutions are not ideal (25, 36). Although plots of $\log k_u$ vs. GdmCl activity yield reasonably straight lines, they produce at the same time upward curvatures in proteins with V-shaped chevrons (34). Such upward curvatures are not easily rationalized by either \ddagger movements or the existence of intermediates. Further, the curvature of U1A's unfolding limb is seen also with the noncharged denaturant urea (data not shown). A third explanation is unfolding intermediates (37)—i.e., the curvatures arise from partial disruption of N prior to global unfolding (38). In the case of U1A, however, we disfavor this scenario also, because the structure in question shows high H/D protection factors (see below). Remember that identical curva-

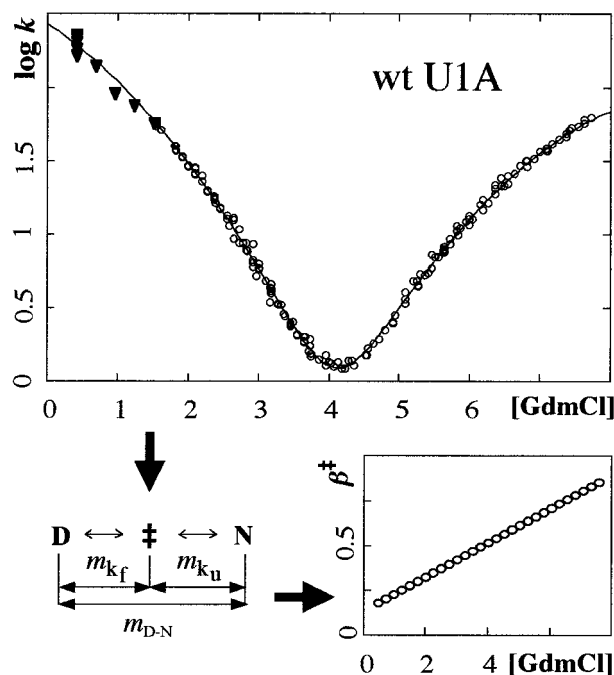


Fig. 1. The symmetrically curved chevron plot of U1A reveals a progressive change of the kinetic m values. The phenomenon is here ascribed to transition state movements along β^\ddagger (Eqs. 1–3). Data below 1.5 M GdmCl were excluded from the fits to eliminate problems with transient aggregation (\blacktriangledown).

ture is seen also on the refolding side, which means that any unfolding intermediate needs to be accompanied by a symmetrically located refolding intermediate. Together with U1A's two-state characteristics, a moving transition state appears to be the simpler explanation. And perhaps most important: moving transition states provide an alternative and unexplored way of accounting for anomalous folding kinetics. For an extensive analysis of U1A's two-state behavior, including D- and N-like dead-time spectra and coinciding GdmCl titrations with CD and fluorescence, see ref. 39.

Construction of the Free-Energy Profile. The movement of \ddagger was used to reconstruct the shape of the folding free-energy profile as follows (30, 31, 33, 34) (Fig. 2).

Step 1. To account for the curvatures, $\log k_f$ and $\log k_u$ were fitted by second-order polynomials the derivatives of which yield the kinetic m values—i.e., $m_f = b_f + 2c_f[\text{GdmCl}]$ and $m_u = b_u + 2c_u[\text{GdmCl}]$ (Eq. 1, Table 1). β^\ddagger was then calculated from Eq. 3, where $m_{\text{D-N}}$ was derived either kinetically from m_f and m_u (Eq. 2) or estimated independently from equilibrium unfolding, and the chevron data were replotted against β^\ddagger (Fig. 2). Data below 1.5 M GdmCl were excluded from the fits to eliminate perturbations from transient aggregation (40).

Step 2. The barrier height ($\text{D} \rightleftharpoons \ddagger$) at each point of β^\ddagger was approximated as $\Delta G^\ddagger = 2.3RT(6 - \log k_f)$, where the pre-factor of 10^6 s^{-1} is from ref. 41. In the unfolding region k_f was obtained from k_u and Eq. 3. Note that each point on this curve is obtained at a different $[\text{GdmCl}]$: early parts are linked to rate constants obtained at low $[\text{GdmCl}]$ in the refolding region, whereas late parts are linked to high $[\text{GdmCl}]$ in the unfolding region. A clearer representation of the free-energy profile is obtained if all points are extrapolated to a common $[\text{GdmCl}]$. This is done next by linear free-energy extrapolations.

Step 3. As β^\ddagger is simply the normalized sensitivity to GdmCl (Eq. 3), each value of ΔG^\ddagger can be extrapolated to a common $[\text{GdmCl}]$ (Y) by standard linear free-energy relations: $\Delta G^\ddagger(\beta^\ddagger,$

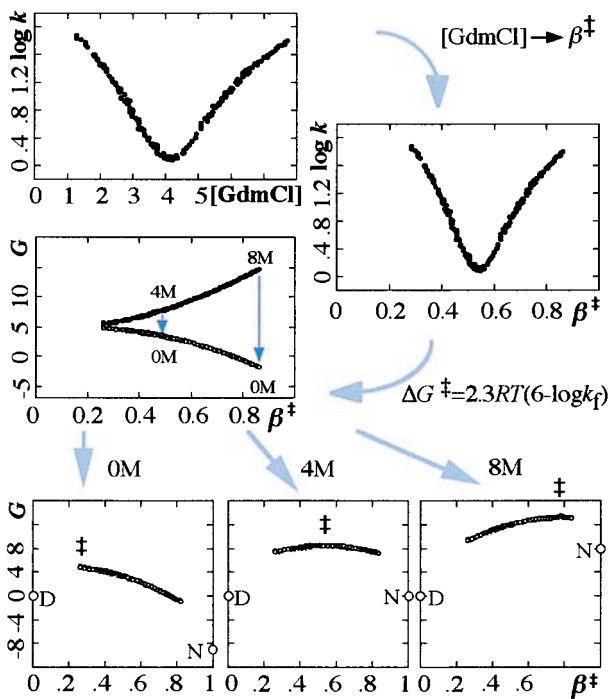


Fig. 2. Derivation of barrier shape from chevron plot. Units are in s^{-1} and $2.3RT$. See steps 1–3 in *Results*. The broad and level barrier profile cause its highest point (\ddagger) to move upon destabilization with GdmCl.

$[Y] = \Delta G^\ddagger(\beta^\ddagger, [X]) + \beta^\ddagger m_{D-N} 2.3RT([Y] - [X])$, where $[X]$ is the GdmCl concentration at which $\log k_f$ was measured. The bottom panels in Fig. 2 show the resulting free-energy profiles at 0, 4, and 8 M GdmCl, and illustrate how \ddagger moves along the top of a broad barrier as $[GdmCl]$ is increased. The key assumption behind the barrier construction is thus that folding proceeds by the same events at all $[GdmCl]$. Consistent motions of the transition state over broad barriers are predicted independently by theoretical models both in the mean-field and capillarity limit (42, 43). Note also that the broad barrier representation provides

a simple rationale for Hammond postulate behavior in protein folding (30, 32).

From Snapshot to Movie. Transition state movements open the possibility to continuously follow the development of interactions along β^\ddagger . We do this by ϕ^\ddagger analysis at each point along the barrier top. To improve the resolution we constrained the system to ideal two-state behavior by the following measures, all of which fall within the experimental error: (i) the transition midpoint and m_{D-N} are derived kinetically, which yields higher precision than equilibrium measurements; (ii) the kinetic curvatures are assumed to be caused by \ddagger changes alone; and (iii) on the basis of data where both limbs of the chevron plot are clearly resolved we have fixed the quadratic terms in the polynomial fits to the same value, i.e. $c_f = c_u$. Data for 10 typical mutants are shown in Fig. 3. Barrier profiles were obtained as for wild type.

For calculation of ϕ vs. β^\ddagger , the wild-type profile ($\Delta G^{\ddagger,wt}(\beta^\ddagger)$) was subtracted from that of the mutant ($\Delta G^{\ddagger,mut}(\beta^\ddagger)$) (Fig. 4), and the difference was divided by the destabilization of N upon mutation ($\Delta\Delta G_{D-N}$)

$$\phi(\beta^\ddagger) = \frac{(\Delta G^{\ddagger,wt}(\beta^\ddagger) - \Delta G^{\ddagger,mut}(\beta^\ddagger))}{(\Delta G_{D-N}^{wt} - \Delta G_{D-N}^{mut})} = \frac{\Delta\Delta G^\ddagger(\beta^\ddagger)}{\Delta\Delta G_{D-N}}, \quad [4]$$

which yields a ϕ^\ddagger at each point of β^\ddagger where the barrier profiles overlap. A series of snapshots are combined to a graph which shows how the protein consolidates at transition state level. The resulting $\phi(\beta^\ddagger)$ graphs are shown in Fig. 5, and their values at $\beta^\ddagger = 0.3$, $\beta^\ddagger = 0.5$, $\beta^\ddagger = 0.7$, $\beta^\ddagger = 0.85$ are listed in Table 1. The graphs are further organized into four groups and color-coded blue, green, yellow, and red according to how early the interactions appear. Note that mutations shift the position of the transition state in the same way as GdmCl. This means that some mutants never expose the earliest parts of the barrier, which prevents calculation of $\phi(\beta^\ddagger)$ at low β^\ddagger , e.g., I40A and F34A.

Errors and Constraints of the Model. The largest errors are in the kinetic m values; the trick is to determine the derivatives without biasing the data. The minimal solution is to use a fixed \ddagger and determine the m values from linear fits (30). Here, we go one step further by assuming a linear dependence between β^\ddagger and

Table 1. Parameters from curve fits to chevron data: $\log k_{obs} = \log(k_f + k_u) = \log(10^{\log k_f} + 10^{\log k_u})$, where k_{obs} is the observed rate constant, and $\log k_f$ and $\log k_u$ were replaced by second-order polynomials $\log k_f = a_f - b_f[GdmCl] - c_f[GdmCl]^2$ and $\log k_u = a_u + b_u[GdmCl] - c_u[GdmCl]^2$

Mutation	Deleted contacts	MP	m_{D-N}	$\Delta\Delta G_{D-N}/2.3RT$	a_f	b_f	a_u	b_u	$c_u = c_f$	ϕ			
										$\beta^\ddagger = 0.3$	$\beta^\ddagger = 0.5$	$\beta^\ddagger = 0.7$	$\beta^\ddagger = 0.85$
Wild type		4.1	1.82	0	2.4 ± 0.0	-0.31 ± 0.02	-5.0 ± 0.1	1.51 ± 0.03	-0.08				
I12V, $\beta 1$	V57 $\beta 1$, A65 $\alpha 1$	3.4	1.86	1.2	2.4 ± 0.0	-0.29 ± 0.03	-4.0 ± 0.1	1.56 ± 0.04	-0.09	0.0	0.1	0.25	0.4
I14A, $\beta 1$	L17 $\beta 1$	2.2	1.81	3.3	2.5 ± 0.1	-0.88 ± 0.06	-1.5 ± 0.0	0.93 ± 0.01	-0.06	—	0.35	0.6	0.75
L17A, $\beta 1$	I14 $\beta 1$, L26 $\alpha 1$	2.6	1.80	2.7	2.6 ± 0.1	-0.56 ± 0.04	-2.1 ± 0.1	1.24 ± 0.04	-0.09	—	0.2	0.45	0.65
L26A, $\alpha 1$	L17 $\beta 1$, I21, L30 $\alpha 1$, Y78	3.1	1.69	1.7	2.0 ± 0.0	-0.38 ± 0.02	-3.3 ± 0.1	1.31 ± 0.04	-0.08	0.1	0.35	0.6	0.8
L30A, $\alpha 1$	L26 $\alpha 1$, K27 $\alpha 1$, I43 $\beta 2$	2.0	1.81	3.8	2.2 ± 0.1	-0.95 ± 0.07	-1.3 ± 0.1	0.86 ± 0.03	-0.05	—	(0.4)	0.65	0.85
F34A, $\alpha 1$	L30, I33, V57 $\beta 3$, M72 $\alpha 2$	2.1	1.55	3.3	2.3 ± 0.0	-0.89 ± 0.04	-1.0 ± 0.1	0.66 ± 0.04	-0.03	—	(0.4)	0.65	0.85
I40A, $\beta 2$	F34 $\alpha 1$, S35 $\alpha 1$, I43 $\beta 2$, V57 $\beta 3$, F59 $\beta 3$	2.0	1.74	3.8	2.5 ± 0.1	-1.20 ± 0.07	-0.9 ± 0.0	0.54 ± 0.02	-0.03	—	(0.4)	0.75	0.95
I43V, $\beta 2$	Y31 $\alpha 1$	3.9	1.86	0.4	2.3 ± 0.0	-0.43 ± 0.03	-4.9 ± 0.1	1.43 ± 0.05	-0.08	1.0	1.2	1.3	1.5
V45A, $\beta 2$	A55 $\beta 3$	3.1	1.71	1.7	1.9 ± 0.0	-0.51 ± 0.03	-3.5 ± 0.1	1.20 ± 0.05	-0.07	0.35	0.6	0.9	1.1
I58A, $\beta 3$	H10 $\beta 1$, L44 $\beta 2$, I94 $\alpha 3$	3.5	1.55	1.0	2.0 ± 0.0	-0.71 ± 0.03	-3.4 ± 0.1	0.84 ± 0.06	-0.03	—	—	—	—
L69A, $\alpha 2$	R70 $\alpha 2$, Q73, Q85 $\beta 4$	3.1	1.56	1.7	2.8 ± 0.1	-0.62 ± 0.06	-2.0 ± 0.2	0.94 ± 0.08	-0.05	<0	0.0	0.5	0.7
I84V, $\beta 4$	Q73	3.6	1.92	1.0	2.3 ± 0.1	-0.16 ± 0.06	-4.6 ± 0.3	1.76 ± 0.10	-0.11	<0	0.0	0.2	0.4
I84A, $\beta 4$	I12 $\beta 1$, A68 $\alpha 2$, Q73	2.7	1.40	2.2	2.4 ± 0.0	-0.37 ± 0.03	-1.4 ± 0.1	1.03 ± 0.03	-0.07	<0	<0	0.25	0.5

MP, GdmCl concentration (M) at the transition midpoint obtained from the intersection of the fitted polynomials; errors $< \pm 0.04$. m_{D-N} , Equilibrium m value derived from kinetic fits—i.e., $b_f + b_u$ (Eqs. 1 and 2). $\Delta\Delta G_{D-N}/2.3RT = \log K_{D-N}$, stability loss upon mutation derived from kinetic fits at $(MP_{wt} + MP_{mut})/2$, assuming two-state behavior $K_{D-N} = k_u/k_f$. $c_u = c_f$. Errors $< \pm 0.01$. ϕ , parentheses indicate that the values are slightly extrapolated—cf. graphs in Fig. 5.

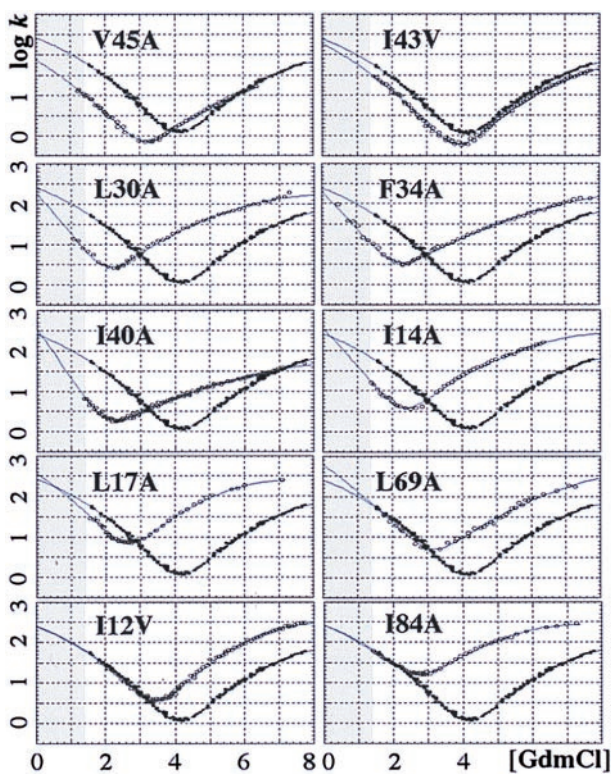


Fig. 3. Chevron plots for the U1A mutants (○) superimposed on wild type data (●).

protein stability. This allows us to obtain m from second-order polynomials. Higher-order polynomials do not significantly change the results; neither do linear fits to “running windows.” The overall features of $\phi(\beta^\ddagger)$, however, may be verified directly from Fig. 3: the extreme of $\phi(\beta^\ddagger) = 1$ offsets only the refolding limb, whereas $\phi(\beta^\ddagger) = 0$ offsets only the unfolding limb. Note further that fractional $\phi(\beta^\ddagger)$ need not always have a mechanistic origin. (i) Reorganizations around the mutation site could underestimate the contact strength. Although such rearrangements are rare in N they could be significant in less ordered transition states. (ii) Statistical spread of \ddagger along the top of shallow barriers may smooth possible steps in $\phi(\beta^\ddagger)$.

Amide Proton Exchange. Fig. 5 shows the amide proton exchange rates color-coded onto the mean coordinates of U1A’s solution structure (44). The measured exchange rates span four orders of magnitude, with values of $-\log k_{ex}$ in the range 2.7–7.4. Several more residues showed observable decays, but these were either too fast or too slow to yield precise measurements. The slowly exchanging residues are those that form early by the $\phi(\beta^\ddagger)$ graphs, implying a connection between equilibrium fluctuations of the native state and the folding reaction (45).

Discussion

The High-Energy Nucleation of U1A. The first ordered contacts in the folding reaction of U1A appear in the hydrophobic core. Nucleation start around I43 and V45 in β strand 2, which are the only residues with clear contacts at $\beta^\ddagger = 0.3$ (Fig. 5, Table 1). Since I43V deletes mainly contacts with helix 1 (L30 and Y31), and V45A deletes an interaction in strand 3 only (A55), this suggests that helix 1 and strand 3 are also involved in the early nucleus. Further evidence for early organization of strand 3 comes from a 10-fold retardation of k_f by mutation I58A. Strands 1 and 4 and helix 2, on the other hand, seem completely unfolded

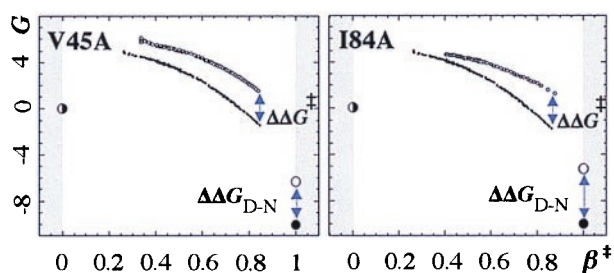


Fig. 4. Barrier profiles for calculation of ϕ^\ddagger graphs (Eq. 4). Units are in $2.3RT$.

at this stage. Next follows an overall consolidation of the hydrophobic interactions linking helix 1 to strands 2 and 3 (L30, F34, I30), and strand 1 begins to close in at the periphery (I14, L17, L26). At $\beta^\ddagger = 0.5$, the folding nucleus has captured strands 2, 3, and 1 and helix 1, whereas the lobe comprising strand 4 and helix 2 is still missing. Despite the expanded nature of the nucleus, this leads to a pronounced polarization of the structure toward the side of helix 1. So far, the transition state structure contains only two sites with ϕ^\ddagger above 0.5, the leading densities at I43 and V45. The exceedingly high $\phi(\beta^\ddagger)$ of V45A implies some degree of nonnative contacts around this site, perhaps arising from less specific collapse of surrounding moieties (35). Overall, however, U1A’s transition state shows a typical nucleation condensation pattern at this point. Finally, at $\beta^\ddagger = 0.7$, the lobe comprising strand 4 and helix 2 has begun to condense at the loose end of the core (I12, L69, I84) and the entire protein is encompassed by the nucleus. Note, however, that the condensation of strand 4 and helix 2 does not appear to propagate from the main nucleus but seems to start on its own around L69. This could show that a secondary nucleation takes place in connection to the main nucleus (6). After this point, the transition state structure undergoes a general consolidation. At $\beta^\ddagger = 0.85$ most of our probes have reached ϕ^\ddagger of 0.8 or more. The condensation of strand 4 and helix 2 remains centered at L69.

A more minimalistic interpretation of $\phi(\beta^\ddagger)$ is how the size of the critical nucleus (5) grows with increasing [GdmCl]. This is because additional interactions are required to turn folding downhill under conditions of low contact energies. The broad barrier view simply assumes that these states are on the folding route at all [GdmCl].

Unfolding Intermediates. As mentioned above, curved unfolding limbs have also been observed in connection with unfolding intermediates (38). In this case, mutants with reduced m_u are those that increase the partial disruption of N—i.e., where \ddagger and the refolding ground state move closer together. Such mutations are located at the interface connecting strands 2 and 3 with helix 1 (Fig. 5). Strands 2 and 3 would thus form a labile flap which folds out from the core prior to global unfolding. This seems inconsistent with the high H/D protection factors found in this region (37). Perhaps molecular dynamics simulations could shed further light on this issue.

Delocalized Versus Condensed Nucleation Patterns. The limits of delocalized and distinct nucleation patterns have been treated theoretically in connection with the capillarity picture of protein folding (43, 46, 47). The capillarity view is based on the nucleation theory of first-order phase transitions and assumes that when a protein is sufficiently large some of the side chains become too distant to interact with one another at any time. Folding then shifts from “mean-field” behavior to a more nucleation-growth-like scenario with distinguishable interfaces between structured and unfolded regions of the polypeptide—cf. nucleation and growth of a droplet (43, 48). Even so, the

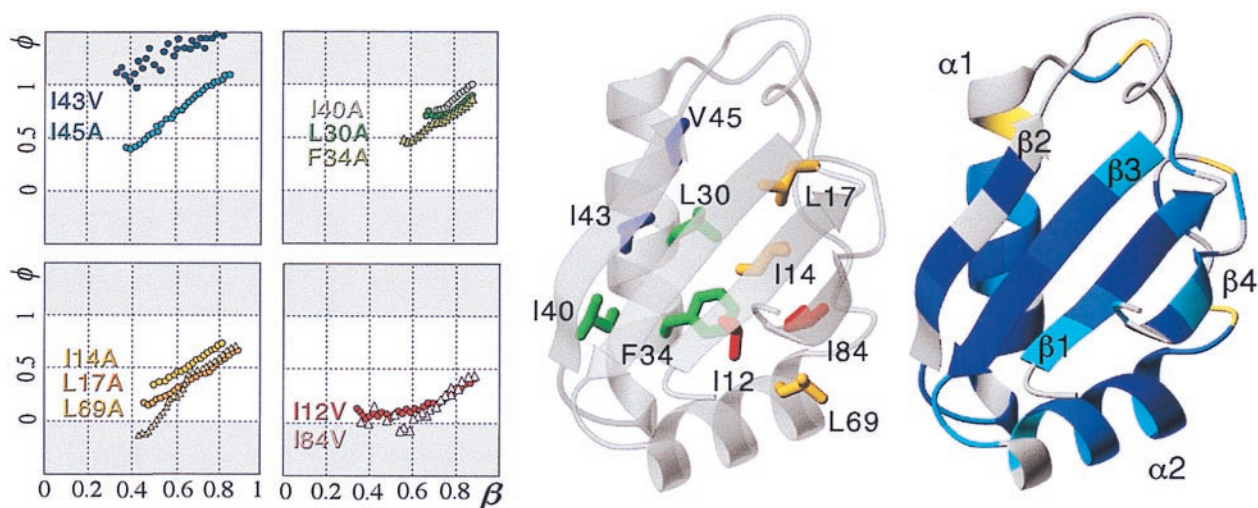


Fig. 5. (Left) ϕ^+ graphs for U1A's transition state. (Center) Location of the mutations in the native structure. (Right) H/D exchange rates color-coded onto the NMR solution structure: $-8 < \log k_{ex} < -3$ (blue), $-3 < \log k_{ex} < -2$ (cyan), $-2 < \log k_{ex}$ (yellow), fast or missing (gray).

capillarity interface need not be as sharp as in a droplet, but could be broad and span over several structural phases depending on the heterogeneity and vectorial distribution of the contact forces (43). For a small protein like U1A, which might be close to the size of a folding nucleus, the interface is expected to diminish (43).

It is thus interesting that the transition state ensemble of U1A still reveals many features consistent with capillarity folding. The interface, however, is very diffuse. We have chosen three structural thresholds to show the propagation of this interface at increasing values of β^+ : fully native contacts ($\phi(\beta^+) = 1$), more than half formed ($\phi(\beta^+) > 0.5$), and noncoil ($\phi(\beta^+) > 0$). Fig. 6 shows the fraction of residues (f) within each of these thresholds as a function of β^+ . Since the structure grows approximately radially from the initiation site (Fig. 5) and f reflects the fraction of structure involved in ordered interactions, the graphs in Fig. 6 provide a minimalistic view of how the structural interface develops and propagates throughout the protein matrix. In the absence of denaturant ($\beta^+ = 0.2$), the critical nucleus (5) consists of only a few fully formed interactions at the initiation site around I43. These first interactions unleash a wave of order through the polypeptide, seen as a steep rise in the number of nonzero ϕ values (Fig. 6). We denote this partial organization “fluctuating order,” and its front constitutes the leading edge of U1A's broad structural interface. At $\beta^+ = 0.8$, the fluctuating order encompasses the entire transition state structure. The final establishment of full native contacts constitute the trailing end of the interface. As full native contacts develop rather late in folding, the interface shows a maximum width at $\beta^+ = 0.5$. As U1A's contact pattern in this point is very similar to that of CI2

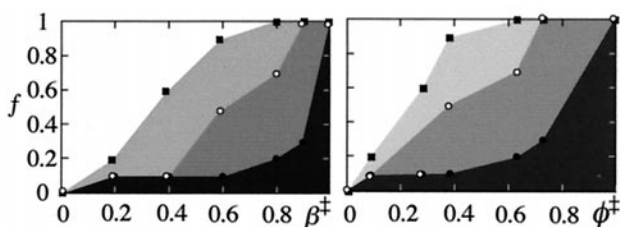


Fig. 6. Development of structure in U1A's transition state versus the empirical reaction coordinate β^+ and average ϕ . Light gray, gray, and black show the fraction of residues (f) with ϕ^+ above 0, 0.5, and 1, respectively.

(8), it is tempting to speculate that nucleation condensation is the experimental manifestation of capillarity behavior near the mean-field limit.

The interface view allows further a simple rationalization of the processes behind delocalized and condensed nucleation patterns. It may be based on only two experimental parameters: the abruptness of local collapse ($C_{local} = \partial\phi(\beta^+)/\partial\beta^+$) and the expansion of the interface ($E_{int} = \partial f/\partial\beta^+$). Both parameters are related to β^+ and given by $\phi(\beta^+)$. If $C_{local} \gg E_{int}$, the local collapse is “faster” than the expansion of order and the nucleus shows a sharp interface [cf. nucleation growth (49)], whereas if $C_{local} \ll E_{int}$ the fluctuating order captures the whole polypeptide before the contacts condense, and the structure appears diffuse. For U1A, the values of C_{local} and E_{int} are all between 1 and 2, which might be typical for nucleation condensation patterns. Proteins accumulating well defined intermediates would be biased toward *local collapse*. Note, however, that U1A's two initiation sites (I43 and L69) seem to collapse more abruptly than the rest of the structure. Perhaps this abrupt collapse is typical for interactions leading the organization of the polypeptide and the hallmark of conformationally restricted nucleation in mesophilic systems.

Chain stiffness and helical propensities may thus affect the nucleation pattern by better transducing the chain order and, at the same time, resist collapse of the contacts. Additional connectivity introduced by partly formed helices, or even slight

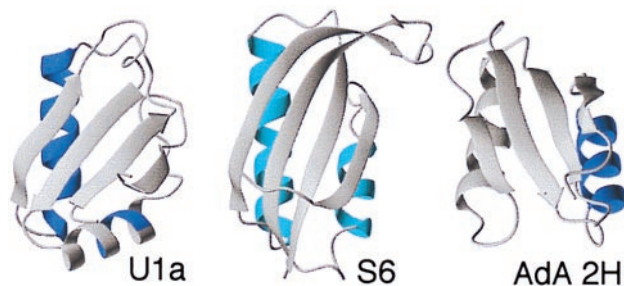


Fig. 7. Structures of U1A, S6, and AdA 2H, indicating in blue the regions where folding starts according to ϕ^+ analysis of the transition state ensemble. It appears that the split β - α - β fold contains two nucleation sites in connection with either helix. U1A nucleates mainly in helix 1 but shows also a secondary nucleation in helix 2. S6 nucleates diffusely in both sites (parallel pathways?), whereas AdA 2H nucleates in helix 2.

helical propensity, links the structure effectively and prevents sharp interfaces to the nucleus—i.e., stiff segments “bend the nucleus apart” (2).

Nucleation and Twist. Interestingly, a structural analogue of U1A, the human procarboxypeptidase A2 (AdA 2H), show a folding nucleus in the opposite part of the structure, namely in connection to helix 2 (50) (Fig. 7). U1A’s nucleation site between strands 2 and 3 and helix 1 is largely unstructured in the transition state of AdA 2H. It is tempting to speculate that the difference arises from the topology of the two structures: U1A and AdA 2H are twisted in opposite directions (Fig. 7). Furthermore, the secondary structure of AdA 2H is arranged in a more parallel fashion than in U1A. The common fold may thus possess two possible nucleation sites whose relative manifestation are determined by the detailed packing and orientation of the secondary structure. The twist of U1A favors the helix 1 site, whereas the straighter arrangement of AdA 2h favors the helix 2 site. Notably, the presence of an alternative nucleus is also

hinted at by U1A’s seemingly independent condensation of helix 2 and strand 4. Interestingly, a third structural analog with an intermediate twist, the ribosomal protein S6 (51), shows a more uniform nucleation pattern with fractional ϕ values at both the helix 1 and helix 2 site (Danie Otzen and M.O., unpublished results). This could show that S6 has equal propensities to nucleate at either site, and that folding occurs by two parallel routes. A second character which also follows the structural twist is the tendency to display transition state movements: U1A shows large movements (31), AdA 2H shows a fixed transition state (50), whereas S6 displays both moving and fixed transition states depending on mutation and experimental conditions (34).

We thank Eugene Shakhnovich, Håkan Wennerström, and Peter Wolynes for valuable discussions, Crafoordska Stiftelsen for donations to instruments, Andreas Muranyi for NMR assistance, and the Swedish NMR Centre and the groups of Martin Billeter and Göran Karlsson for allocation of NMR time. T.T. was supported by a Lawski grant, and M.A. and M.O. by the Swedish Natural Science Research Council.

- Fersht, A. R. (1995) *Curr. Opin. Struct. Biol.* **5**, 79–84.
- Gruebele, M. & Wolynes, P. G. (1998) *Nat. Struct. Biol.* **5**, 662–665.
- Jackson, S. E. (1998) *Fold. Des.* **3**, R81–R91.
- Fersht, A. R. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 10869–10873.
- Shakhnovich, E. (1998) *Fold. Des.* **3**, R108–R111.
- Thirumalai, D. & Klimov, D. K. (1998) *Fold. Des.* **3**, 112–118.
- Matouschek, A., Kellis, J. T., Jr., Serrano, L. & Fersht, A. R. (1989) *Nature (London)* **342**, 122–126.
- Itzhaki, L. S., Daniel, E. O. & Fersht, A. R. (1995) *J. Mol. Biol.* **254**, 260–288.
- Kim, P. S. & Baldwin, R. L. (1990) *Annu. Rev. Biochem.* **59**, 631–660.
- Milla, M. E., Brown, B. M., Waldburger, C. D. & Sauer, R. T. (1995) *Biochemistry* **34**, 13914–13919.
- Lopez-Hernandes, I. & Serrano, L. (1996) *Fold. Des.* **2**, 43–55.
- Burton, R. E., Huang, G. S., Daugherty, M. A., Calderone, T. L. & Oas, T. G. (1997) *Nat. Struct. Biol.* **4**, 305–310.
- Grantcharova, V. P., Riddle, D. S. & Baker, D. (1998) *Nat. Struct. Biol.* **5**, 714–720.
- Martinez, J. C., Pisabarro, T. M. & Serrano, L. (1998) *Nat. Struct. Biol.* **5**, 721–729.
- Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H. & Nagai, K. (1994) *Nature (London)* **372**, 432–438.
- Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A. & Bax, A. (1989) *J. Am. Chem. Soc.* **111**, 1515–1517.
- Shaka, A. J., Lee, C. J. & Pines, A. (1988) *J. Magn. Reson.* **77**, 274–293.
- Cavanagh, J. & Rance, M. (1992) *J. Magn. Reson.* **96**, 670–678.
- Bodenhausen, G. & Ruben, D. J. (1980) *Chem. Phys. Lett.* **69**, 185–189.
- Cavanagh, J., Palmer, A. G., Wright, P. E. & Rance, M. (1991) *J. Magn. Reson.* **91**, 429–436.
- Palmer, A. G., Cavanagh, J., Wright, P. E. & Rance, M. (1991) *J. Magn. Reson.* **93**, 151–170.
- Kay, L. E., Keifer, P. & Saareinen, T. (1992) *J. Am. Chem. Soc.* **114**, 10663–10665.
- Shaka, A. J., Barker, P. B. & Freeman, R. (1985) *J. Magn. Reson.* **64**, 547–552.
- Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1986) *Numerical Recipes. The Art of Scientific Computing* (Cambridge Univ. Press, Cambridge, U.K.).
- Tanford, C. (1970) *Adv. Protein Chem.* **24**, 1–95.
- Privalov, P. L. & Makhatadze, G. I. (1990) *J. Mol. Biol.* **213**, 385–391.
- Tan, Y. J., Oliveberg, M. & Fersht, A. R. (1996) *J. Mol. Biol.* **264**, 377–389.
- Tanford, C. (1968) *Adv. Protein Chem.* **23**, 121–282.
- Oliveberg, M. & Fersht, A. R. (1996) *Biochemistry* **35**, 2726–2737.
- Oliveberg, M. (1998) *Acc. Chem. Res.* **31**, 765–772.
- Silow, M. & Oliveberg, M. (1997) *Biochemistry* **36**, 7633–7637.
- Matouschek, A. & Fersht, A. R. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 7814–7818.
- Oliveberg, M., Tan, Y.-J., Silow, M. & Fersht, A. R. (1998) *J. Mol. Biol.* **277**, 933–943.
- Otzen, D. E., Kristensen, O., Procter, M. & Oliveberg, M. (1999) *Biochemistry* **38**, 6499–6511.
- Bryngelson, J., Onuchic, J. N., Socci, N. D. & Wolynes, P. (1995) *Proteins: Struct. Funct. Genet.* **21**, 167–195.
- Parker, M. J., Spencer, J. & Clarke, A. R. (1995) *J. Mol. Biol.* **253**, 771–786.
- Englander, W. S., Sosnick, T. R., Mayne, L. C., Shtilerman, M., Qi, P. X. & Bai, Y. (1998) *Acc. Chem. Res.* **31**, 737–744.
- Zaidi, N. F., Nath, U. & Udgaonkar, J. B. (1997) *Nat. Struct. Biol.* **4**, 1016–1024.
- Silow, M., Tan, Y. J., Fersht, A. R. & Oliveberg, M. (1999) *Biochemistry* **38**, 13006–13012.
- Silow, M. & Oliveberg, M. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 6084–6086.
- Jones, C. M., Henry, E. R., Hu, Y., Chan, C.-K., Luck, S. D., Bhuyan, A., Roder, H., Hofrichter, J. & Eaton, W. A. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 11860–11864.
- Plotkin, S. S., Wang, J. & Wolynes, P. G. (1997) *J. Chem. Phys.* **106**, 2932–2948.
- Wolynes, P. G. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 6170–6175.
- Avis, J. M., Allain, F. H., Howe, P. W., Varani, G., Nagai, K. & Neuhaus, D. (1996) *J. Mol. Biol.* **257**, 398–411.
- Li, R. & Woodward, C. (1999) *Protein Sci.* **8**, 1571–1590.
- Shoemaker, B. A., Wang, J. & Wolynes, P. G. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 777–782.
- Shoemaker, B. A., Wang, J. & Wolynes, P. G. (1999) *J. Mol. Biol.* **287**, 675–694.
- Finkelstein, A. V. & Badredtinov, A. Y. (1997) *Fold. Des.* **2**, 115–121.
- Lifshitz, E. M. & Pitaevskii, L. P. (1981) *Physical Kinetics* (Pergamon, Oxford).
- Villegas, V., Martinez, J. C., Aviles, F. X. & Serrano, L. (1998) *J. Mol. Biol.* **283**, 1027–1036.
- Lindahl, M., Svensson, L. A., Liljas, A., Sedelnikova, S. E., Eliseikina, I. A., Fomenkova, N. P., Nevskaya, N., Nikonov, S. V., Garber, M. B., Muranova, T. A., et al. (1994) *EMBO J.* **13**, 1249–1254.