Structural Changes in the Transition State of Protein Folding: Alternative Interpretations of Curved Chevron Plots†

Daniel E. Otzen,‡ Ole Kristensen,§ Mark Proctor,|| and Mikael Oliveberg*,‡

Department of Biochemistry and Department of Molecular Biophysics, Chemical Centre, University of Lund, P.O. Box 124, S-22100 Lund, Sweden, and MRC Unit for Protein Function and Design, Cambridge IRC for Protein Engineering, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, U.K.

Received November 30, 1998; Revised Manuscript Received March 8, 1999

ABSTRACT: The interpretation of folding rates is often rationalized within the context of transition state theory. This means that the reaction rate is linked to an activation barrier, the height of which is determined by the free energy difference between a ground state (the starting point) and an apparent transition state. Changes in the folding kinetics are thus caused by effects on either the ground state, the transition state, or both. However, structural changes of the transition state are rarely discussed in connection with experimental data, and kinetic anomalies are commonly ascribed to ground state effects alone, e.g., depletion or accumulation of structural intermediates upon addition of denaturant. In this study, we present kinetic data which are best described by transition state changes. We also show that ground state effects and transition state effects are in general difficult to distinguish kinetically. The analysis is based on the structurally homologous proteins U1A and S6. Both proteins display two-state behavior, but there is a marked difference in their kinetics. S6 exhibits a classical V-shaped chevron plot (log observed rate constant vs denaturant concentration), whereas U1A's chevron plot is symmetrically curved, like an inverted bell curve. However, S6 is readily mutated to display U1A-like kinetics. The seemingly drastic effects of these mutations are readily ascribed to transition state movements where large kinetic differences result from relatively small alterations of a common free energy profile and broad activation barriers.

Thirty years ago, Levinthal suggested that folding intermediates accelerate the rate of protein refolding by reducing the extent of denatured polypeptide's search for the native state through conformational space (*1*). Since then, intermediates have become an integral part of many folding models (*2*, *3*) in which folding is seen as a hierarchical process of consecutive structural consolidation. In these models, certain interactions, such as elements of secondary structure (in the framework model), or a compact semistructured state (in the hydrophobic collapse model), are formed rapidly, often within the dead time of the experiment, before other interactions accrete and the final native conformation is attained. However, in 1991 a 64-residue protein (chymotrypsin inhibitor 2, CI2[1]) was shown to fold without any kinetic intermediate (*4*); several other examples of two-state systems have been reported since then (*4*−*11*). This shows that some proteins do not require populated intermediates to guide folding, and that the conformational search in these cases takes place entirely by labile high-energy states (*11*). In contrast to the hierarchical folding models involving intermediates, two-state folding appears to involve a more cooperative nucleation−condensation process, in which most interactions are already partially formed in the transition state ensemble (TSE), centered around a small folding nucleus (*12*).

A common way to characterize proteins as two-state or three-state is to inspect the shape of their chevron plot (log observed rate vs denaturant concentration). The plots of two-state proteins are V-shaped since the log of both refolding (k_f) and unfolding (k_u) rates depends linearly on denaturant conditions (*4*).

Scheme 1

$$\mathrm{D} \xrightarrow{k_f} \ddagger \xleftarrow{k_u} \mathrm{N}$$

For clarity, the rate-limiting transition state ensemble ‡ is included in the reaction scheme. The equilibrium constant $K_{D-N} = k_f/k_u$. Three-state proteins typically exhibit deviations from linearity (a downward curvature) at low denaturant concentrations (*13*, *14*). This is ascribed to a transient accumulation of intermediate under these conditions, which means that there is a change in the ground state from which folding occurs, i.e., from D to I.

Scheme 2

$$\mathrm{D} \overset{K_I}{\rightleftharpoons} \mathrm{I} \xrightarrow{k_f} \ddagger \xleftarrow{k_u} \mathrm{N}$$

In a similar way, curvature in the unfolding limb is attributed to unfolding intermediates in a scheme which is a mirror image of Scheme 2 (*11*, *15*). Curvature at both low

† D.E.O. gratefully acknowledges support from the Wenner-Grenska Foundation and EMBO for fellowships. This work is supported by the Swedish Natural Science Research Council, the Craafordska Stiftelsen, and S. & E.-C. Hagbergs Stiftelse.

‡ Department of Biochemistry, Chemical Centre, University of Lund.

§ Department of Molecular Biophysics, Chemical Centre, University of Lund.

|| Cambridge IRC for Protein Engineering.

[1] Abbreviations: CI2, chymotrypsin inhibitor 2; D, ensemble of denatured states; I, intermediate; N, native state; TSE, transition state ensemble; ‡, transition state.

10.1021/bi982819j CCC: $18.00
Published on Web 04/22/1999

and high denaturant concentrations (*16*) can be ascribed to the presence of both refolding (I) and unfolding (I*) intermediates.

Scheme 3

$$D \overset{K_I}{\rightleftharpoons} I \overset{k_f}{\rightarrow} \ddagger \overset{k_u}{\leftarrow} I^* \overset{K_{I^*}}{\rightleftharpoons} N$$

However, there are other equally valid ways of accounting for curvatures in chevron plots in addition to changes at the ground state level, namely, movements of the TSE. The broad barrier model of protein folding proposes that much of the folding process takes place rather isoenergetically at the transition state level (*11*) and that as a consequence the position of the TSE is sensitive to denaturant conditions. This idea is inspired by the behavior of the 102-residue human spliceosomal protein U1A (*17*). U1A's chevron plot (log k_{obs} plotted vs denaturant concentration) is symmetrically curved like an inverted bell curve, rather than being V-shaped as it is for a classical two-state folding process. Despite these curvatures, U1A's data are perfectly compatible with a two-state model (the ratio of the refolding and unfolding rate constants correspond within error to the equilibrium constant K_{D-N} at all denaturant concentrations), implying that folding takes place directly from D also at low denaturant concentrations. This has been taken to indicate that the kinetic anomaly (the curvature) is caused by transition state movements, which occur upon addition of denaturant (*11*).

Scheme 1b

$$D \overset{k_f}{\rightarrow} \ddagger \cdots \ddagger' \overset{k_u}{\leftarrow} N$$

$\ddagger$ and $\ddagger'$ symbolize the TSE prevalent at low and high denaturant concentrations, respectively.

The question is whether U1A's behavior exemplifies a general phenomenon. Evidence for transition state movements is also observed with CI2 (*18*), although CI2 is structurally very different from U1A (*17*, *18*). In a few cases, radically destabilized CI2 mutants exhibit a U1A-like curvature in the unfolding limb of the chevron plot. To explore in more detail the mechanism behind such curvature, we have here compared U1A with a close structural analogue, the 101-residue ribosomal protein S6 from *Thermus thermophilus* (*19*). The rationale behind these experiments is that a structural homologue of U1A may exhibit a similar underlying reaction profile despite differences in kinetic behavior, and hence may be relatively easy to provoke into U1A-like behavior. Both S6 and U1A lack disulfide bonds and cis prolyl−peptidyl bonds which could otherwise complicate folding behavior. They have similar folding topologies composed of two split $\beta-\alpha-\beta$ motifs (Figure 1), although they exhibit no homology at the primary level (Figure 2). This is just one of many examples of proteins sharing similar tertiary structure without exhibiting any sequence homology (*20*, *21*).

We now show that although wild-type S6 behaves like a classical two-state protein with a V-shaped chevron plot, curvatures in the unfolding limb are easily induced by destabilizing hydrophobic deletions. This suggests that the wild type's fixed TSE is mobilized in these mutants. The mobilization occurs readily; 15 out of 26 S6 mutants give rise to curvature. In contrast, only about 5 out of 100 CI2 mutants demonstrated appreciable unfolding curvature (*18*). The structural homology between S6 and U1A suggests that the propensity to mobilize TSEs may be a topological property. Homologous structures may be associated with similar folding profiles, although the details of their kinetics are different. There is therefore no firm distinction between broad and sharp activation barriers; apparent sharp activation barriers may represent small energetic "hilltops" in a generally flat barrier surface. An alternative interpretation of the curved chevron plots may be based on the induction of unfolding intermediates, but we favor the more novel broad activation barrier view, since this explains how large kinetic differences may result simply from relatively small alterations of a common free energy profile, rather than from more radical changes of the folding reaction.

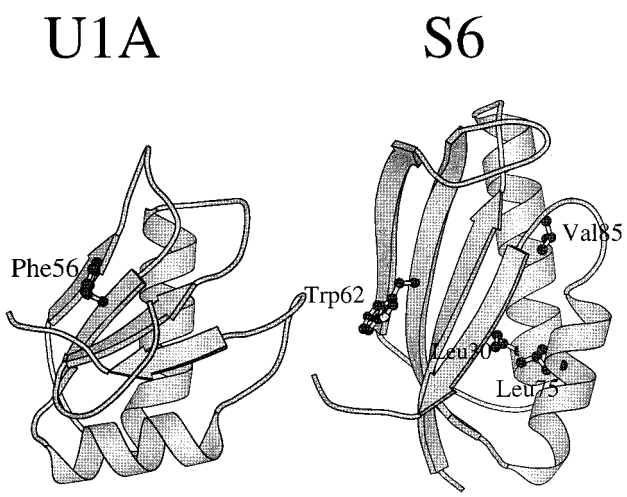

FIGURE 1: Structures of U1A (*62*) and S6 (*19*) generated using the program Molscript (*63*). The PDB file names for U1A and S6 are 1URN and 1RIS, respectively. Phe56 in U1A is substituted with Trp56 in our studies. In S6, side chains of L30, L75, and V85 (mutated in this study) are shown as well as Trp62. The biological function of both U1A and S6 is to interact with RNA, either transcribed RNA (U1A) or ribosomal RNA (S6), in multicomponent protein−RNA complexes. Despite their lack of sequence homology, both proteins have a repeated $\beta-\alpha-\beta$ motif which also occurs in several other nucleic acid binding proteins, such as L30 (*64*), L6 (*65*), L9 (*66*), and the C-terminal domain of the E2 protein from papilloma virus (*67*). In addition, the $\beta-\beta-\beta$ topology is found in other proteins such as ferredoxin (*68*) and merP (*69*).

MATERIALS AND METHODS

Materials

Primers (desalted) and guanidinium chloride (Ultrapure) were from Gibco BRL. The vector prSETA was from Invitrogen. Restriction enzymes were from New England Biolabs. Chromatographic material was from Pharmacia Biotech.

Construction of the Vector

S6 from *T. thermophilus* has the following 101-amino acid sequence (*19*): MRRYEVNIVLNPNLDQSQLALEKEIIQRALENYGARVEKVEELGLRRLAYPIAKDPQGYFLWYQVEMPEDRVNDLARELRIRDNVRRVMVVKSQEPFLANA. However, the gene has not been cloned. Instead, we used the amino acid sequence to derive the following 306 bp DNA sequence that is optimal for expression

```
AA:  1          11          21          31          41
U1A: MAVPETRPNH TIYINNLNEK IKKDELKKSL YAIFSQFGQI LDILVSRSLK
S6:  MRRYEVNIVL NPNLDQSQLA LEKEIIQRAL ENYGARVEKV EELGLRRLAY

AA:  51         61         71          81     91            100
U1A: MRGQAWVIFK EVSSATNALR SMQGFPFYDK PMRIQYAKTD SDIIAKMKGT FV
S6:  PIAKDPQGYF LWYQVEMPED RVNDLARELR IRDNVRRVMV VKSQEPFLAN A
```

FIGURE 2: Amino acid sequences of U1A and S6, showing the lack of homology between the two proteins.

in *Escherichia coli*: ATGCGTCGTTACGAAGTAAACATCGTACTGAACCCGAACCTGGACCAGAGCCAGCTGGCTCTGGAAAAAGAAATCATCCAGCGTGCTCTGGAAAACTACGGCGCTCGTGTTGAAAAAGTTGAAGAACTGGGCCTGCGTCGTCTGGCTTACCCGATCGCTAAAGACCCGCAGGGCTACTTCCTGTGGTACCAGGTTGAAATGCCGGAAGACCGTGTTAACGACCTGGCTCGTGAACTGCGTATCCGTGACAACGTTCGTCGTGTTATGGTTGTTAAAAGCCAGGAACCGTTCCTGGCTAACGCTTGA. The gene was constructed by PCR using the following three overlapping oligonucleotides: fragment 1 (the underlined sequence anneals to the underlined sequence of fragment 2), ATGCGTCGTTACGAAGTAAACATCGTACTGAACCCGAACCTGGACCAGAGCCAGCTGGCTCTGGAAAAAGAAATCATCCAGCGTGCTCTGGAAAACTACGGCGCTCGTGTTGAAAA; fragment 2 (the underlined sequence anneals to the underlined sequence of fragment 1; the italicized sequence overlaps with the italicized sequence of fragment 3), *CGGTCTTCCGGCATTTCAACCTG*GTACCACAGGAAGTAGCCCTGCGGGTCTTTAGCGATCGGGTAAGCCAGACGACGCAGGCCCAGTTCTTCAACTTTTTCAACACGAGCGCCGTAGT; and fragment 3 (the italicized sequence overlaps with the italicized sequence of fragment 2), TCAAGCGTTAGCCAGGAACGGTTCCTGGCTTTTAACAACCATAACACGACGAACGTTGTCACGGATACGCAGTTCACGAGCCAGGTCGTTAACA*CGGTCTTCCGGCATTTCAACCTG*.

The 5′-*Nde*I restriction site was introduced using the following primer (which anneals to the 5′-terminal part of fragment 1): GGACTTCCATATGCGTCGTTACGAAGTAAACATC.

The 3′-*Eco*RI restriction site was introduced with the following primer (which anneals to the 3′-terminal part of fragment 3): CCGGAATTCAAGCGTTAGCCAGGAACGGTTC.

The five DNA fragments were incubated at an oligonucleotide concentration of 10 nM and a primer concentration of 1 μM. After 2 min at 94 °C, 30 cycles of 1 min at 94 °C, 0.5 min at 55 °C, and 0.5 min at 68 °C were carried out, followed by purification on an agarose gel. The gene was subsequently digested with *Nde*I and *Eco*RI and ligated into the expression vector pRSETA after this vector had been opened with the same two restriction enzymes. Digestion of pRSETA with *Nde*I and *Eco*RI removes the polyhistidine tag and enterokinase cleavage site so that S6 is produced in the 101-amino acid sequence form without any fusion protein extension on the N-terminus. Correct insertion was verified by DNA sequencing and expression tests as determined by SDS−PAGE. After the gene was cloned into the pRSETA expression vector, the correct position and primary sequence of the gene were confirmed by sequencing. Following expression in *E. coli*, S6 was purified to >95% purity as estimated by SDS−PAGE. N-Terminal sequencing of the first 12 residues of the S6 wild type, performed on an Applied Biosystems model 477A sequencer, confirmed the sequence. There was no significant amount of nucleic acid present, judging from the high A_{280}/A_{260} ratio (≈2).

Mutagenesis of the S6 Gene

Point mutations in the S6 gene were constructed using Stratagene's QuikChange Site-Directed Mutagenesis kit with *Pfu* DNA polymerase. The following primers, together with their complementary reverse sequence, were used to construct mutants LA30, LA75, and VA85 (mutated codons in bold): LA30, GAAATCATCCAGCGTGCT**GCT**GAAAACTACGGCGCTCG; LA75, GAAGACCGTGTTAACGAC**GCT**GCTCGTGAACTGCGTATC; and VA85, CGTATCCGTGACAAC**GCT**CGTCGTGTTATGGT. The primers were designed to give a melting temperature of 78 °C. Briefly, after PCR with the mutagenic primers, parental wild-type DNA was digested by *Dpn*I (which requires a methylated restriction site). Mutated plasmids were transformed into XL1-Blue supercompetent cells (Stratagene) and checked for expression of S6 by SDS−PAGE before sequencing (ABI Prism 310 Genetic System) to verify the mutation.

Protein Purification

S6. All cell growth was realized at 37 °C. The vector containing the gene for S6 was transformed into BL21/*pLysS* (chloramphenicol-resistant) cells by standard electroporation techniques, and transformants were selected on ampicillin−chloramphenicol LB-Miller agar plates. Twelve hours after transformation, a transformant colony was picked and grown for 8−10 h in a 5−20 mL starter culture of YT medium supplemented with 36 mM K_2HPO_4, 4 mM KH_2PO_4, 4 mM glucose, 100 mg/L ampicillin, and 17 mg/L chloramphenicol, and then transferred to 0.5−4 L of the same medium. The culture was induced with 0.4 mM IPTG at an A_{600} of 0.4−0.6 and grown for 5−6 h before cells were harvested by centrifugation at 5000 rpm for 15 min. RNase A (50 mg/L) and DNase I (50 mg/L) were added to the resuspended cells. Cells were opened by passing them through a Yeda press (Scientific Instruments, Rehovot, Israel) three times at 100 bar, followed by centrifugation at 18 000 rpm for 30 min. To remove nucleic acids, polyethyleneimine P was added to a concentration of 0.5% (w/v) to the supernatant, and after being stirred for 20 min at 4 °C, the solution was centrifuged at 18 000 rpm for 30 min. S6 was precipitated from the supernatant by adding $(NH_4)_2SO_4$ to a concentration of 70% (47.2 g/100 mL), stirring at 4 °C for 1 h, and centrifuging for 45 min at 18 000 rpm. The pellet was resuspended in water and dialyzed against 50 mM Tris-HCl (pH 7.5) for at least 16 h. The solution was centrifuged and filtered before loading it onto a CM-Sepharose column (height of 29 cm,

inside diameter of 1.8 cm) pre-equilibrated with the same buffer. At a flow rate of 90 mL/h, S6 was eluted as the second (major) top in the void buffer and was >95% pure as judged by SDS−PAGE and the A_{280}/A_{260} ratio (ca. 2.0). The molar extinction coefficient of S6 is 12 700 M^{-1} cm^{-1} (due to one Trp and five Tyr residues), and the molecular mass is 11 973 Da, giving an $\epsilon^{1mg/mL}_{280nm}$ of 1.06. The yield of S6 was approximately 40 mg/L of cell culture.

U1A. Naturally occurring human U1A has no Trp residues. However, the F56W mutant of U1A (kindly provided by K. Nagai) gives rise to a large fluorescence change on unfolding and is used for all our folding studies. This mutation occurs naturally in potato (*22*). F56W U1A was expressed and purified as described previously (*17*).

Equilibrium Denaturation Measurements

S6 contains a single Trp residue at position 62 (in the middle of β-strand 3), which is only slightly surface-exposed. Unfolding is accompanied by a 60% reduction of the fluorescence signal and a slight red shift of the emission maximum (from 350 to 353 nm). Human U1A has no Trp residues, so we studied a variant in which the semiburied Phe56 is substituted with Trp (kindly provided by K. Nagai). Like that of S6, unfolding of this variant leads to a 60% reduction of the fluorescence signal.

All equilibrium and kinetic experiments were performed in 50 mM MES (pH 6.3) at 25 °C. S6 was incubated for at least 30 min at 25 °C with 0.2−6 M guanidinium chloride (prepared from an 8 M stock solution) before spectroscopic measurements were carried out in a thermostated cuvette. For fluorescence measurements on an LS-50B Perkin-Elmer luminescence spectrometer, 2 μM S6 was used in a 500 μL cuvette, and an emission spectrum in the range of 300−440 nm was recorded with excitation at 280 nm, an excitation slit width of 10 nm, and an emission slit width of 5 nm. Circular dichroism measurements were performed on a Jasco 720 dichrograph using 10 μM S6 and a 1 mm cuvette. The range of 260−210 nm was scanned three times at 20 mn/min with a bandwidth of 1 nm and a response time of 4 s.

Data were fitted to an equation for two-state unfolding without intermediates (*23*):

$$F = \frac{(\alpha_N + \beta_N[D]) + (\alpha_U + \beta_U[D])e^{[m_{U-F}([D] - [D]^{50\%})]/RT}}{1 + e^{[m_{U-F}([D] - [D]^{50\%})]/RT}} \quad (1)$$

where α_N and α_U are the intercepts and β_N and β_U are the slopes of the fluorescence baselines at low (N) and high (U) denaturant concentrations before and after the transition region of unfolding, respectively, $[D]^{50\%}$ is the denaturant concentration at which the protein is 50% unfolded, and m_{D-N} is the sensitivity of the equilibrium constant of unfolding (K_{D-N}) to denaturant concentration. The same values for $[D]^{50\%}$ and m_{D-N} were obtained independent of whether fluorescence data from one particular emission wavelength were used or data over the range of 290−400 nm were fitted simultaneously using the program Grafit 3.0 (Erithacus Software).

Stopped-Flow Kinetics

All kinetic experiments were performed on an Applied Photophysics SX-18.MV stopped-flow reaction analyzer. The protein solution and denaturant solution were mixed in a 1:10 ratio to give a final protein concentration of 1 μM (unless otherwise stated). Double-jump experiments were carried out with symmetric mixing. Excitation was at 280 nm, and emission was detected above 315 nm using a glass filter. The unfolding kinetics of U1A and S6 describe a single first-order relaxation phase followed by linear drift. Linear drift, reported in the unfolding of several proteins, is often attributed to the slow photobleaching of the Trp fluorophore. Alternatively, in the case of S6, it could arise from a slow proline isomerization in the unfolded state which affects the fluorescence of Trp62 (e.g., involving Pro56 or Pro68). Double-jump experiments, in which S6 is first unfolded in 4.4 M GdmCl and then allowed to refold after different delay periods in 2.2 M GdmCl, enable us to rule out a second unfolding phase. The amplitude of the refolding phase (which indicates how much protein has unfolded during the delay period) increased at 0.071 s^{-1}, close to the rate predicted from the rate of the major unfolding phase in 4.4 M GdmCl (0.08 s^{-1}).

Refolding is described by two relaxation phases. The major phase accounts for at least 90% of the total amplitude and has a rate constant that is at least 10 times faster than that of the minor phase, except in the transition region of unfolding where the rates and amplitudes of the two phases approach each other. The second phase could in principle represent the decay of an intermediate state formed during refolding, as in Scheme 2, but this is disproved by double-jump experiments, in which the unfolded protein (in 4.6 M GdmCl) is allowed to refold for various periods of time in 2.3 M GdmCl before unfolding in 5 M GdmCl. The amplitude of the resulting unfolding phase (which has the expected unfolding rate of 0.17 s^{-1}) increases at a rate of 0.48 s^{-1} versus delay time, in agreement with the expected value of 0.54 s^{-1} for a simple two-state system. Slow refolding phases are often, but not always (*24*), due to cis−trans isomerization of peptidyl−prolyl bonds in the unfolded state (*25*, *26*), where the different conformations interconvert with half-times in the range of tens of seconds. U1A and S6 contain four Pro residues each, but the minor phase is too rapid to be a conventional proline isomerization. For both S6 and U1A, the linear correlation between log k_f^{minor} and GdmCl concentration extrapolates to a k_f^{minor} of 31 s^{-1} in the absence of denaturant (data not shown). This does not rule out proline isomerization as an explanation for the slow phase; it is possible that the fraction of the unfolded population with cis peptidyl−prolyl bonds refolds relatively rapidly to a native-like state in which the cis prolyl bond(s) then slowly interconverts to a trans bond without a change in fluorescence. k_f^{minor} is just as sensitive to GdmCl concentration as k_f^{major}, which also suggests that the minor phase leads to a product that is virtually as structured as the native state. The second fastest phase in the refolding of U1A has been attributed to transient aggregation, since its relative amplitude increases at high concentrations (*27*). This can be ruled out for S6; we saw no change in the relative amplitude of the slow phase between 0.1 and 31 μM S6 (data not shown). We confine ourselves to the rates derived from the major refolding and unfolding phase.

Dead-Time Spectra. Dead-time spectra were obtained by excitation at 280 nm and emission at individual wavelengths between 300 and 440 nm using a second monochromator

Table 1: U1A and S6 Fold in a Two-State Process According to both Equilibrium and Kinetic Data[a]

parameter	U1A	S6
$[D]^{50\%}_{equilibrium}$ (M)[b]	4.07 ± 0.02 (fluorescence) 4.04 ± 0.04 (CD)	3.32 ± 0.03 (fluorescence) 3.33 ± 0.08 (CD)
$\Delta G_{D-N}^{equilibrium}$ (kcal/mol)	9.4 ± 0.2 (fluorescence) 11.6 ± 1.7 (CD)	7.24 ± 0.64 (fluorescence) 8.02 ± 1.8 (CD)
$m_{D-N}^{equilibrium}$ (M^{-1})[c]	1.70 ± 0.01 (fluorescence) 2.1 ± 0.30 (CD)	1.60 ± 0.14 (fluorescence) 1.77 ± 0.40 (CD)
$[D]^{50\%}_{kinetic}$ (M)[d]	4.09 ± 0.05	3.46 ± 0.05
$k_{ref}^{H_2O}$ (s^{-1})	338 ± 3	332 ± 3
m_f (s^{-1} M^{-1})[e]	(−0.40 ± 0.02) − (0.128 ± 0.005)[D]	−1.21 ± 0.01
$k_{unf}^{H_2O}$ (s^{-1})	(19.1 ± 0.4) × 10^{-6}	(309 ± 6) × 10^{-6}
m_u (s^{-1} M^{-1})[f]	(1.42 ± 0.04) − (0.148 ± 0.003)[D]	0.54 ± 0.01
$\Delta G_{D-N}^{kinetic}$ (kcal/mol)[g]	9.9 ± 0.2	8.22 ± 0.18
$m_{D-N}^{kinetic}$ (M^{-1})[h]	(1.81 ± 0.04) − (0.01 ± 0.006)[D]	1.75 ± 0.02

[a] All experiments were performed at 25 °C and pH 6.25. Data are based on fluorescence measurements except where indicated. [b] Concentration of denaturant (GdmCl) at which the equilibrium constant of unfolding $K_{D-N} = 1$ (based on a two-state unfolding model). [c] Dependence of $\log(K_{D-N})$ on GdmCl concentration. [d] GdmCl concentration where $k_f = k_u$. [e] Dependence of $\log(k_f)$ on GdmCl concentration. [f] Dependence of $\log(k_u)$ on GdmCl concentration. [g] Calculated as $-RT \times 1.36 \log(k_f^{H2O}/k_u^{H2O})$. [h] $m_{D-N}^{kinetic} = m_u - m_f$.

for the emitted light. Due to the low intensity of the monochromatic light, the photomultiplier voltage was increased to the maximum level (1000 V). The dead-time fluorescence at each wavelength was calculated by extrapolating the fitted exponentials back to fluorescence at time zero.

Determination of the Crystal Structure of LA30. Crystals were grown by vapor diffusion. Two and one-half microliters of LA30 [12 mg/mL in 10 mM Tris (pH 7.5)] was mixed with 2.5 μL of a precipitant solution and equilibrated with 500 μL of the same precipitant solution. The best crystals were grown using a precipitant composition of 50 mM Tris-HCl (pH 7.25), 0.2 M KF, 0.4 M NaCl, and 15% MPD, similar to the conditions for the S6 wild type (*19*). Crystals appeared after 48 h. X-ray diffraction data were collected on a Mar image plate detector mounted on a Rigaku RU200BEH rotating anode. Reflections were collected with a frame width of 1°. The space group is *C*222 (identical to that of the S6 wild type) with the following unit cell dimensions: $a = 52.555$ Å, $b = 105.805$ Å, and $c = 41.204$ Å with a resolution limit of 1.95 Å. There were 62 306 reflections, 8414 of which were unique, with a completeness of 96.4%. The overall merging *R*-factor was 5.5%, and 31.4% in the 2.02−1.95 Å shell. Data processing was performed with Denzo and Scalepack. The molecular model was fitted into the unit cell by molecular replacement using the program Amore in the CCP4 package (*28*). Refinement was performed by XPLOR (*29*) and REFMAC (*28*). The structure was inspected with the program O (*30*), and 37 water molecules were inserted. The final model had a crystallographic *R*-factor of 19.1% ($R_{free} = 24.7\%$), a bond length rmsd of 0.015 Å, a bond angle rmsd of 2.806°, a dihedral angle rmsd of 26.922°, and an improper dihedral angle rmsd of 2.37°. The coordinates are available from the Brookhaven Protein Data Bank under file name 1LOU.

RESULTS

Both U1A and S6 Obey Two-State Kinetics but Differ at the Transition State Level. A calorimetric study has established that U1A unfolds according to a simple two-state model without any stable intermediates (*31*). We find that S6 and U1A both unfold in denaturant without any stable intermediates under equilibrium conditions (data not shown). The data fit precisely to a two-state model with the following linear relationship:

$$\log K_{D-N} = \log K_{D-N}^{H_2O} + m_{D-N}[D] \quad (2)$$

where $K_{D-N}^{H_2O}$ is the equilibrium constant of unfolding in water and m_{D-N} is a constant related to the change in the solvent-accessible area between the denatured and the native state (*32*). The normalized unfolding transitions measured by fluorescence and far-UV CD are superimposable and give the same stability parameters within error (Table 1). The difference in the free energy of unfolding, ΔG_{D-N}, for U1A calculated by CD and fluorescence is due to the diverging *m* values which are accompanied by large errors. A higher CD *m* value is also difficult to rationalize in terms of an equilibrium intermediate. The denaturation midpoints, which are determined with greater accuracy, are identical. This is further support for a two-state transition. Equilibrium intermediates typically lead to noncoinciding denaturation midpoints for fluorescence and CD data.

In a simple two-state folding model (Scheme 1), where the position of the TSE is independent of denaturant concentration, the following linear relationship holds:

$$\log k_u = \log k_u^{H_2O} + m_u[D] \quad (3a)$$

$$\log k_f = \log k_f^{H_2O} + m_f[D] \quad (3b)$$

where, as above, m_u and m_f are constants which reflect the change in solvent-accessible area in the activation process of unfolding and refolding, respectively (Figure 3). Therefore, a plot of the logarithm of the observed rate constant k_{obs} versus GdmCl concentration is V-shaped:

$$\log k_{obs} = \log(k_u + k_f) = \log(10^{\log k_f^{H2O} + m_f[D]} + 10^{\log k_u^{H2O} + m_u[D]}) \quad (4)$$

S6 conforms to this minimal folding model (Figure 3). $K_{D-N}^{H_2O}$ and m_{D-N} can be calculated from the kinetic data using two-state relationships:

$$K_{D-N} = k_u/k_f \quad (5a)$$

$$m_{D-N} = m_u - m_f \quad (5b)$$

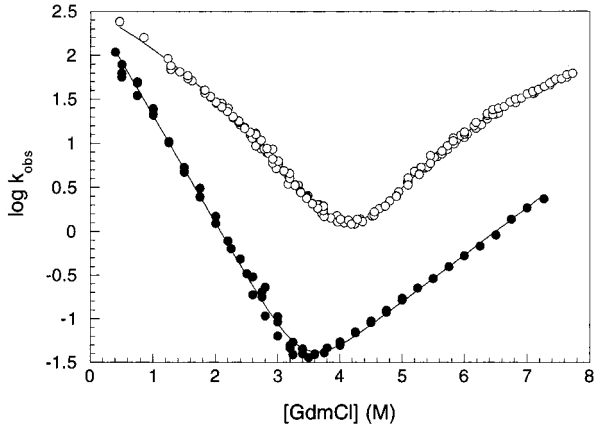

FIGURE 3: Chevron plot of refolding and unfolding rates of U1A (○) and S6 (●). k_{obs} is measured in s^{-1}. The data for S6 conform to a classical two-state folding model with constant values of m_u and m_f, while the corresponding values for U1A are linearly dependent on GdmCl concentration (see Table 1). To increase the accuracy of U1A's fitted parameters for this study, a larger number of data points have been collected than were collected in the initial report (*11*).

where eq 5b follows from eqs 2, 3, and 5a. Equilibrium data and kinetic data yield the same results within error (Table 1).

The position of the TSE ($\beta^{\ddagger}$) on the reaction coordinate between the denatured and native states can be defined in terms of how solvent-exposed the TSE is compared to the denatured state ($\beta_D = 0$) and the native state ($\beta_N = 1$). Since an *m* value for a given transition represents the difference in the extent of solvent exposure between the two states involved in the transition, $\beta^{\ddagger}$ may be calculated as follows (*32*):

$$\beta^{\ddagger} = -m_f/m_{D-N} = 1 - (m_u/m_{D-N}) \qquad (6)$$

S6 has a $\beta^{\ddagger}$ value of 0.7, which is within the normal range for two-state proteins. Other values are 0.58 for CI2 (*4*), 0.39 for the monomeric λ repressor (*33*), and 0.63 for T4 lysozyme (*34*). Note that the $\beta^{\ddagger}$ value reflects the general denaturant accessibility of different transition state ensembles in the folding reaction, and does not necessarily indicate that the TSE is closer to the native state than to the denatured state in terms of side chain structure. Different methods for estimating the degree of structure of a TSE will not a priori give the same value, since they measure different structural parameters which may not be closely correlated. This is beautifully illustrated by protein engineering studies, which make it possible to calculate the degree of structure in the transition state at the level of individual residues (Φ value analysis) (*35*). For the two-state protein chymotrypsin inhibitor 2 (CI2), Φ values of a large number of mutants are significantly lower (generally between 0 and 0.4) than the corresponding $\beta^{\ddagger}$ values (0.58) (*12*). This is reasonable, since collapse of the polypeptide, and hence a decrease in the extent of solvent exposure, will precede the orderly formation of side chain structure.

Interestingly, the behavior of U1A is different from that of S6. A distinct and symmetrical curvature in both limbs of U1A's chevron plot is apparent (Figure 3). Nevertheless, this can still be fitted within the constraints of a two-state folding model, if a quadratic relationship between log k_x and GdmCl concentration is assumed (where x is refolding or unfolding) (*11*):

$$\log k_x = \log k_x^{H_2O} + a^{m_x}[D] + b^{m_x}[D]^2 \qquad (7a)$$

$$\log k_{obs} = \log(10^{\log k_f^{H2O} + a^{mf}[D] + b^{mf}[D]^2} + 10^{\log k_u^{H2O} + a^{mu}[D] + b^{mu}[D]^2}) \qquad (7b)$$

where a^{m_x} and b^{m_x} are constants. The actual *m* values are given by the relationship $\partial(\log k_x)/d[GdmCl] = a^{m_x} + 2b^{m_x}[GdmCl]$. Since b^{m_f} and b^{m_u} are virtually identical (Table 1), they cancel out in eq 5a to give a simple linear GdmCl concentration dependence for log K_{D-N} (eq 2), consistent with equilibrium data. Therefore, we do not need to invoke the existence of folding and unfolding intermediates (Scheme 3), although these states will also give rise to curvature at either end of the denaturant concentration range (*36*).

Broad Activation Barriers in Folding. As is apparent from the chevron plots, there is a crucial difference between S6 and U1A. U1A's $\beta^{\ddagger}$ value (eq 6) is not constant, but increases with denaturant concentration from 0.26 at 0 M GdmCl to 0.84 at 8 M GdmCl (*11*). We have previously rationalized this behavior within the context of a broad activation barrier instead of invoking folding and unfolding intermediates (*11*). Each point on this free energy barrier represents the free energy of an ensemble of conformations, possibly with the same $\beta^{\ddagger}$ value. Under given solvent conditions, the ensemble with the highest free energy constitutes the TSE, whose formation is rate-limiting for the folding process. Since different ensembles exhibit different degrees of compactness, they also have different sensitivities to denaturant. Therefore, changes in denaturant concentration will affect the relative stabilities of the ensembles and alter the position of the TSE. Conformations close to the native state will experience a larger destabilization upon addition of denaturation than conformations close to the denatured state. The result is that the reaction profile tilts as the denaturant concentration increases and the TSE moves closer to the native state.

While U1A's $\beta^{\ddagger}$ value increases by 0.6 over the range of 0−7 M GdmCl, the $\beta^{\ddagger}$ value of S6 remains constant. In other words, the energy level of S6's TSE is so elevated relative to its surroundings that it maintains the highest position on the free energy profile over the entire concentration range, although it may rest on the same broad plateau as U1A. Thus, the detailed features of the free energy profile which S6 and U1A have to traverse are clearly different and are not an intrinsic property of the topology which they share.

The Position of the Transition State of S6 May Be Changed by Mutagenesis. Since wild-type S6 has constant m_u and m_f values, its TSE retains the same $\beta^{\ddagger}$ value over the entire range of denaturant concentrations. Nevertheless, protein engineering could conceivably change the free energy profile sufficiently to delocalize the TSE if this ensemble is perched on top of a minor "bump". We recently described this phenomenon for CI2 (*18*). Wild-type CI2 has a constant $\beta^{\ddagger}$ value, yet when it is destabilized by major hydrophobic deletions, in which Leu, Ile, or Phe is replaced by Ala, distinct downward curvature in the unfolding plot is obtained in five cases. This behavior is even more pronounced for S6. Downward curvature in the unfolding limb is apparent, albeit to different degrees, in the chevron plots of 15 out of 26 S6 mutants (data not shown). Here we concentrate on

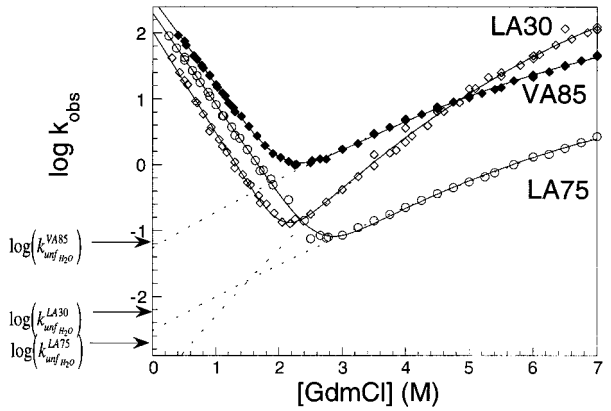

FIGURE 4: Chevron plots of the three S6 mutants LA30 (◇), LA75 (○), and VA85 (◆). Note the curvature in the unfolding limb. Data are fitted to a two-state model with parabolic curvature in the unfolding limb. The dashed lines are curved extrapolations of the unfolding data to water. Also indicated are the theoretical values for log k_u in water, calculated from the equation $\log(k_u^{water}) = \log(K_{D-N}^{water}) - \log(k_f^{water})$, where $\log(k_f^{water})$ is based on the short linear extrapolation from 0.3 M GdmCl and $\log(K_{D-N}^{water})$ is derived from equilibrium experiments. Individual values of m_{D-N} are used to calculate $\log(K_{D-N}^{water})$. The discrepancy between the calculated and extrapolated values for LA30's $\log(k_u^{water})$ is greatly reduced if a linear extrapolation from the region between 2.5 and 4 M GdmCl is used rather than a curved extrapolation.

three mutants, namely, LA30, LA75, and VA85 (chevron plots shown in Figure 4). These mutations truncate hydrophobic side chains which are completely buried in S6's core, destabilizing the protein by ~2−3 kcal/mol. The unfolding data of the S6 mutants fit well to a second-order polynomial, implying a smooth shift through a series of TSEs at different denaturant concentrations (Table 2). We have determined the crystal structure of the most destabilized mutant (LA30) and found it to be essentially identical to that of the wild type (Cα rms of 0.218 Å). There are some minor side chain rearrangements around the site of mutation, but the cavity caused by the truncation of Leu30 does not collapse. Thus, the native state is not significantly perturbed by these mutations.

No Spectroscopic Evidence for Unfolding Intermediates. The existence of unfolding intermediates has previously been suggested (in part) on the basis of burst phases (changes in fluorescence signal within the dead time of mixing) (*37*). Unfolding intermediates would be expected to have a fluorescence signal different from that of the native state because of at least partial loss of tertiary interactions. However, the unfolding amplitudes of the three S6 mutants do not decline with increasing GdmCl concentration, not even in the presence of 1 M NaCl which might otherwise stabilize such an intermediate relative to the transition state (data not shown). Furthermore, the dead-time fluorescence spectra of the species from which LA30 and wild-type S6 unfold are identical, showing that our spectroscopic probes (predominantly Trp62) experience the same apolar environment in the dead-time species of both proteins (Figure 5).

Denaturant Concentration versus Denaturant Activity. The curvature is only visible when log k_{obs} is plotted against denaturant concentration. On the basis of the solvation of nonpolar amino acid side chains at different GdmCl concentrations, concentration may be converted to activity (*32*, *36*). This yields reasonably straight chevron plots for the three mutants, but gives that of wild-type S6 a pronounced upward curvature in the unfolding limb and a downward curvature in the refolding limb (data not shown); the upward curvature in the unfolding limb is not easily rationalized either by broad barriers or by unfolding intermediates. Therefore, we retain concentration rather than activity as a measure of denaturant strength.

Calculating Activation Barriers on an Absolute Energy Scale. The broad barrier of U1A and the three S6 mutants can be reconstructed on an absolute energy scale using the transition state assumption

$$\Delta G_{N-\ddagger} = RT[\ln(10^6) - \ln(k_u)] \quad (8)$$

The term for vibrational frequency ($\kappa k_B T/h \approx 10^{13}$ s^{-1}), which represents the fastest possible rate at which the transition state can decay in simple chemical reactions, is here replaced with the rate 10^6 s^{-1}, which represents the elementary step in tertiary structure formation, namely, closing of a loop (*38*). However, the absolute value of this factor does not affect the general shape of the broad barrier. The free energy is calculated as a function of the reaction coordinate β using the following relationship:

$$\Delta G_i = \Delta G_j + m_f^j([\text{GdmCl}]_i - [\text{GdmCl}]_j) = \Delta G_j + \beta_j^{\ddagger} m_{D-N}([\text{GdmCl}]_i - [\text{GdmCl}]_j) \quad (9)$$

where the *j* values are those at the measured conditions while the *i* values are at a given extrapolated condition. Thus, for example, the m_f value at a given GdmCl concentration allows us to calculate $\beta^{\ddagger}$ (eq 6), while ΔG_j is calculated from k_f using eq 8 and substituting k_u for k_f. The individual steps are described in detail in Figure 9. In all cases, the barriers are broad and relatively flat (Figure 7). Note how the position of the TSE shifts toward the native state as the denaturant concentration increases. The S6 wild type has a constant $\beta^{\ddagger}$ value, i.e., a TSE which corresponds to a point localized on a "sharply" pointed feature.

DISCUSSION

Curved Chevron Plots Are Better Explained by Changes at the Transition State Level Rather Than at the Ground State Level. In this study, we show that mutagenesis of S6 gives rise to visible curvature in the chevron plot's unfolding limb which is not observed for the wild-type protein. Curvature is also seen for U1A, a structural homologue of S6. The downward curvature of the chevron plot of S6 mutants means that they unfold more slowly at high GdmCl concentrations than was predicted from linear extrapolation of unfolding rates at low GdmCl concentrations. One trivial technical explanation is that the rapidity of the unfolding step leads to an underestimate of the unfolding rate; however, the dead time of the instrument (≤6 ms with 1:10 mixing under constant pressure) means that only one reaction half-time is lost within this dead time for the fastest measured unfolding rate (120 s^{-1} for LA30 at 7 M GdmCl). Furthermore, LA75 unfolds much more slowly than LA30, but its plot still exhibits pronounced curvature.

Curvature as such is not a new phenomenon, but it is traditionally ascribed to changes at the ground state level, i.e., the transient accumulation of intermediates within the

Table 2: Curvature in the Chevron Plots for the Three S6 Mutants

	m_{D-N} (M^{-1})	k_f^{H2O} (s^{-1})	m_f (s^{-1} M^{-1})	$k_u^{H_2O}$ ($\times 10^3$ s^{-1})[a]	$k_u^{H_2O}$ ($\times 10^3$ s^{-1})[b]	a^{m_u} (s^{-1} M^{-1})	b^{m_u} (s^{-1} M^{-2})	$\beta^{\ddagger}$ [c]
wild type	1.77 ± 0.05	297 ± 6	−1.24 ± 0.01	0.43 ± 0.003	0.44 ± 0.01	0.51 ± 0.004	0.009 ± 0.004	0.73−0.73
LA30	2.04 ± 0.12	173 ± 3	−1.53 ± 0.03	0.49 ± 0.02	28.9 ± 3.2	1.14 ± 0.05	−0.053 ± 0.005	0.51−0.79
LA75	1.95 ± 0.17	305 ± 6	−1.38 ± 0.01	0.97 ± 0.04	5.76 ± 0.27	0.72 ± 0.04	−0.033 ± 0.004	0.69−0.87
VA85	1.68 ± 0.12	341 ± 3	−1.30 ± 0.01	33.1 ± 1.02	55.7 ± 6.8	0.65 ± 0.01	−0.030 ± 0.002	0.72−0.88

[a] Extrapolated from unfolding data using parameters from a two-order polynomial fit. [b] Calculated from refolding data and equilibrium data according to a two-state system. [c] Position of the TSE on the reaction coordinate over the range of 2−7 M GdmCl.

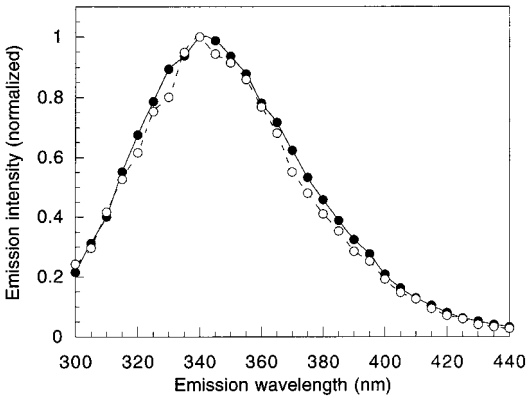

FIGURE 5: Normalized dead-time spectra of the species from which the S6 wild type (●) and the mutant LA30 (○) unfold at 7 M GdmCl. There is no difference between the two proteins; in other words, the curvature of LA30's unfolding plot cannot be explained by the burst phase accumulation of an intermediate with fluorescence properties that are different from those of the native state.

dead time of the experiment (*14*). In the case of S6, this corresponds to a kinetic scheme in which an unfolding intermediate I rapidly equilibrates with the native state in the dead time of the experiment (*11*, *15*) (analogous to Scheme 2). Increasing the denaturant concentration would increase the occupancy of this intermediate because it is less sensitive to denaturant than the native state, since it has a more solvent-exposed surface. This leads to a flattening of the unfolding limb for two interrelated reasons. (1) The change in the extent of solvent exposure of the detected unfolding step is decreased, as it changes from a native-to-transition-state step to an intermediate-to-transition-state step. This leads to a decrease in m_u. (2) As the intermediate is more stable than the native state, the height of the unfolding barrier increases and k_u decreases compared to what it would be in the absence of I. Mutagenesis could increase the stability of I relative to that of N by removing stabilizing interactions which are found in N but not in I; this strategy has been used to stabilize the refolding intermediate of barnase (*39*). An unfolding intermediate is suggested to accumulate during unfolding of the 89-residue barstar and its mutant W38F/W44F (*37*). At 40 °C, unfolding of barstar in urea was also accompanied by a burst phase, as monitored both by CD and by fluorescence (depending on the unfolding conditions), and this was taken as further evidence for rapid accumulation of the intermediate before the subsequent unfolding step.

However, the curvatures of the rate constants reflect solely a change in the extent of solvent exposure of the difference between the ground state and the transition state ensemble for the unfolding reaction, and the very same kinetic patterns could arise from changes in the composition of the TSE instead of changes at the ground state level. These are two

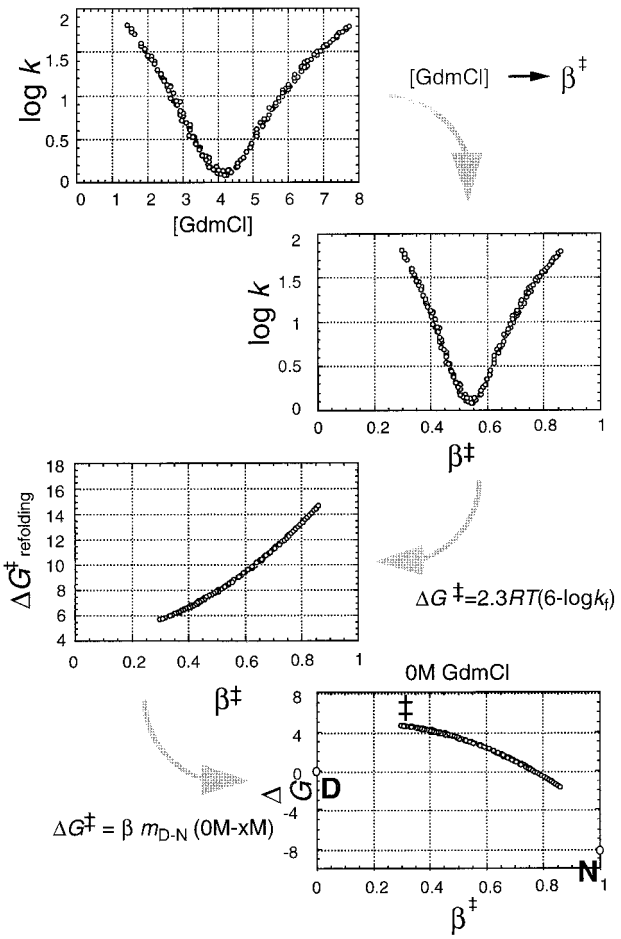

FIGURE 6: Calculation of the activation barriers of wild-type U1A and three S6 mutants. The *x*-axis is first converted to $\beta^{\ddagger}$ values according to eq 6, after which the activation barrier for either folding is calculated from the refolding rates (eq 8). At a given denaturant concentration (0 M GdmCl), the height of the barrier is finally calculated as a function of $\beta^{\ddagger}$ from eq 9.

fundamentally different phenomena, which at least in principle can be distinguished by careful analysis of the dead-time species for the unfolding and refolding reactions. If the ground state changes, this should be seen at the level of the spectroscopic properties of the starting material, whereas if the changes occur at the transition state level, the dead-time species would simply be the denatured state or the native state.

The possible existence of unfolding intermediates in S6 is a crucial point and cannot be ruled out entirely. However, we have several arguments against the presence of an unfolding intermediate for LA30. First, the fluorescence unfolding amplitudes reveal no burst phase accumulation of intermediate between 3 and 7 M GdmCl, not even in the presence of 1 M NaCl (data not shown), and the dead-time fluorescence spectra of the species from which LA30 and

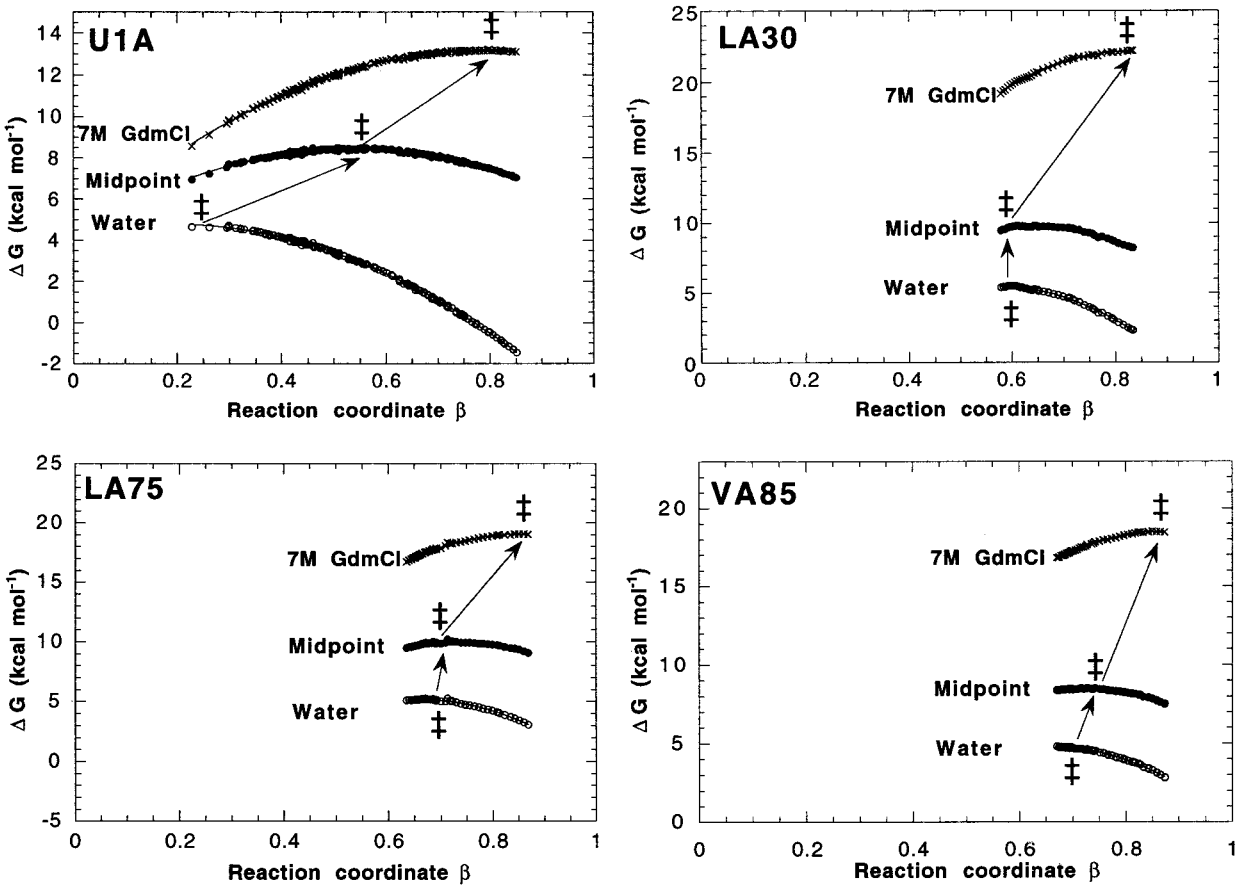

FIGURE 7: Broad barriers of U1A and the three S6 mutants calculated at different denaturant concentrations (constructed using eqs 6 and 7). The denatured state and the native state have reaction coordinates of 0 and 1, respectively. All free energies are relative to the denatured state (set to 0 at all GdmCl concentrations).

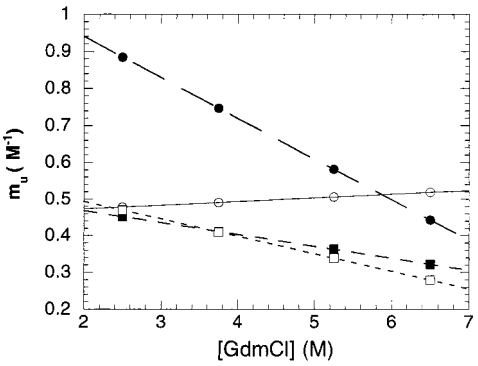

FIGURE 8: m_u plotted vs GdmCl concentration for the S6 wild type (○), LA30 (●), LA75 (■), and VA85 (□). The m_u values are based on the fitted parameters from the plots shown in Figure 6. Note that the S6 wild type has a virtually constant m_u value (there is a small upward curvature in the unfolding plot), while LA30's m_u value is much higher than that of the wild type at low denaturant concentrations but crosses over at 5.9 M GdmCl. The m_u values of LA75 and VA85 are lower than that of the wild type throughout.

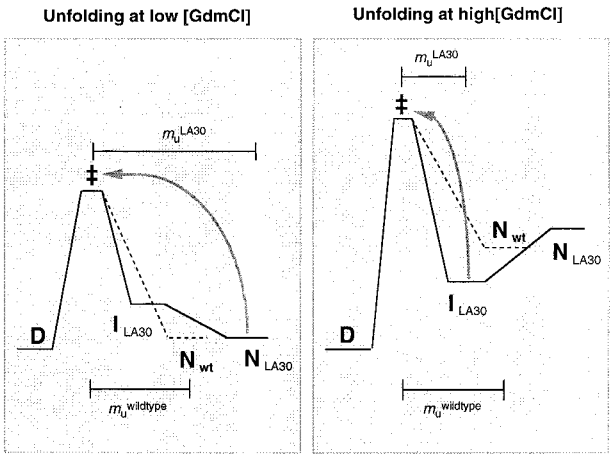

FIGURE 9: Unfolding of LA30 through a putative unfolding intermediate is difficult to reconcile with the high m_u value observed near the transition region of unfolding (low GdmCl concentrations in the figure), since it would require the native state of LA30 to be significantly more compact than the wild-type native state.

wild-type S6 unfold are identical (Figure 8). We have not been able to study these experiments by far-UV CD. Nevertheless, it is not expected that a burst phase signal would be invisible by fluorescence, yet manifest itself by far-UV CD; such an observation would suggest loss of secondary structure before loss of tertiary structure, which is an unlikely order of events. In the unfolding of wild-type barstar and the mutant W38F/W44F, all burst phase changes in the far-UV CD are accompanied by fluorescence burst phase changes, but the converse is not the case (*37*). Note, however, that the lack of a burst phase is not in itself conclusive evidence for an intact native protein at high GdmCl concentrations. For example, the unfolding intermediate of RNase A can only be detected by NMR (*15*).

But, the second, stronger, piece of evidence is the higher m_u value of LA30 at GdmCl concentrations close to the transition region (Figure 8). The m_u value represents the difference in the extent of solvent exposure between the rate-limiting transition state of unfolding and the state of lowest

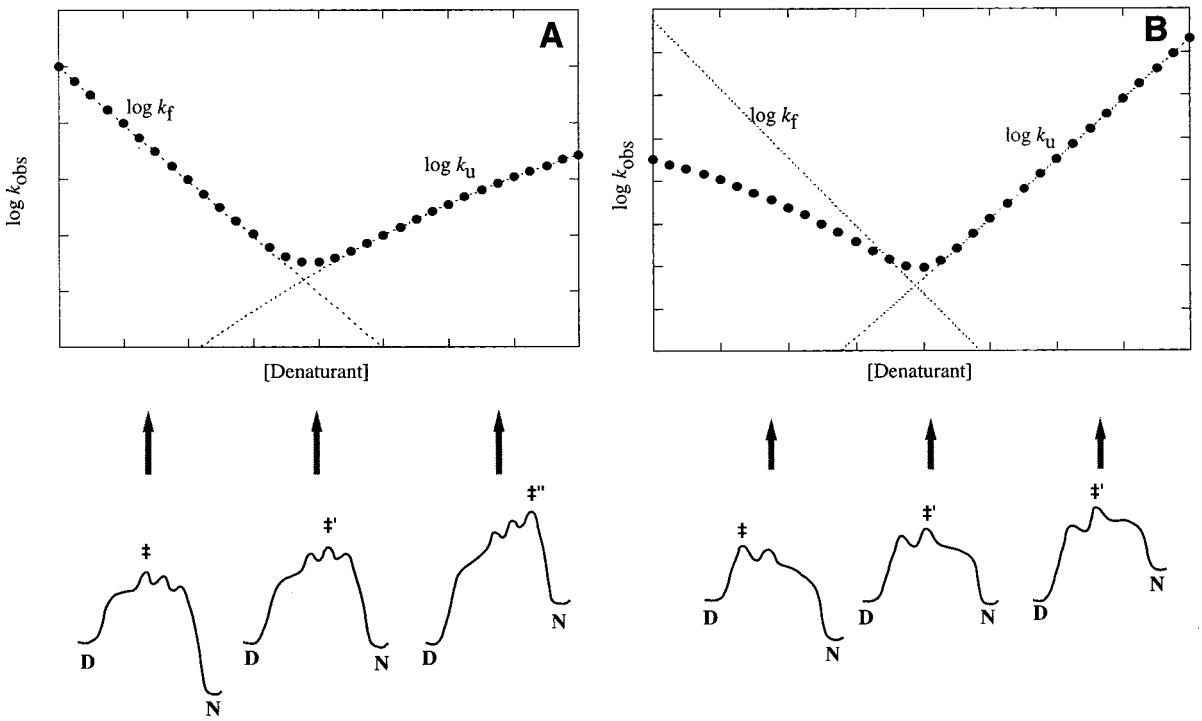

FIGURE 10: (A) Pictorial representation of the shift in the transition state position envisaged to occur in LA30 by a combination of erosion of the wild-type barrier and elevation of other barriers. The bumps in LA30's broad barrier are distributed in such a way that the position of the rate-limiting ensemble (the TSE) only starts to shift at high denaturant concentrations. The TSE remains fixed at low denaturant concentrations because no other ensembles are sufficiently destabilized under these conditions to reach a higher energy level. This means that the chevron plot only becomes curved at high denaturant concentrations, i.e., under unfolding conditions. (B) Broad barriers which may give rise to apparent three-state kinetics. In this case, the position of the TSE only shifts at low denaturant concentrations as the highest point on the free energy profile changes from the first bump to the second between 0 and 4 M denaturant, giving rise to curvature or rollover in the refolding limb of the chevron plot. Rollover is also observed when a refolding intermediate accumulates transiently (*13*, *14*).

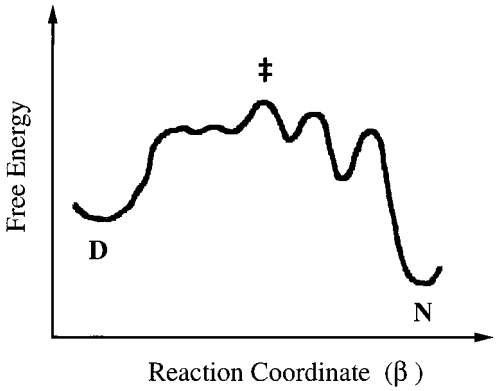

FIGURE 11: The ruggedness of the free energy profile increases as the TSE becomes more native-like and can be trapped in various local minima.

free energy preceding it. The m_u value is expected to decrease below the wild-type value if unfolding takes place from an intermediate, which is less compact than the native state (Figure 9). However, the m_u value of LA30 observed at low denaturant concentrations (0.94 s^{-1} M^{-1} at 2.0 M GdmCl) is twice as high as that of wild-type S6 (0.47 s^{-1} M^{-1}). If a constant position of the TSE is assumed, the high m_u value would have to be rationalized as a compactness of the native state that is significantly greater than that of wild-type S6 (Figure 11). This is unwarranted, as there is no significant difference between the crystal structures of wild-type S6 and LA30. It is simpler to explain the high m_u value as a shift of the TSE so that the transition state is closer to the unfolded state (and therefore gives the unfolding rate a higher m_u value) at low denaturant concentrations than at high denaturant concentrations (Figure 10A).

An attractive feature of this mobile TSE model is that it provides a simple explanation for how small changes induced by mutagenesis can apparently produce radical changes in the folding behavior. A related picture has arisen from recent theoretical work, in which the TSE is considered to be a collection of structurally heterogeneous states with a relatively high degree of native structure in common (*40*−*42*). More specifically, Wolynes and co-workers suggest a TSE with many delocalized nuclei whose relative populations vary with the conditions (*43*).

Only Small Changes of the Free Energy Profile May Turn a Fixed Transition State into a Moving Transition State. The free energy profile of U1A is envisaged to be a broad and smooth plateau, in which conformations with very different reaction coordinates have very similar free energies. The TSE is the ensemble with the highest free energy, but this energy is close to that of other conformations; therefore, the position of the TSE will be perturbed by solvent conditions. Different ensembles have different degrees of compactness and therefore different sensitivities to denaturant. This means that changes in denaturant concentration will affect the relative stabilities of the different ensembles, moving the position of the TSE closer to the native state with increasing denaturant concentration, since more native-like states are destabilized by denaturant to a higher degree than less native-like states. This somewhat geometric interpretation follows directly from mass action and the definition of the $\beta^{\ddagger}$ value. A seemingly analogous phenomenon is described by the

Hammond postulate regarding two neighboring states in a reaction scheme; as the difference in free energy between the two states decreases, they also approach each other in structure.

In contrast to that of U1A, the broad folding barrier of S6 is only manifested through mutagenesis. S6 and U1A exhibit no sequence homology; they may represent convergent evolution toward the same native structure which they appear to reach through a folding free energy profile with different local contours. We suggest that the lack of a unique energetically elevated feature in U1A's profile leads to a mobile TSE, whereas a fixed TSE, such as that of wild-type S6, may result from a small pointed feature projecting from the top of a broad free energy profile. Erosion or relative suppression of this local bump by mutagenesis, through the emergence of other pointed features in the free energy profile, will lead to broad barrier behavior (Figure 10A). Depending on the distribution and shape of these features, the curvature in the chevron plot may be symmetrical (as in the case of U1A) or asymmetrical (as may be the case for LA30, LA75, and VA85). Analogously, the disappearance of bumps close to the denatured state may lead to the rollover at low concentrations, as is usually seen with folding intermediates (*14*) (Figure 10B).

Only a few kilocalories per mole, corresponding to the energy of a few hydrophobic interactions, may separate a fixed and a moving TSE. Therefore, it appears reasonable that individual side chain truncations in the hydrophobic core of S6, when sufficiently destabilizing, can rock the immobility of the transition state. In fact, the three mutants LA30, LA75, and VA85 destabilize S6 by 1.8−2.6 kcal/mol. In contrast, mutations which destabilize S6 to a smaller extent manifest no unfolding curvature (data not shown).

Can Curvature Be Extrapolated to Water? An important question is whether the curvature in the unfolding region of S6 can be justifiably extrapolated to GdmCl concentrations below the unfolding range. Although there is no theoretical justification for a quadratic relationship between log k_u and GdmCl concentration (*32*), the symmetry of U1A's chevron plot indicates nonlinearity throughout the denaturant concentration range (*11*). In the case of S6, it is less obvious whether curvature confines itself to the unfolding limb. The curvature is most visible in the unfolding limb for two reasons: the degree of unfolding is measured over a wider denaturant concentration range than that of refolding, and the overall slope is much shallower. Therefore, it is simply not possible to resolve the curvature in the refolding region, and satisfactory fits are obtained irrespective of whether one uses symmetrical curvature or limits oneself to curvature in the unfolding limb.

In a two-state system where log K_{D-N} depends linearly on GdmCl concentration, any curvature for the unfolding rate constant has to be matched by a similar but opposite curvature for the refolding rate constant over the same denaturant concentration interval. This cannot be verified directly but only by extrapolation. It is very difficult to discern any curvature in the refolding data because of the steep slope in this part of the chevron plot. Extrapolation of LA30's log k_u to water based on the observed curvature leads to a significant deviation from the value predicted from refolding and equilibrium data (Figure 8), but the deviation is greatly diminished if a linear extrapolation is made from the unfolding region close to the transition midpoint. The deviations are smaller in the case of LA75 and VA85, which do not show such a high curvature. If the unfolding rates are not curved below the transition midpoint, there should be no curvature in the refolding data. Consequently, the transition state of S6 mutants, particularly LA30, may only become mobile at high denaturant concentrations. In fact, the coupled dissociation and unfolding of the wild-type Arc repressor dimer are also characterized by curvature at high denaturant concentrations, but a series of elegant energy transfer experiments which determine unfolding rates under folding conditions have shown no deviation from linearity at low denaturant concentrations (*44*). Thus, the Arc repressor dimer's TSE may have rugged features in common with that of S6, so that barrier changes only occur at higher denaturant concentrations.

What Causes the Ruggedness in the Free Energy Folding Profile? As the protein becomes more native-like on the transition state level, structural adjustments involve larger conformational changes and hence larger local energy barriers (*41*). Exactly how this translates to free energy depends on the statistical distribution of microstates in the energy landscape. However, it is likely that pronounced energy barriers, significantly larger than the energy of thermal motion (a few k_BT units), will also be seen on the free energy scale (Figure 11). Such increasing ruggedness of the energy landscape could even lead to localized pathways near the native state, which may change upon mutation, hence causing shifts of TSE. The free energy profiles of U1A and to some extent the three S6 mutants reconstructed in Figure 6 do not exhibit this ruggedness, because the details of the free energy profile cannot be extracted from the kinetic data; these just allow us to calculate the position and energy of the highest points of the TSE (eq 7).

We have previously shown that some destabilizing mutations also cause the two-state protein CI2 to show broad barrier behavior (*18*). However, only 5 out of 100 CI2 mutants had this property; in contrast, 15 out of 26 S6 mutants constructed so far manifest unfolding curvature (data not shown). It is tempting to speculate that the propensity of S6 for unfolding curvature reflects its structural similarity to U1A and that the overall shape of the barrier is linked to protein topology. Nevertheless, it appears that S6 has more pronounced bumps in its free energy profile than U1A. The physical basis for this may be that the folding nucleus of S6 is more specific and less delocalized than that of U1A, making it very sensitive to individual mutations. U1A, on the other hand, may have a more diffuse core which gives the protein freedom to fold in different ways, hence avoiding local barriers. As a consequence, U1A's behavior does not easily respond to mutations, and the only way to change its kinetics to classical two-state behavior (i.e., a V-shaped chevron) is to introduce a more specific core, perhaps via salt links or disulfide bridges.

Broad Activation Barriers Also Seem To Occur in Other Systems. Other experimental and theoretical studies have also shown that highly destabilizing mutations expose new transition states (*42*−*47*). In accord with the Hammond postulate, small changes in the free energy of the TSE lead to small changes in its position on the reaction coordinate for barnase and CI2 (*48*, *49*). These changes are more pronounced than those observed in organic chemistry and

enzyme catalysis. Although virtually all mutants of CI2 retain a fixed position of the transition state, the $\beta^{\ddagger}$ values are distributed over the range of 0.5−0.9 (*49*).

For the dimeric Arc repressor, large changes in the position of the reaction coordinate are induced not only by amino acid substitution (*47*) but also by denaturant (*44*). This was rationalized as a broad activation barrier with a bumpy energy plateau where mutations and denaturant change the positions of the highest bumps (*47*). However, it was later suggested that the different bumps were separated by considerable energy barriers and that the plots of $\log(k_u)$ versus urea should be considered kinked rather than smoothly curved as one unstable intermediate replaces another as the rate-limiting step (*44*), while the actual positions of the kinetic barriers on the free energy profile are not sensitive to denaturant concentration. At this stage, the difference between a series of unstable intermediates and a broad activation barrier is largely semantic, since the unstable intermediates may be positioned between the local bumps on a free energy plateau as previously noted (cf. Figure 6B in ref *44*). Since the broad barrier's free energy profile is expected to become more rugged closer to the native state, unfolding curvature will in fact probably reflect underlying kinks as one bump replaces another as the TSE. However, it is difficult to distinguish smooth curvature from kinks; several kinks in succession will give rise to apparent smooth curvature. Curvature in the refolding region, on the other hand, will according to this view be more likely to reflect smooth movements of the TSE in the vicinity of the denatured state where the ensembles along the reaction coordinate are most delocalized in the conformational space.

With their 101 and 102 residues, respectively, S6 and U1A probably represent the upper size limit for proteins which fold without intermediates; barnase, which folds via a transient intermediate (*13*), consists of 110 residues. It has been suggested that the issue of two-state versus multistate folding is a question of protein size, and that proteins larger than around 100 residues transiently accumulate intermediates during folding (*50*). This may explain why broad activation barriers are revealed readily for two-state proteins in the 100-residue range. Intermediates in this case are too unstable to accumulate even transiently, but they may have equivalents on the transition state level, namely, conformations separated by pronounced energy barriers or bumps, whose heights depend on solvent conditions (*18*).

Structurally Homologous Proteins Exhibit Differences in Folding Behavior. Although S6 and U1A are structurally homologous and exhibit the same broad activation barrier of folding, they exhibit weak evolutionary kinship at the level of the primary sequence. This implies that they will not exhibit the same side chain interactions in the transition state. Even homologous proteins with fixed TSEs differ in folding behavior. For example, human lysozyme deviates from hen lysozyme in 51 of 129 residues, only 27 of which involve changes in charge or size; as a consequence, the α-domain's structural elements do not form so cooperatively, there is no overshoot in the far-UV CD signal, and the burst phase collapsed state contains tertiary as well as secondary interactions, while hen lysozyme only has secondary interactions (*51*). Hen lysozyme also differs from the structurally homologous equine lysozyme and α-lactalbumin in that the latter two unfold in two stages in calorimetric studies (*52*, *53*), whereas lysozyme follows a two-state unfolding transition (*54*). The partially unfolded states which α-lactalbumin and equine lysozyme populate transiently during folding diverge significantly in enthalpy and tertiary structure (*53*). Unlike α-lactalbumin and equine lysozyme, hen lysozyme does not possess a calcium binding site, but this cannot be the major reason for the difference in unfolding behavior, since the introduction of such a site into human lysozyme did not lead to two stages of unfolding (*55*). Another explanation is that equine lysozyme and α-lactalbumin contain two separate hydrophobic cores which merge into one in hen lysozyme, and this may explain the difference (*56*). However, Rothwarf and Scheraga (*57*) have suggested that individual side chains and disulfide bridges in lysozyme may lead to the formation of non-native contacts and thus stabilize partially misfolded intermediates. One such mutation (*57*) seems to have entirely eliminated the direct rapid-folding pathway through which at least 14% of wild-type lysozyme pass (*58*), showing how sensitive folding pathways can be to individual mutations.

Implications for Protein Folding. In the nucleation−condensation model of protein folding (*12*, *59*), the criterion for successful folding is the formation of a critical number of nonlocal and local contacts which stabilize the transition state sufficiently to allow it to descend to the native state rather than back to the unfolded state. These nucleation residues are likely to vary with sequence and in some cases (*60*), but not others (*61*), with chain connectivity, even when the native contact pattern is unchanged.

The picture becomes even more complex when broad activation barriers are introduced, since this couples solvent conditions with the amino acid composition as factors which determine the structure of the TSE. The folding nuclei of several small proteins have been identified by Φ value analysis (*12*, *35*) in which the individual residues are truncated to measure how native-like their interactions are in the transition state. However, this presupposes that such truncations act as probes and do not perturb the system. When individual amino acid substitutions affect the local contours of the free energy profile, as observed with S6, the Φ value approach has to be modified. Nevertheless, this presents an opportunity as well as an added complexity because it opens up new regions of the free energy profile to experimental observation. Thus, one can in principle determine the extent to which different interactions are formed at different $\beta^{\ddagger}$ values, provided the reference protein, as well as the mutants, spans a wide enough range of $\beta^{\ddagger}$ values. We are currently engaged in a protein engineering study of U1A and S6 both to determine how mutations alter this landscape and to identify dynamic folding nuclei in this way.

ACKNOWLEDGMENT

We thank Dr. K. Nagai for the plasmid expressing F56W U1A and Mona Fata for help with development of the expression system.

REFERENCES

1. Levinthal, C. (1968) *J. Chim. Phys. 85*, 44−45.
2. Kim, P. S., and Baldwin, R. L. (1990) *Annu. Rev. Biochem. 59*, 631−660.

3. Karplus, M., and Weaver, D. C. (1976) *Nature 260*, 404−406.
4. Jackson, S. E., and Fersht, A. R. (1991) *Biochemistry 30*, 10428−10435.
5. Alexander, P., Orban, J., and Bryan, P. (1992) *Biochemistry 31*, 7243−7248.
6. Viguera, A. R., Martinez, J. C., Filimonov, V. V., Mateo, P. L., and Serrano, L. (1994) *Biochemistry 33*, 2142−2150.
7. Schindler, T. S., Herrier, M., Marahiel, M. A., and Schmid, F. X. (1995) *Nat. Struct. Biol. 2*, 663−673.
8. Kragelund, B. B., Robinson, C. V., Knudsen, V., Dobson, C. M., and Poulsen, F. M. (1995) *Biochemistry 34*, 7217−7224.
9. Huang, G. S., and Oas, T. G. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 6878−6872.
10. Villegas, V., Azuaga, A., Catasus, L. I., Reverter, D., Mateo, P. L., Aviles, F. X., and Serrano, L. (1995) *Biochemistry 34*, 15105−15110.
11. Silow, M., and Oliveberg, M. (1997) *Biochemistry 36*, 7633−7637.
12. Itzhaki, L. S., Otzen, D. E., and Fersht, A. R. (1995) *J. Mol. Biol. 254*, 260−288.
13. Matouschek, A., Kellis, J. T., Serrano, L., Bycroft, M., and Fersht, A. R. (1990) *Nature 346*, 440−445.
14. Baldwin, R. (1996) *Folding Des. 1*, R1−R8.
15. Kiefhaber, T., Labhardt, A. M., and Baldwin, R. L. (1995) *Nature 375*, 513−515.
16. Sauder, J. M., MacKenzie, M. E., and Roder, H. (1996) *Biochemistry 35*, 16852−16862.
17. Nagai, K., Oubridge, C., Jessen, T.-H., Li, J., and Evans, P. R. (1990) *Nature 348*, 515−520.
18. Oliveberg, M., Tan, Y.-J., Silow, M., and Fersht, A. R. (1998) *J. Mol. Biol. 277*, 933−943.
19. Lindahl, M., et al. (1994) *EMBO J. 13*, 1249−1254.
20. Branden, C. I., and Tooze, J. (1991) *Introduction to Protein Structure*, Garland, New York.
21. Chothia, C., and Finkelstein, A. V. (1990) *Annu. Rev. Biochem. 59*, 1007−1039.
22. Simpson, G. G., Clark, G. P., Rothnie, H. M., Boelens, W., van Venrooij, W., and Brown, J. W. S. (1995) *EMBO J. 14*, 4540−4550.
23. Clarke, J., and Fersht, A. R. (1993) *Biochemistry 32*, 4322−4329.
24. Herning, T., Yutani, K., Taniyama, Y., and Kikuchi, M. (1991) *Biochemistry 30*, 9882−9891.
25. Brandts, J. F., Halvorson, H. R., and Brennan, M. (1975) *Biochemistry 14*, 4953−4963.
26. Schmid, F. X., and Baldwin, R. L. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4764−4768.
27. Silow, M., and Oliveberg, M. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 6084−6086.
28. *The SERC (UK) Collaborative Computing Project No. 4* (1979) Daresbury Laboratory, Warrington, U.K.
29. Brünger, A. T. (1987) *Science 238*, 1403−1406.
30. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A47*, 110−119.
31. Lu, J., and Hall, K. B. (1997) *Biophys. Chem. 64*, 111−119.
32. Tanford, C. (1970) *Adv. Protein Chem. 24*, 1−95.
33. Burton, R. E., Huang, G. S., Daugherty, M. A., Fullbright, P. W., and Oas, T. G. (1996) *J. Mol. Biol. 263*, 311−322.
34. Chen, B., Baase, W. A., Nicholson, H., and Schellman, J. A. (1992) *Biochemistry 31*, 1464−1476.
35. Fersht, A. R., Matouschek, A., and Serrano, L. (1992) *J. Mol. Biol. 224*, 771−782.
36. Parker, M. J., Spencer, J., and Clarke, A. R. (1995) *J. Mol. Biol. 253*, 771−786.
37. Zaidi, F. N., Nath, U., and Udgaonkar, J. B. (1997) *Nat. Struct. Biol. 4*, 1016−1023.
38. Hagen, S. J., Hofrichter, J., Szabo, A., and Eaton, W. A. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 11615−11617.
39. Sanz, J. M., and Fersht, A. R. (1993) *Biochemistry 32*, 13584−13592.
40. Onuchic, J. N., Socci, N. D., Luthey-Schulten, Z., and Wolynes, P. G. (1996) *Folding Des. 1*, 441−450.
41. Bryngelson, J., Onuchic, J. N., Socci, N. D., and Wolynes, P. (1995) *Proteins: Struct., Funct., Genet. 21*, 167−195.
42. Sali, A., Schaknovich, E. I., and Karplus, M. (1994) *Nature 369*, 248−251.
43. Plotkin, S. S., Wang, J., and Wolynes, P. G. (1997) *J. Chem. Phys. 106*, 1−17.
44. Jonsson, T., Waldburger, C. D., and Sauer, T. (1996) *Biochemistry 35*, 4795−4802.
45. Finkelstein, A. V., and Badredtinov, A. Y. (1997) *Folding Des. 2*, 115−121.
46. Pande, V. S., Grosberg, A. Y., and Tanaka, T. (1997) *Folding Des. 2*, 109−114.
47. Milla, M. E., Brown, B. M., Waldburger, C. D., and Sauer, R. T. (1995) *Biochemistry 34*, 13914−13919.
48. Matouschek, A., and Fersht, A. R. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 7814−7818.
49. Matouschek, A., Otzen, D. E., Itzhaki, L. S., Jackson, S. E., and Fersht, A. R. (1995) *Biochemistry 34*, 13656−13662.
50. Muñoz, V., Lopez, E. M., Jager, M., and Serrano, L. (1994) *Biochemistry 33*, 5858−5866.
51. Hooke, S. D., Radford, S. E., and Dobson, C. M. (1994) *Biochemistry 33*, 5867−5876.
52. Griko, Y. V., Freire, E., and Privalov, P. I. (1994) *Biochemistry 33*, 1889−1899.
53. Van Dael, H., Haezebrouck, P., Morozova, L., Aricomuendel, C., and Dobson, C. M. (1993) *Biochemistry 32*, 11886−11894.
54. Privalov, P. L., and Khechinashvili, N. N. (1974) *J. Mol. Biol. 86*, 665−684.
55. Kuroki, R., Kawakita, S., Nakamura, H., and Yutani, K. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 6803−6807.
56. Privalov, P. L. (1996) *J. Mol. Biol. 258*, 707−725.
57. Rothwarf, D. M., and Scheraga, H. A. (1996) *Biochemistry 35*, 13797−13807.
58. Kiefhaber, T. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 9029−9033.
59. Fersht, A. R. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 10869−10873.
60. Viguera, A. R., Serrano, L., and Wilmanns, M. (1996) *Nat. Struct. Biol. 3*, 874−880.
61. Otzen, D. E., and Fersht, A. R. (1998) *Biochemistry 37*, 8139−8146.
62. Avis, J. M., Allain, F. H., Howe, P. W., Varani, G., Nagai, K., and Neuhaus, D. (1996) *J. Mol. Biol. 257*, 398−411.
63. Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
64. Wilson, K. S., Appelt, K., Badger, J., Tanaka, I., and White, S. W. (1986) *J. Mol. Biol. 83*, 7251−7255.
65. Golden, B. L., Ramakrishnan, V., and White, S. W. (1993) *EMBO J. 12*, 4901−4908.
66. Hoffmann, D. W., Wuery, C. C., Golden, B. L., White, S., and Keene, J. D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2495−2499.
67. Leijonmarck, M., Eriksson, S., and Liljas, A. (1980) *Nature 286*, 824−826.
68. Adman, E. T., Sieker, J. C., and Jensen, L. H. (1973) *J. Biol. Chem. 248*, 3987−3996.
69. Steele, R. A., and Opella, S. J. (1997) *Biochemistry 36*, 6885−6895.

BI982819J