# Kinetic Evidence for Folding and Unfolding Intermediates in Staphylococcal Nuclease†

William F. Walkenhorst,‡ Susan M. Green,§,|| and Heinrich Roder*,‡,⊥

*Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111, Department of Biophysics & Biophysical Chemistry, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, and Department of Biochemistry & Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6059*

*Received January 8, 1997*⊗

ABSTRACT: The complex kinetic behavior commonly observed in protein folding studies suggests that a heterogeneous population of molecules exists in solution and that a number of discrete steps are involved in the conversion of unfolded molecules to the fully native form. A central issue in protein folding is whether any of these kinetic events represent conformational steps important for efficient folding rather than side reactions caused by slow steps such as proline isomerization or misfolding of the polypeptide chain. In order to address this question, we used stopped-flow fluorescence techniques to characterize the kinetic mechanism of folding and unfolding for a Pro$^-$ variant of SNase in which all six proline residues were replaced by glycines or alanines. Compared to the wild-type protein, which exhibits a series of proline-dependent slow folding phases, the folding kinetics of Pro$^-$ SNase were much simpler, which made quantitative kinetic analysis possible. Despite the absence of prolines or other complicating factors, the folding kinetics still contain several phases and exhibit a complex denaturant dependence. The GuHCl dependence of the major observable folding phase and a distinct lag in the appearance of the native state provide clear evidence for an early folding intermediate. The fluorescence of Trp140 in the α-helical domain is insensitive to the formation of this early intermediate, which is consistent with a partially folded state with a stable β-domain and a largely disordered α-helical region. A second intermediate is required to model the kinetics of unfolding for the Pro$^-$ variant, which shows evidence for a denaturant-induced change in the rate-limiting unfolding step. With the inclusion of these two intermediates, we are able to completely model the major phase(s) in both folding and unfolding across a wide range of denaturant concentrations using a sequential four-state folding mechanism. In order to model the minor slow phase observed for the Pro$^-$ mutant, a six-state scheme containing a parallel pathway originating from a distinct unfolded state was required. The properties of this alternate unfolded conformation are consistent with those expected due to the presence of a non-prolyl cis peptide bond. To test the kinetic model, we used simulations based on the six-state scheme and were able to completely reproduce the folding kinetics for Pro$^-$ SNase across a range of denaturant concentrations.

Many proteins exhibit complex folding kinetics with multiple steps and a complicated denaturant dependence of rates and amplitudes. It is well-known that intrinsically slow conformational steps, such as isomerization of peptide bonds (Lin & Brandts, 1978; Schmid, 1992; Nall, 1994), disulfide bond formation or rearrangements (Creighton, 1990; Weissman & Kim, 1992), and metal ligand exchange (Sosnick et la., 1994; Elöve et al., 1994) can give rise to conformational heterogeneity and multiple parallel folding (or unfolding) pathways. While cis−trans isomerization of X−Pro peptide bonds is the most common source of slow folding events, cis isomers for non-proline peptide bonds have also been reported (Odefey et al., 1995; Vanhove et al., 1996; Dodge & Scheraga, 1996). If such complications can be avoided, folding is generally more efficient (i.e., a larger fraction of unfolded chains fold rapidly) and, in some cases, occurs without detectable intermediates [see, e.g., Jackson and Fersht (1991), Huang and Oas (1995), and Schindler et al. (1995)]. These observations have led to the generalization that accumulation of folding intermediates is not necessary for a protein to fold rapidly or may even slow down formation of the native structure (Sosnick et al., 1994; Creighton, 1994; Kiefhaber, 1995a; Fersht, 1995). On the other hand, formation of compact folding intermediates with high secondary structure content within the first few milliseconds of refolding is a common observation, even for some small single-domain proteins [reviewed in Matthews (1993), Ptitsyn (1995), and Roder and Colón (1997)]. Quantitative kinetic studies on several small proteins have shown that the effects of solution conditions or mutational perturbations on the kinetics of folding are consistent with a sequential folding mechanism

† This work was supported by NIH Grant GM35926 to H.R. and by NIH Grant CA06927 and an appropriation from the Commonwealth of Pennsylvania to the Institute for Cancer Research. W.F.W. was supported by an NIH Training Grant (T32 CA09035) and by an individual NIH NRSA Postdoctoral Fellowship (GM15971).

* Correspondence should be addressed to this author at the Fox Chase Cancer Center. Phone: (215) 728-3123. Fax: (215) 728-3574. E-mail: H_Roder@fccc.edu.

‡ Fox Chase Cancer Center.

§ Johns Hopkins School of Medicine.

|| Present address: Department of Macromolecular Sciences, SmithKline Beecham Pharmaceuticals, King of Prussia, PA 19406.

⊥ University of Pennsylvania.

⊗ Abstract published in *Advance ACS Abstracts,* April 15, 1997.

S0006-2960(97)00047-0 CCC: $14.00 

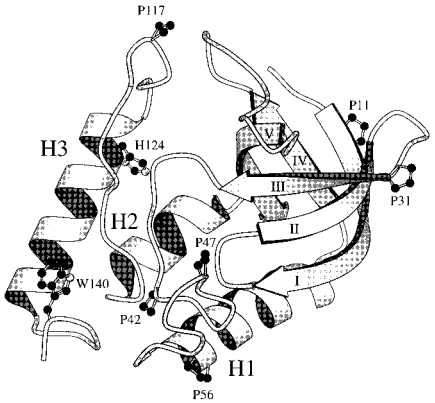



FIGURE 1: Ribbon diagram of WT staphylococcal nuclease, based on the crystal structure (Hynes & Fox, 1991), oriented to show the two subdomains and their connecting loops. The side chains of Trp140 (the sole tryptophan used as a fluorescence probe), the six proline residues, and His124 are shown explicitly. The five $\beta$-strands are numbered with Roman numerals, and the three α-helices are designated H1−H3. The diagram was prepared using the program Molscript (Kraulis, 1991).

containing at least one early intermediate in rapid equilibrium with the unfolded state (Houry et al., 1995; Parker et al., 1995; Khorasanizadeh et al., 1996). The observation of native-like structural features in some of these early intermediates or their equilibrium analogues [reviewed in Roder and Colón (1997)] provides further evidence for their productive role in folding.

In order to focus on the intrinsic conformational events in folding without the complications due to slow proline isomerization steps, we chose for this study a variant of SNase[1] in which all six proline residues had been replaced by alanine or glycine (Figure 1). SNase also lacks disulfide bonds or free cysteines, and although binding of a calcium ion is necessary for nuclease activity, metal binding is not required for folding (Sugawara et al., 1991). Although SNase exhibits the cooperative equilibrium unfolding transition characteristic of single-domain proteins (Shortle, 1996), its structure can be divided into two subdomains (Figure 1): an N-terminal domain consisting of five antiparallel $\beta$-strands folded into a barrel-like structure and a C-terminal domain containing the three α-helices of SNase (Loll & Lattman, 1989; Hynes & Fox, 1991). Recent equilibrium studies showed that, under certain conditions, the two domains can be decoupled by mutations at the helix−sheet interface (Gittis et al., 1993; Carra et al., 1994; Carra & Privalov, 1995, 1996) or deletion of some C-terminal residues (Wang & Shortle, 1995). In these SNase variants, the less stable α-helical domain unfolds more readily, while the $\beta$-sheet region remains partly structured. Using pulsed hydrogen exchange and NMR methods, Jacobs and Fox (1994) detected an early kinetic intermediate with similar structural characteristics. The kinetic consequences of amino acid changes at the periphery of the $\beta$-barrel (Kalnin & Kuwajima, 1995) are also consistent with the presence of a folding intermediate containing native-like structure in the N-terminal domain.

SNase was one of the first proteins to be studied by stopped-flow fluorescence (Anfinsen, 1973) and has continued to be a model protein for folding and stability studies [for recent reviews, see Hinck et al. (1996) and Shortle et al. (1996)]. Many of the previous studies on SNase have focused on the kinetic effects of proline isomerization. The Lys116−Pro117 peptide bond of SNase is found predominately in the cis configuration in the native state (Evans et al., 1987), and the isomerization of this peptide bond is strongly coupled to structural events (Alexandrescu et al., 1990). H124L, a naturally occurring variant of SNase found in the V8 strain of *Staphylococcus aureus*, has enhanced stability compared to the WT protein isolated from the Foggi strain (Shortle, 1986; Tuckses et al., 1996). The crystal structures of the two forms are very similar, except for the side chain of residue 124, which is solvent accessible in both variants (Tuckses et al., 1996). The folding and unfolding kinetics of both the wild-type protein (Chen et al., 1991; Sugawara et al., 1991; Nakano et al., 1993) and H124L (see below) are multiphasic and show a complex denaturant dependence indicating the presence of multiple intermediates with conformational heterogeneity in both native and denatured states.

How much of this kinetic complexity reflects intrinsic conformational changes in the folding of SNase, and how much is due to slow side reactions that may be peripheral to understanding the mechanism of folding for this protein? We addressed these issues by comparing the kinetics of folding and unfolding for wild-type (H124L) SNase with that of the proline-free variant, using the fluorescence of Trp140 as a conformational probe. The absence of proline-dependent slow folding phases in Pro$^-$ SNase made it possible to derive kinetic schemes consistent with the experimental data by using quantitative kinetic modeling. Despite the absence of prolines and other complicating factors, the Pro$^-$ variant exhibits multiple folding phases and complex denaturant dependence, indicative of a mechanism with structural intermediates both in refolding and unfolding.

## EXPERIMENTAL PROCEDURES

*Mutagenesis.* A proline-free variant of SNase from the Foggi strain (Pro$^-$) was generated through oligonucleotide-directed mutagenesis according to the method of Kunkel (1985) and cloned into the M13 phage vector MF09. Each Pro codon was replaced by either Ala (GCN) or Gly (GGN) codons, depending on the stability of the corresponding single-proline variants measured in earlier denaturation studies (Green et al., 1992). After transformation into competent DH5αF′ cells, plaques were picked and the phage DNA was sequenced. These steps were repeated until all the proline residues were substituted. The nuclease gene was then moved into the expression plasmid pL12 which contains unique *Spe*I and *Sph*I sites (Shortle et al., 1990). To ensure that no spurious secondary mutation had been introduced, the entire gene was sequenced. The final protein product contains the substitutions P11A, P31A, P42A, P47G, P56A, and P117G.

*Purification.* H124L nuclease was prepared as described previously (Wang et al., 1990), except that the cells were grown in LB medium. The strain containing the Pro$^-$ mutant was grown in LB media at 30 °C to minimize plasmid loss. Upon reaching late log phase, the temperature was raised to 37 °C and protein production induced by the addition of nalidixic acid (Shortle & Meeker, 1989). The protein was purified according to Shortle and Meeker (1989), except that reversed-phase HPLC (10 × 250 mm, 5 $\mu$m Bakerbond $C_8$;

[1] Abbreviations: SNase, staphylococcal nuclease; GuHCl, guanidine hydrochloride; H124L, histidine to leucine substitution at position 124 of wild-type SNase; Pro$^-$, SNase variant (Foggi strain) containing P11A, P31A, P42A, P47G, P56A, and P117G substitutions; WT, wild-type.

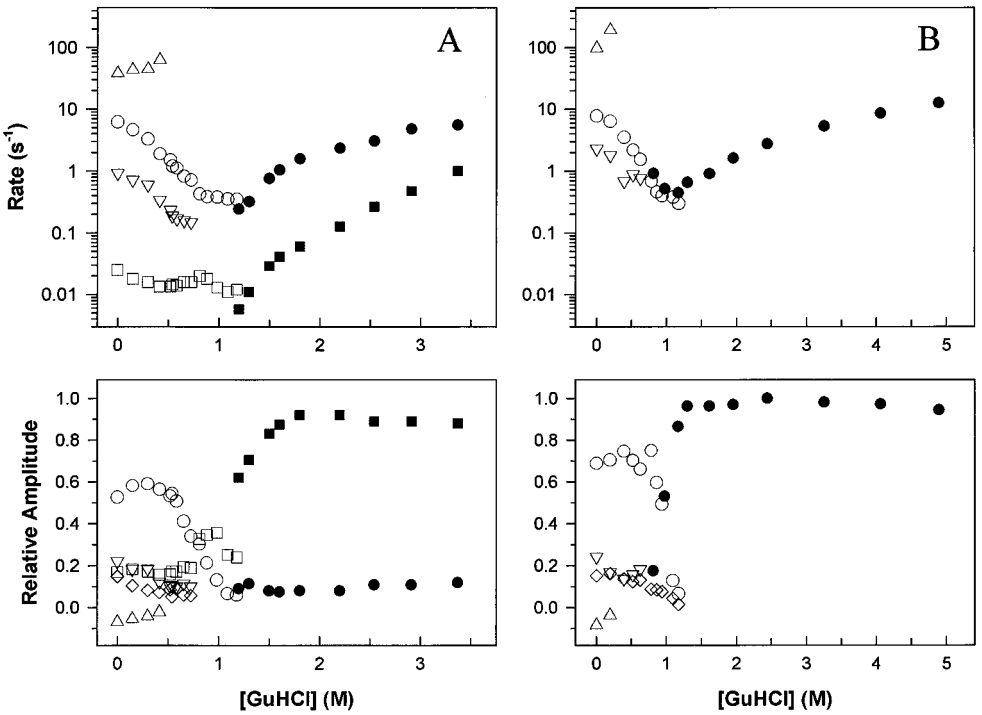



FIGURE 2: Denaturant dependence of the folding and unfolding rates and amplitudes for H124L SNase (A) and the Pro$^-$ variant (B) as detected by stopped-flow fluorescence. The top panel in each case shows the GuHCl dependence of the rates for both the unfolding phase(s) (solid symbols) and refolding phases (open symbols). The bottom panels show the corresponding GuHCl dependence of the relative amplitudes for each phase in folding and unfolding. The lag phase is represented by triangles, the major phase by circles, the intermediate phase by inverted triangles, and the slow phase (for H124L) by squares. The missing amplitude at each GuHCl concentration is indicated by open diamonds.

J. T. Baker, Phillipsburg, NJ) was substituted for the final chromatographic step. The final purity was better than 95% by HPLC using a linear gradient of acetonitrile in water (15% to 60% acetonitrile; 0.1% TFA). Amino acid composition was confirmed by amino acid analysis (PICO·TAG method; Waters Company, Milford, MA).

*Fluorescence-Detected Stopped-Flow Kinetics.* The kinetic experiments on the Pro$^-$ variant were performed on a Bio-Logic SFM-4/QS stopped-flow instrument (Molecular Kinetics, Pullman, WA), as described previously (Colón et al., 1996), except that an excitation wavelength of 286 nm was used. The kinetic experiments on H124L were performed on a PQ/SF-53 stopped-flow instrument (Hi-Tech Scientific, Salisbury, England), as described in an earlier paper (Elöve et al., 1994). The dead times of mixing were 1.5 ms (Bio-Logic) and 2.4 ms (Hi-Tech). On both instruments, the fluorescence changes from 64 $\mu$s to ~60 s were recorded in a single kinetic trace, using logarithmic averaging with a sampling time of 49 $\mu$s (Khorasanizadeh et al., 1996). A typical 60 s time trace was measured with about 200 000 data points, but displayed and fit as ~100 logarithmically averaged points. The data at each GuHCl concentration were typically the average of three to five traces.

The kinetics of refolding was measured at 15 °C following 6-fold dilution of an acid-denatured solution of SNase (20 mM phosphate, pH 2.2), into various final GuHCl concentrations (from 0 to about 1.2 M) at pH 5.3 in 100 mM sodium acetate. Final protein concentrations were 10 $\mu$M. The kinetics of unfolding was measured by 6-fold dilution of native SNase (100 mM sodium acetate, pH 5.3) into concentrations of GuHCl within or above the unfolding transition.

The amplitudes of the unresolved process occurring in the dead time (burst phase) and of the observable kinetic phases were normalized with respect to the total fluorescence changes expected at each final GuHCl concentration, as described in Colón et al. (1996). The data from different experiments were normalized with respect to the signal of the unfolded protein in 2.4 M GuHCl. The equilibrium denaturation curve for the Pro$^-$ mutant was reconstructed from the baseline amplitudes of the kinetic experiments at long refolding and unfolding times. The kinetic data were fit as a sum of two or three exponentials, using Sigma Plot for Windows 2.0 (Jandel Scientific, San Rafael, CA). The fitting errors for the rates and amplitudes were less than 10% unless otherwise indicated by error bars (cf. Figure 4).

*Kinetic Modeling.* A variety of kinetic schemes were tested for compatibility with the experimental rate and amplitude data for the Pro$^-$ mutant. A program written in the ASYST programming language (Keithly Metrobyte, Rochester, NY) was used to calculate the observable rate constants and the corresponding amplitudes as described in Khorasanizadeh et al. (1996). For each scheme, the microscopic rate constants and their dependence on denaturant concentration, and the relative fluorescence values of each species were used as input to calculate the eigenvalues and eigenvectors for the associated rate matrix. The logarithm of each rate constant, $k_{ij}$, was assumed to depend linearly on denaturant concentration, $C$, according to

$$\ln k_{ij} = \ln k_{ij}^{\circ} + (m^{\ddagger}_{ij}/RT)C \qquad (1)$$

where $k_{ij}^{\circ}$ is the rate constant in the absence of denaturant and $m^{\ddagger}_{ij}/RT$ is the corresponding slope (Tanford, 1970). The results are described below for a sequential four-state scheme

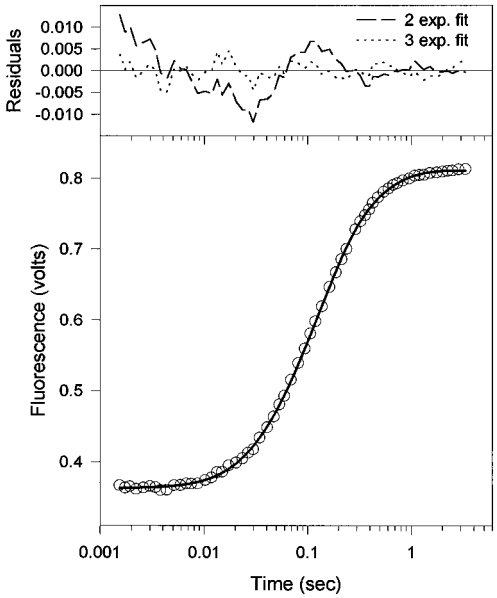



FIGURE 3: Multiexponential fits to the time course of refolding measured by stopped-flow fluorescence for Pro$^-$ SNase at 0 M GuHCl. The solid line in the main panel shows the triexponential fit to the data, using nonlinear least-squares analysis (Sigma-Plot). In the panel above, the corresponding residuals (dotted line) are compared with those for a biexponential fit (dashed line), indicating that inclusion of a lag phase substantially improves the fit.

(see Scheme 1) and an expanded six-state scheme containing a parallel pathway (see Scheme 2). In both schemes, the rates and amplitudes of the lag phase and of the major phase in folding and unfolding were modeled while neglecting a minor burst phase. To accomplish this, the fractional amplitudes for each observable phase, $A_i$, were renormalized such that the sum of all $A_i$ plus the fraction unfolded, $f_u$, is equal to 1. In modeling the data with the four-state scheme, the amplitudes $A_2$ and $A_3$ were combined to obtain a composite amplitude (see Figure 4A). For the six-state scheme, all observable phases were modeled explicitly (Figure 4B).

*Time-Trace Simulations.* The time courses at several GuHCl concentrations were predicted by kinetic simulation using the parameters from the six-state model. The raw fluorescence data were normalized to a value of 1.0 for the native state, and the initial fluorescence of the unfolded state was varied in the simulations to account for the proper amount of burst phase (missing amplitude) at each GuHCl concentration.

*Free Energy Diagrams.* Free energy diagrams were calculated using the kinetic parameters based on Scheme 1 (see Table 2). The activation energy required to cross the barrier between each state was calculated as follows:

$$\Delta G^{\ddagger}_{ij} = -RT\ln(k_{ij}^{\circ}/A_o) - m^{\ddagger}_{ij}[\text{GuHCl}] \qquad (2)$$

The Arrhenius preexponential factor ($A_o$) was arbitrarily chosen to be $1 \times 10^9$ M$^{-1}$ s$^{-1}$ for these calculations.

## RESULTS

*Denaturant Dependence of the Folding Kinetics of H124L SNase.* The folding kinetics of H124L SNase and the Pro$^-$ mutant were monitored using tryptophan fluorescence as a probe of conformational changes. The fluorescence of Trp140 (Figure 1) is partially quenched in the denatured state, presumably by interacting with the side chain of Lys134 (Royer et al., 1993), and a large increase in fluorescence (4−8-fold increase in quantum yield) is observed upon refolding (Anfinsen et al., 1972; Shortle, 1986). The

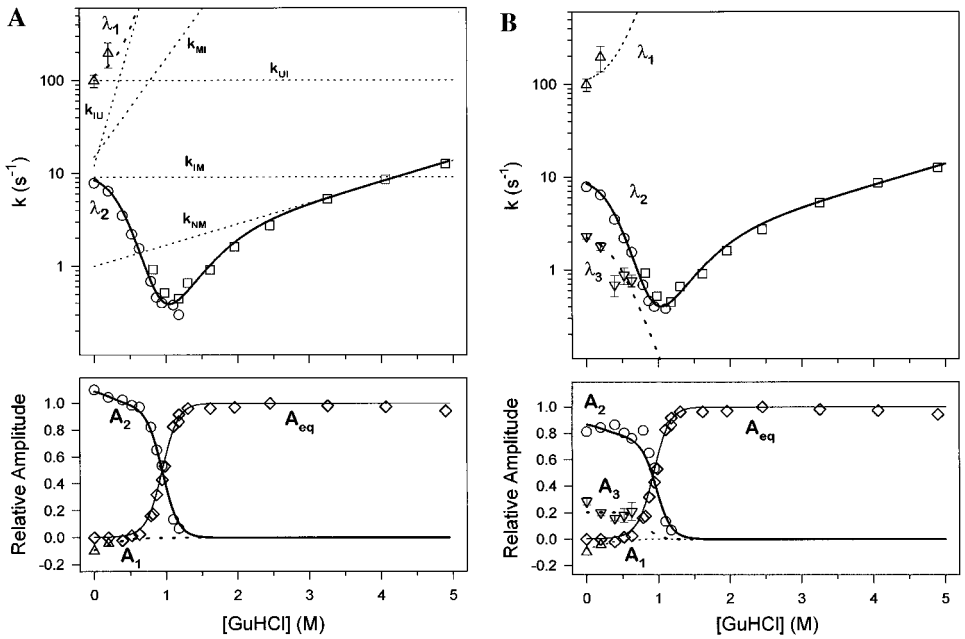



FIGURE 4: (A) Four-state kinetic analysis, based on Scheme 1, of the major kinetic phases in folding and unfolding for Pro$^-$ SNase. The upper panel shows the unfolding rates (squares) and folding rates for the main phase (circles) and the lag phase (triangles) as a function of GuHCl concentration; the bottom panel shows the corresponding relative amplitudes. The equilibrium denaturation curve derived from the kinetic baseline amplitudes is plotted in the lower panel (diamonds). Thick solid lines represent the predicted rates of folding and unfolding ($\lambda_2$, top) and the corresponding amplitudes ($A_2$, bottom) based on the four-state scheme. Dashed lines (- - -) indicate the predicted rates ($\lambda_1$) and amplitudes ($A_1$) of the lag phase associated with the formation of the first intermediate. The predicted equilibrium unfolding curve ($A_{eq}$) is represented by a thin solid line in the lower panel. Thin dotted lines indicate the microscopic rate constants, $k_{ij}$, used in the modeling (Table 2). (B) Six-state kinetic analysis, based on Scheme 2, of all phases in folding and unfolding, using the modeling parameters in Table 2. Symbols are identical to those in the four-state model except that the observed (inverted triangles) and predicted rates and amplitudes (- - -) for the minor third phase ($\lambda_3$) are included in addition to the lag phase (thin dashed lines) and main phase (solid lines).

Table 1: Comparison of the SNase Refolding Kinetics with and without Prolines[a]

| | $\lambda_0$ ($s^{-1}$) | $A_0$ | $\lambda_1$ ($s^{-1}$) | $A_1$ | $\lambda_2$ ($s^{-1}$) | $A_2$ | $\lambda_3$ ($s^{-1}$) | $A_3$ | $\lambda_4$ ($s^{-1}$) | $A_4$ |
|---|---|---|---|---|---|---|---|---|---|---|
| H124L | >1000[b] | 0.15 ± 0.03[c] | 38 ± 9 | −0.07 ± 0.01 | 6.2 ± 0.3 | 0.53 ± 0.01 | 0.93 ± 0.07 | 0.22 ± 0.01 | 0.025 ± 0.001 | 0.17 ± 0.003 |
| Pro$^-$ | >1000 | 0.15 ± 0.02 | 99 ± 15 | −0.08 ± 0.01 | 7.8 ± 0.4 | 0.69 ± 0.03 | 2.3 ± 0.2 | 0.24 ± 0.03 | na[d] | na |

[a] Rate constants and normalized amplitudes of the kinetic phases observed in fluorescence-detected stopped-flow refolding experiments at 15 °C. Refolding conditions: 0.1 M sodium acetate buffer, pH 5.3 (no denaturant). [b] Estimate for rate of the burst phase process which occurs during the dead-time of mixing. [c] Error estimates from nonlinear least-squares fitting (1 standard deviation). [d] Not applicable.

Table 2: Kinetic Modeling Parameters[a]

| $k_{ij}$ | rate constant ($s^{-1}$)[c] | $m^{\ddagger}_{ij}$ (kcal $mol^{-1}$ $M^{-1}$)[d] |
|---|---|---|
| $k_{UI}$ | 100 | 0 |
| $k_{IU}$ | 12 | 3.5 |
| $k_{IM}$ | 9.7 | 0 |
| $k_{MI}$[b] | 15 | 1.4 |
| $k_{MN}$[b] | 1000 | 0 |
| $k_{NM}$ | 1 | 0.3 |
| $k_{UU'}$ | 0.025 | 0 |
| $k_{U'U}$ | 0.1 | 0 |
| $k_{U'I'}$ | 100 | 0 |
| $k_{I'U'}$ | 12 | 3.0 |
| $k_{I'M}$ | 2.5 | 0 |
| $k_{MI'}$ | 1 | 1.4 |

[a] The relative fluorescence yields used in the simulations are 1 for U, U′, I, and I′, 0.1 for M, and 0 for N. [b] Only the ratio $k_{MN}/k_{MI}$ is uniquely determined (see text). [c] Rate constant for a given process in the absence of GuHCl. [d] Kinetic *m*-values according to eq 1.

denaturant dependence of the unfolding and refolding kinetics of SNase and its mutant(s) was investigated by collecting stopped-flow fluorescence data in the presence of various concentrations of GuHCl. In Figure 2, the rates of folding and unfolding for H124L (Figure 2A) and the Pro$^-$ mutant (Figure 2B) are plotted on a logarithmic scale *vs* GuHCl concentration (upper panel), and the corresponding normalized amplitudes are plotted in the lower panel. The unfolding phases are represented by solid symbols and refolding phases by open symbols.

Four kinetic phases were detected in the refolding and two in the unfolding of H124L by stopped-flow fluorescence in the experimentally accessible time range between 2.4 ms and 600 s. The rates and amplitudes for the refolding phases, measured in the absence of denaturant, are listed in Table 1. Phases 2−4 exhibit rather typical behavior in that all three phases have positive amplitudes ($A_2$−$A_4$) and that the apparent rate constant of each phase ($\lambda_2$−$\lambda_4$) decreases monotonically with increasing denaturant concentration (Figure 2A). The rate of the fast phase ($\lambda_2$) is about 6 $s^{-1}$ in the absence of denaturant and decreases by more than an order of magnitude as the concentration of GuHCl is increased to values near the midpoint of the unfolding transition ($C_m$). At low concentrations of denaturant, the plot of log($\lambda_2$) *vs* GuHCl concentration levels off, indicating accumulation of an intermediate under stabilizing conditions (Matouschek et al., 1990; Khorasanizadeh et al., 1996). By contrast, in proteins for which the two-state approximation holds, log($\lambda$) increases linearly with decreasing denaturant concentration [e.g., Jackson and Fersht (1991)]. The fast phase (phase 2, open circles) is the dominant observable process below 0.8 M GuHCl, while phase 4 is the dominant process near $C_m$. The denaturant dependence of the intermediate phase rate (phase 3, inverted triangles) parallels that of the fast phase. Its rate is about 1 $s^{-1}$ at 0 M GuHCl and decreases with increasing denaturant. The corresponding amplitude ($A_3$) accounts for about 20% of the total fluorescence change upon folding. The slow phase detected by stopped-flow fluorescence (phase 4) with a rate of about 0.025 $s^{-1}$ at 0 M GuHCl is relatively insensitive to addition of denaturant, which is often diagnostic for a step involving proline isomerization (Kiefhaber et al., 1992). The small increase in the rate of phase 4 near 0.8 M GuHCl (open squares) is due to a fitting artifact as the number of exponentials used to fit the data drops from three to two. The amplitude of this slow process, $A_4$, is small at low concentrations of denaturant, but increases with additional GuHCl to make $A_4$ the major phase in the transition region. This is typically observed as proline isomerization becomes coupled to conformational events (Kiefhaber et al., 1992).

*Observation of a Lag Phase in Refolding.* H124L shows an apparent lag in the refolding kinetics at low GuHCl concentrations, which requires the use of an additional exponential term to completely fit the data (see Figure 3). This term gives rise to the fastest observable process (phase 1) in the refolding of H124L (Figure 2A). The rates and amplitudes of this phase (open triangles) exhibit different behavior than those of the other phases discussed above. The apparent rate constant ($\lambda_1$) is initially about 40 $s^{-1}$ but *increases* with increasing denaturant concentration to a rate of about 60 $s^{-1}$ at 0.4 M GuHCl. The lag phase is also unusual in that it has a negative amplitude ($A_1$). Although $A_1$ accounts for less than 10% of the total amplitude, the residuals in Figure 3 (data for Pro$^-$ mutant) show that inclusion of this lag phase substantially improves the quality of the fit. $A_1$ decreases with increasing GuHCl before becoming negligible at about 0.4 M GuHCl. This behavior is consistent with the transient accumulation of an intermediate that is not directly detected (because its fluorescence is indistinguishable from the initial unfolded state), but results in delayed formation of the folded state.

*Burst Phase.* In addition to the observable phases described above, H124L exhibits a small burst phase in refolding at concentrations of GuHCl below the $C_m$ (open diamonds in Figure 2A), indicating that an unresolved process occurs during the dead time of mixing. This "missing amplitude" ($A_0$ in Table 1) reaches a maximum of about 15% of the total amplitude during refolding in the absence of denaturant and decreases steadily with increasing GuHCl until it becomes negligible at about 0.8 M GuHCl. No attempt was made in our modeling (see below) to account for this small burst phase.

*Unfolding Kinetics.* Several groups have reported that the unfolding kinetics of SNase are biphasic under most conditions (Sugawara et al., 1991; Chen et al., 1991). This behavior can be attributed to the presence of two distinct populations of native-like molecules that differ mainly by cis−trans isomerism about the Lys116−Pro117 peptide bond. When Pro117 is mutated, the unfolding kinetics of SNase becomes monophasic (Kuwajima et al., 1991; Nakano et al., 1993). The more stable cis form (filled squares in Figure 2A) predominates and unfolds more slowly than does the

minor trans form (filled circles). For both unfolding phases, the log($\lambda_i$) *vs* [GuHCl] plot shows a regime with a steep slope in the transition region and a shallower denaturant dependence above 2 M GuHCl.

*Effect of Proline 117 Substitutions on Slow Phases in Refolding.* A slower phase in refolding ($k_5 \approx 0.002$ s$^{-1}$ at 4 °C, data not shown) is observable by fluorescence only when using manual mixing. This phase has been previously attributed to trans−cis isomerization of the Lys116−Pro117 peptide bond (Kuwajima et al., 1991; Nakano et al., 1993). In related studies, however, we observed that the P117G substitution (in an H124L background) causes the amplitudes of both of the two slowest phases ($A_4$ and $A_5$) to decrease by a factor of 2 (rather than disappear entirely), while the rate $\lambda_4$ (∼0.025 s$^{-1}$ at 15 °C) is accelerated approximately 4-fold (W. F. Walkenhorst, unpublished data). The double proline mutations, P117G/P47G and P117G/P42G, had a similar effect, with that containing the P47G substitution causing the largest increase in $\lambda_4$ [unpublished data; see also Hinck et al. (1996)].

*Effect of Global Substitution of Prolines on Stability and Complexity of Folding.* Preliminary studies indicated that the Pro$^-$ mutant is about 0.4 kcal/mol more stable than the WT protein (Green et al., 1992). Our present results confirm that the Pro$^-$ mutant is more stable to denaturation by GuHCl ($C_m = 0.98$ M; $m = 6.1$ kcal mol$^{-1}$ M$^{-1}$) than the WT protein ($C_m = 0.82$ M, $m = 6.85$ kcal mol$^{-1}$ M$^{-1}$; Shortle, 1986), although it is still less stable than H124L ($C_m = 1.08$ M; $m = 6.4$ kcal mol$^{-1}$ M$^{-1}$; Shortle, 1986). In general, Pro to Gly substitutions are expected to destabilize the protein by increasing the entropy of the unfolded state (Herning et al., 1992). In SNase, however, the P117G substitution has been found to be energetically favorable due either to a gain in favorable interactions or to a reduction in strain (Hodel et al., 1993; Hinck et al., 1993; Hinck et al., 1996).

Although only small effects on stability and on some of the kinetic rate constants (Table 1) are observed, elimination of all proline residues results in much simpler kinetic behavior (Figure 2B). In particular, the two slowest refolding phases (phases 4 and 5) and the slower of the two unfolding phases are no longer observed. Nevertheless, the fact that the Pro$^-$ mutant retains two phases in refolding and exhibits curved rate profiles, both at low GuHCl and at high GuHCl, shows that the kinetics cannot be described by a simple two-state mechanism. At low concentrations of denaturant, the Pro$^-$ mutant also exhibits an apparent lag phase (phase 1) similar to that observed for H124L.

*Quantitative Kinetic Modeling.* In order to simplify the analysis, the initial modeling considered only the major refolding phase together with the single phase in unfolding ($\lambda_2$) and the apparent lag phase ($\lambda_1$, see Figure 4A). We found that the minimal scheme able to reproduce both the rate ($\lambda_1$, $\lambda_2$) and amplitude data ($A_1$, $A_2$) contains at least four states. The kinetic behavior of the primary phase depicted in Figure 4A is consistent with a sequential mechanism containing at least one intermediate in refolding, necessary to account for the curvature at low denaturant concentrations, and an unfolding intermediate responsible for the curvature at high denaturant concentrations (Scheme 1). The incorporation of the lag phase into the analysis allows limits to be placed on some of the microscopic rate constants in the sequential four-state scheme described below.

The model assumes four thermodynamically distinct states interconverting according to the following mechanism:

Scheme 1

$$\mathrm{U} \underset{k_{\mathrm{IU}}}{\overset{k_{\mathrm{UI}}}{\rightleftharpoons}} \mathrm{I} \underset{k_{\mathrm{MI}}}{\overset{k_{\mathrm{IM}}}{\rightleftharpoons}} \mathrm{M} \underset{k_{\mathrm{NM}}}{\overset{k_{\mathrm{MN}}}{\rightleftharpoons}} \mathrm{N}$$

where U represents the ensemble of fully unfolded conformations, I and M represent dynamic ensembles of partially folded intermediate states, and N represents the native state. For modeling purposes, U and N were assigned relative fluorescence values of 1.0 and 0.0, respectively (the actual fluorescence of Trp140 in SNase is partially quenched in the unfolded state and has a much higher yield in the native state). We assigned the first intermediate, I, a relative fluorescence value of 1.0, identical to that of U, while the unfolding intermediate M was assigned a fluorescence value of 0.1 (this choice turned out to be irrelevant, since M is always poorly populated). The model is further characterized by six microscopic rate constants ($k_{ij}°$) and their denaturant dependence ($m^{‡}_{ij}$; see eq 2), which are listed in Table 2 and plotted as a function of GuHCl in Figure 4A (dotted lines). To reproduce the curvature in the logarithm of the rate for the main folding phase at low concentrations of GuHCl, it was necessary to set the equilibrium constant between U and I ($K_{\mathrm{UI}} = k_{\mathrm{UI}}/k_{\mathrm{IU}}$) to a value of about 8, favoring the accumulation of *I*. Since our fastest measurable process (lag phase) has a rate, $\lambda_1$, of about 100 s$^{-1}$ (excluding any processes occurring in the dead time), we chose values for $k_{\mathrm{UI}}$ and $k_{\mathrm{IU}}$ of 100 and 12 s$^{-1}$, respectively, such that the curvature in refolding and the initial value of $\lambda_1$ are satisfied simultaneously. The negative amplitude ($A_1$) associated with the lag phase can be modeled by assigning U and I identical or nearly identical relative fluorescence values. The magnitude of $A_1$ is governed by the ratio $k_{\mathrm{IU}}/k_{\mathrm{IM}}$. As this ratio decreases in magnitude and becomes less than 1, $A_1$ becomes progressively more negative.

While the general solution of a four-state scheme is complex and was determined numerically, it is instructive to consider the following approximation for the major observable folding/unfolding rate, $\lambda_2 = k_{\mathrm{f}} + k_{\mathrm{u}}$ [cf. Sauder et al. (1996)]:

$$k_{\mathrm{f}} = \frac{k_{\mathrm{UI}}}{k_{\mathrm{UI}} + k_{\mathrm{IU}}} k_{\mathrm{IM}} \tag{3}$$

$$k_{\mathrm{u}} = \frac{k_{\mathrm{MI}}}{k_{\mathrm{MI}} + k_{\mathrm{MN}}} k_{\mathrm{NM}} \tag{4}$$

which is valid as long as the kinetic barrier between I and M is much larger than the barriers separating U from I and N from M, respectively (see Figure 6). These expressions can be further simplified in the following limiting cases. At low denaturant concentrations (<0.6 M GuHCl), where $k_{\mathrm{UI}} \gg k_{\mathrm{IU}}$, I is well-populated in the pre-equilibrium between U and I ($K_{\mathrm{UI}} > 1$), and the observed rate of folding approaches the microscopic rate constant $k_{\mathrm{IM}}$ (∼10 s$^{-1}$ at 0 M GuHCl). At higher denaturant concentration, but still below the $C_m$, under conditions where $k_{\mathrm{IU}} \gg k_{\mathrm{UI}}$, the observed rate is given by the product $K_{\mathrm{UI}}k_{\mathrm{IM}}$, where $K_{\mathrm{UI}}$ is the equilibrium constant for formation of I ($K_{\mathrm{UI}} = k_{\mathrm{UI}}/k_{\mathrm{IU}}$). In this limit, the plot of log($\lambda_2$) *vs* [GuHCl] is linear with a negative slope dominated by the denaturant dependence of the I → U transition ($m^{‡}_{\mathrm{app}} = m^{‡}_{\mathrm{UI}} - m^{‡}_{\mathrm{IU}} + m^{‡}_{\mathrm{IM}} \approx -m^{‡}_{\mathrm{IU}}$). Near $C_m$, the intermediate is highly destabilized and its population is low ($K_{\mathrm{UI}} \ll 1$), resulting in a greatly reduced rate of folding.

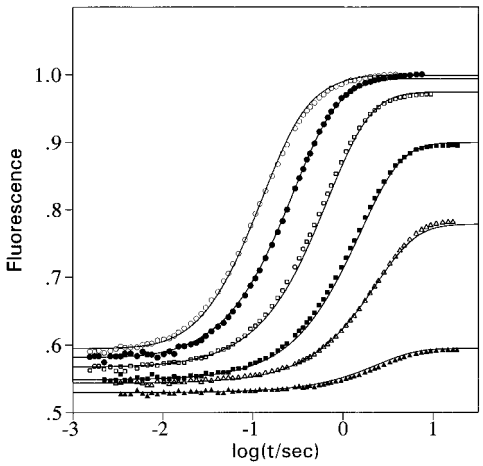



FIGURE 5: Global analysis of the kinetics of refolding for Pro$^-$ SNase as a function of denaturant concentration. The symbols indicate the fluorescence changes observed in a series of stopped-flow measurements at final GuHCl concentrations of 0 M (open circles), 0.39 M (filled circles), 0.6 M (open squares), 0.78 M (filled squares), 0.94 M (open triangles), and 1.13 M (filled triangles). The corresponding time courses predicted by kinetic simulation, using Scheme 2 and the parameters in Table 2, are plotted as solid lines. The raw fluorescence data (measured under identical instrumental conditions) were normalized such that the native state has a value of 1.0. Note that the whole family of curves was calculated using a single set of rate constants and relative fluorescence values for the native state and intermediates (minor adjustments in the fluorescence of the unfolded state were necessary to account for the dependence of initial values on GuHCl concentration).

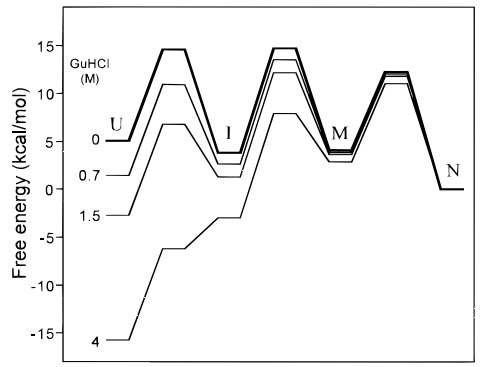
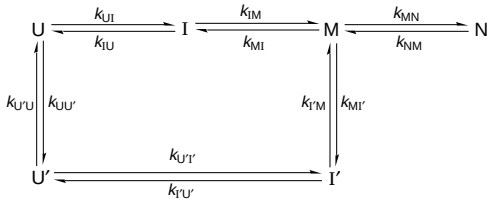



FIGURE 6: Free energy diagrams showing the effect of GuHCl on the energy levels and transition states predicted by quantitative analysis of the folding/unfolding kinetics for Pro$^-$ SNase, based on the four-state mechanism (Scheme 1). The native state was assigned a free energy of 0 kcal/mol. The activation energies were calculated according to eq 2, using the appropriate modeling parameters from Table 2.

The change in slope in the denaturant dependence of the unfolding rate observed between 2 and 3 M GuHCl (Figure 4A) is attributed to a change in the rate-determining barrier encountered in unfolding. The rate constant with the most pronounced denaturant dependence is $k_{MI}$, which represents the main conformational unfolding event. In the transition region, but still above $C_m$, $k_{MI} \ll k_{MN}$, and the rate of unfolding (eq 4) can be written as $k_u = K_{NM}k_{MI}$, where $K_{NM} \ll 1$ and only weakly dependent on denaturant concentration. At high denaturant concentrations (>3 M GuHCl), $k_{MI} \gg k_{MN}$, and the observed rate of unfolding is given by $k_u = k_{NM}$. Since the corresponding slope, $m^{\ddagger}_{NM}$, is smaller than that of the main unfolding step, $m^{\ddagger}_{MI}$ (Table 2), this gives rise to a pronounced kink in the rate profile near 2.5 M GuHCl (Figure 4A).

Thus, eqs 3 and 4 qualitatively describe the denaturant dependence of the main observable kinetic process, including

Scheme 2




the curvature in the log($\lambda_2$) *vs* GuHCl plot observed both at low and high denaturant concentrations (Figure 4A). In contrast to other proteins [reviewed in Roder and Colón (1997)], the observation of a lag phase allows us to determine not only the pre-equilibrium constant, $K_{UI}$, but also the individual rate constants, $k_{UI}$ and $k_{IU}$. However, the rates given in Table 2 for $k_{MI}$ and $k_{MN}$ are not uniquely determined by the observed rates and amplitudes (Figure 4), since we observed no faster unfolding process to provide additional constraints. While $k_{NM}$ is directly determined by the observed unfolding rate in the high-GuHCl limit, eq 4 shows that only the ratio $k_{MI}/k_{MN}$ (partitioning constant) is uniquely determined by the kinetic behavior in the transition region, as long as $k_{MN}$ remains the fastest forward rate ($k_{MN} > k_{UI}$).

*Six-State Parallel Scheme.* The minor phase in refolding observed for the Pro$^-$ mutant ($\lambda_3$, Figure 4B) requires inclusion of additional states in the analysis. Extensive modeling showed that the simplest scheme able to account for the data contains a parallel pathway with at least one structural intermediate originating from a distinct unfolded population (Scheme 2). The additional kinetic parameters needed for the six-state model are listed in the bottom half of Table 2. All other parameters are identical to those in the four-state model. To properly distribute the amplitudes between $\lambda_2$ and $\lambda_3$, an equilibrium was modeled in the unfolded state between U and U′ with an equilibrium constant ($K_{UU'} = k_{UU'}/k_{U'U}$) of approximately 0.25. The rates for these steps were assumed to be denaturant independent as would be the case for two states with a similar degree of exposed hydrophobic surface area. The rates of interconversion between U and U′ must be relatively slow in comparison to later folding steps in order to produce a satisfactory fit of the amplitude data throughout the range of denaturant concentrations, with upper limits of 0.1 and 0.025 s$^{-1}$ (Table 2). The equilibrium between U′ and I′, as well as the rate constants defining this equilibrium, were assigned values identical to those of the corresponding U ⇌ I equilibrium, reflecting the population of an intermediate with properties similar to I. To avoid possible complications in modeling the unfolding kinetics, we chose to connect I′ to M instead of to N, thus allowing us to retain the important features already established in the four-state model. The rate constant for the conversion of I′ to M ($k_{I'M}$) gives rise to the limiting refolding rate for $\lambda_3$ of about 2.5 s$^{-1}$ observed in the absence of denaturant. As in the case of the four-state model, the partitioning of M in a constant ratio ($k_{MI'}/k_{MN} = 0.001$) between N and I′ is required to fit the data. A lower limit of about 0.2 s$^{-1}$ can be placed on $k_{MI'}$ as long as $k_{MN}$ and $k_{MI}$ are decreased proportionally.

As a final test of our kinetic model (Scheme 2), we simulated the time courses for refolding of SNase at various concentrations of denaturant. The fluorescence data for refolding at a number of GuHCl concentrations are plotted in Figure 5 along with a family of time traces calculated on the basis of Scheme 2, using a single set of kinetic parameters

(global analysis). Starting with the kinetic parameters derived from empirically fitted rate and amplitude data *vs* GuHCl concentration (Figure 4B), only minor adjustments in a few rate constants were sufficient to simultaneously reproduce the kinetic traces for refolding (Figure 5) and unfolding experiments (data not shown) at all GuHCl concentrations measured. The final parameters for the simulations are those listed in Table 2 (and Figure 4B), except that $k_{UI}$ and $k_{IU}$ were lowered to 80 and 9.6 $s^{-1}$, respectively, to give a better fit for the lag phase observed at the beginning of the time traces. This global analysis procedure provides a rigorous test of the kinetic model by directly comparing its predictions with the raw data. Another advantage over the previous two-step procedure is that it avoids the sometimes difficult choice of the number of exponential terms to be used in fitting individual kinetic traces.

## DISCUSSION

*Early Folding Events.* A small burst phase indicative of unresolved folding events occurring during the dead time of mixing has been observed previously by both fluorescence and CD for SNase and some of its mutants (Sugawara et al., 1991; Chen & Tsong, 1994; Kalnin & Kuwajima, 1995). Although the amplitude for this process is small, this may be more a reflection of the particular probes of secondary structure used. In SNase, Trp140 is positioned near the C-terminus at the periphery of the α-helical domain (Figure 1) and thus is not a good probe for conformational changes in the $\beta$-sheet region of the protein. Likewise, the CD signal at 222 nm primarily reports on the α-helical portion of the protein and is rather insensitive to $\beta$-sheet formation. Because of its small amplitude, we chose not to model the burst phase process at this time. However, the amide protection results of Jacobs and Fox (1994) clearly indicate the rapid formation (within 10 ms) of an intermediate containing hydrogen-bonded structure in parts of the $\beta$-sheet, which apparently escapes detection by tryptophan fluorescence or CD at 222 nm. Thus, we expect that introduction of a tryptophan residue into the $\beta$-sheet by mutagenesis methods should result in a much larger burst phase signal [cf. Gittis et al. (1993)].

A lag phase or induction period can sometimes be observed kinetically in the early stages of a reaction [e.g., Benson (1952]. During the induction period the concentration of intermediates increases before substantial accumulation of product occurs. In protein folding, an initial lag phase reflecting a delay in the appearance of the native state can be observed if the steps in a sequential mechanism have similar rate constants and if the intermediates are spectroscopically indistingushable from the initial state. Observation of a lag phase can thus provide evidence for an intermediate even if it lacks a specific spectroscopic signature. However, a lag phase cannot be resolved if intermediates are formed during the experimental dead time, or it can be obscured by the signal change associated with the formation of early intermediates (Kiefhaber, 1995b). In our investigation of SNase, we found that the fastest observable process ($\lambda_1$) has the properties expected for a lag phase, including the sign of its amplitude ($A_1$), which is opposite to that of the main folding phase (Figure 4). An induction period with an apparent time constant of 40−200 ms was observed not only for the $Pro^-$ variant (Figure 3) but also for H124L SNase (Figure 2) and several other mutant forms studied (unpublished results). The lag is most pronounced in the absence of denaturant where the intermediate I is well populated, but a lag is no longer observed when I is destabilized by addition of GuHCl. In order to reproduce this phase in the simulations, it was necessary to set the rate constants that govern the partitioning of I between U and M ($k_{IU}$ and $k_{IM}$) to comparable values and also to assign U and I the same relative fluorescence value (see Experimental Procedures). Thus, the results provide clear evidence for the presence of an intermediate on the 10−100 ms time scale despite the fact that the U → I transition does not give rise to a change in fluorescence.

Our mechanism with a well-populated folding intermediate is further confirmed by the dependence of the major folding phase $\lambda_2$ on denaturant concentration (Figure 4). As in previous studies on ubiquitin and cytochrome $c_2$ (Khorasanizadeh et al., 1996; Sauder et al., 1996), the logarithm of the folding rate levels off at low denaturant concentrations, indicating that a denaturant-insensitive process becomes rate-limiting (Figure 4A). In the present four-state model, this process corresponds to the formation of a second intermediate, M, from which the native state is formed rapidly. M has no effect on the kinetics of refolding, but is necessary to model the denaturant dependence of the rate of unfolding (see below). In a recent stopped-flow CD study on WT SNase (Foggi strain), Su et al. (1996) observed a similar curvature in the logarithm of the main folding rate as a function of GuHCl concentration, but no explanation was offered for this phenomenon. Our kinetic analysis shows that a sequential folding mechanism with an obligatory intermediate (Scheme 1) is fully consistent with the kinetic data, including both the variation of rates with denaturant concentration (Figure 4) and the apparent lag in individual kinetic traces (Figures 3 and 5).

*Physical Properties of I.* The major folding intermediate I is well-populated during refolding at low denaturant concentrations. In contrast to many other small proteins [reviewed in Roder and Colón (1997)] where burst intermediates are formed during the experimental dead time, the major folding intermediate of SNase is formed on an observable time scale (∼10 ms). While previous hydrogen exchange labeling (Jacobs & Fox, 1994) and stopped-flow CD (Kalnin & Kuwajima, 1995) studies found evidence for a burst phase in SNase folding, their dead times were too long (5 and 10 ms, respectively) to resolve the formation of this intermediate. However, some of the amide protection data presented by Jacobs and Fox (1994) show a systematic increase in protection at the earliest times measured, consistent with a 10 ms process preceding the main folding phase. The lack of a significant fluorescence change associated with the U → I transition is explained by the position of Trp140 at the C-terminus of the protein, far from the $\beta$-sheet domain. This is corroborated by evidence from thermodynamic studies showing that fluorescence is not as sensitive to changes in this region as other methods, such as calorimetry, which reveal the presence of an equilibrium intermediate under certain conditions (Carra et al., 1994; Carra & Privalov, 1995). In fact, Gittis et al. (1993) previously observed a biphasic equilibrium transition indicative of a three-state mechanism for the V66W variant, which contains a second fluorescence probe at the interface between α-helix H1 and the $\beta$-barrel (Figure 1). In our four-state

kinetic mechanism, M is not populated, and therefore would not be expected to give rise to an additional equilibrium transition.

The effect of denaturant concentration on individual rate constants (Table 2) provides qualitative information on the average change in solvent-exposed surface area associated with each transition (Tanford, 1970; Chen et al., 1992a). In particular, the fraction of buried surface area in I relative to U, $\alpha_{IU}$, is given by $(m^{\ddagger}_{IU} - m^{\ddagger}_{UI})/m_{eq}$, where $m_{eq}$ is the overall $m$ value obtained from the modeled equilibrium data (Figure 4A; $m_{eq} = 5.2$ kcal mol$^{-1}$ M$^{-1}$). The value obtained, $\alpha_{IU} = 0.67$, using the $m$ values from Table 2, indicates that I represents an ensemble of relatively compact states with about 2/3 as much solvent-excluded surface area as the native state. In contrast, the change in solvent exposure associated with the next step, I → M, appears to be much smaller (although the slope of $k_{IM}$ is not precisely determined by the data in Figure 4, it is clear that the magnitude of $m^{\ddagger}_{IM}$ is much smaller than $m^{\ddagger}_{IU}$, which determines the slope of the folding rate profile in the transition region). Thus, the main transition state for folding is comparable to I in terms of solvent-accessible surface area, which is a characteristic property of early intermediates in a number of proteins [summarized in Roder and Colón (1997)].

*Minor Folding Phases.* Our kinetic results on H124L SNase (Figure 2A) are largely consistent with several earlier studies on WT SNase (Foggi strain), which detected one or two folding phases with time constants slower than 10 s, using far-UV CD (Sugawara et al., 1991), Trp140 fluorescence (Chen et al., 1991; Su et al., 1996), or size-exclusion chromatography (Shalongo et al., 1992). The slowest process with a denaturant-independent time constant of 500−800 s (depending on temperature) was no longer observed for SNase variants lacking Pro117 (Kuwajima et al., 1991; Nakano et al., 1993), indicating that it is due to isomerization about the Lys116−Pro117 peptide bond, which is predominantly cis under stabilizing conditions (Evans et al., 1987). The second-slowest phase, however, remained. In contrast, the folding kinetics of the Pro$^-$ variant (Figure 2B) show no sign of any fluorescence-detected folding events slower than a few seconds. Thus, the minor folding phase in H124L SNase with a time constant of about 40 s in the absence of denaturant (open squares in Figure 2A) can be attributed to a population of denatured molecules containing non-native isomers for at least one of the remaining five prolyl peptide bonds, other than Pro117.

If proline isomerization were the only source of kinetic complexity, one would expect the folding kinetics of the Pro$^-$ variant to be monophasic. However, a satisfactory fit of the kinetic traces at low GuHCl concentration requires three exponential phases (Figure 4B), including a minor process ($\lambda_3$) with a time constant of about 0.5 s and a relative amplitude of about 20%, in addition to the lag phase ($\lambda_1$) and the main phase ($\lambda_2$). The corresponding process in H124L (Figure 2A) has a 2-fold slower rate, and its denaturant dependence parallels that of the main folding phase. Despite the small amplitude of $\lambda_3$, a process in the 1 s time range has been consistently observed, by both fluorescence and CD, for several SNase variants under various solution conditions (Schechter et al., 1970; Sugawara et al., 1991; Chen et al., 1991; Su et al., 1996).

Quantitative kinetic analysis of the data for the Pro$^-$ variant (Figure 4B) showed that the minor phase can be attributed to a distinct population of unfolded molecules giving rise to a parallel pathway (Scheme 2). The most likely origin of this heterogeneity is the presence of a cis peptide bond at one or several non-proline residues, which has been shown to be responsible for slow folding events in other proteins (Mayr et al., 1994b; Odefey et al., 1995; Vanhove et al., 1996). Odefey et al. (1995) found that a cis peptide bond preceding Pro39 in the native structure of ribonuclease $T_1$ persisted after mutation of Pro39 to Ala. Isomerization of this cis peptide bond to the energetically much more favorable trans form after unfolding (in 6 M GuHCl, pH 1.6 at 25 °C) showed a time constant of 730 ms, whereas the inverse process, trans−cis isomerization, is a much slower process (∼500 s) that limits the rate of refolding. The large difference between the two rate constants is consistent with theoretical estimates and model compound data indicating that the trans form is favored by 2−3 orders of magnitude (Jorgensen & Gao, 1988; Perrin & Dwyer, 1990). In a protein like SNase that does not contain non-prolyl cis peptide bonds in the native state (Loll & Lattman, 1989; Hynes & Fox, 1991), even in the presence of a Pro → Gly substitution at position 117 (Hynes et al., 1994), a small population of molecules may nevertheless acquire cis isomers at any of the non-proline residues in the denatured state. While these are not likely to give rise to prominent slow folding phases, the rate of the cis−trans transition, 1.4 s$^{-1}$, reported by Odefey et al. (1995) is similar to the $k_{I'M}$ value of 2.5 s$^{-1}$ that we found for Pro$^-$ SNase (Table 2). Thus, we tentatively assign the state U′ and I′ in Scheme 2 to a minor (∼20%) population of unfolded and compact forms containing at least one cis peptide bond whose conversion into the native trans form limits the rate of folding along the parallel pathway (I′ → M step). The presence of an intermediate, I′, is not only physically reasonable (a single non-native peptide isomer is not expected to prevent collapse and partial folding) but also necessary to reproduce the denaturant dependence of $\lambda_3$, which is indicative of a tight coupling between isomerization and structure formation. This mechanism may be more general, since minor folding phases with similar time constants (0.1−1 s) and denaturant dependence have been observed for several other proteins, including ubiquitin (Khorasanizadeh et al., 1993), dihydrofolate reductase (Jennings et al., 1993), and cytochrome $c_2$ (Sauder et al., 1996).

Our kinetic mechanism for SNase folding (Scheme 2) departs from previous proposals as it contains obligatory structural intermediates, I and I′, on a direct path toward the native state for each population of unfolded states. For WT SNase with all six proline residues, one would have to include at least two additional parallel pathways with possible structural intermediates to account for the two slower folding phases. By contrast, Chen et al. (1991, 1992b,c) initially proposed a model consisting of a series of denatured states followed by a cooperative conformational step ($D_3 \rightleftharpoons D_2 \rightleftharpoons D_1 \rightleftharpoons N$), which could account qualitatively for the triphasic folding kinetics of WT SNase, but does not explain the complex denaturant dependence of the rates and amplitudes (cf. Figure 2A).

*Unfolding Mechanism.* The unfolding mechanism of SNase is complicated by cis−trans isomerism about the Lys116−Pro117 peptide bond, which gives rise to biphasic unfolding kinetics [Figure 2A; cf. Sugawara et al. (1991) and Su et al. (1996)]. As in previous work on SNase variants lacking Pro117 (Kuwajima et al., 1991; Nakano et al., 1993), the proline-free variant exhibits a single unfolding phase with

a GuHCl-dependent rate (Figure 2B) that closely matches the faster of the two phases observed for the wild-type protein (Figure 2A). However, while elimination of prolines renders the native state kinetically homogeneous, the curved rate profile indicates that unfolding is not a simple two-state process. Our results rule out proline isomerization as a possible explanation and indicate that a structural intermediate is responsible for the curvature in the rate profile at high denaturant concentrations (Figure 2B). Our kinetic simulations show that a sequential mechanism with a single intermediate, M, located directly on the pathway of unfolding (Scheme 1) can fully account for the observed unfolding behavior. At high GuHCl concentrations, the strongly denaturant-dependent main unfolding step (M → I) is so fast that formation of M becomes rate-limiting, which causes the rate profile to level off. This situation is illustrated by the free energy diagram in Figure 6, which shows that at relatively low denaturant concentrations (<2 M GuHCl) the rate-limiting barrier is that between M and I, while at higher denaturant concentrations the transition from N to M becomes the highest point in the free energy profile. The weak dependence of this process on GuHCl concentration ($m^{\ddagger}_{NM}/m_{eq} = 0.06$) indicates that M is nearly as compact as the native state. In contrast to the folding intermediate I, the unfolding intermediate M is always poorly populated, since it decays more rapidly than it is formed ($k_{MI} + k_{MN} \gg k_{NM}$; see Table 2). Nevertheless, the switch in rate-determining barriers in unfolding in going from moderate (1−1.5 M) to high (>1.5 M) GuHCl concentrations (Figure 6) clearly indicates the presence of a structural intermediate.

Evidence for a similar unfolding mechanism has previously been presented for horse cytochrome *c* (Elöve et al., 1994; Colón et al., 1996) and *Rhodobacter capsulatus* cytochrome $c_2$ (Sauder et al., 1996). In these cases, the nonlinear unfolding rate profile can be attributed to the dissociation of the Met80 ligand from the heme iron, which limits the rate of unfolding under strongly denaturing conditions. Jonsson et al. (1996) observed a pronounced downward curvature in logarithmic plots of the unfolding rate *vs* GuHCl and urea concentration for the dimeric arc repressor protein, which they also attributed to a change in rate-limiting barriers. As discussed by Jonsson et al. (1996) and Sauder et al. (1996), possible alternative models, such as intrinsic denaturant effects [cf. Parker et al. (1995)] and denaturant-induced shifts in the position of the transition state (Matouschek & Fersht, 1993) would predict rate profiles with a more gradual change in slope, in contrast to the relatively sharp kink observed in all these cases. Although the structural origin is uncertain, a possible explanation for the unfolding behavior of the arc repressor is that dissociation of the native dimer may precede the major structural unfolding step and become rate-limiting under sufficiently destabilizing conditions.

In contrast to the cytochromes *c* and arc repressor, there is no obvious structural feature to explain the presence of an unfolding intermediate in SNase. Since M is not well-populated, we have no direct information on its physical properties. What we know is that its solvent-accessible surface area is only slightly (6%) larger than that of the native structure. Since small single-domain proteins generally show linear unfolding rate-profiles up to high denaturant concentrations [e.g., Jackson and Fersht (1991), Khorasanizadeh et al. (1993), and Kragelund et al. (1995)], it is reasonable to assume that the complex unfolding behavior of SNase is related to its two-domain structure. One possible scenario is that some critical long-range contacts [especially charged and/or H-bonded interactions; see Hinck et al. (1993, 1996)] between the two subdomains may have to be disrupted before further unfolding can occur. This proposal is supported by the observation that binding of a $Ca^{2+}$ ion and an inhibitor (deoxythymidine 3′,5′-bisphosphate), linking two loop regions across the active-site cleft (Loll & Lattman, 1989), results in a dramatic reduction in the rate of unfolding without affecting the kinetics of folding (Sugawara et al., 1991). This has also been observed in other proteins, where cross-linking of the polypeptide chain *via* disulfide bonds (Mayr et al., 1994a) or metal ligands (Sauder et al., 1996) can greatly reduce the rate of unfolding. Thus, M appears to be a slightly expanded form of the native state with intact subdomains, but lacks some of the specific contacts between the domains, which is consistent with the small *m* value associated with the N → M transition and with previous evidence for two-domain behavior in SNase (Carra et al., 1994; Carra & Privalov, 1995, 1996; Hinck et al., 1996). Further unfolding of this loosened state would then occur in a major structural transition characterized by a strongly denaturant-dependent rate.

*Summary of Kinetic Mechanism.* Contrary to earlier predictions (Sosnick et al., 1994), elimination of potential kinetic complications due to proline isomerization did not result in a simple two-state folding/unfolding process. Observation of a lag in the refolding kinetics and nonlinear denaturant dependence of the folding rate at low GuHCl concentrations provides evidence that a partly folded intermediate is populated on the 10 ms time scale. The kinetic behavior and fluorescence properties associated with this intermediate are consistent with previous amide protection results (Jacobs & Fox, 1994) and kinetic analysis of core mutations (Kalnin & Kuwajima, 1995), which indicate that this early intermediate contains stable structure primarily in the $\beta$-domain. Consistent with earlier evidence for a three-state unfolding equilibrium for the V66W mutant of SNase (Gittis et al., 1993), recent calorimetric studies have indicated that SNase contains two subdomains of differing stability linked by the pairing of two loops across the active site cleft (Carra et al., 1994; Carra & Privalov, 1995). Together with our observation of a kinetic unfolding intermediate that is only slightly expanded in comparison to the native state, this suggests the following structural interpretation of the SNase folding pathway (Scheme 3):

Scheme 3

$$U \rightleftharpoons I_{\beta} \rightleftharpoons I_{\alpha\beta} \rightleftharpoons N$$

where folding begins by formation of the N-terminal $\beta$-sheet domain ($I_{\beta}$) and continues with acquisition of helical structure in the rest of the molecule, resulting in a native-like structure with each of the domains present ($I_{\alpha\beta}$), followed by the docking of the two-domains as a rapid final step.

## ACKNOWLEDGMENT

We thank W. Colón, J. Feng, S.-H. Park, J. M. Sauder, and M. C. R. Shastry for helpful discussions, J. M. Sauder for help with computer graphics, and A. Pomenti and the Protein Analysis Facility for their services. J. L. Markley and members of his laboratory generously provided us with

some nuclease variants used at early stages of this investigation.

## REFERENCES


Alexandrescu, A. T., Hinck, A. P., & Markley, J. L. (1990) *Biochemistry 29*, 4516−4525.

Anfinsen, C. B. (1973) *Science 181*, 223−229.

Anfinsen, C. B., Schecter, A. N., & Taniuchi, H. (1972) in *Cold Spring Harbor Symposia on Quantitative Biology*, pp 249−255, Cold Spring Harbor Laboratory Press, Plainview, New York.

Benson, S. W. (1952) *J. Chem. Phys. 20*, 1605−1612.

Carra, J. H., & Privalov, P. L. (1995) *Biochemistry 34*, 2034−2041.

Carra, J. H., & Privalov, P. L. (1996) *FASEB J. 10*, 67−74.

Carra, J. H., Anderson, E. A., & Privalov, P. L. (1994) *Biochemistry 33*, 10842−10850.

Chen, B.-L., Baase, W. A., Nicholson, H., & Schellman, J. A. (1992a) *Biochemistry 31*, 1464−1476.

Chen, H. M., & Tsong, T. Y. (1994) *Biophys. J. 66*, 40−45.

Chen, H. M., You, J. L., Markin, V. S., & Tsong, T. Y. (1991) *J. Mol. Biol. 220*, 771−778.

Chen, H. M., Markin, V. S., & Tsong, T. Y. (1992b) *Biochemistry 31*, 12369−12375.

Chen, H. M., Markin, V. S., & Tsong, T. Y. (1992c) *Biochemistry 31*, 1483−1491.

Colón, W., Elöve, G. A., Wakem, L. P., Sherman, F., & Roder, H. (1996) *Biochemistry 35*, 5538−5549.

Creighton, T. E. (1990) *Biochem. J. 270*, 1−16.

Creighton, T. E. (1994) *Nat. Struct. Biol. 1*, 135−138.

Dodge, R. W., & Scheraga, H. A. (1996) *Biochemistry 35*, 1548−1559.

Elöve, G. A., Bhuyan, A. K., & Roder, H. (1994) *Biochemistry 33*, 6925−6935.

Evans, P. A., Dobson, C. M., Kautz, R. A., Hatfull, G., & Fox, R. O. (1987) *Nature 329*, 266−268.

Fersht, A. R. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 10869−10873.

Gittis, A. G., Stites, W. E., & Lattman, E. E. (1993) *J. Mol. Biol. 232*, 718−724.

Green, S. M., Meeker, A. K., & Shortle, D. (1992) *Biochemistry 31*, 5717−5728.

Herning, T., Yutani, K., Inaka, K., Kuroki, R., Matsushima, M., & Kikuchi, M. (1992) *Biochemistry 31*, 7077−7085.

Hinck, A. P., Eberhardt, E. S., & Markley, J. L. (1993) *Biochemistry 32*, 11810−11818.

Hinck, A. P., Walkenhorst, W. F., Truckses, D. M., & Markley, J. L. (1996) in *Biological NMR Spectroscopy* (Markley, J. L., & Opella, S. J., Eds.) pp 133−160, Oxford University Press.

Hodel, A., Kautz, R. A., Jacobs, M. D., & Fox, R. O. (1993) *Protein Sci. 2*, 838−850.

Houry, W. A., Rothwarf, D. M., & Scheraga, H. A. (1995) *Nat. Struct. Biol. 2*, 495−503.

Huang, G. S., & Oas, T. G. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 6878−6882.

Hynes, T. R., & Fox, R. O. (1991) *Proteins 10*, 92−105.

Hynes, T. R., Hodel, A., & Fox, R. O. (1994) *Biochemistry 33*, 5021−5030.

Jackson, S. E., & Fersht, A. R. (1991) *Biochemistry 30*, 10428−10435.

Jacobs, M. D., & Fox, R. O. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 449−453.

Jennings, P. A., Finn, B. A., Jones, B. E., & Matthews, C. R. (1993) *Biochemistry 32*, 3783−3789.

Jonsson, T., Waldburger, C. D., & Sauer, R. T. (1996) *Biochemistry 35*, 4795−4802.

Jorgensen, W. L., & Gao, J. (1988) *J. Am. Chem. Soc. 110*, 4212−4216.

Kalnin, N. N., & Kuwajima, K. (1995) *Proteins 23*, 163−176.

Khorasanizadeh, S., Peters, I. D., Butt, T. R., & Roder, H. (1993) *Biochemistry 32*, 7054−7063.

Khorasanizadeh, S., Peters, I. D., & Roder, H. (1996) *Nat. Struct. Biol. 3*, 193−205.

Kiefhaber, T. (1995a) *Proc. Natl. Acad. Sci. U.S.A. 92*, 9029−9033.

Kiefhaber, T. (1995b) in *Methods in Molecular Biology: Protein Stability and Folding: Theory and Practice* (Shirley, B. A., Ed.) pp 313−343, Humana Press, Totowa, NJ.

Kiefhaber, T., Kohler, H., & Schmid, F. X. (1992) *J. Mol. Biol. 224*, 217−229.

Kim, P. S., & Baldwin, R. L. (1990) *Annu. Rev. Biochem. 59*, 631−660.

Kragelund, B. B., Robinson, C. V., Knudsen, J., Dobson, C. M., & Poulsen, F. M. (1995) *Biochemistry 34*, 7217−7224.

Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.

Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 488−492.

Kuwajima, K., Okayama, N., Yamamoto, K., Ishihara, T., & Sugai, S. (1991) *FEBS Lett. 290*, 135−138.

Lin, L.-N., & Brandts, J. F. (1978) *Biochemistry 17*, 4102−4110.

Loll, P. J., & Lattman, E. E. (1989) *Proteins 5*, 183−201.

Matouschek, A., & Fersht, A. R. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 7814−7818.

Matouschek, A., Kellis, J. T., Jr., Serrano, L., Bycroft, M., & Fersht, A. R. (1990) *Nature 346*, 440−445.

Matthews, C. R. (1993) *Annu. Rev. Biochem. 62*, 653−683.

Mayr, L. M., Willbold, D., Landt, O., & Schmid, F. X. (1994a) *Protein Sci. 3*, 227−239.

Mayr, L. M., Willbold, D., Rosch, P., & Schmid, F. X. (1994b) *J. Mol. Biol. 240*, 288−293.

Nakano, T., Antonino, L. C., Fox, R. O., & Fink, A. L. (1993) *Biochemistry 32*, 2534−2541.

Nall, B. T. (1994) in *Mechanisms of Protein Folding: Frontiers in Molecular Biology* (Pain, R. H., Ed.) pp 80−103, Oxford University Press, New York.

Odefey, C., Mayr, L. M., & Schmid, F. X. (1995) *J. Mol. Biol. 245*, 69−78.

Parker, M. J., Spencer, J., & Clarke, A. R. (1995) *J. Mol. Biol. 253*, 771−786.

Perrin, C. L., & Dwyer, T. J. (1990) *Chem. Rev. 90*, 935−967.

Ptitsyn, O. B. (1995) *Adv. Protein Chem. 47*, 83−229.

Roder, H., & Colón, W. (1997) *Curr. Opin. Struct. Biol. 7*, 15−28.

Royer, C. A., Hinck, A. P., Loh, S. N., Prehoda, K. E., Peng, X., Jonas, J., & Markley, J. L. (1993) *Biochemistry 32*, 5222−5232.

Sauder, J. M., MacKenzie, N. E., & Roder, H. (1996) *Biochemistry 35*, 16852−16862.

Schechter, A. N., Chen, R. F., & Anfinsen, C. B. (1970) *Science 167*, 886−887.

Schindler, T., Herrler, M., Marahiel, M. A., & Schmid, F. X. (1995) *Nat. Struct. Biol. 2*, 663−673.

Schmid, F. X. (1992) in *Protein Folding* (Creighton, T. E., Ed.) pp 197−241, W. H. Freeman & Co., New York.

Shalongo, W., Jagannadham, M. V., Heid, P., & Stellwagen, E. (1992) *Biochemistry 31*, 11390−11396.

Shortle, D. (1986) *J. Cell. Biochem. 30*, 281−289.

Shortle, D., & Meeker, A. K. (1989) *Biochemistry 28*, 936−944.

Shortle, D., Stites, W. E., & Meeker, A. K. (1990) *Biochemistry 29*, 8033−8041.

Shortle, D., Wang, Y., Gillespie, J. R., & Wrabl, J. O. (1996) *Protein Sci. 5*, 991−1000.

Sosnick, T. R., Mayne, L., Hiller, R., & Englander, S. W. (1994) *Nat. Struct. Biol. 1*, 149−156.

Su, Z.-D., Arooz, M. T., Chen, H. M., Gross, C. J., & Tsong, T. Y. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 2539−2544.

Sugawara, T., Kuwajima, K., & Sugai, S. (1991) *Biochemistry 30*, 2698−2706.

Tanford, C. (1970) *Adv. Protein Chem. 24*, 1−95.

Tuckses, D. M., Somoza, J. R., Prehoda, K. E., Miller, S. C., & Markley, J. L. (1996) *Protein Sci. 5*, 1907−1916.

Vanhove, M., Raquet, X., Palzkill, T., Pain, R. H., & Frère, J. M. (1996) *Proteins 25*, 104−111.

Wang, J., Hinck, A. P., Loh, S. N., & Markley, J. L. (1990) *Biochemistry 29*, 102−113.

Wang, Y., & Shortle, D. (1995) *Biochemistry 34*, 15895−15905.

Weissman, J. S., & Kim, P. S. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 9900−9904.


BI9700476