Characterization of the Hydrodynamic Properties of the Folding Transition State of an SH3 Domain by Magnetization Transfer NMR Spectroscopy†

Martin Tollinger,[‡,§,||] Chris Neale,[‡,§] Lewis E. Kay,[§,⊥] and Julie D. Forman-Kay*[,‡,§]

Structural Biology and Biochemistry, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada, Department of Biochemistry and Protein Engineering Network Centres of Excellence, University of Toronto, Toronto, Ontario M5S 1A8, Canada, and Departments of Medical Genetics and Chemistry, University of Toronto, Toronto, Ontario M5S 1A8, Canada

Received February 8, 2006; Revised Manuscript Received March 22, 2006

ABSTRACT: Protein folding kinetic data have been obtained for the marginally stable N-terminal Src homology 3 domain of the *Drosophila* protein drk (drkN SH3) in an investigation of the hydrodynamic properties of its folding transition state. Due to the presence of NMR resonances of both folded and unfolded states at equilibrium, kinetic data can be derived from NMR magnetization transfer techniques under equilibrium conditions. Kinetic analysis as a function of urea (less than ~1 M) and glycerol enables determination of α values, measures of the energetic sensitivity of the transition state to the perturbation relative to the end states of the protein folding reaction (the folded and unfolded states). Both end states have previously been studied experimentally by NMR spectroscopic and other biophysical methods in great detail and under nondenaturing conditions. Combining these results with the kinetic folding data obtained here, we can characterize the folding transition state without requiring empirical models for the unfolded state structure. We are thus able to give a reliable measure of the solvent-accessible surface area of the transition state of the drkN SH3 domain (4730 ± 360 Å^2) based on urea titration data. Glycerol titration data give similar results and additionally demonstrate that folding of this SH3 domain is dependent on solvent viscosity, which is indicative of at least partial hydration of the transition state. Because SH3 domains appear to fold by a common folding mechanism, the data presented here provide valuable insight into the transition states of the drkN and other SH3 domains.

Small proteins that fold without a significant accumulation of folding intermediates (nominally "two-state" folding proteins) have been the focus of intense experimental and computational studies in recent years (*1*−*3*). For these proteins, interconversion between the two end (folded and unfolded) states is limited by the formation of a transition state, representing the ensemble of conformations of highest free energy along the protein folding pathway (*4*). Together with detailed analyses of the structures and biophysical properties of folded and unfolded proteins, the characterization of the transition state ensembles (TSEs)[1] is vital in obtaining a comprehensive view of folding pathways and understanding the mechanisms by which proteins fold.

It is difficult to analyze the protein folding transition state by conventional experimental techniques, because the TSE is only transiently populated. The protein engineering method, pioneered by Fersht and co-workers (*5*), can provide detailed structural information about the rate-limiting step of folding. In this approach, changes in kinetic and thermodynamic parameters upon mutation of residues in different regions of a protein are used to calculate ϕ_f values, which provide residue-specific structural information about interactions that are present in the folding transition state. Complementary information about global parameters of the TSE, such as its compactness, thermodynamic parameters, or hydration, has been obtained by monitoring changes in kinetic folding or unfolding rate constants upon perturbation of the folding−unfolding equilibrium with denaturing cosolvents, temperature, or pressure employing (linear) rate−equilibrium free energy relationships (*6*, *7*).

Recently, experimental ϕ_f values have been used as input restraints for the calculation of ensembles of structures that represent the folding transition state of various two-state folding proteins (*8*−*20*). From these calculations, it appears that the TSE structures of most proteins have global fold characteristics similar to those of the fully folded protein and that the topology of the folding nucleus that is present at the rate-limiting step of folding is defined by a network of interactions between a subset of residues. It was concluded that these "key residues" represent the minimal set of interactions that are required to attain the overall topology of the TSE (*8*, *10*, *15*) and enable rapid folding without

† The research was supported by grants from the Canadian Institutes of Health Research (J.D.F.-K. and L.E.K.). M.T. is a recipient of an E. Schrödinger Return Fellowship of the Austrian Science Fund (R18).

* To whom correspondence should be addressed. Phone: (416) 813-5358. Fax: (416) 813-5022. E-mail: forman@sickkids.ca.

‡ Hospital for Sick Children.

§ Department of Biochemistry and Protein Engineering Network Centres of Excellence, University of Toronto.

|| Current address: Department of Biomolecular Structural Chemistry, University of Vienna, Campus Vienna Biocenter 5, A-1030 Vienna, Austria.

⊥ Departments of Medical Genetics and Chemistry, University of Toronto.

[1] Abbreviations: SH3, Src homology 3; drkN SH3 domain, N-terminal SH3 domain of the *Drosophila* protein drk; TSE, transition state ensemble; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; SASA, solvent-accessible surface area; MD, molecular dynamics; MC, Monte Carlo; 2D, two-dimensional.

10.1021/bi060268o CCC: $33.50
Published on Web 04/27/2006

requiring a large number of specific interactions in the conformational search process.

Src homology 3 (SH3) domains are among the most extensively investigated models for protein folding, using both experimental and computational methods and combinations thereof. In the folded state, SH3 domains are small globular proteins that are composed of two β-sheets, which pack orthogonally to each other to form a hydrophobic core (*21*). Experimental studies on SH3 domain folding have focused on ϕ_f value analysis and have provided a wealth of information about the structure of the TSE (*22*−*29*). These studies have revealed a structurally polarized folding transition state in which part of the native (folded) structure is almost completely formed, whereas a number of residues are considerably less structured. This is consistent with a nucleation−condensation mechanism for folding, where early nucleation appears to be the rate-limiting step and formation of the hydrophobic core occurs after the folding transition state (*30*). Experimental ϕ_f values have been used as input restraints for the calculation of structures that represent the TSE of SH3 domains (*11*, *14*, *17*, *20*), showing that the TSE exhibits a high degree of nativelike topology. It was concluded that solvent water is only partially excluded from the folding transition state, indicating a partially hydrated TSE (*14*, *20*). In addition, computational folding studies performed on SH3 domains (*31*, *32*) also suggest a common folding mechanism for SH3 domains, where the fully solvated unfolded polypeptide chain undergoes an initial structural collapse to form a fairly compact folding TSE, followed by further rearrangements to complete the folded state structure, including the formation of the hydrophobic core, and the expulsion of water from the core region only late in the folding process. It was again concluded that, despite its compactness, the SH3 domain TSE is partially hydrated.

Here, we endeavor to complement the experimental data on SH3 domain folding transition states by studying the folding and unfolding kinetics of the N-terminal SH3 domain (residues 1−59) of the *Drosophila* protein drk (drkN SH3). For a number of reasons, the drkN SH3 domain is a model particularly well-suited for folding kinetic studies and investigation of transition state properties. Due to the low thermodynamic stability of the drkN SH3 domain, the unfolded state of this protein is significantly populated under nondenaturing conditions and can be studied in detail by standard NMR and other spectroscopic methods. Previous experimental data that have been collected for the unfolded drkN SH3 domain showed it to be an ensemble of rapidly interconverting, structurally and dynamically diverse conformers that exhibit considerable amounts of nonrandom structure. Much of this residual structure becomes disrupted when denaturant is added (*33*−*38*). The structure of the drkN SH3 domain folded state has also been determined (*21*), and its two-state folding mechanism has been experimentally established (*39*). Since this SH3 domain is thermodynamically marginally stable under nondenaturing conditions [the unfolding free energy ($\Delta G_{F\rightarrow U}$) = 0.4 kcal/mol at pH 6.0 and 15 °C], NMR spectroscopic relaxation methods can be used to obtain accurate kinetic data under equilibrium conditions that do not require the perturbation of the folding−unfolding equilibrium. (Note that kinetic analysis by conventional stopped-flow kinetic experiments is not possible without high-salt stabilizing conditions.) Because experimental and computational studies of various SH3 domains suggest a common folding pathway and because TSEs of different SH3 domains appear to have very similar topologies (*14*), our results can be readily transferred to other SH3 domains.

In this article, information about the folding transition state under nondenaturing conditions is derived from rate−equilibrium free energy relationships by variation of solvent conditions (*6*). We have previously characterized a specific electrostatic interaction in the TSE of the drkN SH3 domain that is conserved in SH3 domains (*40*) from pH-dependent measurements (*40*, *41*) using this approach. Very recently, we have investigated the effect of moderate pressure to probe the volume of the transition state (*42*). We now focus on complementary hydrodynamic properties of the TSE by means of denaturant and glycerol titrations. Due to the low thermodynamic stability of the drkN SH3 domain, which renders the unfolded state under nondenaturing conditions amenable to detailed spectroscopic analysis, the solvent-accessible surface area of the unfolded state can be determined. Utilizing this information in conjunction with the kinetic data reported here, we obtain a reliable estimate of the size of the solvent-accessible surface area of the folding transition state that is based on existing structural ensembles of the end states of the protein folding reaction and does not require the use of empirical relationships to estimate the size of the unfolded state solvent-accessible surface area. In addition, we have obtained information about the viscosity dependence of folding, providing information about the nature of the rate-limiting step(s) involving the transition state, which is consistent with results from computational SH3 domain folding studies.

MATERIALS AND METHODS

The drkN SH3 domain was expressed and purified as described previously (*33*). NMR experiments were performed on 1.0 mM samples of ^{15}N-labeled SH3 domain containing 50 mM sodium phosphate in a 92% H_2O/8% D_2O mixture at 15 °C on a Varian UNITY spectrometer with a room-temperature triple-resonance probehead at 500 MHz (^{1}H frequency). Folding and unfolding rate constants of the drkN SH3 domain were determined by simultaneously fitting the experimental data from a series of two-dimensional ^{1}H−^{15}N exchange correlation spectra with mixing times, T, ranging from 11 to 451 ms to analytical expressions describing the time course of longitudinal magnetization, including chemical exchange between two sites (*43*). The dependence of the auto (ff,uu) and exchange (fu,uf) cross-peak volumes on the variable mixing time is given by

$$I_{ff}(T)/I_{ff}(0) = A_f[-(\lambda_2 - a_{11})e^{-\lambda_1 T} + (\lambda_1 - a_{11})e^{-\lambda_2 T}]/(\lambda_1 - \lambda_2)$$

$$I_{uu}(T)/I_{uu}(0) = A_u[-(\lambda_2 - a_{22})e^{-\lambda_1 T} + (\lambda_1 - a_{22})e^{-\lambda_2 T}]/(\lambda_1 - \lambda_2)$$

$$I_{fu}(T)/I_{ff}(0) = A_u(a_{21}e^{-\lambda_1 T} - a_{21}e^{-\lambda_2 T})/(\lambda_1 - \lambda_2)$$

$$I_{uf}(T)/I_{uu}(0) = A_f(a_{12}e^{-\lambda_1 T} - a_{12}e^{-\lambda_2 T})/(\lambda_1 - \lambda_2) \quad (1)$$

where $\lambda_{1,2} = {}^1/_2\{(a_{11} + a_{22}) \pm [(a_{11} - a_{22})^2 + 4k_f k_u]^{1/2}\}$, $a_{11} = R_{1f} + k_u$, $a_{12} = -k_f$, $a_{22} = R_{1u} + k_f$, and $a_{21} = -k_u$. R_{1u} and R_{1f} are the longitudinal relaxation rate constants of

magnetization in sites U and F, respectively, and $I_{uu}(0)$ and $I_{ff}(0)$ denote the longitudinal magnetization associated with the unfolded and folded states, respectively, at $T = 0$. Factors A_f and A_u represent the efficiency of coherence transfer after the mixing period T, and were determined as described previously (*44*). Values of k_u and k_f, along with R_{1u} and R_{1f}, were extracted by least-squares fitting the experimental data, $I_{uu}(T)$, $I_{ff}(T)$, $I_{fu}(T)$, and $I_{uf}(T)$, simultaneously to the expressions given above, using in-house written software (available on request). The longitudinal exchange experiment yielded residue-specific rate constants that were uniform for well-resolved residues within 10%. Averages of residue-specific values of k_f and k_u were employed in all analyses and errors in rate constants estimated on the basis of the standard deviation of the values over the residues that were included in the average.

The urea titration data were recorded at pH 6.0 and eight urea concentrations (0, 0.19, 0.28, 0.40, 0.57, 0.73, 0.89, and 1.04 M). Analytical grade urea was purified with Amberlite MB-150 (Sigma) mixed bed ion-exchange resin, and concentrations were verified by refractometry. The glycerol titration data were recorded at pH 6.0 in the absence of glycerol as well as at five glycerol concentrations [2.5, 5.0, 7.5, 10.0, and 12.5% (w/v)] using d_8-glycerol (Cambridge Isotope Laboratories) without further purification. The protein concentration and the pH were kept constant throughout the titration experiments.

Relative solvent viscosities, η/η_0 (where η_0 is the viscosity of the sample with no cosolvent present), for the urea and glycerol titrations were determined using a pulsed-field-gradient NMR experiment designed to measure translational diffusion coefficients (*36*) from the ratio of diffusion constants measured at different viscogen concentrations. The analysis assumes that the hydrodynamic radius of the protein is invariant with viscogen concentration, which was confirmed experimentally for CspB in glycerol (*45*) and is likely true for drkN SH3 considering the relatively low concentrations of viscogen used. Only diffusion data for residues from the folded state were analyzed, and averages over 17 well-resolved residues were taken. For the urea titration (up to 1.04 M urea), only small viscosity changes were observed in this experiment ($\eta/\eta_0 = 1.07$ at 1.04 M urea), in agreement with ref *46*, due to the low concentrations of urea that were used. For the glycerol titration, relative solvent viscosities (η/η_0) of 1.10, 1.15, 1.31, 1.49, and 1.62 were obtained.

Analysis of the Kinetic Data. For the analysis of kinetic data, errors in k_f and k_u were estimated as described above and subsequently propagated to errors in $\Delta G_{F\rightarrow U}$, $-RT \ln k_f$, and $-RT \ln k_u$. Linear least-squares fits were employed to extract equilibrium and kinetic m values and α values. Experimental uncertainties in m values were estimated via a Monte Carlo approach (*47*) in which 1000 synthetic data sets were generated using the extracted m values and k_f and k_u at zero denaturant concentration along with the experimental error in the rates, and the fit was repeated for all data sets. The errors quoted in the paper are standard deviations in fitted m values that were obtained in this procedure. A similar approach was employed for the reported α values.

Calculation of Solvent-Accessible Surface Areas. For the calculation of the solvent-accessible surface area of the unfolded state, structural ensembles consistent with unfolded state experimental restraints (*33*, *36*−*38*, *48*) were generated using two conformer pools and three ENSEMBLE (*49*) pseudo-energy minimization algorithms (*50*). The two conformer pools were (1) random structures generated from TraDES (*51*) using GOR3 (*52*) sampling and (2) a mixture of input structures including (a) those from multiple CNS (*53*) structure calculations using a randomly chosen 80% of total unfolded state NOEs, (b) those generated from AMBER (*54*) high-temperature unfolding trajectories from the folded drkN SH3 domain (M. Philippopoulos, J. D. Forman-Kay, and R. Pomès, unpublished data), (c) structures generated within ROSETTA (*55*), (d) random and compact structures from TraDES with GOR3, amino acid-biased or uniform ϕ and ψ sampling, (e) those from TraDES unfolding trajectories from the folded drkN SH3 domain, and (f) those from TRADES unfolding trajectories of the drkN SH3 domain sequence misthreaded onto dissimilar folds using ESyPred3D (*56*). Following minimization with three different algorithms of the population weights for each of the 100000−300000 conformers within these two large sets of structures to best fit experimental data, a total of six ensembles of more highly populated conformers were obtained, each consisting of 60−150 structures. The total solvent-accessible surface area (SASA) of heavy atoms was calculated for the resulting unfolded state ensembles and for the salt-stabilized folded structure (*21*) using CNS with a water probe radius of 1.4 Å. Prior to ENSEMBLE minimization, hydrodynamic radii were predicted using HYDROPRO (*51*) as outlined in ref *36*. The calculated SASA was consistent among the six ensembles (with a standard deviation of only 260 Å^2), likely due to a strong correlation between SASA and the experimental hydrodynamic restraints utilized in the minimization. Error bounds for the folded state were obtained from the 20 best structures of the recently calculated folded state structure (*21*). In addition, upper and lower boundaries for the solvent-accessible surface area of the unfolded state derived from a database were determined as described previously (*57*).

RESULTS AND DISCUSSION

Rate−Equilibrium Free Energy Relationships. Protein folding transition states are routinely investigated via rate−equilibrium free energy relationships (*58*) in which the thermodynamic stability of a protein is perturbed and changes in kinetic folding and unfolding rate constants upon perturbation are monitored (reviewed in refs *6* and *7*). A proportionality constant α can be defined by relating changes in activation free energy for folding, $\Delta\Delta G_{U\rightarrow\ddagger}$, to changes in the equilibrium free energy (folding free energy), $\Delta\Delta G_{U\rightarrow F}$, upon perturbation as

$$\alpha = \frac{\Delta\Delta G_{U\rightarrow\ddagger}}{\Delta\Delta G_{U\rightarrow F}} \qquad (2)$$

Within the framework of transition state theory, α values provide a measure of the energetic sensitivity of the TSE with respect to a perturbation relative to the sensitivity of the folded and unfolded states. The most frequently used method for altering the thermodynamic stability of a protein for the purpose of studying the TSE is site-directed mutagenesis (*59*). Typically, single mutations are made at various positions in the protein to yield ϕ_f values, which represent the change in stability of the TSE relative to the change in stability of the folded state by taking the unfolded state as a reference, and which are interpreted as a measure of the extent of nativelike structure in the TSE. Thus, a ϕ_f

of 1 indicates that a particular residue is involved in a highly structured region in the TSE, whereas a ϕ_f of 0 suggests that a residue forms few interactions with other residues in the TSE. Intermediate ϕ_f values are more difficult to interpret because they can be caused by partial formation of contacts in a single dominant folding pathway or by distinctive parallel pathways in which interactions are either fully formed or absent (*60*).

Other frequently used methods of altering the stability of a protein include changes in solvent conditions, such as the addition of chemical denaturants. For example, variation of the denaturant concentration and calculation of the corresponding α value from kinetic data yield information about the compactness of the TSE relative to the folded and unfolded states (commonly termed the Tanford β_T value). Changes in kinetic data with changes in solvent conditions are usually measured under a number of different solvent conditions, allowing for the determination of accurate α values from plots of $\Delta\Delta G_{U\rightarrow\ddagger}$ versus $\Delta\Delta G_{U\rightarrow F}$. In addition, nonlinearities in rate−equilibrium free energy relationships which are indicative of changes in the folding mechanism or in the rate-limiting step can be detected (*6*, *60*, *61*).

Equilibrium NMR Kinetic Experiments. Unfolding and refolding kinetics of proteins are typically studied by transient methods in which the thermodynamic equilibrium between folded and unfolded states is rapidly perturbed by chemical denaturants, pH, temperature, or pressure shocks (*62*, *63*). Subsequently, the recovery of equilibrium is monitored by observing optical properties as a function of time. For a two-state folding protein

$$U \underset{k_u}{\overset{k_f}{\rightleftharpoons}} F$$

k_f and k_u are the microscopic folding and unfolding rate constants, respectively. The rate by which equilibrium is approached, k_{obs}, is given by the sum of the folding and unfolding rate constants ($k_{obs} = k_f + k_u$). Microscopic rate constants k_f and k_u cannot be separately determined but are obtained in an indirect manner from a plot of the dependence of the logarithm of k_{obs} on denaturant concentration (chevron plot), for example, by extrapolation of the refolding and unfolding limbs to 0 M denaturant, respectively. For thermodynamically unstable proteins such as the drkN SH3 domain, however, measurement of refolding kinetics by transient experiments is not possible because the refolding limb of the chevron plot cannot be sampled sufficiently to determine k_f by extrapolation.

NMR equilibrium techniques are ideally suited for kinetic measurements of marginally stable (or unstable) proteins. The relative values of k_{obs} and the change in resonance frequency, $\Delta\omega$, between the two states define whether an exchange process is termed fast ($k_{obs} \gg \Delta\omega$), intermediate ($k_{obs} \approx \Delta\omega$), or slow ($k_{obs} \ll \Delta\omega$) on the NMR chemical shift time scale. For the great majority of residues in the drkN SH3 domain, the interconversion between the folded and unfolded states is slow on the NMR chemical shift time scale (k_f and k_u, at pH 6.0 and 15 °C, are 0.99 ± 0.05 and 0.47 ± 0.04 s^{-1}, respectively, for a k_{obs} of 1.5 s^{-1}), giving rise to separate sets of resonances for more than 90% of residues of the two protein states in NMR spectra. The minimum difference in resonance frequency for which separate peaks are observed in the ^1H or ^{15}N dimension of the HSQC spectrum at 500 MHz is ∼5 Hz, leading to $\Delta\omega$ values of >30 s^{-1} (much greater than k_{obs}). Such slow processes can be measured using NMR spectroscopic relaxation techniques that monitor the exchange of longitudinal nitrogen magnetization (*43*, *64*). These experiments do not require a perturbation from equilibrium but directly monitor the interconversion of the two states under equilibrium conditions. Exchange cross-peaks are observed, which result from the transfer of magnetization from the folded to the unfolded state (and vice versa) due to the dynamic nature of the equilibrium. Microscopic folding and unfolding rate constants can be obtained simultaneously without requiring a denaturant and extrapolation. We have recently used this method to characterize a specific electrostatic interaction in the TSE of the drkN SH3 domain (*41*).

This method can be applied to systems where exchange between the folded and unfolded states is slow enough to observe a discrete set for each state in NMR spectra yet fast enough to give exchange cross-peaks, and the populations of both states are large enough to generate observable signals for both states. It is therefore particularly useful where the unfolding free energy $|\Delta G_{F\rightarrow U}| < \sim 1$ kcal/mol. Apart from marginally stable proteins such as the drkN SH3 domain, for various proteins exchange between the folded and unfolded states is slow on the NMR chemical shift time scale at or near the denaturation midpoint or near the melting point (*45*, *65*, *66*). Alternatively, line shape analysis can be employed to extract folding and unfolding rate constants from proton resonance frequencies and line broadening that results from the contribution of the interconversion of species to NMR spin relaxation (*67*). In addition, NMR spin relaxation dispersion techniques have been used to study protein folding under equilibrium conditions (*64*). Spin relaxation dispersion methods are inherently more sensitive than line shape analysis and can even be employed in cases where the populations of the involved states are low ($|\Delta G_{F\rightarrow U}| < \sim 3$ kcal/mol).

For a number of reasons, kinetic experiments that do not require a perturbation of the folding−unfolding equilibrium are an attractive alternative to transient methods. For example, structural differences between unfolded states of proteins under nondenaturing and denaturing conditions have been observed experimentally (*33*, *37*, *68*), indicating that unfolded states in the absence of denaturants are generally more structured than unfolded proteins under denaturing conditions. Structural propensities such as transiently formed nativelike or non-nativelike interactions that are present in the unfolded state only under nondenaturing conditions can potentially affect the folding kinetics of a protein by biasing the folding process (*69*, *70*). In fact, the unfolded drkN SH3 domain under nondenaturing conditions contains rather compact conformations that exhibit transiently populated secondary structure and tertiary contacts, whereas under denaturing conditions, the unfolded state ensemble appears to comprise more extended conformations and displays significantly fewer intramolecular interactions (*35*, *38*). Some of the interactions that are present in the unfolded state under nondenaturing conditions are cooperatively disrupted upon addition of denaturants (*37*). Since residual structure in unfolded states of proteins can be relevant for protein folding, spectroscopic methods that do not require denaturing of the protein can potentially be more informative for providing a description of protein folding pathways.

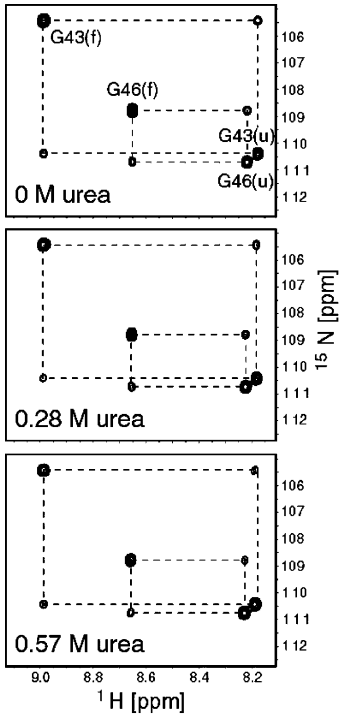

FIGURE 1: Longitudinal exchange NMR experiment at three different urea concentrations (0, 0.28, and 0.57 M) showing for each backbone amide a pair of auto peaks corresponding to the folded and unfolded states as well as a pair of exchange cross-peaks resulting from the transfer of magnetization due to folding and unfolding transitions. Only the high-field (^{15}N) portions of 2D spectra with the two glycine residues of the drkN SH3 domain (Gly43 and Gly46) are shown.

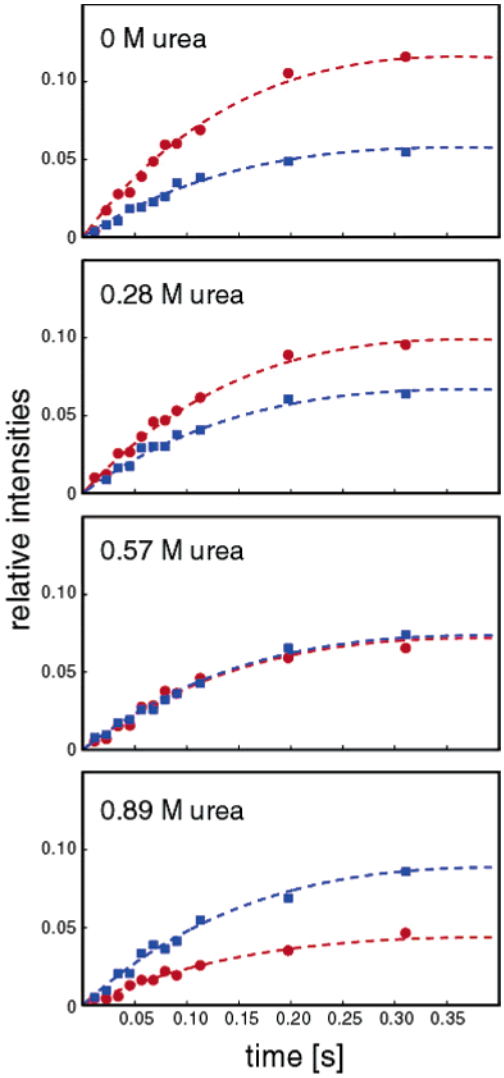

FIGURE 2: Relative intensities of the folding (red) and unfolding (blue) cross-peaks for residue Gly43 as a function of time (normalized to the intensity of the corresponding auto peak) together with best-fit curves at four different urea concentrations (0, 0.28, 0.57, and 0.89 M).

In addition, the assumption is frequently made that the folding and unfolding pathways are identical so that, for example, the TSE for protein folding can be studied via unfolding experiments. As dictated by the principle of microscopic reversibility, this is necessarily true if both folding and unfolding experiments are performed under the same conditions. Since equilibrium experiments do not require a perturbation of the folding−unfolding equilibrium, the kinetic parameters for folding and unfolding that are obtained simultaneously and under identical experimental conditions are indeed determined by identical pathways. However, by employing transient methods, folding experiments are usually performed by refolding of a denatured state into conditions under which the protein is stable, whereas kinetic unfolding experiments are performed under denaturing conditions, in which case unfolding does not necessarily occur by the same pathways as folding (*71*, *72*).

Urea Titration Experiment. Urea was added to the drkN SH3 domain (from 0 to 1.04 M) and magnetization transfer between the folded and unfolded states monitored (Figures 1 and 2). For a thermodynamically marginally stable protein (such as the drkN SH3 domain), only low concentrations of denaturant are required to produce a sufficiently large change in stability to determine the denaturant-related α value by our equilibrium experimental approach (leading to an effect on the equilibrium ratio between the folded and unfolded states and intensities of the folded and unfolded state peaks). Note that the highest concentration of urea that was used (1.04 M) is lower than the guanidinium chloride denaturant concentrations for which a cooperative melt of residual structure in the unfolded state was reported (*37*), consistent with minimal perturbation of the properties of the unfolded state. Folding and unfolding rate constants were determined by analysis of the intensities of the folded, unfolded, and exchange cross-peaks as a function of urea concentration (Figure 3). Figure 3a shows a plot of the drkN SH3 domain unfolding free energy [$\Delta G_{F\rightarrow U} = -RT \ln(k_u/k_f)$] versus urea concentration. The equilibrium m value for urea denaturation (m_{eq}) determined from these data is -0.85 ± 0.03 kcal mol^{-1} M^{-1}, consistent with urea denaturation data from other SH3 domains (m_{eq} ranges from -0.75 to -0.90) (*28*, *29*, *73*). A significant curvature in Figure 3a is not apparent, which confirms the validity of the empirical linear extrapolation method (*74*) for the drkN SH3 domain even at low denaturant concentrations.

By plotting ln k_f and ln k_u as a function of urea concentration, we can reconstitute the folding and unfolding limbs of the chevron plot as shown in Figure 3b, and kinetic m values m_{kf} and m_{ku} can be derived from the slopes ($m_{kf} = -1.13 \pm 0.04$ M^{-1} and $m_{ku} = 0.35 \pm 0.03$ M^{-1}). The urea-related α value is given by $\alpha_{urea} = |m_{kf}|/(|m_{kf}| + |m_{ku}|)$ and is 0.76 ± 0.04 for the drkN SH3 domain. Kinetic m values and the urea-related α value are similar to those of other

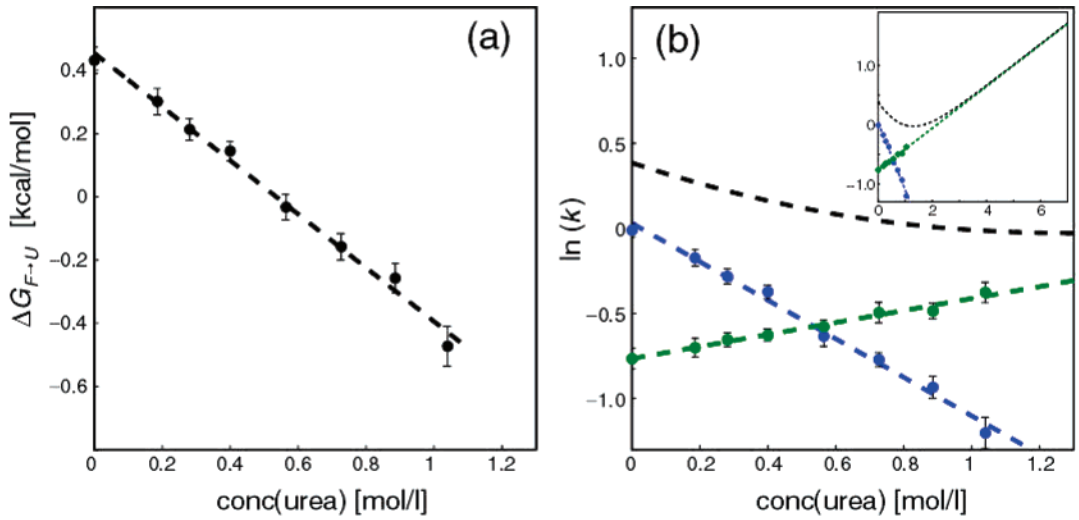

FIGURE 3: (a) Plot of $\Delta G_{F\rightarrow U} = -RT \ln(k_u/k_f)$ vs urea concentration. The equilibrium m value (m_{eq}) as determined from the slope in a linear least-squares fit (dashed line) equals -0.85 ± 0.03 kcal mol^{-1} M^{-1}. (b) Kinetic folding and unfolding data of the drkN SH3 domain for urea concentrations between 0 and 1.04 M. Plot of ln k_f (blue circles) and ln k_u (green circles) vs urea concentration. At each urea concentration, both microscopic folding and unfolding rate constants, k_f and k_u, respectively, were obtained simultaneously from the longitudinal exchange experiment. Experimental rate constants were corrected for the very small (at the concentrations used) effect of urea concentration on viscosity using eq 4. Linear least-squares fits of the folding (blue dashed line) and unfolding (green dashed line) experimental data give the following: $m_{kf} = -1.13 \pm 0.04$ M^{-1} and $m_{ku} = 0.35 \pm 0.03$ M^{-1}. The black dashed line shows the urea dependence of log k_{obs} as calculated from k_f and k_u at 0 M urea using m_{kf} and m_{ku}. This corresponds to data that would be observed in stopped-flow transient experiments (reconstituted chevron plot). The inset shows the expansion of the chevron plot.

SH3 domains (0.7−0.8) (*26*−*29*, *73*). It should be noted that using NMR magnetization transfer techniques both k_f and k_u are obtained independently and at each urea concentration. This is conceptually different from line shape analysis, where data are commonly analyzed by fitting shapes and frequencies of resonances in multiple spectra that are obtained from a denaturation series (*67*, *75*, *76*).

Of note, no significant deviations of either the folding or the unfolding limb of the chevron plot from linearity are observed in Figure 3b. Curvatures in chevron plots can arise due to formation of intermediates along the folding pathway (kinetic rollover), as well as other effects, including solvent viscosity changes (*77*, *78*). Because NMR equilibrium techniques can sample both the folding and the unfolding limb of the chevron plot simultaneously, intermediates that affect unfolding rates at low denaturant concentrations can potentially be detected, whereas for most proteins, transient stopped-flow data (log k_{obs}) at low denaturant concentrations (left of the minimum in the chevron plot) are dominated by folding rates. Considering that intermediates appear to be most populated at low denaturant concentrations, this is a particularly useful feature of NMR equilibrium kinetic experiments. For the drkN SH3 domain, the linearity of the unfolding limb at low urea concentrations is, however, clearly consistent with two-state folding.

Denaturant α values are a measure of the "compactness" of the folding TSE as they reflect the difference in solvent-accessible surface area (SASA) between the TSE and the unfolded state scaled by the increase in solvent-accessible surface area upon unfolding, as given by

$$\alpha_{urea} = \frac{SASA_{\ddagger} - SASA_U}{SASA_F - SASA_U} \qquad (3)$$

where ‡ represents the transition state. The α_{urea} value of 0.76 ± 0.04 for the drkN SH3 domain indicates that ∼25% of the surface area that is exposed to solvent upon unfolding is solvent-accessible in the transition state. To relate this relative number into an absolute estimate of the solvent-accessible surface area of the folding transition state, the solvent-accessible surface areas of both folded and unfolded states have to be known.

From pulsed-field gradient NMR spectroscopic experiments (yielding translational diffusion coefficients) in combination with small-angle X-ray scattering (SAXS) data, the size distribution of the unfolded state ensemble of the drkN SH3 domain has been studied previously, revealing that the unfolded state of this protein under nondenaturing conditions contains fairly compact conformations with the average size of conformers within the unfolded state ensemble only ∼30−40% larger than the folded state structure (*36*). Using this information together with detailed site-specific structural information based on tryptophan solvent accessibility (*37*), experimental NOE (*38*), chemical shift (*33*), *J* coupling (*33*), and paramagnetic relaxation enhancement data (*50*), ensembles of representative structures of the drkN SH3 unfolded state were calculated that are consistent with the experimental data, using the software program ENSEMBLE (*49*). From these structures and from the NMR solution structure of the folded drkN SH3 domain (*21*), we calculate the solvent-accessible surface area of the unfolded and folded states: $SASA_U = 6650 \pm 260$ Å^2 and $SASA_F = 4120 \pm 130$ Å^2. Note that the $SASA_U$ is largely restrained by the hydrodynamic data (translational diffusion coefficient ratio and tryptophan solvent accessibility) such that there was minimal variability depending on input structures biased either by nativelike conformations or by highly random structures. On the basis of these figures, we obtain a measure of the average solvent-accessible surface area of the transition state ensemble from eq 3 ($SASA_{\ddagger} = 4730 \pm 360$ Å^2).

Viscosity Dependence of Folding and Unfolding Kinetics. For two-state folding proteins, transition state theory yields simple monoexponential rate laws for a reaction across an energy barrier of the form

$$k_f = \kappa \exp\left(\frac{-\Delta G_{U\rightarrow\ddagger}}{RT}\right) \quad (4)$$

where k_f is the folding rate constant, $\Delta G_{U\rightarrow\ddagger}$ is the activation free energy for folding, and κ is a transmission coefficient, related to the probability that the reaction proceeds to product from the transition state. The corresponding equation for k_u is obtained by replacing U with F. An alternative theory that was developed by Kramers for reactions in solution describes Brownian motion over potential energy barriers in viscous solvents (*79*). Kramers' formalism takes viscosity effects into account, as solvent collisions can impede the movement of a reactant molecule along a reaction coordinate in high-friction media (such as aqueous solutions), rendering the pre-exponential term solvent viscosity-dependent (inversely proportional to the solvent viscosity η).

$$k_f \sim \frac{1}{\eta} \exp\left(\frac{-\Delta G_{U\rightarrow\ddagger}}{RT}\right) \quad (5)$$

The principle of microscopic reversibility dictates that the viscosity dependence of a barrier-crossing process is independent of the direction of the crossing (i.e., both folding and unfolding are equally affected by viscosity, and an analogous expression holds for k_u). For protein folding studies, Kramers' formalism appears to be more appropriate than the original transition state formalism because protein folding necessarily involves diffusional events, as the expanded and hydrated unfolded state polypeptide chain collapses toward a more compact and conformationally restricted folded protein. During this compaction, solvent water is being displaced relative to the polypeptide chain as water is expelled from the interior of the protein. According to eq 5, in cases where the rate-limiting step involves such diffusive processes, a $1/\eta$ dependence of k_f (and k_u) on solvent viscosity is expected, whereas rate constants will be independent of solvent viscosity if the rate-limiting step involves only rearrangements that are limited by the internal friction of the protein.

The viscosity dependence of protein folding can be tested experimentally by comparing changes in folding or unfolding rate constants with changes in solvent viscosity mediated by the addition of viscogenic cosolvents. Because viscosity affects both folding and unfolding to the same extent, $1/\eta$, it does not change the equilibrium of the folding reaction (i.e., protein stability). However, viscogenic agents do affect protein stability due to differential interactions with the folded and unfolded state ensembles. While denaturants (urea and guanidine) as cosolvents are known to favorably interact with the protein surface and therefore result in a destabilization of the folded state relative to the unfolded state, polyols such as sugars or glycerol, which are frequently used as viscogenic cosolvents, stabilize folded proteins due to preferential exclusion from the protein surface (*80*). The effect of a cosolvent, whether it is stabilizing or destabilizating, is defined by the balance between the affinities of the protein for water and the cosolvent. Because preferential exclusion is thermodynamically unfavorable, the system will tend to minimize the effect by decreasing the area of protein−solvent contact and the equilibrium will shift to the smaller (folded) protein (*81*). To a first approximation, both viscogen- and denaturant-related α values therefore reflect the relative solvent exposure of the folding transition state and should be similar in magnitude.

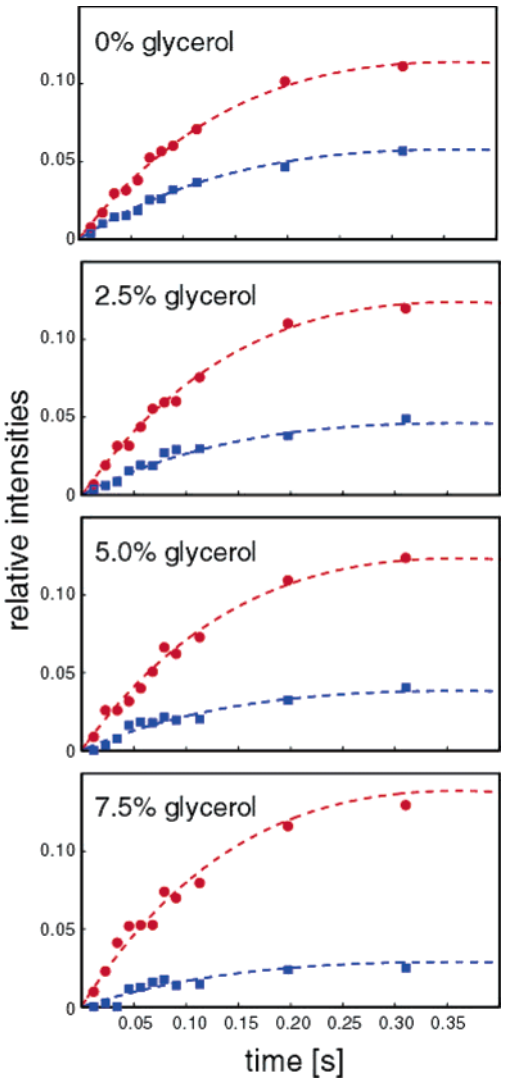

FIGURE 4: Relative intensities of the folding (red) and unfolding (blue) cross-peaks in the longitudinal exchange NMR experiment for residue Gly43 as a function of time for four glycerol concentrations (0, 2.5, 5.0, and 7.5%, w/v), together with best-fit curves.

The commonly used approach to deconvoluting the dual effects of viscogenic agents is the isostability technique, which assumes that the effect of viscogenic cosolvents on folding and unfolding rate constants due to the stabilizing effect can be exactly counterbalanced by denaturant (i.e., viscogen-related and denaturant-related α values are identical). Folding rate constants are determined under conditions that produce equivalent stabilities, by counteracting the effect of viscogenic agents on stability by adding destabilizing cosolvents such as denaturants. Proteins with viscosity-dependent folding will then show a $1/\eta$ dependence of folding rate constants on solvent viscosity (*82*).

Glycerol Titration Experiment. To measure folding kinetics as a function of viscogen, we added glycerol to the drkN SH3 domain (0−12.5%, w/v) and recorded magnetization transfer between the folded and unfolded states (Figure 4). In our approach, we do not add denaturant to counterbalance the effect of viscogen on stability, but we simply determine folding and unfolding rate constants with different amounts of glycerol in the absence of denaturant. To separate the dual effects of glycerol on the kinetic data, we take advantage of the fact that folding and unfolding rate constants can be used to calculate $\Delta G_{F\rightarrow U}$ at different viscogen concentrations.

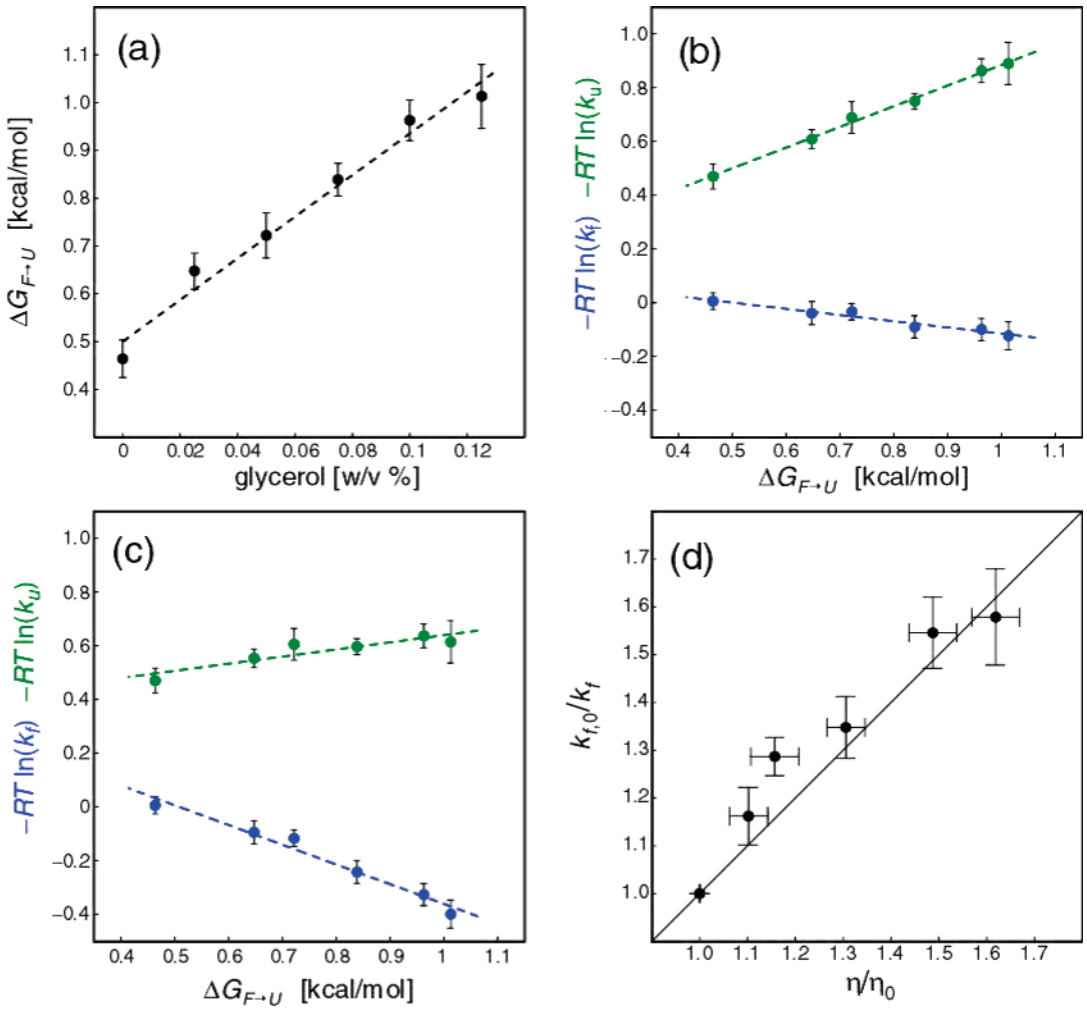

FIGURE 5: (a) Plot of $\Delta G_{F\rightarrow U} = -RT\ln(k_u/k_f)$ vs glycerol concentration. The dashed line represents a linear least-squares fit to the experimental data. (b) Kinetic folding and unfolding data of the drkN SH3 domain for various glycerol concentrations vs $\Delta G_{F\rightarrow U}$. Experimentally determined values of $-RT\ln k_f$ (blue circles) and $-RT\ln k_u$ (green circles) are shown along with results obtained from linear least-squares fits of the experimental data (dashed blue and green lines). (c) Plot of $-RT\ln k_f$ and $-RT\ln k_u$ vs $\Delta G_{F\rightarrow U}$ after correction for the effect of viscosity on k_f and k_u using eq 5. The viscosity-corrected glycerol-related α value derived from these data is 0.73 ± 0.05. (d) Dependence of the inverse of the folding rate constant, $1/k_f$, on solvent viscosity, η, where each axis in the plot is normalized to values in the absence of glycerol ($k_{f,0}$ and η_0, respectively). Experimental values of k_f were corrected for the effect of glycerol on $\Delta G_{F\rightarrow U}$ as described in the text. A line with a slope of 1 is drawn to indicate the $k_{f,0}/k_f$ vs η/η_0 dependence that is expected for a viscosity-dependent folding−unfolding transition.

Upon addition of viscogen, the folding−unfolding equilibrium for the drkN SH3 domain is shifted toward the folded state and $\Delta G_{F\rightarrow U}$ is linearly dependent on glycerol concentration (Figure 5a), consistent with other proteins (*83*−*85*). Plotting the kinetic data (Figure 5b) as $-RT\ln k_f$ and $-RT\ln k_u$ versus $\Delta G_{F\rightarrow U}$, according to linear rate−equilibrium free energy relationships (eq 2), leads to α values for folding and unfolding given by the slopes of 0.23 and 0.77, respectively, but which do not take into account any potential effect of viscosity. (Note that α values for folding are most commonly reported in the literature, and $\alpha_{folding} + \alpha_{unfolding} = 1$.) The fact that the α value for folding is 0.23, significantly different from the α_{urea} value of 0.76, is indicative of an additional contribution to the folding kinetics apart from the stabilizing effect of viscogen.

Assuming that folding and unfolding of the drkN SH3 domain are indeed viscosity-dependent, we can account for the effect of viscosity using eq 5 and correct the experimental values of k_f and k_u by multiplication with η/η_0 at each viscosity. This leads to a glycerol-related (and viscosity-corrected) α value from the linear rate−equilibrium free energy relationship (α_{glyc}) of 0.73 ± 0.06 (Figure 5c), which is equivalent to α_{urea} within experimental uncertainty, consistent with the effects of both glycerol and urea in modulating protein stability by interactions with the protein surface. We can thus use the α_{glyc} value of 0.73 to correct experimental rate constants for the glycerol-induced stability change by multiplying k_f values with $\exp[-\alpha_{glyc} \times \Delta\Delta G_{F\rightarrow U}/RT]$ at each concentration of glycerol (where $\Delta\Delta G_{F\rightarrow U}$ is the change in folding free energy caused by the addition of that concentration of viscogen) and obtain a near $1/\eta$ dependence of folding rate constants on solvent viscosity as shown in Figure 5d. Such a correlation is consistent with a viscosity dependence of folding. Our approach is conceptually similar to the method proposed by Plaxco and Baker (*85*), who assumed an explicit model for the effect of viscogen concentration on the transition state energy to show the correspondence of α values and viscosity dependence of refolding.

It should be noted that underlying these approaches is the assumption that the addition of viscogen does not alter the folding pathway. The linearity of plots of $-RT\ln k_f$ and

$-RT \ln k_u$ versus $\Delta G_{F\rightarrow U}$ (Figure 5c) is in agreement with the assumption (but not conclusive proof) that alternative folding pathways are not significantly populated upon addition of viscogen.

Structural Models of SH3 Domain Transition States. The measures presented here complement ϕ_f value analysis obtained by the protein engineering method. Comprehensive experimental studies by ϕ_f value analysis have provided a fairly detailed structural characterization of the folding TSEs of various SH3 domains, revealing a markedly polarized folding transition state. A number of residues with ϕ_f values close to 1 have been identified that correspond to a folding nucleus, and a large number of residues displaying intermediate ϕ_f values are participating in the folding nucleus to a lesser extent (*22*−*27*). These data suggest that in the folding TSE of SH3 domains part of the final folded structure is preformed, whereas other parts are considerably less structured, corresponding to a nucleation−condensation mechanism of protein folding (*30*). The most structured regions include the well-ordered three-stranded central β-sheet and a particularly well-structured β-turn region that is commonly called the distal loop. Experimental studies on the α-spectrin, src, and drkN SH3 domain folding kinetics employing multiple amino acid substitutions indicate that intermediate ϕ_f values arise from fractional contact formation in a single dominant folding pathway and that parallel folding pathways involving distinct transition states are not significantly populated (*24*, *26*, *39*, *73*).

These experimental studies of SH3 domain folding have been complemented by computational studies employing a variety of different techniques, including high-temperature molecular dynamics (MD) unfolding trajectories (*86*), MD simulations employing Gô models and subsequent identification of the folding TSE (*87*), native-centric models (*88*), and MD folding and unfolding studies in explicit solvent (*89*, *90*). Generally, these studies have been able to reproduce the experimental data and have provided valuable insight into the (un)folding process, including the order of events along the (un)folding pathway.

More recently, experimental ϕ_f values have been used as input restraints for the calculation of ensembles of structures that represent the transition state of the src, fyn, and α-spectrin SH3 domains (*11*, *14*, *20*). In this approach, MD or Monte Carlo (MC) unfolding simulations are performed using the structure of the folded state as a starting point where, for residues whose experimental ϕ_f values are available, the fraction of native contacts (assuming no contacts in the unfolded state) is restrained to remain close to the experimental ϕ_f values. A large number of conformations are generated, and representative structures are chosen by clustering techniques, which involve the selection of structures with small mean deviations between experimental and calculated ϕ_f values as putative members of the TSE (*10*, *11*).

Because the determination of TSE structures depends critically on the reliability and quality of experimental input restraints, calculations can be significantly enhanced by incorporation of experimental data in addition to ϕ_f values. Since denaturant-related α values are related to the solvent-accessible surface area of the TSE, such data can complement information that is derived from ϕ_f values and serve to correctly define the compactness of the ensemble of structures that represent the folding TSE. In TSE structure calculations that have been published to date, denaturant-related α values were employed in only a few cases to tune the compactness of the structures (*8*, *14*). To translate denaturant-related α values into an absolute estimate of the solvent-accessible surface area of the folding transition state, the solvent-accessible surface areas of both folded and unfolded states have to be known. Because structures of the unfolded states are not generally known, the solvent-accessible surface areas of the unfolded states of these proteins have been estimated by modeling the unfolded state as random-coil ensembles (*8*) or based on models that yield upper and lower boundaries for the solvent-accessible surface area of the unfolded state derived from databases (*57*) of extended and compact conformers, respectively, and the solvent-accessible surface area of the unfolded state was assumed to lie between those two values (*14*). It is well established in the literature that a number of proteins that are unfolded under nondenaturing conditions possess transiently populated secondary structure and tertiary contacts, whereas under denaturing conditions, these proteins can (cooperatively) lose such residual structure and approach random-coil-like behavior (*33*, *37*, *65*, *66*, *68*, *91*). Generally, unfolded states in the absence of denaturants are more compact (and more structured) than extended or random-coil-like polypeptides. Comparing the solvent-accessible surface area of the unfolded state of the drkN SH3 domain that we derive from the its representative structure ensemble determined previously (*21*) with boundary values that are obtained from the commonly used model database (*57*) reveals that the experimental solvent-accessible surface area (6650 ± 260 Å^2) is practically identical to the lower boundary (6653 Å^2) and significantly smaller than the upper boundary of 8629 Å^2.

In computational studies aimed at the determination of TSE structures, the choice of parameters that determine the compactness is crucial for obtaining a reliable set of structures that correctly represents the transition state ensemble, just as for the calculation of conformers representing the unfolded state of the drkN SH3 domain. In our experience with computational studies on unfolded states, the size distribution and the level of compactness of the calculated conformers depend critically on experimental restraints that report on the molecular size of the unfolded state ensemble (refs *36* and *49*, as well as unpublished results of C. Neale, W. Y. Choy, and J. D. Forman-Kay). The estimation of the size of the solvent-accessible surface area of the folding TSE of the drkN SH3 domain that we present here is based on an extensive set of experimental data on both folded and unfolded SH3 domains that have been collected over the past several years and does not rely on empirical models for the unfolded state structure. Given the fact that the high compactness of the unfolded drkN SH3 domain appears not to be a unique feature of this particular SH3 domain [the unfolded state of the α-spectrin SH3 domain is equally compact, as was shown by pulsed-field gradient NMR diffusion experiments (F. A. A. Mulder et al., unpublished results)], we suggest that the lower bound for the solvent-accessible surface area of the unfolded state should be used as the input restraint in TSE structure calculations of SH3 domains.

Role of Water in SH3 Domain Folding. Our kinetic data on the drkN SH3 domain strongly suggest that the rate-limiting step(s) for the protein folding−unfolding process

is dependent on solvent water viscosity, i.e., that solvent water is being displaced relative to the polypeptide chain and/or side chains as folding passes through the TSE. A viscosity dependence of protein folding and unfolding implies that the rate-limiting steps that are encompassing the formation of the transition state are indeed opposed predominantly by solvent-derived rather than internal frictional forces. On a microscopic level, the precise nature of diffusional processes that can lead to a solvent viscosity dependence of protein folding is poorly understood, and it is not known how extensive motions have to be to produce a solvent viscosity dependence of folding. Note that complementary information supportive of partial hydration of the TSE of the drkN SH3 domain has been very recently obtained from pressure-dependent folding studies (*42*).

CONCLUSIONS

To obtain physically meaningful estimates of the solvent-accessible surface area of folding transition states, it is vitally important to obtain estimates of the solvent-accessible surface area of unfolded proteins under nondenaturing conditions, and it is favorable to study protein folding kinetics without denaturant or with as little denaturant present as possible. The information about the solvent-accessible surface area of the folding TSE of the drkN SH3 domain that is presented here is derived from experimental data and provides important insights into the transition state of the protein. In addition, it constitutes a reliable input restraint for future TSE structure calculations and other computational studies of the folding pathway(s) of SH3 domains. Our kinetic data on the drkN SH3 domain folding are consistent with a folding mechanism that involves formation of a (partially) hydrated folding transition state. Given the fact that the folding pathways of SH3 domains appear to be conserved within the SH3 domain family and computational studies suggest a common folding mechanism for SH3 domains, these results are likely a common feature of SH3 domains.

ACKNOWLEDGMENT

We thank Dr. H. S. Chan for insightful discussions and reading the manuscript.

REFERENCES

1. Vendruscolo, M., and Paci, E. (2003) Protein folding: Bringing theory and experiment closer together, *Curr. Opin. Struct. Biol. 13*, 82−87.
2. Vendruscolo, M., and Dobson, C. M. (2005) Towards complete descriptions of the free-energy landscapes of proteins, *Philos. Trans. R. Soc. London, Ser. A 363*, 433−450.
3. Lindorff-Larsen, K., Rogen, P., Paci, E., Vendruscolo, M., and Dobson, C. M. (2005) Protein folding and the organization of the protein topology universe, *Trends Biochem. Sci. 30*, 13−19.
4. Fersht, A. (1998) *Structure and Mechanism in Protein Science*, W. H. Freeman and Company, New York.
5. Matouschek, A., Kellis, J. T., Jr., Serrano, L., and Fersht, A. R. (1989) Mapping the transition state and pathway of protein folding by protein engineering, *Nature 340*, 122−126.
6. Sanchez, I. E., and Kiefhaber, T. (2003) Non-linear rate-equilibrium free energy relationships and Hammond behavior in protein folding, *Biophys. Chem. 100*, 397−407.
7. Fersht, A. R. (2004) Relationship of Leffler (Bronsted) α-values and protein folding ϕ-values to position of transition-state structures on reaction coordinates, *Proc. Natl. Acad. Sci. U.S.A. 101*, 14338−14342.
8. Vendruscolo, M., Paci, E., Dobson, C. M., and Karplus, M. (2001) Three key residues form a critical contact network in a protein folding transition state, *Nature 409*, 641−645.
9. Li, L., and Shakhnovich, E. I. (2001) Constructing, verifying, and dissecting the folding transition state of chymotrypsin inhibitor 2 with all-atom simulations, *Proc. Natl. Acad. Sci. U.S.A. 98*, 13014−13018.
10. Paci, E., Vendruscolo, M., Dobson, C. M., and Karplus, M. (2002) Determination of a transition state at atomic resolution from protein engineering data, *J. Mol. Biol. 324*, 151−163.
11. Gsponer, J., and Caflisch, A. (2002) Molecular dynamics simulations of protein folding from the transition state, *Proc. Natl. Acad. Sci. U.S.A. 99*, 6719−6724.
12. Paci, E., Clarke, J., Steward, A., Vendruscolo, M., and Karplus, M. (2003) Self-consistent determination of the transition state for protein folding: Application to a fibronectin type III domain, *Proc. Natl. Acad. Sci. U.S.A. 100*, 394−399.
13. Paci, E., Friel, C. T., Lindorff-Larsen, K., Radford, S. E., Karplus, M., and Vendruscolo, M. (2004) Comparison of the transition state ensembles for folding of Im7 and Im9 determined using all-atom molecular dynamics simulations with ϕ value restraints, *Proteins 54*, 513−525.
14. Lindorff-Larsen, K., Vendruscolo, M., Paci, E., and Dobson, C. M. (2004) Transition states for protein folding have native topologies despite high structural variability, *Nat. Struct. Mol. Biol. 11*, 443−449.
15. Geierhaas, C. D., Paci, E., Vendruscolo, M., and Clarke, J. (2004) Comparison of the transition states for folding of two Ig-like proteins from different superfamilies, *J. Mol. Biol. 343*, 1111−1123.
16. Hubner, I. A., Shimada, J., and Shakhnovich, E. I. (2004) Commitment and nucleation in the protein G transition state, *J. Mol. Biol. 336*, 745−761.
17. Settanni, G., Gsponer, J., and Caflisch, A. (2004) Formation of the folding nucleus of an SH3 domain investigated by loosely coupled molecular dynamics simulations, *Biophys. J. 86*, 1691−1701.
18. Hubner, I. A., Oliveberg, M., and Shakhnovich, E. I. (2004) Simulation, experiment, and evolution: Understanding nucleation in protein S6 folding, *Proc. Natl. Acad. Sci. U.S.A. 101*, 8354−8359.
19. Settanni, G., Rao, F., and Caflisch, A. (2005) ϕ-value analysis by molecular dynamics simulations of reversible folding, *Proc. Natl. Acad. Sci. U.S.A. 102*, 628−633.
20. Hubner, I. A., Edmonds, K. A., and Shakhnovich, E. I. (2005) Nucleation and the Transition State of the SH3 Domain, *J. Mol. Biol. 349*, 424−434.
21. Bezsonova, I., Singer, A., Choy, W. Y., Tollinger, M., and Forman-Kay, J. D. (2005) Structural Comparison of the Unstable drkN SH3 Domain and a Stable Mutant, *Biochemistry 44*, 15550−15560.
22. Grantcharova, V. P., Riddle, D. S., Santiago, J. V., and Baker, D. (1998) Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain, *Nat. Struct. Biol. 5*, 714−720.
23. Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I., and Baker, D. (1999) Experiment and theory highlight role of native state topology in SH3 folding, *Nat. Struct. Biol. 6*, 1016−1024.
24. Martinez, J. C., and Serrano, L. (1999) The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved, *Nat. Struct. Biol. 6*, 1010−1016.
25. Grantcharova, V. P., Riddle, D. S., and Baker, D. (2000) Long-range order in the src SH3 folding transition state, *Proc. Natl. Acad. Sci. U.S.A. 97*, 7084−7089.
26. Northey, J. G., Maxwell, K. L., and Davidson, A. R. (2002) Protein folding kinetics beyond the ϕ value: Using multiple amino acid substitutions to investigate the structure of the SH3 domain folding transition state, *J. Mol. Biol. 320*, 389−402.
27. Northey, J. G., Di Nardo, A. A., and Davidson, A. R. (2002) Hydrophobic core packing in the SH3 domain folding transition state, *Nat. Struct. Biol. 9*, 126−130.
28. Cobos, E. S., Filimonov, V. V., Vega, M. C., Mateo, P. L., Serrano, L., and Martinez, J. C. (2003) A thermodynamic and kinetic analysis of the folding pathway of an SH3 domain entropically stabilised by a redesigned hydrophobic core, *J. Mol. Biol. 329*, 185−187.
29. Fernandez-Escamilla, A. M., Cheung, M. S., Vega, M. C., Wilmanns, M., Onuchic, J. N., and Serrano, L. (2004) Solvation

in protein folding analysis: Combination of theoretical and experimental approaches, *Proc. Natl. Acad. Sci. U.S.A. 101*, 2834−2839.
30. Fersht, A. R. (1995) Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications, *Proc. Natl. Acad. Sci. U.S.A. 92*, 10869−10873.
31. Cheung, M. S., Garcia, A. E., and Onuchic, J. N. (2002) Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse, *Proc. Natl. Acad. Sci. U.S.A. 99*, 685−690.
32. Shea, J. E., Onuchic, J. N., and Brooks, C. L., III (2002) Probing the folding free energy landscape of the Src-SH3 protein domain, *Proc. Natl. Acad. Sci. U.S.A. 99*, 16064−16068.
33. Zhang, O., and Forman-Kay, J. D. (1997) NMR studies of unfolded states of an SH3 domain in aqueous solution and denaturing conditions, *Biochemistry 36*, 3959−3970.
34. Yang, D., Mok, Y. K., Muhandiram, D. R., Forman-Kay, J. D., and Kay, L. E. (1999) ^{1}H-^{13}C dipole−dipole cross-correlated spin relaxation as a probe of dynamics in unfolded proteins: Application to the drkN SH3 domain, *J. Am. Chem. Soc. 121*, 3555−3556.
35. Mok, Y. K., Kay, C. M., Kay, L. E., and Forman-Kay, J. D. (1999) NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions, *J. Mol. Biol. 289*, 619−638.
36. Choy, W. Y., Mulder, F. A., Crowhurst, K. A., Muhandiram, D. R., Millett, I. S., Doniach, S., Forman-Kay, J. D., and Kay, L. E. (2002) Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques, *J. Mol. Biol. 316*, 101−112.
37. Crowhurst, K. A., Tollinger, M., and Forman-Kay, J. D. (2002) Cooperative interactions and a non-native buried Trp in the unfolded state of an SH3 domain, *J. Mol. Biol. 322*, 163−178.
38. Crowhurst, K. A., and Forman-Kay, J. D. (2003) Aromatic and methyl NOEs highlight hydrophobic clustering in the unfolded state of an SH3 domain, *Biochemistry 42*, 8687−8695.
39. Mok, Y. K., Elisseeva, E. L., Davidson, A. R., and Forman-Kay, J. D. (2001) Dramatic stabilization of an SH3 domain by a single substitution: Roles of the folded and unfolded states, *J. Mol. Biol. 307*, 913−928.
40. Larson, S. M., and Davidson, A. R. (2000) The identification of conserved interactions within the SH3 domain by alignment of sequences and structures, *Protein Sci. 9*, 2170−2180.
41. Tollinger, M., Kay, L. E., and Forman-Kay, J. D. (2005) Measuring pK_a values in protein folding transition state ensembles by NMR spectroscopy, *J. Am. Chem. Soc. 127*, 8904−8905.
42. Bezsonova, I., Korzhnev, D. M., Prosser, R. S., Forman-Kay, J. D., and Kay, L. E. (2006) Hydration and packing along the folding pathway of SH3 domains by pressure dependent NMR, *Biochemistry 45*, 4711−4719.
43. Farrow, N. A., Zhang, O., Forman-Kay, J. D., and Kay, L. E. (1994) A heteronuclear correlation experiment for simultaneous determination of ^{15}N longitudinal decay and chemical exchange rates of systems in slow equilibrium, *J. Biomol. NMR 4*, 727−734.
44. Tollinger, M., Skrynnikov, N. R., Mulder, F. A. A., Forman-Kay, J. D., and Kay, L. E. (2001) Slow dynamics in folded and unfolded states of an SH3 domain, *J. Am. Chem. Soc. 123*, 11341−11352.
45. Zeeb, M., Jacob, M. H., Schindler, T., and Balbach, J. (2003) ^{15}N relaxation study of the cold shock protein CspB at various solvent viscosities, *J. Biomol. NMR 27*, 221−234.
46. Perl, D., Jacob, M., Bano, M., Stupak, M., Antalik, M., and Schmid, F. X. (2002) Thermodynamics of a diffusional protein folding reaction, *Biophys. Chem. 96*, 173−190.
47. Press, W. H. (1997) *Numerical Recipes in C: The Art of Scientific Computing*, University Press, Cambridge.
48. Crowhurst, K. A., Choy, W. Y., Mok, Y. K., and Forman-Kay, J. D. (2003) Corrigendum to the paper by Mok et al. (1999) NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions, *J. Mol. Biol. 329*, 185−187.
49. Choy, W. Y., and Forman-Kay, J. D. (2001) Calculation of ensembles of structures representing the unfolded state of an SH3 domain, *J. Mol. Biol. 308*, 1011−1032.
50. Neale, C., Marsh, J. A., Jack, F., Choy, W.-Y., Lee, A., Crowhurst, K., and Forman-Kay, J. D. (2006) manuscript in preparation.
51. Feldman, H. J., and Hogue, C. W. (2000) A fast method to sample real protein conformational space, *Proteins 39*, 112−131.
52. Garnier, J., Gibrat, J. F., and Robson, B. (1996) GOR method for predicting protein secondary structure from amino acid sequence, *Methods Enzymol. 266*, 540−553.
53. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography and NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D54* (Part 5), 905−921.
54. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. R., Cheatham, T. E., III, DeBolt, S., Ferguson, D., Seibel, G., and Kollman, P. (1995) AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules, *Comput. Phys. Commun. 91*, 1−41.
55. Simons, K. T., Kooperberg, C., Huang, E., and Baker, D. (1997) Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions, *J. Mol. Biol. 268*, 209−225.
56. Lambert, C., Leonard, N., De Bolle, X., and Depiereux, E. (2002) ESyPred3D: Prediction of proteins 3D structures, *Bioinformatics 18*, 1250−1256.
57. Creamer, T. P., Srinivasan, R., and Rose, G. D. (1997) Modeling unfolded states of proteins and peptides. II. Backbone solvent accessibility, *Biochemistry 36*, 2832−2835.
58. Leffler, J. E. (1953) Parameters for the description of transition states, *Science 117*, 340−341.
59. Fersht, A. R., Matouschek, A., and Serrano, L. (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding, *J. Mol. Biol. 224*, 771−782.
60. Fersht, A. R., Itzhaki, L. S., elMasry, N. F., Matthews, J. M., and Otzen, D. E. (1994) Single versus parallel pathways of protein folding and fractional formation of structure in the transition state, *Proc. Natl. Acad. Sci. U.S.A. 91*, 10426−10429.
61. Wright, C. F., Lindorff-Larsen, K., Randles, L. G., and Clarke, J. (2003) Parallel protein-unfolding pathways revealed and mapped, *Nat. Struct. Biol. 10*, 658−662.
62. Myers, J. K., and Oas, T. G. (2002) Mechanism of fast protein folding, *Annu. Rev. Biochem. 71*, 783−815.
63. Zarrine-Afsar, A., and Davidson, A. R. (2004) The analysis of protein folding kinetic data produced in protein engineering experiments, *Methods 34*, 41−50.
64. Palmer, A. G., III, Kroenke, C. D., and Loria, J. P. (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules, *Methods Enzymol. 339*, 204−238.
65. Lindorff-Larsen, K., Kristjansdottir, S., Teilum, K., Fieber, W., Dobson, C. M., Poulsen, F. M., and Vendruscolo, M. (2004) Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme a binding protein, *J. Am. Chem. Soc. 126*, 3291−3299.
66. Blanco, F. J., Serrano, L., and Forman-Kay, J. D. (1998) High populations of non-native structures in the denatured state are compatible with the formation of the native folded state, *J. Mol. Biol. 284*, 1153−1164.
67. Huang, G. S., and Oas, T. G. (1995) Submillisecond folding of monomeric lambda repressor, *Proc. Natl. Acad. Sci. U.S.A. 92*, 6878−6882.
68. Mayor, U., Grossmann, J. G., Foster, N. W., Freund, S. M., and Fersht, A. R. (2003) The denatured state of engrailed homeodomain under denaturing and native conditions, *J. Mol. Biol. 333*, 977−991.
69. van Gunsteren, W. F., Burgi, R., Peter, C., and Daura, X. (2001) The key to solving the protein-folding problem lies in an accurate description of the denatured state, *Angew. Chem., Int. Ed. 40*, 351−355.
70. Daggett, V., and Fersht, A. (2003) The present view of the mechanism of protein folding, *Nat. Rev. Mol. Cell Biol. 4*, 497−502.
71. Dinner, A. R., and Karplus, M. (1999) Is protein unfolding the reverse of protein folding? A lattice simulation analysis, *J. Mol. Biol. 292*, 403−419.
72. Shea, J. E., and Brooks, C. L., III (2001) From folding theories to folding proteins: A review and assessment of simulation studies of protein folding and unfolding, *Annu. Rev. Phys. Chem. 52*, 499−535.
73. Martinez, J. C., Pisabarro, M. T., and Serrano, L. (1998) Obligatory steps in protein folding and the conformational diversity of the transition state, *Nat. Struct. Biol. 5*, 721−729.

74. Pace, C. N., and Shaw, K. L. (2000) Linear extrapolation method of analyzing solvent denaturation curves, *Proteins Suppl. 4*, 1−7.
75. Burton, R. E., Huang, G. S., Daugherty, M. A., Fullbright, P. W., and Oas, T. G. (1996) Microsecond protein folding through a compact transition state, *J. Mol. Biol. 263*, 311−322.
76. Burton, R. E., Busby, R. S., and Oas, T. G. (1998) ALASKA: A mathematica package for two-state analysis of protein folding reactions, *J. Biomol. NMR 11*, 355−360.
77. Jacob, M., and Schmid, F. X. (1999) Protein folding as a diffusional process, *Biochemistry 38*, 13773−13779.
78. Maxwell, K. L., Wildes, D., Zarrine-Afsar, A., De Los Rios, M. A., Brown, A. G., Friel, C. T., Hedberg, L., Horng, J. C., Bona, D., Miller, E. J., Vallee-Belisle, A., Main, E. R., Bemporad, F., Qiu, L., Teilum, K., Vu, N. D., Edwards, A. M., Ruczinski, I., Poulsen, F. M., Kragelund, B. B., Michnick, S. W., Chiti, F., Bai, Y., Hagen, S. J., Serrano, L., Oliveberg, M., Raleigh, D. P., Wittung-Stafshede, P., Radford, S. E., Jackson, S. E., Sosnick, T. R., Marqusee, S., Davidson, A. R., and Plaxco, K. W. (2005) Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins, *Protein Sci. 14*, 602−616.
79. Kramers, H. A. (1940) Brownian motion in a field of force and the diffusion model of chemical reactions, *Physica 7*, 284−304.
80. Timasheff, S. N. (1993) The control of protein stability and association by weak interactions with water: How do solvents affect these processes? *Annu. Rev. Biophys. Biomol. Struct. 22*, 67−97.
81. Gekko, K., and Timasheff, S. N. (1981) Thermodynamic and kinetic examination of protein stabilization by glycerol, *Biochemistry 20*, 4677−4786.
82. Jacob, M., Schindler, T., Balbach, J., and Schmid, F. X. (1997) Diffusion control in an elementary protein folding reaction, *Proc. Natl. Acad. Sci. U.S.A. 94*, 5622−5627.
83. Xie, G., and Timasheff, S. N. (1997) Mechanism of the stabilization of ribonuclease A by sorbitol: Preferential hydration is greater for the denatured then for the native protein, *Protein Sci. 6*, 211−221.
84. Xie, G., and Timasheff, S. N. (1997) The thermodynamic mechanism of protein stabilization by trehalose, *Biophys. Chem. 64*, 25−43.
85. Plaxco, K. W., and Baker, D. (1998) Limited internal friction in the rate-limiting step of a two-state protein folding reaction, *Proc. Natl. Acad. Sci. U.S.A. 95*, 13591−13596.
86. Tsai, J., Levitt, M., and Baker, D. (1999) Hierarchy of structure loss in MD simulations of src SH3 domain unfolding, *J. Mol. Biol. 291*, 215−225.
87. Ding, F., Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E., and Shakhnovich, E. I. (2002) Direct molecular dynamics observation of protein folding transition state ensemble, *Biophys. J. 83*, 3525−3532.
88. Ollerenshaw, J. E., Kaya, H., Chan, H. S., and Kay, L. E. (2004) Sparsely populated folding intermediates of the Fyn SH3 domain: Matching native-centric essential dynamics and experiment, *Proc. Natl. Acad. Sci. U.S.A. 101*, 14748−14753.
89. Guo, W., Lampoudi, S., and Shea, J. E. (2003) Posttransition state desolvation of the hydrophobic core of the src-SH3 protein domain, *Biophys. J. 85*, 61−69.
90. Guo, W., Lampoudi, S., and Shea, J. E. (2004) Temperature dependence of the free energy landscape of the src-SH3 protein domain, *Proteins 55*, 395−406.
91. Kristjansdottir, S., Lindorff-Larsen, K., Fieber, W., Dobson, C. M., Vendruscolo, M., and Poulsen, F. M. (2005) Formation of native and non-native interactions in ensembles of denatured ACBP molecules from paramagnetic relaxation enhancement studies, *J. Mol. Biol. 347*, 1053−1062.

BI060268O