Biochemistry

 Volume 38, Number 42 October 19, 1999

Perspectives in Biochemistry

Protein Folding as a Diffusional Process

Maik Jacob and Franz X. Schmid*

Biochemisches Laboratorium, Universität Bayreuth, D-95440 Bayreuth, Germany

Received June 29, 1999; Revised Manuscript Received August 25, 1999

ABSTRACT: A protein chain must move relative to the solvent molecules and explore many conformations when it folds from the extended unfolded state to the compact native state. Experimental and theoretical approaches suggest that diffusional processes in fact contribute to the kinetics of protein folding. We describe here how variations of the solvent viscosity can be employed to uncover the diffusional contributions to a folding reaction and assess the use of transition state theory and Kramers' rate theory for the analysis of protein folding reactions.

Experiments by Pohl in the late 1960s (*1*) suggested that proteins can fold to the native state in apparently simple one-step reactions in the time range of only a few minutes. Then, however, Levinthal pointed out that even small proteins would need an astronomic time for folding, if they had to search randomly through all possible conformations to find the native conformation (*2*). Although overly simplistic, his calculation deeply influenced the field and motivated the search for partially folded intermediates that could direct folding rapidly to one route or a few productive routes. By using fast reaction techniques, the Tanford (*3*−*5*) and Baldwin groups (*6*, *7*) discovered that the refolding reactions of several small proteins were in fact multiphasic processes. Soon it became clear, however, that this complexity was in large part due to the parallel fast and slow refolding of protein molecules with different prolyl isomers (*8*−*10*). Later, multimixing kinetic studies (*11*, *12*) and, in particular, the combination of amide NH protection and NMR spectroscopy (*13*, *14*) showed that partially folded intermediates indeed exist. Some of these studies benefited from the presence of incorrect prolyl isomers, which retard the final step of folding and thus increase the time window for characterizing the transient folding intermediates (*14*, *15*).

To understand a protein folding reaction, information is needed not only about the intermediates but also about the activated states that control the rates of individual steps. The physical properties of activated states cannot easily be obtained for proteins that rapidly form intermediates before the rate-limiting step (as in $U^{1} \rightleftharpoons I \rightarrow N$ reactions). It is very difficult to discriminate between the changes that occur between the unfolded state U and the intermediate I and those between I and the activated state X that controls the reaction from I to N. This has been clearly illustrated by alternative analyses of the kinetics of refolding of ubiquitin (*16*).

In the past years, several small proteins have been found to fold very rapidly in single $N \rightleftharpoons U$ two-state reactions (*17*−*24*). These proteins provide excellent systems with which to investigate protein folding at a very elementary level and,

* To whom correspondence should be addressed: Universität Bayreuth, D-95440 Bayreuth, Germany. Telephone: ++49 921 55 36 60. Fax: ++49 921 55 36 61. E-mail: FX.Schmid@uni-bayreuth.de.

[1] Abbreviations: N and U, native and unfolded forms of a protein, respectively; I, folding intermediate; X, activated state of a chemical reaction; GdmCl, guanidinium chloride; TST, transition state theory; E_A, experimental (Arrhenius) activation enthalpy; ΔU, activation energy in Kramers' relation.

10.1021/bi991503o CCC: $18.00
Published on Web 09/23/1999

in particular, to characterize the activated states of folding in the absence of the complications that originate from the presence of intermediates. In particular, protein engineering has been used with great success to probe such folding reactions (*24*, *25*).

RATE THEORIES FOR FOLDING REACTIONS

Which formalism or theory should be used to analyze the time courses of protein folding reactions? Many single-domain proteins unfold and refold in monoexponential processes. Therefore, the simplest approach is to model folding as a chemical reaction in which the substrate molecules must cross a potential energy barrier to be transformed to the product. Virtually all treatments of chemical kinetics have in common the fact that the reaction rate depends on the height of this energy barrier E_A and on a probability factor that accounts for the number of states that are accessible for the molecule in the ground and activated states. Thermal motion is required to cross an energy barrier; i.e., the reaction rate k is proportional to a Boltzmann factor of the type $\exp(-E_A/RT)$, as in the Arrhenius relation (eq 1)

$$k = A \exp(-E_A/RT) \quad (1)$$

where A is simply a proportionality constant. In transition state theory (TST), it is assumed that the state with the highest energy (the activated state) represents a quasi-thermodynamic state (the transition state) that is in an equilibrium with the reactant state (eq 2).

$$k = (k_B T/h) \exp(-\Delta G^{\ddagger}/RT) \quad (2)$$

$$k = (k_B T/h) \exp(\Delta S^{\ddagger}/R) \exp(-\Delta H^{\ddagger}/RT) \quad (3)$$

In eq 2, k_B and h are Boltzmann's and Planck's constants, respectively, and $\Delta G^{\ddagger}$ is the Gibbs free energy of activation. TST is often used because the reaction rate in the absence of a barrier can be calculated from the preexponential factor and because all thermodynamic activation parameters (including the activation entropy $\Delta S^{\ddagger}$) can be calculated from the measured rate and its temperature dependence (eqs 2 and 3). TST was developed for reactions of small molecules in the gas phase, which move along a simple linear reaction coordinate from the substrate to the product, and it has been questioned whether it is well-suited to describing reactions that occur in solution and involve the simultaneous breakage and formation of many weak intra- and intermolecular interactions, as in protein folding. Absolute rates calculated for chemical reactions from TST are unrealistically high, and therefore, a correction factor, the transmission coefficient r ($0 < r \leq 1$), is often included in eq 2. Because of these problems, it is difficult to derive reliable values for $\Delta G^{\ddagger}$ and $\Delta S^{\ddagger}$ from measured rate constants.

Kramers' theory provides an alternative formalism specifically for reactions in solution. In this theory, a chemical reaction is modeled as a diffusive passage over a potential energy barrier. The mathematical formulation of the theory in Kramers' original paper (*26*) is difficult to read, but excellent discussions of his theory are provided in refs *27*−*30*. Landauer and Swanson extended the theory to the multidimensional case (*31*). A reaction theory that accounts for diffusional motion should be well-suited to treating protein reactions in aqueous solvents. In fact, Kramers' theory forms the basis for the diffusion−collision model for protein folding (*32*), and it was used to describe protein−ligand interactions that are coupled with conformational changes of the protein (*33*−*36*).

For a reaction from the substrate S over the activated state X to the product P ($S \rightarrow X \rightarrow P$), Kramers' relation can be written as in eq 4 (*29*)

$$k = \tau_{SX}^{-1} \exp(-\Delta U/RT) \quad (4)$$

where k is the rate constant for the reaction from S to P, ΔU is the height of the potential energy barrier, and τ_{SX} is the time constant of the reaction in the absence of the barrier. τ_{SX} depends on the frequency of the system in the ground state S and the frequency of escape from the activated state X, which reflects the local mobility of the system while being in the activated state (for a review, see ref *29*). In viscous solvents such as water, this mobility should depend on the friction with the solvent, which is usually modeled by the macroscopic viscosity η. Whether a reaction in solution follows Kramers' relation can thus be assessed by experiment. In the simplest case, the time constant τ ($\tau = k^{-1}$) of a reaction that involves diffusional motion over a potential energy barrier should depend linearly on the solvent viscosity η as in eq 5.

$$\tau \sim \eta \quad (5)$$

VISCOSITY DEPENDENCE OF PROTEIN FOLDING REACTIONS

In the 1970s and 1980s, the viscosity dependence of protein folding reactions was investigated (i) to identify domain docking reactions in the folding of large proteins (*37*−*39*) and (ii) to test Karplus and Weaver's diffusion-collision model (*40*). Folding reactions of proteins with several domains, such as the α subunit of tryptophan synthase, aspartokinase homoserine dehydrogenase, and octopine dehydrogenase, were indeed found to depend on solvent viscosity (*37*−*39*, *41*). This suggests that these folding reactions are controlled by the movement of domains. In contrast, the folding rate of ribonuclease A remained unchanged when the solvent viscosity was increased by adding the viscogenic agents sucrose and glycerol. Tsong and Baldwin concluded from this result that segmental diffusion, as assumed in the diffusion−collision model and as measured for peptides by fluorescence energy transfer (*42*), does not contribute to the rate-limiting process in the folding of this protein (*40*). It is possible that in these experiments the viscosity dependence was masked by the stabilizing effect of the viscogen. Recently, Goldberg and Baldwin showed that the unfolding and refolding reactions of pancreatic ribonuclease S depend on viscosity and that the viscosity-dependent step of refolding occurs after the encounter complex between S-peptide and S-protein has formed (*43*, *44*).

The discovery that several small proteins fold extremely fast in the absence of intermediates sparked renewed interest in the role of diffusional processes in the kinetics of protein folding, and in 1997, we found that both refolding and unfolding of the cold shock protein CspB from *Bacillus*

subtilis were decelerated in the presence of sucrose or ethylene glycol (*45*). Equation 5 predicts that for a folding reaction that involves diffusional barrier crossing the relative time constants of both unfolding, $k_{NU}^{-1}/k_{NU,0}^{-1}$, and refolding, $k_{UN}^{-1}/k_{UN,0}^{-1}$, should increase linearly with the relative solvent viscosity η/η_0 (the subscript 0 refers to the measurements in the absence of the viscogenic agent) with a slope of 1. Such simple relationships were not observed in the initial experiments with CspB. Rather, this slope was smaller than unity for refolding and larger than unity for unfolding.

Viscogenic additives, such as ethylene glycol, glycerol, or sucrose, increase not only solvent viscosity but also protein stability (*46*). For a two-state protein folding reaction, an increase in the stability constant ($K_{stab} = [N]/[U] = k_{UN}/k_{NU}$) is necessarily correlated with an increase of the rate constant of refolding k_{UN} relative to the rate constant of unfolding k_{NU}. In refolding, this stabilizing effect thus antagonizes and weakens the viscosity effect; in unfolding, however, it adds to and thus enforces a potential viscosity effect, just as observed for CspB (*45*). We circumvented this problem by measuring the viscosity dependence of the folding kinetics under conditions of identical stability ("isostability" conditions), i.e., by adding a denaturant to neutralize the stabilizing effects of the viscogenic agents. After this correction, unfolding and refolding indeed exhibited identical dependencies on viscosity with slopes close to 1 when the relative viscosity was varied between 1 and 2.5. The isostability approach was originally introduced by Matthews and co-workers when they investigated domain pairing in the folding of the α subunit of tryptophan synthase (*39*). It was also used recently by Plaxco and Baker when they characterized the viscosity dependence of the folding of a fragment of the IgG-binding protein L. They found that, as for CspB, the relative time constant of folding of the L fragment increased linearly with the relative viscosity (*23*). In these studies (*23*, *45*, *47*), as well as in the folding of tryptophan synthase (*39*), the extent of retardation was found to be independent of the chemical nature of the viscogenic agent. Bhattacharyya and Sosnick accounted for the stabilizing effect of the viscogen by comparing the folding kinetics of the monomeric and dimeric forms of a protein (*48*).

The isostability approach was questioned by Ladurner and Fersht (*49*). They argued that the increase in stability by the viscogenic agents should not be counterbalanced by adding guanindinium chloride or urea, because the denaturants and the viscogenic agents might affect protein stability by different mechanisms. They found experimentally that the rate of refolding of chymotrypsin inhibitor 2 (CI2) was increased in the presence of sucrose and concluded from this finding that the refolding of this protein cannot involve diffusional motion. Unfortunately, the unfolding kinetics and the dynamics of equilibration near the transition midpoint were not measured.

The validity of the isostability approach could be confirmed experimentally because conditions exist under which the viscogenic agent ethylene glycol has no effect on protein stability (*47*). Like sucrose or glycerol, ethylene glycol strongly stabilizes proteins at room temperature. This stability, however, strongly decreases with increasing temperature, because these polyols increase the enthalpy and thus the cooperativity of thermal unfolding. This has been shown in a calorimetric study by Gekko (*50*) for hen lysozyme, and it holds true also for CspB. Ethylene glycol strongly stabilizes this protein against denaturant-induced unfolding at 25 °C (*45*), but near 50 °C, the stability of CspB is independent of the ethylene glycol concentration between 0 and 50% (*47*). Therefore, a potential viscosity dependence of the CspB folding kinetics is no longer superimposed by a stabilizing effect of the viscogenic agent near this temperature. Moreover, 50 °C is close to the midpoint of the thermal unfolding transition, and therefore, the kinetics of the N ⇌ U equilibration reaction could be measured as a function of viscosity in the absence of a denaturant by using a very fast pressure-jump method (*51*). The microscopic rate constants of unfolding and refolding, k_{NU} and k_{UN}, could thus be determined under identical solvent conditions and with high precision from the combined equilibrium and kinetic data. It was found that the microscopic rate constants k_{NU} and k_{UN} depended linearly on viscosity with slopes near unity (*47*). These results, obtained at 50 °C under conditions where the viscogenic additive does not change the stability of the folding protein molecules, confirmed that the folding of CspB is very well described by Kramers' model. They also validate the conclusions drawn from the previous more indirect experiments at 25 °C in which the increases in protein stability had to be balanced by adding a denaturant (*23*, *39*, *45*).

The dual effects of the polyols on solvent viscosity and protein stability highlight a general difficulty in the interpretation of experimental results. Ideally, one would like to vary a single *physical property* of the solvent, such as viscosity, to examine the validity of a particular model, such as Kramers' model in this case. Experimentally, however, only the *composition* of the solvent can be varied, and inevitably, this leads to multiple changes in the properties of the solvent. Polyols increase not only viscosity but also protein stability. They also modulate the polarity and surface tension. Both properties can affect the protein folding kinetics. Here experiments with different viscogens such as ethylene glycol and sucrose are helpful because they differ strongly in their effects on surface tension. The effects of changes in protein stability and solvent polarity can be examined in control experiments with nonviscogenic cosolvents that change these properties, such as Hofmeister salts and methanol (*45*). Typically, they affect unfolding and refolding differently (e.g., they accelerate unfolding and decelerate refolding), unlike viscosity, which dampens the dynamic motions in the activated state region and thus decelerates unfolding and refolding to the same extent.

At relative viscosities between 1 and 3, the increases in the folding and unfolding times of several proteins scale linearly with viscosity (*23*, *45*), and the retardation is independent of the chemical nature of the viscogen, provided that the viscogen is small, relative to the protein. These are strong arguments in support of a diffusional, Kramers type model for protein folding. It provides the simplest explanation for the observed retardations in folding. The dependence on viscosity becomes stronger in the presence of high concentrations of ethylene glycol or sucrose (*45*, *47*). This is difficult to interpret, because it is not known whether the measured macroscopic viscosity and the microscopic viscosity at the protein−solvent interface are linearly related over the entire range of viscogen concentrations.

VISCOSITY DEPENDENCE AND EXPERIMENTAL ACTIVATION ENERGIES

The viscosity of water depends on temperature, and it decreases by about 20% when the temperature is increased by 10 °C. Therefore, the temperature dependence of a diffusional reaction originates from two sources: (i) from the true activation energy ΔU and (ii) from the decrease in the solvent viscosity with temperature W (cf. eqs 6 and 7). The apparent activation energy E_A, as derived from an Arrhenius plot, is thus larger than the true activation energy ΔU. For a reaction that obeys Kramers' relation, the rate increases 1.24-fold between 25 and 35 °C because the viscosity of water decreases by the same factor. The "true" activation energy ΔU of both unfolding and refolding is therefore 16.4 kJ/mol smaller than the apparent (Arrhenius) activation energy E_A.

$$E_A = \Delta U + W \tag{6}$$

$$W = R\frac{\partial(\ln \eta)}{\partial(1/T)} \tag{7}$$

The values for the activation heat capacities of unfolding and refolding do not require correction because the viscosity of water depends almost linearly on temperature. An analysis between 10 and 70 °C indicates that the nonlinearity in the temperature dependence of the viscosity of water would add only 0.1 kJ K^{-1} mol^{-1} to the activation heat capacities of unfolding and refolding. This value is smaller than the confidence limits of the experimental activation heat capacities.

Denaturant-dependent folding kinetics must also be corrected for viscosity effects, because the solvent viscosity increases with the concentration of denaturants, such as urea and guanidinium chloride (GdmCl), in a nonlinear fashion (*52*). The viscosity correction remains small at ≤2 M denaturant, but becomes more pronounced at high denaturant concentrations (Figure 1). Therefore, unfolding, which is assessed at high denaturant concentrations, is affected more strongly than refolding, which is assessed at low denaturant concentrations, and the kinetic m value of unfolding increases slightly relative to the kinetic m value of refolding. Figure 1 shows the urea- and GdmCl-dependent unfolding and refolding kinetics of CspB before and after the correction for the viscosity effect. This correction leads to a decrease in the fractional m value of refolding from 0.90 to 0.87 (GdmCl) and from 0.86 to 0.81 (urea). In other words, the activated state becomes slightly "less native" after this correction, and thus, the value more closely resembles the value that is derived from the fractional changes in the activation heat capacities. This confirms a suggestion made earlier by Plaxco and Baker (*23*). The slight nonlinearity in the unfolding data between 3 and 7 M GdmCl in Figure 1B is much less apparent after the correction for the viscosity effect. It is possible that some of the observed nonlinearities in kinetic chevron plots at high GdmCl concentrations originate from the increase in solvent viscosity.

It seems hard to reconcile the picture of a dynamic, viscosity-dependent folding process with the experimental finding that the activated state of folding of CspB is unusually native-like, as judged by its heat capacity and by the m value analysis (*18*, *53*, *54*). Both these criteria suggest that the activated molecules resemble the native molecules in their interactions with the aqueous solvent. It is important to note that the m values and the heat capacities of activation are pseudoequilibrium properties, which yield a time- and ensemble-averaged picture of the activated state. The viscosity dependence gives a different type of information. It reports on the dynamics that are necessary to explore and proceed through the particular part of the energy landscape that represents the region of the activated state. It is not clear how large such motions must be to become sensitive to friction with the solvent molecules. Possibly, they range over the same order of magnitude as the size of the viscogenic molecules.

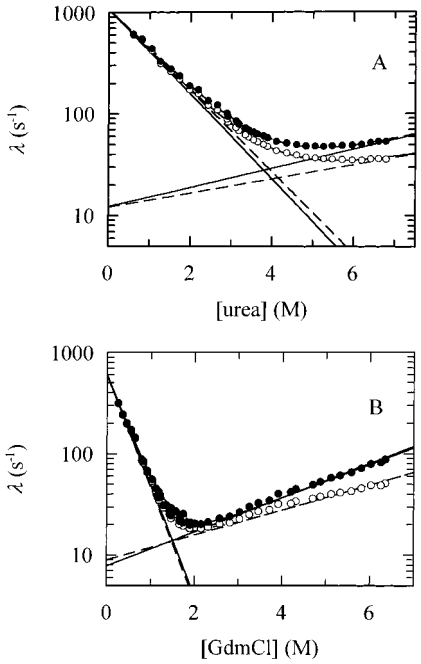

FIGURE 1: Correction of the (A) urea-dependent and (B) GdmCl-dependent unfolding and refolding kinetics of CspB for the increase in viscosity. The apparent rate constants λ are shown as a function of denaturant concentration before (○, broken lines) and after (●, continuous lines) correction. Fits of the data based on a two-state mechanism (*53*) are shown for the λ values and for the microscopic rate constants. The curves were corrected by using the equation $\ln \lambda_{corr} = \ln \lambda + \ln(\eta/\eta_0)$, where λ is the measured rate constant, λ_{corr} is the apparent rate constant after the viscosity correction, and η and η_0 are the viscosities in the presence and in the absence, respectively, of the denaturant. The relative viscosities of urea and GdmCl were taken from ref *52*. The analysis of the urea data depicted in panel A gave the following values: $m_{NU} = 0.16$ M^{-1} and $m_{UN} = -0.96$ M^{-1} before correction and $m_{NU} = 0.22$ M^{-1} and $m_{UN} = -0.92$ M^{-1} after correction. The analysis of the GdmCl data depicted in panel B gave the following values: $m_{NU} = 0.28$ M^{-1} and $m_{UN} = -2.6$ M^{-1} before correction and $m_{NU} = 0.38$ M^{-1} and $m_{UN} = -2.5$ M^{-1} after correction. The kinetics were measured in 0.1 M sodium cacodylate at pH 7.0 and 25 °C. The data depicted in panel A were taken from ref *53*, and the data depicted in panel B were measured by D. Perl (unpublished results).

The viscosity effects on protein folding are still difficult to understand at the molecular level. Part of this difficulty arises, because a concept (such as viscosity) that is well-defined for macroscopic systems is used to describe phenomena at the molecular level at protein−solvent interfaces. In this context, it is interesting to note that even limited conformational changes in native proteins can be sensitive to solvent viscosity. This has been shown for carboxypeptidase A (*33*) and for myoglobin (*34*, *36*). In both cases, the

viscosity dependence is assumed not to originate from the binding reaction itself, but from the coupled conformational rearrangements of the protein−ligand complexes.

COMPARISON OF KRAMERS' AND TST THEORIES FOR PROTEIN FOLDING

Both TST and Kramers' theories are descendants of the original Arrhenius relation (eq 1), and therefore, both contain an exponential factor which carries the information about the height of the activation energy barrier. This barrier is related to the slope when ln k is plotted as a function of $1/RT$. In TST, this slope is equal to the activation enthalpy $\Delta H^{\ddagger}$ (for controversies about this interpretation, see ref *29*). In Kramers' theory, the slope, as derived from an Arrhenius plot, carries contributions from both the energy barrier ΔU and the temperature dependence of η (eqs 6 and 7). In Kramers' formalism, the true activation energy ΔU is thus always smaller than the Arrhenius activation energy E_A. As outlined above, in aqueous solutions near 25 °C, this difference (W) amounts to about 16 kJ/mol.

The rates of chemical reactions, and particularly of protein folding reactions, depend not only on the height of the energy barrier but also on a statistical factor that accounts for the difference in the number of microscopic states that can be sampled by the system in the ground and activated states. In TST, this factor is $\exp(\Delta S^{\ddagger}/R)$ (cf. eq 3), which is thus immediately related to the activation entropy. It can, in principle, be calculated from eq 8 when $\Delta H^{\ddagger}$ is known from the temperature dependence of the rate constant.

$$\Delta S^{\ddagger} = (\Delta H^{\ddagger} - \Delta G^{\ddagger})/T \qquad (8)$$

In Kramers' theory, the statistical factor is ascribed to the frequency factors for the ground and activated states (*29*, *31*), which are located in the preexponential term. These frequency factors are less appealing, because, unlike $\Delta S^{\ddagger}$ in TST, they are not immediately related to the standard parameters that are used in equilibrium thermodynamics. This might have been one reason for the reluctancy to use Kramers' theory.

TST therefore appears to be much more attractive for answering questions about the entropy changes during the course of a protein folding reaction. It is, however, important to keep in mind that the $\Delta S^{\ddagger}$ values, as calculated from eq 8, are unreliable, because the $\Delta G^{\ddagger}$ values that are computed by TST are ambiguous and because the temperature dependence of the viscosity contributes to the apparent activation enthalpies (eq 6). Both theories are thus unable to provide quantitative values for entropic barriers. Basically, this difficulty relates to the fact that, unlike heat, entropy cannot be measured directly.

Another problem that is common to both theories and to all experimental work is the fact that only the overall changes in the activation parameters can be determined. It is impossible to measure the changes in chain entropy ΔS_{conf} separately from those in solvent entropy, ΔS_{solv}. Therefore, it is still very difficult to relate experimental data with theoretical folding scenarios that are based on funnel type models. In these models, the width of the funnel is usually defined by ΔS_{conf} and the depth by the free energy minus $T\Delta S_{conf}$ (*55*).

INTERNAL AND EXTERNAL FRICTION

Not all protein folding reactions are viscosity-dependent. This could be true for several reasons, with the first obviously being that Kramers' model is inadequate. As outlined above, we and others (*23*, *27*, *28*, *30*) consider this to be less likely. A folding reaction could also be independent of solvent viscosity when the movements of the protein chain in the activated state region are so small that they do not require the displacement of solvent molecules. This is also quite unlikely, because even conformational changes in folded proteins, such as those linked with ligand binding, are retarded when the viscosity of the solvent is increased (*33*−*36*).

It is more likely that a protein folds independently of solvent viscosity when the internal friction of the protein chain in the activated state region is much higher than the friction with the solvent molecules. This internal friction is certainly not constant. It starts from a presumably low value as in unstructured peptides (see ref *42*) and then increases during folding. For proteins such as protein L (*23*) and CspB (*45*, *47*), the internal viscosity seems to be still lower than the viscosity of the surrounding solvent when the activation barrier is crossed, and thus, it is not relevant as a rate-determining factor. This is probably the exception rather than the rule. In many protein folding reactions, the internal friction increases early in folding, in particular when intermediates are formed, and becomes higher than the friction with the solvent when the activated state region is traversed. In this case, folding will be independent of the viscosity of the solvent, as found for α-lactalbumin (*45*).

This adds another aspect to the picture of kinetic traps in folding. Such traps occur not only when high energy barriers are encountered but also when the protein chain becomes stuck because of a strong increase in internal viscosity. Proteins with low internal viscosity avoid such kinetic traps. A rapid collapse of the protein chain early in folding may be an advantage because it rapidly reduces the number of possible conformations. On the other hand, it can reduce the folding rate because the diffusional motions that are necessary to reach the native conformation are more difficult in a restricted space (see also the discussion in ref *56*).

Intuitively, the concept of an internal friction or internal viscosity appears to be attractive, but how can it be understood at the molecular level? Internal friction could originate from the sliding of chain segments relative to one another, similar to the relative sliding of layers of solvent molecules in viscous liquids. This might be a good representation of low internal friction when the protein chain can move rather freely. When, however, the protein becomes more densely packed, the chain segments can no longer slide smoothly along energy trajectories with many very small bumps, but they jump between discrete conformational substates, which are separated by sizable energy barriers. In such a case where the internal friction dominates over the friction with the solvent, the internal viscosity effect and the size of the energy barrier may become correlated, because they both orignate from the hindered movement of chain segments relative to one another in the activated state region. Therefore, it might not be adequate to simply replace the solvent viscosity in the preexponential factor of Kramers' relation with the internal viscosity. Further theoretical and

experimental work is needed to understand internal friction and to extend Kramers' model to describing folding reactions in which hindered diffusion contributes to both the preexponential factor and the barrier height.

The solvent is very important for folding, but its role is not clear at the energetic and structural level. Most proteins fold only in aqueous environments, and the folding dynamics and energetics are determined not only by the changes in the intraprotein interactions but also by the changes in the protein−solvent and solvent−solvent interactions. For a deeper understanding of protein folding, it would be necessary to dissect and quantitate these contributions individually. Experimentally, this is almost impossible because protein and solvent form a single thermodynamic system. Theoretical modeling of the folding process in an aqeous buffer is a formidable task because it is not yet possible to calculate the changes that occur in complex solvents when a protein chain folds in this solvent.

THEORETICAL APPROACHES TO MODELING PROTEIN FOLDING AS A DIFFUSIONAL PROCESS

Kramers' formalism was used by Karplus and Weaver when they developed the diffusion−collision model for protein folding (*32*), and on the basis of this model, the folding kinetics of myoglobin (*57*, *58*) and of the λ repressor headpiece (*59*) were analyzed. A Kramers type approach has also been used by Wolynes and co-workers to model the folding kinetics of a simplified on-lattice 27-mer protein as a diffusive reaction (*55*, *60*). Klimov and Thirumalai calculated the folding rates for a small off-lattice protein model by using the Langevin equation and found that Kramers' theory provided a quantitative fit of their numerical results (*61*). Thus, both theory and experiment suggest that diffusional processes determine the elementary folding of very small proteins as well as the coalescence of domains in the folding of large proteins. The recent review by Alm and Baker (*56*) provides an excellent discussion of the interrelation between theory and experiment in protein folding.

OUTLOOK

TST and Kramers' theory both have strengths and weaknesses. Kramers' theory should provide a good model for protein folding reactions because it is based on diffusional motion in a viscous solvent and thus accounts for the chain dynamics that are required for exploring and crossing the activation barrier. It can be tested by experiment, and it specifies how experimental activation energies must be corrected for the temperature dependence of viscosity to obtain the true activation enthalpies.

TST was developed for the reactions of small molecules in the gas phase. It is assumed that the molecules cross the activated state only once and that the ground and activated states exist in a pseudothermodynamic equilibrium. These assumptions are probably not met by protein folding reactions in solution. TST has remained attractive because all activation parameters, including the entropy of activation, can be calculated from the kinetic data. The absolute values for $\Delta S^{\ddagger}$ and for $\Delta G^{\ddagger}$ from a TST treatment of folding data are unreliable, but nevertheless, differences ($\Delta\Delta G^{\ddagger}$ and $\Delta\Delta S^{\ddagger}$), caused, for example, by a mutation, can be useful and interpretable in cases where the mutation does not change the structure of the ground and activated states. For such an analysis, it must be ascertained that the mutation leaves the overall folding mechanism unchanged and does not affect parameters such as the activation *m* values and the activation heat capacities. Analyses based on both theories are thus valuable and not mutually exclusive. TST can be used with the appropriate caution, but there is little doubt that Kramers' theory is better suited to modeling experimental kinetics and to simulating the folding kinetics of protein models.

ACKNOWLEDGMENT

We thank D. Perl for allowing us to reproduce his data in Figure 1 and acknowledge many discussions on viscosity effects with J. Balbach, W. Doster, E. Haas, R. Maier, A. Martin, D. Perl, C. Scholz, and D. Thirumalai. We also thank H. Gutfreund for writing an excellent book on the kinetics for the life sciences, which inspired much of our own recent work on protein folding.

REFERENCES

1. Pohl, F. M. (1976) *FEBS Lett. 65*, 293−296.
2. Levinthal, C. J. (1968) *J. Chim. Phys. 65*, 44−45.
3. Ikai, A., and Tanford, C. (1973) *J. Mol. Biol. 73*, 145−163.
4. Ikai, A., Fish, W. W., and Tanford, C. (1973) *J. Mol. Biol. 73*, 165−184.
5. Tanford, C., Aune, K. C., and Ikai, A. (1973) *J. Mol. Biol. 73*, 185−197.
6. Tsong, T.-Y., Baldwin, R. L., and Elson, E. L. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1809−1812.
7. Garel, J. R., and Baldwin, R. L. (1973) *Proc. Natl. Acad. Sci. U.S.A. 70*, 3347−3351.
8. Brandts, J. F., Halvorson, H. R., and Brennan, M. (1975) *Biochemistry 14*, 4953−4963.
9. Nall, B. T., Garel, J.-R., and Baldwin, R. L. (1978) *J. Mol. Biol. 118*, 317−330.
10. Ridge, J. A., Baldwin, R. L., and Labhardt, A. M. (1981) *Biochemistry 20*, 1622−1630.
11. Cook, K. H., Schmid, F. X., and Baldwin, R. L. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 6157−6161.
12. Schmid, F. X., and Blaschek, H. (1981) *Eur. J. Biochem. 114*, 111−117.
13. Roder, H., Elöve, G. A., and Englander, S. W. (1988) *Nature 335*, 700−704.
14. Udgaonkar, J. B., and Baldwin, R. L. (1988) *Nature 335*, 694−699.
15. Balbach, J., Steegborn, C., Schindler, T., and Schmid, F. X. (1999) *J. Mol. Biol. 285*, 829−842.
16. Roder, H., and Colon, W. (1997) *Curr. Opin. Struct. Biol. 7*, 15−28.
17. Jackson, S. E., and Fersht, A. R. (1991) *Biochemistry 30*, 10428−10435.
18. Schindler, T., Herrler, M., Marahiel, M. A., and Schmid, F. X. (1995) *Nat. Struct. Biol. 2*, 663−673.
19. Kragelund, B. B., Robinson, C. V., Knudsen, J., Dobson, C. M., and Poulsen, F. M. (1995) *Biochemistry 34*, 7217−7224.
20. Huang, G. S., and Oas, T. G. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 6878−6882.
21. Waldburger, C. D., Jonsson, T., and Sauer, R. T. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 2629−2634.
22. Viguera, A. R., and Serrano, L. (1997) *Nat. Struct. Biol. 4*, 939−946.
23. Plaxco, K. W., and Baker, D. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 13591−13596.
24. Grantcharova, V. P., Riddle, D. S., Santiago, J. V., and Baker, D. (1998) *Nat. Struct. Biol. 5*, 714−720.
25. Fersht, A. R. (1995) *Curr. Opin. Struct. Biol. 5*, 79−84.
26. Kramers, H. A. (1940) *Physica 7*, 284−304.

27. Hynes, J. T. (1985) *Annu. Rev. Phys. Chem. 36*, 573−597.
28. Frauenfelder, H., and Wolynes, P. G. (1985) *Science 229*, 337−345.
29. Gavish, B. (1986) in *The fluctuating enzyme* (Welch, G. R., Ed.) pp 263−339, Wiley-Interscience, New York.
30. Gutfreund, H. (1995) *Kinetics for the life sciences*, Cambridge University Press, Cambridge, U.K.
31. Landauer, R., and Swanson, J. A. (1961) *Phys. Rev. 121*, 1668−1674.
32. Karplus, M., and Weaver, D. L. (1979) *Biopolymers 18*, 1421−1437.
33. Gavish, B., and Werber, M. M. (1979) *Biochemistry 18*, 1269−1275.
34. Ansari, A., Jones, C. M., Henry, E. R., Hofrichter, J., and Eaton, W. A. (1992) *Science 256*, 1796−1798.
35. Beece, D., Eisenstein, L., Frauenfelder, H., Good, D., Marden, M. C., Reinisch, L., Reynolds, A. H., Sorensen, L. B., and Yue, K. T. (1980) *Biochemistry 19*, 5147−5157.
36. Kleinert, T., Doster, W., Leyser, H., Petry, W., Schwarz, V., and Settles, M. (1998) *Biochemistry 37*, 717−733.
37. Teschner, W., Rudolph, R., and Garel, J.-R. (1987) *Biochemistry 26*, 2791−2796.
38. Vaucheret, H., Signon, L., Lebras, G., and Garel, J.-R. (1987) *Biochemistry 26*, 2785−2790.
39. Chrunyk, B. A., and Matthews, C. R. (1990) *Biochemistry 29*, 2149−2154.
40. Tsong, T. Y., and Baldwin, R. L. (1978) *Bioploymers 17*, 1669−1678.
41. Garel, J.-R. (1992) in *Protein Folding* (Creighton, T. E., Ed.) pp 405−454, Freeman, New York.
42. Haas, E., Katchalski-Katzir, E., and Steinberg, I. Z. (1978) *Biopolymers 17*, 11−31.
43. Goldberg, J. M., and Baldwin, R. L. (1998) *Biochemistry 37*, 2556−2563.
44. Goldberg, J. M., and Baldwin, R. L. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 2019−2024.
45. Jacob, M., Schindler, T., Balbach, J., and Schmid, F. X. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 5622−5627.
46. Timasheff, S. N. (1993) *Annu. Rev. Biophys. Biomol. Struct. 22*, 67−97.
47. Jacob, M., Geeves, M., Holtermann, G., and Schmid, F. X. (1999) *Nat. Struct. Biol.* (in press).
48. Bhattacharyya, R. P., and Sosnick, T. R. (1999) *Biochemistry 38*, 2601−2609.
49. Ladurner, A. G., and Fersht, A. R. (1999) *Nat. Struct. Biol. 6*, 28−31.
50. Gekko, K. (1982) *J. Biochem. (Tokyo) 91*, 1197−1204.
51. Jacob, M., Holtermann, G., Perl, D., Reinstein, J., Schindler, T., Geeves, M. A., and Schmid, F. X. (1999) *Biochemistry 38*, 2882−2891.
52. Kawahara, K., and Tanford, C. (1966) *J. Biol. Chem. 241*, 3228−3232.
53. Schindler, T., and Schmid, F. X. (1996) *Biochemistry 35*, 16833−16842.
54. Perl, D., Welker, C., Schindler, T., Schröder, K., Marahiel, M. A., Jaenicke, R., and Schmid, F. X. (1998) *Nat. Struct. Biol. 5*, 229−235.
55. Onuchic, J. N., Socci, N. D., Luthey-Schulten, Z., and Wolynes, P. G. (1996) *Folding Des. 1*, 441−450.
56. Alm, E., and Baker, D. (1999) *Curr. Opin. Struct. Biol. 9*, 189−196.
57. Bashford, D., Cohen, F. E., Karplus, M., Kuntz, I. D., and Weaver, D. L. (1988) *Proteins: Struct., Funct., Genet. 4*, 211−227.
58. Pappu, R. V., and Weaver, D. L. (1998) *Protein Sci. 7*, 480−490.
59. Burton, R. E., Myers, J. K., and Oas, T. G. (1998) *Biochemistry 37*, 5337−5343.
60. Socci, N. D., Onuchic, J. N., and Wolynes, P. G. (1996) *J. Chem. Phys. 104*, 5860−5868.
61. Klimov, D. K., and Thirumalai, D. (1997) *Phys. Rev. Lett. 79*, 317−320.

BI991503O